

В. ФЕЛЛЕР

**ВВЕДЕНИЕ
В ТЕОРИЮ ВЕРОЯТНОСТЕЙ
И ЕЕ ПРИЛОЖЕНИЯ**

В 2-х томах

Том 2

Перевод со второго
английского издания
Ю. В. Прохорова

МОСКВА «МИР» 1984

AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS

WILLIAM FELLER (1906—1970)

**Eugene Higgins Professor of Mathematics
Princeton University**

VOLUME II

Second edition

**John Wiley & Sons, Inc.
New York · London · Sydney · Toronto
1971**

ББК 22.171

Ф 30

УДК 519.24

Феллер В.

Ф 30 Введение в теорию вероятностей и ее приложения.
В 2-х томах. Т. 2. Пер. с англ.— М.: Мир, 1984.— 738 с., ил.

Второй том всемирно известного двухтомного курса теории вероятностей, написанного выдающимся американским математиком. Классическое учебное руководство, оказавшее значительное влияние на развитие современной теории вероятностей и подготовку специалистов. Перевод заново выполнен со второго переработанного автором издания. Предыдущее издание выходило в русском переводе (М.: Мир, 1967).

Для математиков — от студентов до специалистов по теории вероятностей, для физиков и инженеров, применяющих вероятностные методы.

Ф $\frac{1702060000-113}{041(01)-84}$ 32—84, ч. 1

ББК 22.171
517.8

Редакция литературы по математическим наукам

ИЗ ПРЕДИСЛОВИЯ К РУССКОМУ ИЗДАНИЮ 1967 г.

Второй том книги известного американского математика В. Феллера „Введение в теорию вероятностей и ее приложения“ вышел в США в 1966 г. Несмотря на большой промежуток времени между выходом первого и второго томов, оба тома имеют общий замысел и составляют единое целое. Книга дает строгое изложение теории вероятностей как самостоятельного раздела математики и в то же время знакомит читателя с опытными основаниями теории и различными применениями. Последнее достигается включением большого числа примеров и задач.

Книга очень богата содержанием. Многочисленные отступления от основного текста содержат сведения, интересные и специалистам. Большая работа, проведенная автором при подготовке второго тома, позволила существенно упростить изложение целых разделов (например, вывод асимптотических формул в теории случайных блужданий и задачах о разорении и многое другое).

Изложение тщательно продумано. Оно часто дополняется замечаниями, отражающими отношение автора к приводимым фактам (см., например, замечание в гл. VI, 7, о том, что рассмотрение в теории очередей потоков вызовов с независимыми промежутками между вызовами и с законом распределения длины этих промежутков, отличным от вырожденного или показательного, создает лишь иллюзию общности; трудно, говорит автор, найти примеры таких процессов, кроме разве движения автобуса по кольцевому маршруту без расписания). Напомню, что в первом томе автор удачно продемонстрировал тот факт, что сравнительно простые модели позволяют, хотя бы в первом приближении, правильно описать широкий круг практических задач (такими моделями в первом томе являются, например, размещения частиц по ячейкам, урновые схемы и случайные блуждания). Во многих случаях, где интуиция не подсказывает правильного порядка соответствующих вероятностей, автор приводил численные результаты. Подчеркивался ряд свойств случайности, идущих вразрез с интуитивными представлениями (закон арксинуса, пропорциональность времени до n -го возвращения величине n^2 и т. п.). Эти тенденции сохраняются и во втором томе (см., в частности, неоднократное обсуждение парадокса инспекции и сходных тем).

После того как разобран дискретный случай, элементарная теория непрерывных распределений требует лишь нескольких слов дополнительного объяснения. Поэтому до всякой общей теории, в первых трех главах, автор излагает задачи, связанные с тремя важнейшими распределениями — равномерным, показательным и нормальным. При этом автор затрагивает и ряд глубоких вопросов (случайные разбиения и теоремы о покрытиях, отклонение эмпи-

рического распределения от теоретического, характеристика нормального распределения независимостью статистик, структура некоторых стационарных нормальных последовательностей и т. п.).

Необходимые сведения из теории меры сообщаются в четвертой главе. Автор подчеркивает вспомогательный характер этих сведений (что отличает книгу Феллера от ряда других современных руководств, где до трети объема уходит на теорию меры и интеграла).

Важную роль играет гл. VI, являющаяся, как замечает автор, собранием введений во все последующие главы.

В настоящем томе, по-видимому, впервые излагается так обстоятельно теория преобразований Лапласа в применении к вероятностным проблемам (гл. VII, XIII, XIV). Много места отведено вопросам теории восстановления и случайным блужданиям (гл. VI, XI, XII, XIV и XVIII). Предельные теоремы для сумм независимых величин излагаются дважды: сначала как иллюстрация операторных методов (гл. IX), а затем в общей форме в гл. XVII.

Можно надеяться, что выход в свет русского издания второго тома книги Феллера окажет заметное воздействие на многие стороны развития теории вероятностей, в частности сильно повлияет на характер преподавания теории вероятностей, позволив, наконец, привести его в соответствие с современными требованиями.

Профессор Феллер, узнав о подготовке перевода второго тома, любезно прислал список ряда необходимых исправлений, которые были внесены в текст. Я весьма благодарен ему за эту любезность. В работе над переводом существенную помощь мне оказали А. В. Прохоров и В. В. Сазонов. Они внимательно прочли рукопись перевода и сделали ряд ценных замечаний, которыми я воспользовался. Они также указали на ряд неточностей в оригинале. Пользуюсь случаем выразить им свою глубокую благодарность.

9 мая 1967 г.

Ю. Прохоров

ОТ ПЕРЕВОДЧИКА

План второго английского издания данного тома (с которого выполнен перевод) в целом повторяет план его первого издания, и в то же время автором внесено в текст большое число изменений. Подробно о них сказано в предисловии автора.

При подготовке перевода второго тома неоценимую помощь оказал мне А. Прохоров, а в работе над отдельными частями текста мне помогали А. Архангельский, А. Бряндинская, В. Матвеев, А. Ушакова, В. Ушаков, Н. Ушаков. Всем им я выражаю свою глубокую благодарность.

20 апреля 1983 г.

Ю. Прохоров

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

В то время когда писался первый том этой книги (между 1941 и 1948 гг.), интерес к теории вероятностей еще не был столь широким. Преподавание велось очень ограниченно, и такие вопросы, как цепи Маркова, которые теперь интенсивно используются в самых различных дисциплинах, рассматривались лишь как весьма специальные главы чистой математики. Поэтому первый том можно уподобить годному на все случаи жизни путеводителю для человека, отправляющегося в чужую страну. Чтобы отразить сущность теории вероятностей, в первом томе нужно было подчеркнуть как математическое содержание этой науки, так и удивительное разнообразие потенциальных применений. Некоторые предсказывали, что неизбежно возникающие из этой задачи колебания степени трудности изложения ограничат полезность книги. На самом деле первым томом книги широко пользуются и сейчас, когда новизна этой книги пропала, а ее материал (включая способ изложения) можно почерпнуть из более новых книг, написанных для специальных целей. Книга, как нам кажется, все еще завоевывает новых друзей. Тот факт, что неспециалисты не застряли в местах, оказавшихся трудными для студентов-математиков, показывает невозможность объективного измерения уровня трудности; этот уровень зависит от типа информации, которую ищет читатель, и от тех деталей, которые он готов пропустить. Так, путешественник часто стоит перед выбором: взбираться ли ему на гору самому или использовать подъемник.

Ввиду сказанного второй том написан в том же самом стиле, что и первый. Он включает более трудные разделы математики, но значительная часть текста может восприниматься с различной степенью трудности. Поясним это на примере способа изложения теории меры. Гл. IV содержит неформальное введение в основные идеи теории меры, а также дает основные понятия теории вероятностей. В этой же главе перечисляются некоторые факты теории меры, применяемые в последующих главах для того, чтобы сформулировать аналитические теоремы в их простейшей форме и избежать праздной дискуссии об условиях регулярности. Основное назначение теории меры в этой связи состоит в том, чтобы оправ-

¹⁾ Отто Нейгебауэр (р. 1899 г., с 1939 г. — в США) — австрийский математик, историк математики и астрономии. На русском языке опубликованы его книги «Лекции по истории античных математических наук. Том. I. Догреческая математика». — М. — Л.: ОНТИ, 1937 и «Точные науки древности». — М., 1968.

²⁾ О отрада моя, честь и прибежище! — К Меценату («Славный вкус, Меценат»), перевод А. Семенова-Тян-Шанского. — В кн.: Квинт Гораций Флакк. Оды, Эподы, Сатиры, Послания. — М.; Художественная литература, 1970.

дать формальные операции и переходы к пределу, чего никогда не потребовал бы нематематик. Поэтому читатели, заинтересованные в первую очередь в практических результатах, никогда не почувствуют какой-либо необходимости в теории меры.

Чтобы облегчить доступ к отдельным темам, изложение во всех главах сделано столь замкнутым, сколь это было возможно. Иногда частные случаи разбираются отдельно от общей теории. Некоторые темы (такие, как устойчивые законы или теория восстановления) обсуждаются в нескольких местах с различных позиций. Во избежание повторений определения и иллюстративные примеры собраны в гл. VI, которую можно описать как собрание введений к последующим главам. Остов книги образуют гл. V, VIII и XV. Читатель сам решит, сколько подготовительных глав нужно ему прочитать и какие экскурсы произвести.

Специалисты найдут в этом томе новые результаты и доказательства, но более важной представляется попытка сделать единой общую методологию. В самом деле, некоторые части теории вероятностей страдают от того, что они недостаточно внутренне согласованы, а также от того, что группировка материала и способ изложения в большой степени определяются случайностями исторического развития. В образующейся путанице тесно связанные между собой проблемы выступают разобщенными, а простые вещи затемняются усложненными методами. Значительные упрощения были достигнуты за счет систематического применения и развития наилучших доступных сейчас методов. Это относится, в частности, к такой беспорядочной области, как предельные теоремы (гл. XVI, XVII). В других местах упрощения были достигнуты за счет трактовки задач в их естественном контексте. Например, элементарное исследование специального случайного блуждания привело к обобщению асимптотической оценки, которая ранее была выведена тяжелыми, трудоемкими методами в математической теории страхования (и независимо, при более ограничительных условиях, в теории очередей).

Я пытался достичь математической строгости, не впадая в педантизм. Например, утверждение, что $1/(1+\xi^2)$ является характеристической функцией для $\frac{1}{2}e^{-|x|}$, представляется мне желательным и законным сокращением для логически корректного варианта: функция, которая в точке ξ принимает значение $1/(1+\xi^2)$, является характеристической функцией для функции, которая в точке x принимает значение $\frac{1}{2}e^{-|x|}$.

Боюсь, что краткие исторические замечания и ссылки не отдают должного многим авторам, внесшим свой вклад в теорию вероятностей. Однако всюду, где было возможно, я старался сделать это. Первоначальные работы во многих случаях перек-

рыты более новыми исследованиями, и, как правило, полные ссылки даются только на статьи, к которым читатель может обратиться для получения дальнейшей информации. Например, нет ссылок на мои собственные работы по предельным теоремам, в то же время какая-либо статья, описывающая результаты наблюдений или идеи, лежащие в основе примера, цитируется, даже если она совсем не содержит математики¹⁾. В этих обстоятельствах авторский указатель не дает никаких сведений о важности тех или иных исследований для теории вероятностей. Другая трудность состояла в справедливой оценке работ, которым мы обязаны новыми направлениями исследований, новыми подходами, новыми методами. Некоторые теоремы, рассматривавшиеся в свое время как поразительно оригинальные и глубокие, теперь получили простые доказательства и выступают в окружении более тонких результатов. Трудно воспринимать такие теоремы в их исторической перспективе и нелегко понять, что здесь, как и в других случаях, первый шаг значит очень многое.

Я благодарен Исследовательскому бюро армии США за поддержку моей работы по теории вероятностей в Принстонском университете, где мне помогали Дж. Голдмэн, Л. Питт, М. Силверстейн. Они устранили значительное число неточностей и неясных мест. Все главы переписывались много раз, и первоначальные варианты предшествующих глав распространялись среди моих друзей. Таким образом, благодаря замечаниям Дж. Эллиота, Р. С. Пинхема и Л. Дж. Сэвиджа возник ряд улучшений. Я особенно благодарен Дж. Л. Дубу и Дж. Волфовицу за советы и критику. График случайного блуждания Коши подготовлен Г. Тротером. За печатанием наблюдала г-жа Х. Мак-Дугл, и внешний вид книги многим обязан ей.

Вильям Феллер

Октябрь 1965 г.

¹⁾ Эта система была принята и в первом томе, однако она была неправильно понята некоторыми последующими авторами; они теперь приписывают методы, использованные в книге, работавшим до этого ученым, которые не могли знать их.

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Характер и структура книги остались без изменений, но весь текст подвергся основательному пересмотру. Многие разделы были переписаны полностью (в частности, гл. XVII) и были добавлены некоторые новые параграфы. В ряде мест изложение упрощено благодаря усовершенствованным (а иногда и новым) рассуждениям. В текст был также включен новый материал.

При подготовке первого издания меня преследовал страх, что объем получится чрезмерно большим. Из-за этого, к сожалению, я провел несколько бесплодных месяцев, сокращая первоначальный текст и выделяя отдельные места в мелкий шрифт. Этот ущерб теперь возмещен, и много усилий было потрачено на то, чтобы облегчить чтение книги. Встречающиеся время от времени повторения упрощают независимое чтение отдельных глав и позволяют связать некоторые части этой книги с материалом первого тома.

Расположение материала описано в предисловии к первому изданию, воспроизведенному здесь (см. второй абзац и далее).

Я благодарен многим читателям за указание ошибок и упущений. Особо я благодарю Хаджала (D. A. Hejhal) из Чикаго за весьма полный и впечатляющий список опечаток и за замечания, касающиеся многих мест книги.

Январь 1970 г., Принстон

Вильям Феллер

К моменту кончины автора работа над рукописью уже была завершена, но корректуры получены не были. Я благодарю издателей, выделивших сотрудника для тщательного сличения корректуры с рукописью и за составление указателя. Проверку математического содержания книги произвели сообща Голдмэн (J. Goldman), Грюнбаум (A. Grunbaum), Маккин (H. McKean), Питт (L. Pitt) и Питтенджер (A. Pittenger). Каждый математик понимает, какой невероятный объем работы стоит за этим. *Я выражаю свою глубокую признательность этим ученым и обращаю к ним слова искренней благодарности за их бескорыстный труд.*

Клара Н. Феллер

Май 1970 г.

ОБОЗНАЧЕНИЯ

Интервалы обозначаются чертой сверху:

$\overline{a, b}$ — открытый, $\overline{\overline{a, b}}$ — замкнутый интервал; полуоткрытые интервалы обозначаются $\overline{\overline{a, b}}$ и $\overline{\overline{a, b}}$. Эти обозначения применяются также и в многомерном случае. Соответствующие соглашения о векторных обозначениях и упорядочении см. в гл. V, 1 (а также в гл. IV, 2). Символ (a, b) резервирован для пар и для точек.

$\mathcal{R}^1, \mathcal{R}^2, \mathcal{R}^r$ обозначают соответственно прямую, плоскость и r -мерное евклидово пространство.

1 означает ссылку на первый том; римские цифры указывают номер главы. Так, I, гл. XI, (3.6) указывает на отношение к § 3 гл. XI первого тома.

► указывает конец доказательства или серии примеров.
и и $\mathfrak{N}$ обозначают соответственно нормальную плотность и нормальное распределение с нулевым математическим ожиданием и единичной дисперсией.

O, o и $\sim$ Пусть u и v зависят от параметра x , который стремится, скажем, к a . Предполагая, что v положительно, мы пишем

$$\left. \begin{array}{l} u = O(v) \\ u = o(v) \\ u \sim v \end{array} \right\} \text{ если } \frac{u}{v} \left\{ \begin{array}{l} \text{ограничено} \\ \rightarrow 0 \\ \rightarrow 1 \end{array} \right.$$

$f(x)U\{dx\}$ об этом сокращении см. гл. V, 3.

Относительно борелевских множеств и бэровских функций см. введение к гл. V.

§ 1. ВВЕДЕНИЕ

В первом томе мы неоднократно имели дело с вероятностями, определяемыми суммами многих малых слагаемых, и пользовались приближениями вида

$$P\{a < X < b\} \approx \int_a^b f(x) dx. \quad (1.1)$$

Основным примером служит нормальное приближение к биномиальному распределению¹⁾. Приближение этого типа обыкновенно устанавливается в форме предельной теоремы, включающей ряд все более и более «тонких» дискретных вероятностных моделей. Во многих случаях этот переход к пределу приводит, вообще говоря, к новому выборочному пространству, и последнее может быть интуитивно проще, чем первоначальная дискретная модель.

Примеры. а) *Показательные времена ожидания.* Для того чтобы описать времена ожидания дискретной моделью, мы должны сделать время дискретным и условиться, что изменения могут осуществляться только в моменты времени δ , 2δ , ... Простейшее время ожидания T есть время ожидания первого успеха в последовательности испытаний Бернулли с вероятностью успеха p_δ . Тогда $P\{T > n\delta\} = (1 - p_\delta)^n$ и среднее время ожидания равно $E(T) = \delta/p_\delta$. Эту модель можно усовершенствовать, если предположить δ убывающим так, что математическое ожидание $\delta/p_\delta = \alpha$ остается фиксированным. Любому фиксированному интервалу времени t соответствует $n \approx t/\delta$ испытаний, и, следовательно, при малом δ мы имеем

$$P\{T > t\} \approx \left(1 - \frac{\delta}{\alpha}\right)^{t/\delta} \approx e^{-t/\alpha}, \quad (1.2)$$

что доказывается переходом к логарифмам. Эта модель описывает время ожидания как геометрически распределенную дискретную случайную величину, а (1.2) утверждает, что «в пределе» получается показательное распределение. Интуиция подсказывает, что

¹⁾ Другие примеры из первого тома: распределение арксинуса, гл. III,4; распределение числа возвращений в начальное состояние и времен первого прохождения, гл. III,7; предельные теоремы для случайных блужданий, гл. XIV; равномерное распределение, задача 20 в гл. XI,7.

более естественно начать с выборочного пространства, точками которого служат действительные числа, и ввести показательное распределение непосредственно.

б) *Случайный выбор*. «Выбор точки наудачу» в интервале ¹⁾ $\overline{0,1}$ является воображаемым экспериментом с очевидным интуитивным смыслом. Этот эксперимент может быть описан дискретными приближениями, но проще взять в качестве выборочного пространства весь интервал и приписать каждому подынтервалу в качестве вероятности его длину. Воображаемый эксперимент, заключающийся в том, что производятся два независимых выбора точки в $\overline{0,1}$, дает пару вещественных чисел, и поэтому естественным выборочным пространством служит единичный квадрат. В этом выборочном пространстве «вероятность» почти инстинктивно приравнивается «площади». Этого вполне достаточно для некоторых элементарных целей, но рано или поздно возникает вопрос о том, что же в действительности означает слово «площадь». ►

Как показывают эти примеры, непрерывное выборочное пространство может быть проще для восприятия, чем дискретная модель, но определение вероятностей в нем зависит от такого аппарата, как интегрирование и теория меры. В счетных выборочных пространствах можно было приписать вероятности *всем* вообще событиям, в то время как в пространствах произвольной природы эта простая процедура ведет к логическим противоречиям, и наша интуиция должна приспособливаться к крайностям формальной логики. Мы вскоре увидим, что наивный подход может привести к затруднениям даже в сравнительно простых задачах. Ради справедливости стоит сказать, что в то же время многие значительные вероятностные задачи не требуют четкого определения понятия вероятности. Иногда они имеют аналитический характер, а вероятностное содержание служит лишь опорой для нашей интуиции. Более существен тот факт, что при изучении сложных случайных процессов (и соответственно усложненных выборочных пространств) могут возникнуть важные и понятные задачи, не связанные с теми тонкими средствами, которые используются при анализе процесса в целом. Типичное рассуждение может иметь следующий вид: если процесс вообще допускает математическое описание, то случайная величина Z должна иметь такие-то и такие-то свойства, а ее распределение должно в силу этого удовлетворять такому-то уравнению. Хотя вероятностные соображения могут сильно влиять на ход исследования этого уравнения, последнее в принципе не зависит от аксиом теории вероятностей. Специалисты различных прикладных областей настолько привыкли иметь дело с подобными

¹⁾ Для обозначения интервалов используется черта. Символ (a, b) сохраняется для координатной записи точек на плоскости, см. обозначения в начале книги.

проблемами, что они отрицают необходимость теории меры: они не знакомы с задачами другого типа и с ситуациями, где нечеткие рассуждения приводили к неверным результатам¹⁾.

Эти замечания постепенно разъясняются на протяжении этой главы, которая предназначена служить неформальным введением в теорию. В ней описываются некоторые аналитические свойства двух важных распределений, которые будут постоянно использоваться в этой книге. Специальные темы затрагиваются отчасти по причине важных применений, отчасти для демонстрации новых задач, требующих привлечения новых методов. Нет необходимости изучать их систематически или в том порядке, как они представлены.

Всюду в этой главе вероятности понимаются как элементарные интегралы со всеми вытекающими отсюда ограничениями. Использование вероятностного «жаргона» и таких терминов, как случайные величины и математические ожидания, может быть оправдано с двух точек зрения. Можно рассматривать их как техническое вспомогательное средство для нашей интуиции, имеющее в основе формальную аналогию с изложенным в первом томе. С другой стороны, все сказанное в этой главе может быть формально безупречно обосновано предельным переходом от дискретной модели, описанной в примере 2, а). Последняя процедура, хотя и не является ни необходимой, ни желательной, может быть хорошим упражнением для начинающих.

§ 2. ПЛОТНОСТИ. СВЕРТКИ

Плотностью вероятности на прямой (или в $\mathcal{R}^1$) называется функция f , такая, что

$$f(x) \geq 0, \quad \int_{-\infty}^{+\infty} f(x) dx = 1. \quad (2.1)$$

Для начала мы рассмотрим только кусочно-непрерывные плотности (общий случай см. в гл. V,3). Каждой плотности f мы

¹⁾ Роли строгости и интуиции часто понимаются неправильно. Как уже отмечалось в первом томе, природная интуиция и естественный способ мышления дают мало, но они становятся сильнее по мере развития математической теории. Сегодняшняя интуиция и применения опираются на сложные теории вчерашнего дня. Кроме того, строгая теория не роскошь, а способ экономии мышления. В самом деле, наблюдения показывают, что в приложениях многие авторы полагаются на длинные вычисления, а не на простые рассуждения, которые им кажутся рискованными [ближайшая иллюстрация — пример 5,а)].

поставим в соответствие *функцию распределения*¹⁾ F , определенную равенством

$$F(x) = \int_{-\infty}^x f(y) dy. \quad (2.2)$$

Это есть монотонная непрерывная функция, возрастающая от 0 до 1. Мы скажем, что f и F *сосредоточены на интервале* $a \leq x \leq b$, если f равна нулю вне этого интервала. Плотность f будет рассматриваться как функция, задающая вероятности на интервалах прямой; интервалу $\overline{a, b} = \{a < x < b\}$ соответствует вероятность

$$F(b) - F(a) = \int_a^b f(x) dx. \quad (2.3)$$

Иногда эта вероятность будет обозначаться $P\{\overline{a, b}\}$. При таком распределении отдельной точке соответствует вероятность 0, а замкнутому интервалу $\overline{a, b}$ — та же вероятность, что и $\overline{a, b}$.

В простейшей ситуации действительная прямая служит *«выборочным пространством»*, т. е. результат воображаемого эксперимента представляется числом. (Точно так же, как и в первом томе, это лишь первый шаг в построении выборочных пространств, представляющих последовательности экспериментов.) *Случайные величины* являются функциями, определенными на выборочном пространстве. Для простоты мы договоримся на некоторое время считать функцию U случайной величиной, если для каждого t событие $\{U \leq t\}$ состоит из конечного множества интервалов. Тогда функция

$$G(t) = P\{U \leq t\} \quad (2.4)$$

может быть определена как интеграл от f по объединению этих интервалов. Функция G , заданная равенством (2.4), называется *функцией распределения* величины U . Если G является интегралом от некоторой функции g , то g называется *плотностью* распределения G или (что равнозначно) *плотностью* величины U .

Основная случайная величина — это, конечно, координатная величина²⁾ X сама по себе, а все другие случайные величины

¹⁾ Мы напоминаем, что под «функцией распределения» понимается непрерывная справа неубывающая функция с пределами 0 и 1 на $-\infty$ и $+\infty$. В первом томе мы преимущественно имели дело со ступенчатыми функциями. Теперь сосредоточим наше внимание на функциях распределения, задаваемых интегралами. Более подробно мы будем изучать функции распределения в гл. V.

²⁾ Всюду, где это возможно, мы будем обозначать случайные величины (т. е. функции на выборочном пространстве) заглавными полужирными буквами, оставляя малые буквы для обозначения чисел или масштабных параметров. Это справедливо, в частности, для координатной величины X , а именно функции, определенной равенством $X(x) = x$.

являются функциями X . Функция распределения X тождественна функции F , посредством которой задаются вероятности. Необходимо сказать, что любая случайная величина $Y = g(X)$ может рассматриваться как координатная величина на новой прямой.

Как сказано выше, употребление этих терминов может быть оправдано аналогией с рассмотренным в первом томе. Следующий пример показывает, что наша модель может быть получена из дискретных моделей предельным переходом.

Примеры. а) *Группировка данных.* Пусть F — заданная функция распределения. Выберем фиксированное $\delta > 0$ и рассмотрим дискретную случайную величину X_δ , которая в интервале $(n-1)\delta < x \leq n\delta$ принимает постоянное значение $n\delta$. Здесь $n = 0, \pm 1, \pm 2, \dots$. В первом томе мы использовали совокупность чисел, кратных δ , как выборочное пространство и описывали распределение вероятностей X_δ равенством

$$P\{X_\delta = n\delta\} = F(n\delta) - F((n-1)\delta). \quad (2.5)$$

Теперь X_δ — случайная величина в расширенном выборочном пространстве, и ее функция распределения — это функция, которая в интервале $n\delta \leq x < (n+1)\delta$ равна $F(n\delta)$. В непрерывной модели X_δ служит приближением к X , получаемым отождествлением наших интервалов с их концами (процедура, известная статистикам как группировка данных). В духе первого тома мы будем рассматривать X_δ как основную случайную величину, и δ — как свободный параметр. Полагая $\delta \rightarrow 0$, мы получим предельные теоремы, устанавливающие, например, что F служит предельным распределением для X_δ .

б) При $x > 0$ событие $\{X^2 \leq x\}$ равносильно событию $\{-\sqrt{x} \leq X \leq \sqrt{x}\}$; случайная величина X^2 имеет распределение, сосредоточенное на $0, \infty$ с функцией распределения $F(\sqrt{x}) - F(-\sqrt{x})$. После дифференцирования видно, что плотность g величины X^2 равна $g(x) = \frac{1}{2}[f(\sqrt{x}) + f(-\sqrt{x})]/\sqrt{x}$ при $x > 0$ и $g(x) = 0$ при $x < 0$. Аналогично функция распределения X^3 равна $F(\sqrt[3]{x})$ и имеет плотность $\frac{1}{3}f(\sqrt[3]{x})/\sqrt[3]{x^2}$. ►

Математическое ожидание X определяется формулой

$$E(X) = \int_{-\infty}^{+\infty} xf(x) dx \quad (2.6)$$

при условии, что интеграл сходится абсолютно. Математические ожидания аппроксимирующих дискретных величин X_δ из примера а) совпадают с римановыми суммами этого интеграла, и поэтому $E(X_\delta) \rightarrow E(X)$. Если μ есть ограниченная непрерывная функция, то те же самые рассуждения применимы к случайной величине

$u(\mathbf{X})$, и соотношение $E(u(\mathbf{X}_k)) \rightarrow E(u(\mathbf{X}))$ влечет за собой

$$E(u(\mathbf{X})) = \int_{-\infty}^{+\infty} u(x) f(x) dx. \quad (2.7)$$

Здесь следует отметить, что в этой формуле явно не используется распределение $u(\mathbf{X})$. Таким образом, для вычисления математического ожидания функций от $\mathbf{X}$ достаточно знать распределение случайной величины $\mathbf{X}$.

Второй момент $\mathbf{X}$ определяется равенством

$$E(\mathbf{X}^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx \quad (2.8)$$

при условии, что интеграл сходится. Обозначим $\mu = E(\mathbf{X})$. Тогда дисперсия $\mathbf{X}$ определяется как

$$\text{Var}(\mathbf{X}) = E((\mathbf{X} - \mu)^2) = E(\mathbf{X}^2) - \mu^2. \quad (2.9)$$

Замечание. Если величина $\mathbf{X}$ положительна (т. е. если плотность f сосредоточена на $[0, \infty)$) и если интеграл в выражении (2.6) расходится, то удобно говорить, что $\mathbf{X}$ имеет бесконечное математическое ожидание, и писать $E(\mathbf{X}) = \infty$. Точно так же говорят, что $\mathbf{X}$ имеет бесконечную дисперсию, когда интеграл в (2.8) расходится. Для величин, принимающих положительные и отрицательные значения, математическое ожидание остается неопределенным, когда интеграл (2.6) расходится. Типичный пример доставляет плотность $x^{-1}(1+x^2)^{-1}$. ►

Понятие плотности переносится на размерности более высоких порядков, но общее обсуждение мы отложим до гл. III. Пока же мы будем рассматривать только аналог произведения вероятностей, введенного определением 2 в I, гл. V, 4, для описания комбинаций независимых экспериментов. Другими словами, в этой главе мы будем иметь дело только с плотностями, представляемыми произведениями вида $f(x)g(y)$, $f(x)g(y)h(z)$ и т. д., где $f, g, \dots$ — плотности на прямой. Задание плотности вида $f(x)g(y)$ на плоскости $\mathcal{R}^2$ означает отождествление «вероятностей» с интегралами

$$\mathbf{P}\{A\} = \iint_A f(x)g(y) dx dy. \quad (2.10)$$

Когда мы говорим о «двух независимых случайных величинах $\mathbf{X}$ и $\mathbf{Y}$ с плотностями f и g », то это есть сокращение высказывания о том, что вероятности на $(\mathbf{X}, \mathbf{Y})$ -плоскости определяются в соответствии с формулой (2.10). Отсюда следует правило умножения для интервалов, например $\mathbf{P}\{\mathbf{X} > a, \mathbf{Y} > b\} = \mathbf{P}\{\mathbf{X} > a\} \times \mathbf{P}\{\mathbf{Y} > b\}$. Аналогия с дискретным случаем настолько очевидна, что дальнейшие разъяснения не требуются.

Можно ввести много новых случайных величин, рассматривая функции X и Y , но наиболее важную роль играет сумма $S = X + Y$. Событие $A = \{S \leq s\}$ изображается полуплоскостью точек (x, y) , таких, что $x + y \leq s$. Обозначим функцию распределения Y через G , так что $g(y) = G'(y)$. Чтобы найти функцию распределения $X + Y$, мы интегрируем в (2.10) по области $y \leq s - x$. В результате получаем

$$P\{X + Y \leq s\} = \int_{-\infty}^{+\infty} G(s-x) f(x) dx. \quad (2.11)$$

Ввиду симметрии можно поменять F и G ролями без влияния на результат. После дифференцирования видно, что плотность $X + Y$ равна любому из двух интегралов

$$\int_{-\infty}^{+\infty} f(s-y) g(y) dy = \int_{-\infty}^{+\infty} f(y) g(s-y) dy. \quad (2.12)$$

Операция, определенная формулой (2.12), является частным случаем свертки, которая будет введена в гл. V, 4. До этого времени мы будем использовать термин «свертка» только для плотностей: *свертка двух плотностей f и g есть функция, определенная в (2.12). Свертка будет обозначаться $f * g$.*

Повсюду в первом томе мы имели дело со свертками дискретных распределений. Соответствующие правила справедливы и здесь. В силу (2.12) мы имеем $f * g = g * f$. Если задана третья плотность h , мы можем составить свертку $(f * g) * h$. Это есть плотность суммы $X + Y + Z$ трех независимых величин с плотностями f, g, h . Тот факт, что суммирование коммутативно и ассоциативно, влечет те же свойства для сверток, и поэтому результат $f * g * h$ не зависит от порядка операций.

Положительные случайные величины играют важнейшую роль, и потому полезно отметить, что если плотности f и g сосредоточены на $\overline{0, \infty}$, то свертка $f * g$ равна

$$f * g(s) = \int_0^s f(s-y) g(y) dy = \int_0^s f(x) g(s-x) dx. \quad (2.13) \blacktriangleright$$

Пример. в) Пусть f и g сосредоточены на $\overline{0, \infty}$ и определены здесь равенствами $f(x) = \alpha e^{-\alpha x}$ и $g(x) = \beta e^{-\beta x}$. Тогда

$$f * g(x) = \alpha \beta \frac{e^{-\alpha x} - e^{-\beta x}}{\beta - \alpha}, \quad x > 0. \quad (2.14) \blacktriangleright$$

(Продолжение в задаче 12.)

Замечание о понятии случайной величины. Использование прямой линии или евклидовых пространств $\mathcal{R}^n$ как выборочных пространств иногда затемняет различие между случайными вели-

чинами и «обычными» функциями одной или большего числа переменных. В первом томе случайная величина X могла принимать лишь счетное множество значений, и в этом случае было ясно, говорили ли мы о функции, например квадрате или экспоненте, определенной на прямой, или о случайной величине X^2 или e^X , определенной на выборочном пространстве. Даже по внешнему виду эти функции совершенно различны, поскольку «обычная» показательная функция принимает все положительные значения, тогда как e^X имеет счетную область значений. Для того чтобы увидеть изменение в этой ситуации, рассмотрим теперь «две независимые случайные величины X и Y с одинаковой плотностью f ». Другими словами, выборочным пространством служит плоскость и вероятности определяются как интегралы от $f(x)f(y)$. Теперь любая функция двух переменных *может* быть определена на выборочном пространстве, и тогда она становится случайной величиной, однако следует иметь в виду, что функция двух переменных может быть также определена независимо от нашего выборочного пространства. Например, некоторые статистические задачи требуют введения случайной величины $f(X)f(Y)$ [см. пример 12, г) гл. VI]. С другой стороны, вводя наше выборочное пространство $\mathcal{A}^2$, мы очевидным образом упоминали «обычную» функцию f , определенную независимо от выборочного пространства. Эта «обычная» функция задает множество случайных величин, например $f(X)$, $f(Y)$, $f(X \pm Y)$ и т. д. Таким образом, одна и та же функция f может служить или случайной величиной, или «обычной» функцией.

Как правило (и в каждом отдельном случае), будет ясно, имеем ли мы дело со случайной величиной или нет. Тем не менее в общей теории возникают ситуации, при которых функции (такие, как условные вероятности и условные математические ожидания) могут рассматриваться как «свободные» функции или как случайные величины, и это отчасти приводит к путанице, если свобода выбора не понимается должным образом.

Замечание о терминологии и обозначениях. Чтобы избежать громоздких формулировок, условимся называть $E(X)$ математическим ожиданием величины X , или плотности f , или распределения F . Подобные вольности будут введены и для других терминов. Например, «свертка» фактически означает операцию, но этот термин применяют и к результату операции, и функция $f * g$ называется «сверткой».

В старой литературе термины «распределение» и «функция частот» применялись к тому, что мы теперь называем плотностями; наши функции распределения описывались как «кумулятивные», а сокращение c.d.f.¹⁾ употребительно до сих пор.

¹⁾ Cumulative distribution function — кумулятивная функция распределения. — Прим. перев.

§ 3. ПОКАЗАТЕЛЬНАЯ ПЛОТНОСТЬ

Для произвольного, но фиксированного $\alpha > 0$ положим

$$f(x) = \alpha e^{-\alpha x}, \quad F(x) = 1 - e^{-\alpha x} \quad \text{при } x \geq 0 \quad (3.1)$$

и $F(x) = f(x) = 0$ при $x < 0$. Тогда f — показательная плотность, F — ее функция распределения. Простые вычисления показывают, что математическое ожидание равно α^{-1} , дисперсия — α^{-2} .

В примере 1, а) показательное распределение было выведено как предел геометрических распределений. Метод примера 2, а) приводит к тому же результату. Напомним, что в случайных процессах геометрическое распределение часто имеют времена ожидания или продолжительность жизни и что это обусловлено «отсутствием последствия», описанным в 1, гл. XIII, 9: *каков бы ни был настоящий возраст, оставшееся время жизни не зависит от прошлого и имеет то же самое распределение, что и само время жизни*. Теперь будет показано, что это свойство переносится на показательное распределение и только на показательное распределение.

Пусть T — произвольная положительная величина, которую мы будем интерпретировать как время жизни или как время ожидания. Удобно заменить функцию распределения T «хвостом распределения»

$$U(t) = P\{T > t\}. \quad (3.2)$$

С наглядной точки зрения $U(t)$ есть «вероятность того, что время жизни (от момента рождения) превзойдет t ». При данном возрасте s событие, состоящее в том, что оставшееся время жизни превосходит t , записывается как $\{T > s + t\}$, и условная вероятность этого события (при данном возрасте s) равна отношению $U(s+t)/U(s)$. Это есть распределение оставшегося времени жизни. Оно совпадает с распределением общего времени жизни тогда и только тогда, когда

$$U(s+t) = U(s)U(t), \quad s, t > 0. \quad (3.3)$$

В 1; гл. XVII, 6 было показано, что положительное решение этого уравнения должно иметь вид $U(t) = e^{-\alpha t}$, и, следовательно, *отсутствие последствия, выделенное выше, имеет место тогда и только тогда, когда распределение времени жизни показательное*.

Мы будем называть это свойство отсутствия последствия (lack of memory) *марковским свойством* показательного распределения. Аналитически марковское свойство сводится к утверждению, что только для показательного распределения F хвосты $U = 1 - F$ удовлетворяют равенству (3.3). Этим объясняется постоянное присутствие показательного распределения в марковских процессах. (Усиленный вариант марковского свойства будет описан

в § 6.) Наше описание относится к временным процессам, но наши рассуждения имеют общий характер, и марковское свойство остается осмысленным, когда время заменяется каким-нибудь другим параметром.

Примеры. а) *Прочность на разрыв.* Чтобы получить непрерывный аналог общеизвестной конечной цепи, прочность которой равна прочности ее самого слабого звена, обозначим $U(t)$ вероятность того, что нить длины t (данного материала) может выдержать некоторый фиксированный груз. Нить длины $s+t$ не рвется, если два отрезка по отдельности выдерживают данный груз. В предположении, что взаимодействие отсутствует, эти два события должны рассматриваться как независимые и U должно удовлетворять (3.3). Здесь длина нити выполняет роль временного параметра, а длина, при которой нить порвется, является показателем распределенной случайной величиной.

б) *Случайные ансамбли точек в пространстве* существенны во многих отношениях, поэтому важно иметь подходящее определение этого понятия. Интуиция подсказывает, что первым свойством, характеризующим «чистую случайность», должно быть отсутствие взаимодействия между различными областями: если области A_1 и A_2 не перекрываются, то наблюдения за положением в области A_1 не позволяют сделать вывод о положении в области A_2 . Конкретно, вероятность p того, что области A_1 и A_2 одновременно не содержат точек ансамбля, должна быть равна произведению вероятностей p_1 и p_2 того, что каждая область A_1 и A_2 соответственно пуста. Правдоподобно, что это правило произведения справедливо не для всех разбиений на области, но в том случае, когда вероятность p зависит лишь от объема области A , а не от ее формы. Предполагая, что последнее выполнено, обозначим через $U(t)$ вероятность того, что область, имеющая объем t , является пустой. Тогда вероятности $U(t)$ удовлетворяют равенству (3.3) и поэтому $U(t) = e^{-\alpha t}$; постоянная α зависит от плотности точек ансамбля, или, что то же самое, от единицы измерения. В следующем разделе будет показано, что знание U позволяет вычислять вероятности $p_n(t)$ того, что область объема t содержит ровно n точек ансамбля; эти вероятности определяются распределением Пуассона: $p_n(t) = e^{-\alpha t} (\alpha t)^n / n!$. В этом случае мы говорим о *пуассоновских ансамблях точек*, и этот термин является более определенным, нежели термин случайный ансамбль, который может иметь и другие значения.

в) *Ансамбли кругов и шаров.* Случайные ансамбли частиц ставят перед нами более сложную задачу. Предположим для простоты, что частицы имеют круговую или сферическую форму фиксированного радиуса r . Тогда положение частиц полностью определяется их центрами и заманчиво считать, что эти центры образуют пуассоновский ансамбль. При строгом подходе, однако, это неверно, поскольку взаимные расстояния между центрами необходимо

превосходят 2ρ . Кажется тем не менее правдоподобным, что при малом радиусе ρ пренебрежение размерами рассматриваемых фигур незначительно сказывается на практике и поэтому модель пуассоновского ансамбля центров можно считать приемлемой в качестве разумного приближения.

В соответствующей математической модели мы постулируем, что центры образуют пуассоновский ансамбль, и допускаем в связи с этим возможность того, что круги или шары пересекаются. Эта идеализация не должна иметь практических последствий, если радиусы ρ малы, так как вероятность пересечений незначительно мала. Астрономы обращаются со звездными системами как с пуассоновскими ансамблями, и достигаемое при этом приближение к действительности оказывается превосходным. Следующие два примера демонстрируют использование этой модели на практике.

г) *Ближайшие соседи*. Рассмотрим пуассоновский ансамбль шаров (звезд) с плотностью α . Вероятность того, что область объема t не содержит центров, равна $e^{-\alpha t}$. Высказывание о том, что расстояние до ближайшего к началу координат соседа больше r , сводится к высказыванию, что внутренность шара радиуса r не содержит ни одного центра из ансамбля. Объем такого шара равен $\frac{4}{3}\pi r^3$, и поэтому для пуассоновского ансамбля звезд вероятность того, что ближайший сосед находится на расстоянии, большем r , равна $e^{-\frac{4}{3}\pi\alpha r^3}$. Тот факт, что полученное выражение не зависит от радиусов ρ звезд, указывает на приближенный характер модели и ее ограниченность.

На плоскости шары заменяются кругами, а распределение расстояний между ближайшими соседями задается функцией $1 - e^{-\alpha\pi r^2}$.

д) *Продолжение: задача свободного пробега*. Для упрощения описания мы начнем с двумерной модели. Представим себе случайный ансамбль кругов (круговых дисков) как поперечные сечения деревьев (круговых цилиндров) в достаточно редком лесу. Я стою в начале координат, которое не содержится ни в одном из дисков, и смотрю в положительном направлении оси x . Длина самого длинного интервала $[0, t]$, не пересекающего ни одного диска, характеризует свободный пробег, или *видимость*, в x -направлении. Эту случайную величину обозначим L .

Обозначим через A область, образованную множеством точек, расстояния которых от точек интервала $[0, t]$ на оси x не превосходят ρ . Граница области A состоит из двух сегментов $0 \leq x \leq t$ на прямых $y = \pm \rho$ и двух полуокружностей радиуса ρ с центрами в точках 0 и t на оси x . Таким образом, площадь A равна $2\rho t + \pi\rho^2$. Событие $\{L > t\}$ происходит тогда и только тогда, когда внутри области A нет ни одного центра диска, но уже известно заранее, что круг радиуса ρ с центром в начале координат не содержит центров диска. Оставшаяся область имеет

площадь $2\pi t$, и можно сделать вывод, что *распределение случайной величины L (видимости в x -направлении) является показательным*:

$$P\{L > t\} = e^{-2\alpha\pi t}.$$

Проведенные нами рассуждения остаются верными и в пространстве $\mathcal{H}^3$, а соответствующая этому случаю область образуется вращением A около оси x . Прямоугольник $0 < x < t, |y| < \rho$ заменяется цилиндром объема $\pi\rho^2 t$. Мы получаем, что для пуассоновского ансамбля шаровых звезд *величина L свободного пробега в любом направлении имеет показательное распределение*: $P\{L > t\} = e^{-\pi\alpha\rho^2 t}$. Средняя величина свободного пробега равна $E(L) = 1/(\pi\alpha\rho^2)$. ►

Следующая теорема будет использоваться неоднократно.

Теорема. Если $X_1, \dots, X_n$ — взаимно независимые случайные величины с показательным распределением (3.1), то сумма $X_1 + \dots + X_n$ имеет плотность g_n и функцию распределения G_n , задаваемые формулами

$$g_n(x) = \alpha \frac{(\alpha x)^{n-1}}{(n-1)!} e^{-\alpha x}, \quad x > 0, \quad (3.4)$$

$$G_n(x) = 1 - e^{-\alpha x} \left(1 + \frac{\alpha x}{1!} + \dots + \frac{(\alpha x)^{n-1}}{(n-1)!} \right), \quad x > 0. \quad (3.5)$$

Доказательство. При $n=1$ утверждение сводится к определению (3.1). Плотность g_{n+1} определяется сверткой

$$g_{n+1}(t) = \int_0^t g_n(t-x) g_1(x) dx, \quad (3.6)$$

и в предположении справедливости (3.4) это сводится к

$$g_{n+1}(t) = \frac{\alpha^{n+1}}{(n-1)!} e^{-\alpha t} \int_0^t x^{n-1} dx = \alpha \frac{(\alpha t)^n}{n!} e^{-\alpha t}. \quad (3.7)$$

Таким образом, (3.4) выполняется по индукции для всех n . Справедливость равенства (3.5) доказывается дифференцированием. ►

Плотности g_n входят в семейство *гамма-плотностей*, которое будет введено в гл. II, 2. Они представляют собой непрерывный аналог отрицательного биномиального распределения, определенного в I, гл. VI, 8, как распределение суммы n случайных величин с одинаковым геометрическим распределением. (См. задачу 6.)

§ 4. ПАРАДОКСЫ. СВЯЗАННЫЕ С ВРЕМЕНЕМ ОЖИДАНИЯ. ПУАССОНОВСКИЙ ПРОЦЕСС

Обозначим $X_1, X_2, \dots$ взаимно независимые случайные величины с одинаковым показательным распределением (3.1). Пусть

$$S_0 = 0,$$

$$S_n = X_1 + \dots + X_n, \quad n = 1, 2, \dots \quad (4.1)$$

Мы вводим семейство новых случайных величин $N(t)$ следующим образом: $N(t)$ есть число индексов $k \geq 1$, таких, что $S_k \leq t$. Событие $\{N(t) = n\}$ происходит тогда и только тогда, когда $S_n \leq t$, а $S_{n+1} > t$. Так как S_n имеет распределение G_n , то вероятность этого события равна $G_n(t) - G_{n+1}(t)$ или

$$P\{N(t) = n\} = e^{-\alpha t} \frac{(\alpha t)^n}{n!}. \quad (4.2)$$

Иначе говоря, случайная величина $N(t)$ имеет пуассоновское распределение с математическим ожиданием αt .

Эта аргументация выглядит как новый вывод пуассоновского распределения, но на самом деле это только перефразировка первоначального вывода из I, гл. VI, 6, в терминах случайных величин. Для наглядного описания рассмотрим случайно наступающие события (такие, как вспышки космических лучей или телефонные вызовы), которые мы называем «поступлениями». Допустим, что последствие отсутствует, т. е. прошлое не дает возможности сделать заключение в отношении будущего. Как мы видели, это условие требует, чтобы время ожидания X_1 первого поступления было распределено показательным. Но при каждом поступлении процесс начинается снова как вероятностная копия всего процесса: времена ожидания X_k между последовательными поступлениями должны быть независимыми и должны иметь одно и то же распределение. Сумма S_n изображает момент n -го поступления, а $N(t)$ — число поступлений в интервале $0, t$. В такой форме доводы отличаются от первоначального вывода пуассоновского распределения лишь использованием лучших специальных терминов.

(По терминологии случайных процессов последовательность $\{S_n\}$ есть процесс восстановления с показательными временами X_k между поступлениями; общее понятие процесса восстановления см. в гл. VI, 6.)

Даже эта простая ситуация приводит к кажущимся противоречиям, которые иллюстрируют необходимость более сложного подхода. Мы начнем с несколько наивной формулировки.

Пример. Парадокс времени ожидания. Автобусы прибывают в согласии с пуассоновским процессом, причем среднее время между последовательными автобусами равно α^{-1} . Я прихожу в момент t . Каково математическое ожидание $E(W_t)$ времени W_t , в течение которого я жду следующий автобус? (Понятно, что момент t моего прибытия не зависит от автобусов — скажем, я прихожу ровно в полдень.) Очевидно, имеются два противоречивых ответа.

а) Отсутствие последствия, свойственное пуассоновскому процессу, влечет за собой независимость времени ожидания авто-

буса от момента моего прибытия. В этом случае

$$E(W_t) = E(W_0) = \alpha^{-1}.$$

б) момент моего прибытия «выбран наудачу» в интервале между двумя последовательными автобусами, и по соображениям симметрии мое среднее время ожидания должно быть равно половине среднего промежутка времени между двумя последовательными автобусами, т. е.

$$E(W_t) = \frac{1}{2} \alpha^{-1}.$$

Оба рассуждения кажутся приемлемыми, и оба использовались на практике. Что же делать с противоречиями? Легчайший выход у формалиста, который отказывается рассматривать задачу, если она не сформулирована строго. Но задачи не решаются их игнорированием.

Теперь мы покажем, что *оба* рассуждения по существу, если не формально, *корректны*. Ошибка лежит в непредвиденном пункте, и мы приступаем теперь к ее объяснению¹⁾. ►

Мы рассмотрим последовательность времен между прибытиями

$$X_1 = S_1, X_2 = S_2 - S_1, \dots$$

По предположению X_k имеют одинаковое показательное распределение с математическим ожиданием α^{-1} . Выбирая «любое» отдельное X_k , мы получаем случайную величину, и интуитивно ожидается, что ее математическое ожидание будет равно α^{-1} при условии, что выбор сделан без знания последовательности $X_1, X_2, \dots$. Однако это неверно. В нашем примере мы взяли тот элемент X_k , для которого $S_{k-1} < t \leq S_k$, где t фиксированно. Этот выбор сделан без учета действительного процесса, но оказывается, что так выбранное X_k имеет двойное математическое ожидание $2\alpha^{-1}$. С учетом этого факта противоречие исчезает, так как в пункте б) нашего примера постулируется, что среднее время ожидания равно α^{-1} .

Это разрешение парадокса вызвало шок у опытных работников, однако оно становится интуитивно ясным, когда должным образом подбирается способ рассуждения. Грубо говоря, длинный интервал имеет больше шансов накрыть точку t , нежели короткий. Это приближительное утверждение подтверждается следующим предложением.

Предложение. Пусть $X_1, X_2, \dots$ взаимно независимы и одинаково показательно распределены с математическим ожиданием α^{-1} .

¹⁾ Вариант парадокса см. в примере VI, 7, а) Этот парадокс известен в общей теории восстановления, где он вызвал серьезные затруднения и противоречия, прежде чем был правильно понят (см. по этому поводу XI, 4).

Пусть $t > 0$ фиксированно, но произвольно. Элемент X_k , удовлетворяющий условию $S_{k-1} < t \leq S_k$, имеет плотность

$$v_t(x) = \begin{cases} \alpha^2 x e^{-\alpha x} & \text{для } 0 < x \leq t, \\ \alpha(1 + \alpha t) e^{-\alpha x} & \text{для } x > t. \end{cases} \quad (4.3)$$

Здесь главное в том, что плотность (4.3) не является общей плотностью для X_k . Явный вид этой плотности представляет второстепенный интерес. [Аналог произвольного распределения времени ожидания содержится в XI, (4.16).]

Доказательство. Пусть k — такой (зависящий от случая) индекс, что $S_{k-1} < t \leq S_k$, и пусть L_t равно $S_k - S_{k-1}$. Мы должны доказать, что L_t имеет плотность (4.3). Допустим сначала, что $x < t$. Событие $\{L_t \leq x\}$ происходит, если и только если $S_n = y$ и $t - y < X_{n+1} \leq x$ при некоторой комбинации n, y . Отсюда следует неравенство $t - x \leq y < t$. Суммируя по всем возможным n и y , получаем

$$P\{L_t \leq x\} = \sum_{n=1}^{\infty} \int_{t-x}^t g_n(y) \cdot [e^{-\alpha(t-y)} - e^{-\alpha x}] dy. \quad (4.4)$$

Но $g_1(y) + g_2(y) + \dots = \alpha$ тождественно, и поэтому

$$P\{L_t \leq x\} = 1 - e^{-\alpha x} - \alpha x e^{-\alpha x}. \quad (4.5)$$

Дифференцируя, мы получим (4.3) для $x < t$. Для $x > t$ применимо это же доказательство с той разницей, что y меняется от 0 до x , и мы должны прибавить к правой части равенства (4.4) вероятность $e^{-\alpha t} - e^{-\alpha x}$ того, что $0 < t < S_1 < x$. Этим доказательство завершается. ►

Скачок функции (4.3) при $x = t$ обусловлен специальной ролью начала как исходной точки процесса. Очевидно,

$$\lim_{t \rightarrow \infty} v_t(x) = \alpha^2 x e^{-\alpha x}, \quad (4.6)$$

откуда видно, что со временем роль начала исчезает и для «старого» процесса распределение L_t почти не зависит от t . Удобнее выразить это, сказав, что правая часть (4.6) задает «стационарную» плотность L_t .

В обозначениях, принятых в доказательстве, время ожидания W_t , рассмотренное в примере, есть случайная величина $W_t = S_k - t$. Из доказательства теоремы следует также, что

$$P\{W_t \leq x\} = e^{-\alpha t} - e^{-\alpha(x+t)} + \sum_{n=1}^{\infty} \int_0^t g_n(y) [e^{-\alpha(t-y)} - e^{-\alpha(x+t-y)}] dy = 1 - e^{-\alpha x}. \quad (4.7)$$

Таким образом, W_t имеет то же самое показательное распределение, что и X_k , в согласии с доводами пункта а). (См. задачу 7.)

Наконец, несколько слов о пуассоновском процессе. Пуассоновские величины $N(t)$ были введены как функции, определенные на выборочном пространстве бесконечной последовательности случайных величин $X_1, X_2, \dots$. Эта процедура удовлетворительна для многих целей, но более естественно другое выборочное пространство. Мысленный эксперимент «регистрации числа поступивших вызовов вплоть до момента t » дает при каждом положительном t целое число, и результатом поэтому является ступенчатая функция с единичными скачками. Эти ступенчатые функции служат точками выборочного пространства; выборочное пространство является функциональным пространством — пространством всех возможных «траекторий». В этом пространстве $N(t)$ определяется как значение ординаты в момент t , а S_n — как координата n -го скачка и т. д. Теперь могут быть рассмотрены события, которые невыразимы в терминах первоначальных случайных величин. Типичный пример, имеющий практический интерес (см. задачу о разорении в гл. VI, 5), доставляет событие, заключающееся в том, что $N(t) > a + bt$ при некотором t . Отдельная траектория (точно так же, как отдельная бесконечная последовательность из ± 1 в биномиальных испытаниях) представляет собой естественный и неизбежный объект вероятностного исследования. Стоит только привыкнуть к новой терминологии, как пространство всех траекторий становится наиболее наглядным выборочным пространством.

К сожалению, введение вероятностей в пространстве выборочных траекторий — дело далеко не простое. Для сравнения заметим, что переход от дискретных выборочных пространств к прямой, плоскости и т. д. и даже к бесконечным последовательностям случайных величин ни мысленно, ни технически не труден. В связи с функциональными пространствами возникают проблемы нового типа, и мы предупреждаем читателя, что не будем иметь с ними дела в этом томе. Мы удовлетворимся частным рассмотрением выборочных пространств последовательностей (счетного множества координатных величин). Упоминание о случайных процессах вообще и о пуассоновском процессе в частности мы будем делать свободно, но только для того, чтобы подготовить интуитивную основу или привлечь интерес к нашим проблемам.

Пуассоновские ансамбли точек

Как показано в 1; гл. VI, 6, пуассоновский закон управляет не только «точками, распределенными случайно на оси времени», но также ансамблями точек (такими, как дефекты в изделиях или изюминки в булках), распределенных случайно на плоскости или в пространстве при условии, что t интерпретируется как площадь или объем. Основное предположение состоит в том, что вероятность нахождения k точек в заданной области зависит только от

площади или объема области, но не от ее формы и что явления в неперекрывающихся областях независимы.

При этом же самом предположении в примере 3, б) было показано, что вероятность отсутствия точек в области объема t равна e^{-at} . Это соответствует показательному распределению времени ожидания первого события, и мы теперь видим, что распределение Пуассона для числа событий является простым следствием этого. Те же доводы применимы к случайным ансамблям точек в пространстве, и мы приходим к возможности нового доказательства того, что число точек ансамбля, содержащихся в данной области, есть случайная величина, распределенная по закону Пуассона. Путем несложных формальных выкладок можно прийти к интересным результатам, касающимся таких случайных ансамблей точек, но замечания относительно пуассоновского процесса одинаково применимы и к пуассоновским ансамблям; полное вероятностное описание сложно и лежит вне рамок настоящего тома.

§ 5. УСТОЙЧИВОСТЬ НЕУДАЧ

Общеизвестно, что тот, кто встает в очередь, случается, вынужден ждать неопределенно долгое время. Подобные неудачи преследуют нас во многих положениях. Как может способствовать объяснению этого теория вероятностей? Для ответа на этот вопрос мы рассмотрим три примера, типичные для множества ситуаций. Они иллюстрируют неожиданные общие свойства случайных флуктуаций.

Примеры. а) *Рекордные значения.* Обозначим X_0 мое время ожидания (или размер финансовых потерь) некоторого случайного события. Предположим, что мои друзья подвергли себя опыту того же типа. Обозначим их результаты $X_1, X_2, \dots$. Чтобы отразить «беспристрастность», предположим, что $X_0, X_1, \dots$ — взаимно независимые случайные величины с одним и тем же распределением. Природа последнего в действительности не имеет значения, но, так как показательное распределение служит моделью случайности, мы предположим, что X_i показательно распределены в соответствии с (3.1). Для простоты описания мы предполагаем последовательность $\{X_i\}$ неограниченной.

Чтобы оценить размер моей неудачи, я спрашиваю, как много времени должно пройти, прежде чем один из моих друзей испытает большую неудачу (мы пренебрегаем событием $X_k = X_0$, вероятность которого равна нулю)? Более формально, мы вводим время ожидания N как значение первого индекса n , такого, что $X_n > X_0$. Событие $\{N > n-1\}$ происходит, если и только если максимальный член строки $X_0, X_1, \dots, X_{n-1}$ является начальным; по соображениям симметрии вероятность этого события равна n^{-1} . Событие $\{N = n\}$ — это то же, что $\{N > n-1\} - \{N > n\}$, и, следо-

вательно, при $n = 1, 2, \dots$

$$P\{N=n\} = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}. \quad (5.1)$$

Этот результат полностью подтверждает, что мне действительно ужасно не везет: случайная величина N имеет бесконечное математическое ожидание! Было бы достаточно плохо, если нужно было бы провести в среднем 1000 испытаний, чтобы побить «рекорд» моих неудач, но действительное время ожидания имеет бесконечное математическое ожидание.

Отметим, что наше рассуждение не зависит от условия, что X_k показательны распределены. В действительности всякий раз, когда величины X_j независимы и имеют одну и ту же непрерывную функцию распределения F , первое «рекордное» значение имеет распределение (5.1). Тот факт, что это распределение не зависит от F , используется статистиками для проверки независимости. (См. также задачи 8—11.)

Замечательный и общий характер результата (5.1) в сочетании с простотой доказательства способен внушить подозрение. На самом деле доказательство безупречно (за исключением неформальных рассуждений), но те, кто предпочитает потягаться на надежные вычисления, могут легко проверить правильность (5.1), используя непосредственное определение рассматриваемой вероятности как $(n+1)$ -кратного интеграла от $\alpha^{n+1}e^{-\alpha(x_0 + \dots + x_n)}$ по области, определенной неравенствами $0 < x_0 < x_n$ и $0 < x_j < x_0$ при $j = 1, \dots, n-1$.

Дадим другой вывод формулы (5.1), представляющий собой поучительное упражнение с условными вероятностями; он менее прост, но приводит к дополнительным результатам (задача 8). При условии, что $X_0 = x$, вероятность большого значения и более поздних испытаний равна $p = e^{-\alpha x}$, и мы имеем дело с временем ожидания первого «успеха» в испытаниях Бернулли с вероятностью успеха p . Условная вероятность, что $N=n$ при условии $X_0=x$, равна поэтому $p(1-p)^{n-1}$. Чтобы получить $P\{N=n\}$, мы должны умножить это на плотность $\alpha e^{-\alpha x}$ и проинтегрировать по x . Подстановка $1 - e^{-\alpha x} = t$ сводит подынтегральную функцию к функции $t^{n-1}(1-t)$, интеграл от которой равен $n^{-1} - (n+1)^{-1}$, что соответствует (5.1).

б) *Отношения*. Если X и Y — две независимые случайные величины с одинаковым показательным распределением, то отношение Y/X является новой случайной величиной. Ее функция распределения получается интегрированием выражения $\alpha^2 e^{-\alpha(x+y)}$ по области $0 < y < tx$, $0 < x < \infty$. Интегрирование по y приводит к результату

$$P\left\{\frac{Y}{X} \leq t\right\} = \int_0^{\infty} \alpha e^{-\alpha x} (1 - e^{-\alpha tx}) dx = \frac{t}{1+t}. \quad (5.2)$$

Соответствующая плотность равна $(1+t)^{-2}$. Заслуживает внима-

ния то, что величина Y/X имеет бесконечное математическое ожидание.

Мы находим здесь новое подтверждение устойчивости неудач. Конечно, Петр имеет основание для недовольства, если ему пришлось ждать в три раза больше, чем ждал Павел. Однако распределение (5.2) приписывает этому событию вероятность $1/4$. Отсюда следует, что в среднем в одном из двух случаев или Павел, или Петр имеют основания для недовольства. Наблюдаемая частота практически возрастает, так как очень короткие времена ожидания, естественно, проходят незамеченными.

в) *Параллельные очереди.* Я приезжаю в своем автомобиле на автомобильную инспекционную станцию (или ко въезду в туннель, или на автомобильный паром и т. д.). Имеются на выбор две очереди, но раз уж я встал в одну очередь, я должен стоять в ней. Мистер Смит, который приехал вслед за мной, занимает место в другой очереди, которое я мог бы выбрать, и я теперь слежу за тем, продвигается ли он быстрее или медленнее меня. Большее время мы стоим неподвижно, но время от времени то одна, то другая очередь передвигается на один автомобиль вперед. Чтобы увеличить влияние чистого случая, мы предположим две очереди стохастически независимыми; интервалы времени между последовательными продвижениями также являются независимыми величинами с одинаковым показательным распределением. При данных обстоятельствах последовательные продвижения представляют собой испытания Бернулли, где «успех» — это то, что продвинулся вперед я, «неудача» — что продвинулся мистер Смит. Так как вероятности успеха равны $1/2$, мы в сущности имеем дело с симметричным случайным блужданием, и любопытные свойства флуктуаций при случайных блужданиях находят неожиданную интерпретацию. (Для простоты описания мы отвлечемся от того, что имеется только ограниченное число автомобилей.) Смогу ли я когда-нибудь обогнать мистера Смита? В переводе на язык случайных блужданий этот вопрос звучит так: будет ли когда-нибудь иметь место первое прохождение через $+1$? Как известно, это событие имеет вероятность, равную единице. Однако среднее время ожидания для него бесконечно. Такое ожидание дает достаточные возможности для оплакивания моих неудач, причем раздражение мое только возрастает от того, что мистер Смит приводит те же самые доводы. ►

§ 6. ВРЕМЕНА ОЖИДАНИЯ И ПОРЯДКОВЫЕ СТАТИСТИКИ

Пусть дана упорядоченная строка $(x_1, \dots, x_n)$ из n действительных чисел. Переставляя их в порядке возрастания величины, получим новую строку

$$(x_{(1)}, x_{(2)}, \dots, x_{(n)}), \text{ где } x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Эта операция, примененная ко всем точкам пространства $\mathfrak{R}^n$, дает нам n вполне определенных функций, которые будут обозначаться $X_{(1)}, \dots, X_{(n)}$. Если в $\mathfrak{R}^n$ определены вероятности, то эти функции становятся случайными величинами. Мы говорим, что $(X_{(1)}, \dots, X_{(n)})$ получается перестановкой из $X_1, \dots, X_n$ в соответствии с возрастанием величины. Величина $X_{(k)}$ называется k -й *порядковой статистикой*¹⁾ данной выборки $X_1, \dots, X_n$, в частности, $X_{(1)}$ и $X_{(n)}$ — это *выборочные экстремумы*; если $n = 2v + 1$ нечетно, то $X_{(v+1)}$ есть *выборочная медиана*.

Мы применим эти понятия к частному случаю независимых случайных величин $X_1, \dots, X_n$ с одинаковым показательным распределением с плотностью $\alpha e^{-\alpha x}$.

Пример. а) *Параллельные очереди.* Будем интерпретировать $X_1, \dots, X_n$ как длительности n времен обслуживания, начинающегося с момента 0 в почтовом отделении с n окошками. Порядковые статистики изображают последовательные моменты окончания обслуживаний, или, как говорится, *моменты последовательных разгрузок* («выходной процесс»). В частности, $X_{(1)}$ представляет собой время ожидания первой разгрузки. Теперь если предположение об отсутствии последствия осмысленно, то время ожидания $X_{(1)}$ должно обладать марковским свойством, т. е. $X_{(1)}$ должно быть распределено показательно. Фактически событие $\{X_{(1)} > t\}$ есть одновременная реализация n событий $\{X_k > t\}$, каждое из которых имеет вероятность $e^{-\alpha t}$; ввиду предположения о независимости вероятности перемножаются, и мы имеем

$$P\{X_{(1)} > t\} = e^{-n\alpha t}. \quad (6.1)$$

Сделаем теперь следующий шаг и рассмотрим положение в момент $X_{(1)}$. Предположение об отсутствии памяти, по-видимому, подразумевает, что восстанавливается первоначальное положение с той разницей, что теперь в операциях участвует $n-1$ окошек; продолжение процесса не должно зависеть от $X_{(1)}$ и должно быть копией всего процесса. В частности, время ожидания следующей разгрузки, а именно $X_{(2)} - X_{(1)}$, должно иметь распределение

$$P\{X_{(2)} - X_{(1)} > t\} = e^{-(n-1)\alpha t}, \quad (6.2)$$

аналогичное (6.1). Это рассуждение приводит к следующей общей теореме, касающейся порядковых статистик для независимых величин с одинаковым показательным распределением:

¹⁾ Строго говоря, термин «выборочная статистика» является синонимом термина «функция от выборочных величин», т. е. термина «случайная величина». Он используется для лингвистического выделения разницы в ролях, играемых в *данном контексте* первоначальными величинами (выборкой) и некоторыми полученными из них величинами. Например, «выборочное среднее» $(X_1 + \dots + X_n)/n$ называется статистикой. Порядковые статистики часто встречаются в статистической литературе. Мы придерживаемся стандартной терминологии, за исключением того, что экстремумы обычно называются экстремальными значениями.

Предложение 1). Случайные величины $X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}$ являются независимыми, и плотность $X_{(k+1)} - X_{(k)}$ равна $(n-k)ae^{-(n-k)at}$.

Прежде чем проверить это утверждение формально, рассмотрим его следствия. Если $n=2$, то разность $X_{(2)} - X_{(1)}$ означает «остаточное время ожидания» после истечения более короткого из двух времен. Теорема утверждает, что это «остаточное время ожидания» имеет то же самое показательное распределение, что и первоначальное время ожидания, и не зависит от $X_{(1)}$. Это есть расширение марковского свойства, сформулированного для фиксированных моментов t , на зависящее от случая время остановки $X_{(1)}$. Это свойство называется *строгим марковским свойством*. (Так как мы имеем дело лишь с конечным множеством величин, мы в состоянии вывести строго марковское свойство из обычного, но в более сложных случайных процессах различие существенно.)

Доказательство теоремы служит примером формальных действий с интегралами. Для упрощения типографского набора формул рассмотрим $n=3$. Так же как и в других подобных ситуациях, мы используем соображения симметрии. С вероятностью 1 никакие две из величин X_i не равны. Пренебрегая событием вероятности нуль, делаем вывод, что шесть возможных упорядоченных расположений X_1, X_2, X_3 представляют собой шесть взаимно исключающих равновероятных событий. Поэтому для того, чтобы найти распределение порядковых статистик, достаточно рассмотреть случай $X_1 < X_2 < X_3$. Таким образом,

$$P\{X_{(1)} > t_1, X_{(2)} - X_{(1)} > t_2, X_{(3)} - X_{(2)} > t_3\} = 6P\{X_1 > t_1, X_2 - X_1 > t_2, X_3 - X_2 > t_3\}. \quad (6.3)$$

(Аналитически, пространство $\mathbb{R}^3$ разбивается на шесть частей, конгруэнтных области, определенной неравенствами $x_1 < x_2 < x_3$. Каждая из них вносит одну и ту же величину в соответствующий интеграл. Границы, где две или более координат равны, имеют вероятность нуль и не играют роли.) Чтобы получить правую часть в (6.3), мы должны проинтегрировать $a^3 e^{-a(x_1 + x_2 + x_3)}$ по области, определенной неравенствами

$$x_1 > t_1, \quad x_2 - x_1 > t_2, \quad x_3 - x_2 > t_3.$$

Интегрирование по x_3 приводит к

$$\begin{aligned} 6e^{-at_1} \int_{t_1}^{\infty} ae^{-ax_1} dx_1 \int_{x_1+t_2}^{\infty} ae^{-2ax_2} dx_2 = \\ = 3e^{-at_1 - 2at_2} \int_{t_1}^{\infty} ae^{-3ax_1} dx_1 = e^{-at_1 - 2at_2 - 3at_3}. \end{aligned} \quad (6.4)$$

1) Эту теорему неоднократно открывали в связи с задачами статистического оценивания, но обычные ее доказательства сводились к вычислениям вместо обращения к марковскому свойству. См. также задачу 13.

Таким образом, совместное распределение трех величин $X_{(1)}$, $X_{(2)}$ — $X_{(1)}$, $X_{(3)}$ — $X_{(2)}$ есть произведение трех показательных распределений, и это доказывает наше утверждение.

Отсюда, в частности, следует, что $E(X_{(k+1)} - X_{(k)}) = 1/(n-k)\alpha$. Суммируя по $k=0, 1, \dots, v-1$, мы получаем

$$E(X_{(v)}) = \frac{1}{\alpha} \left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-v+1} \right). \quad (6.5)$$

Отметим, что это математическое ожидание было вычислено без знания распределения $X_{(v)}$. Итак, мы имеем еще один пример преимущества представления случайной величины как суммы других величин. (См. I, гл. IX, 3.)

б) *Использование строго марковского свойства.* В качестве образного примера рассмотрим следующую ситуацию: в момент 0 три лица A , B и C прибывают в почтовое отделение и находят два окна свободными. Три времени обслуживания представляют собой независимые случайные величины X , Y , Z с одинаковым показательным распределением. Обслуживание A и B начинается сразу же, а обслуживание C начинается в момент $X_{(1)}$, когда или A , или B уходит. Мы покажем, что марковское свойство приводит к простым ответам на различные вопросы.

(i) Какова вероятность, что C не последним уйдет из почтового отделения? Ответ: $1/2$, так как момент $X_{(1)}$ первого ухода устанавливает симметрию между C и другим обслуживаемым лицом.

(ii) Каково распределение времени T , проведенного лицом C в почтовом отделении? Ясно, что величина $T = X_{(1)} + Z$ есть сумма двух независимых величин, которые распределены показательно с параметрами 2α и α . Свертка любых двух показательных распределений найдена в (2.14), и мы получаем, что T имеет плотность $u(t) = 2\alpha(e^{-\alpha t} - e^{-2\alpha t})$ и что $E(T) = \frac{3}{2\alpha}$.

(iii) Каково распределение момента *последнего* ухода? Обозначим моменты последовательных уходов $X_{(1)}$, $X_{(2)}$, $X_{(3)}$. Разность $X_{(3)} - X_{(1)}$ представляет собой сумму двух величин $X_{(3)} - X_{(2)}$ и $X_{(2)} - X_{(1)}$. Мы видели в предыдущем примере, что эти величины независимы и имеют показательные распределения с параметрами 2α и α . Отсюда следует, что $X_{(3)} - X_{(1)}$ имеет ту же плотность u , что и величина T . Теперь $X_{(1)}$ не зависит от $X_{(3)} - X_{(1)}$ и имеет плотность $2\alpha e^{-2\alpha t}$. Формула свертки, использованная в (ii), показывает, что $X_{(3)}$ имеет плотность $4\alpha[e^{-\alpha t} - e^{-2\alpha t} - \alpha t e^{-2\alpha t}]$ и что $E(X_{(3)}) = 2/\alpha$.

Преимущества этого метода становятся очевидными при сравнении с непосредственным вычислением, но последнее применимо и для других распределений времени обслуживания (задача 19).

в) *Распределение порядковых статистик.* В качестве заключительного упражнения мы выведем распределение $X_{(k)}$. Событие $\{X_{(k)} \leq t\}$ означает, что по крайней мере k из n величин X_i не

превосходят t . Это равносильно наступлению по крайней мере k «успехов» в n независимых испытаниях, и поэтому

$$P\{X_{(k)} \leq t\} = \sum_{j=k}^n \binom{n}{j} (1 - e^{-\alpha t})^j e^{-(n-j)\alpha t}. \quad (6.6)$$

После дифференцирования видно, что плотность $X_{(k)}$ равна

$$n \binom{n-1}{k-1} (1 - e^{-\alpha t})^{k-1} e^{-(n-k)\alpha t} \cdot \alpha e^{-\alpha t}. \quad (6.7)$$

Этот результат может быть получен непосредственно следующими нестрогими рассуждениями. Мы разыскиваем (вплоть до членов пренебрежимо малых в пределе при $h \rightarrow 0$) вероятность события, заключающегося в том, что одна из величин X_j лежит между t и $t+h$, а $k-1$ величин из оставшихся $n-1$ величин не превосходят t , в то время как другие $n-k$ величин строго больше $t+h$. Перемножая число наборов и соответствующие вероятности, приходим к (6.7). Начиная вывод рекомендуется формализовать это рассуждение, а также вывести формулу (6.7) из дискретной модели. (Продолжение в задачах 13, 17). ►

§ 7. РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ

Случайная величина X *распределена равномерно в интервале a, b* , если ее плотность постоянна — равна $(b-a)^{-1}$ при $a < x < b$ и равна 0 вне этого интервала. В этом случае величина $(X-a)(b-a)^{-1}$ распределена равномерно в $\overline{0,1}$, и мы будем обычно использовать этот интервал как стандартный. Из-за внешнего вида графиков плотности равномерные распределения называются «прямоугольными».

При равномерном распределении интервал $\overline{0,1}$ становится выборочным пространством, в котором вероятности интервалов тождественны их длинам. Выборочное пространство, соответствующее двум независимым величинам X и Y , которые равномерно распределены на $\overline{0,1}$, представляет собой единичный квадрат в $\mathbb{R}^2$, а вероятности в нем определяются как площади. Те же идеи применимы к трем и большему числу величин.

Про равномерно распределенную случайную величину часто говорят: «точка X выбрана наудачу». Результат мысленного эксперимента « n независимых случайных выборов точки в $\overline{0,1}$ » требует для своего вероятностного описания n -мерного гиперкуба, но эксперимент сам по себе дает n точек $X_1, \dots, X_n$ в том же самом интервале. С вероятностью единица никакие две из этих точек не совпадают, и, следовательно, они разбивают интервал $\overline{0,1}$ на $n+1$ подынтервалов. Переставляя n точек $X_1, \dots, X_n$ в их естественном порядке слева направо, мы получаем n новых

случайных величин, которые обозначим $X_{(1)}, \dots, X_{(n)}$. Это и есть *порядковые статистики*, определенные в предыдущем параграфе. Подынтервалы разбиения теперь имеют вид $0, \overline{X_{(1)}}, \overline{X_{(2)}}, \overline{X_{(3)}} \dots$ и т. д.

Понятие *точки, выбранной наудачу на окружности*, говорит само за себя. Чтобы отчетливо представить себе исход n независимых выборов точки на окружности, мы вообразим окружность, *ориентированную* против часовой стрелки, так что интервалы имеют левый и правый концы и могут быть изображены в форме a, b . Две точки X_1 и X_2 , выбранные независимо и наудачу, делят окружность на два интервала $\overline{X_1}, \overline{X_2}$ и $\overline{X_2}, \overline{X_1}$. (Мы снова пренебрегаем событием $X_1 = X_2$, имеющим вероятность нуль.)

Примеры. а) *Опытные интерпретации.* Колесо рулетки обычно приводится как средство осуществления «случайного выбора» на окружности. *Ошибка округления* при вычислениях с шестью десятичными знаками обычно рассматривается как случайная величина, распределенная равномерно в интервале длиной 10^{-6} . (Для ошибок, совершенных при отбрасывании двух последних десятичных знаков, дискретная модель со 100 возможными значениями более уместна, хотя менее практична.) Время ожидания пассажира, прибывшего на автобусную станцию без учета расписания, может рассматриваться как случайная величина, равномерно распределенная в интервале между последовательными отъездами автобусов. Большой теоретический интерес представляют применения к *случайным разбиениям*, обсуждаемым в § 8. Во многих задачах математической статистики (таких, как непараметрические критерии) равномерное распределение входит косвенным образом: при условии, что произвольная случайная величина X имеет непрерывную функцию распределения F , случайная величина $\overline{F(X)}$ *распределена равномерно на $0, 1$* (см. § 12).

б) *Индукцированное разбиение.* Мы докажем следующее утверждение: n *независимо и случайно выбранных на $0, 1$ точек $X_1, \dots, X_n$ разбивают $0, 1$ на $n+1$ интервалов, длины которых имеют одно и то же распределение*

$$P\{L > t\} = (1-t)^n, \quad 0 < t < 1. \quad (7.1)$$

Это неожиданно, так как интуитивно можно было бы ожидать, что по крайней мере два концевых интервала должны иметь другое распределение. То, что все интервалы будут иметь одинаковое распределение, становится ясным при рассмотрении эквивалентной ситуации на (ориентированной) окружности единичной длины¹⁾. Здесь $n+1$ точек $X_1, \dots, X_{n+1}$, выбранных независимо

¹⁾ При численной проверке следует иметь в виду, что вероятность события $\{X_{(k+1)} - X_{(k)} > t\}$ равна интегралу от постоянной 1 по объединению nt конгруэнтных областей, определенных или цепью неравенств $x_1 < \dots < x_k < x_{k+1} - t < x_{k+1} < \dots < x_n$, или аналогичными цепями, полученными пере-

и наудачу, разбивают окружность на $n+1$ интервалов, и по соображениям симметрии эти интервалы должны иметь одинаковое распределение. Представим теперь, что окружность разрезана в точке X_{n+1} для того, чтобы получить интервал, в котором $X_1, \dots, X_n$ выбраны независимо и наудачу. Длины всех $n+1$ интервалов индуцированного разбиения имеют одно и то же распределение. Рассматривая самый левый интервал $0, \overline{X_{(1)}}$, можно видеть, что это распределение задается формулой (7.1). Длина этого интервала превосходит t тогда и только тогда, когда все n точек $X_1, \dots, X_n$ лежат в $\overline{t, 1}$. Вероятность этого равна $(1-t)^n$.

Хорошим упражнением является проверка утверждения в частном случае $n=2$ путем рассмотрения трех событий в единичном квадрате, изображающем выборочное пространство. (Продолжение см. в задачах 22—26.)

в) *Парадокс* (связанный с парадоксом времени ожидания из § 4). Пусть две точки X_1 и X_2 выбраны независимо и наудачу на окружности единичной длины. Тогда *длины двух интервалов $\overline{X_1, X_2}$ и $\overline{X_2, X_1}$ распределены равномерно, а длина λ интервала, содержащего произвольную точку P , имеет другое распределение (с плотностью $2x$)*.

В частности, средняя длина каждого из двух интервалов равна $1/2$, а средняя длина интервала, содержащего P , равна $1/3$. Точка P берется фиксированной, но произвольной. Поэтому можно ожидать, что интервал, покрывающий P , выбран (заимствуя фразу у философов теории вероятности) «без предварительного знания его свойств». Наивная интуиция не подготовлена, конечно, воспринять большое различие между покрытием и непокрытием произвольной точки, но после должного размышления это различие становится «интуитивно очевидным». На самом деле, однако, даже довольно опытные авторы попадали в ловушку.

Для доказательства представим себе, что окружность разъединена в точке P . Нам остаются две точки, выбранные независимо и наудачу в $0, \overline{1}$. Используя те же обозначения, что и раньше, мы находим, что событие $\{\lambda < t\}$ происходит тогда и только тогда, когда $X_{(2)} - X_{(1)} > 1-t$. По формуле (7.1) вероятность этого равна t^2 . Величина λ имеет поэтому плотность $2t$, как утверждалось. (Начинающим рекомендуется провести *непосредственную* вычислительную проверку.)

г) *Распределение порядковых статистик*. Если $X_1, \dots, X_n$ независимы и распределены равномерно в интервале $0, \overline{1}$, то число величин, удовлетворяющих неравенству $0 < X_j \leq t < 1$, имеет биномиальное распределение с вероятностью «успеха», равной t . Теперь событие $\{X_{(k)} \leq t\}$ происходит, если и только если по крайней

становкой индексов. Другой способ вычисления, приводящий к более сильному результату, содержится в примере III, 3, в),

мере k из величин X_j не превосходят t , и, следовательно,

$$P\{X_{(k)} \leq t\} = \sum_{j=k}^n \binom{n}{j} t^j (1-t)^{n-j}. \quad (7.2)$$

Это дает нам функцию распределения k -й порядковой статистики. После дифференцирования находим, что плотность $X_{(k)}$ равна

$$n \binom{n-1}{k-1} t^{k-1} (1-t)^{n-k}. \quad (7.3)$$

Это можно увидеть непосредственно из следующего: вероятность того, что одна из величин X_j лежит между t и $t+h$ и что $k-1$ из оставшихся величин меньше t , в то время как $n-k$ величин больше $t+h$, равна

$$n \binom{n-1}{k-1} t^{k-1} (1-t-h)^{n-k} h.$$

Деля последнее выражение на h и устремляя h к 0, получим (7.3).

д) *Предельные теоремы.* Чтобы понять характер распределения $X_{(1)}$, когда n велико, лучше ввести $E(X_{(1)}) = (n+1)^{-1}$ как новую единицу измерения. Тогда при $n \rightarrow \infty$ мы получим для хвоста функции распределения

$$P\{nX_{(1)} > t\} = \left(1 - \frac{t}{n}\right)^n \rightarrow e^{-t}. \quad (7.4)$$

Словами это соотношение обычно описывают так: *в пределе $X_{(1)}$ показательно распределена с математическим ожиданием n^{-1} .* Аналогично

$$P\{nX_{(2)} > t\} = \left(1 - \frac{t}{n}\right)^n + \binom{n}{1} \frac{t}{n} \left(1 - \frac{t}{n}\right)^{n-1} \rightarrow e^{-t} + te^{-t}, \quad (7.5)$$

и в правой части мы узнаем хвост гамма-распределения G_k из формулы (3.5). Подобным способом легко проверить, что при каждом фиксированном k при $n \rightarrow \infty$ распределение $nX_{(k)}$ стремится к гамма-распределению G_k (см. задачу 33).

Теперь G_k есть распределение суммы k независимых показательно распределенных величин, тогда как $X_{(k)}$ есть сумма первых k интервалов, рассмотренных в примере б). Мы можем поэтому сказать, что длины последовательных интервалов нашего разбиения ведут себя в пределе так, как если бы они были взаимно независимыми показательно распределенными величинами.

Ввиду очевидной связи (7.2) с биномиальным распределением для получения приближений к распределению $X_{(k)}$, когда n и k велики, может быть использована центральная предельная теорема. (См. задачу 34.)

е) *Отношения.* Пусть X выбрано наудачу в $\overline{0, 1}$. Обозначим через U длину более короткого из интервалов $\overline{0, X}$ и $\overline{X, 1}$ и через

$V = 1 - U$ длину более длинного. Случайная величина U равномерно распределена между 0 и $1/2$, так как событие $\{U < t < 1/2\}$ происходит тогда и только тогда, когда или $X < t$, или $1 - X < t$, и потому имеет вероятность $2t$. По соображениям симметрии величина V распределена равномерно между $1/2$ и 1, и отсюда $E(U) = 1/4$, $E(V) = 3/4$. Что мы можем сказать об отношении V/U ? Оно необходимо превосходит 1 и лежит между 1 и t тогда и только тогда, когда или $\frac{1}{1+t} \leq X \leq \frac{1}{2}$, или $\frac{1}{2} \leq X \leq \frac{t}{1+t}$. При $t > 1$ отсюда следует, что

$$P\left\{\frac{V}{U} \leq t\right\} = \frac{t-1}{t+1}, \quad (7.6)$$

и плотность этого распределения равна $2(t+1)^{-2}$. Мы видим, что отношение V/U имеет бесконечное математическое ожидание. Этот пример показывает, сколь малая информация содержится в наблюдении $E(V)/E(U) = 3$. ►

§ 8. СЛУЧАЙНЫЕ РАЗБИЕНИЯ

Задача этого параграфа завершает предыдущий ряд примеров и выделена из них отчасти ввиду ее важности в физике, отчасти ввиду того, что она дает прототип обычных марковских цепей.

Формально мы будем иметь дело с произведениями вида $Z_n = X_1 X_2 \dots X_n$, где $X_1, \dots, X_n$ — взаимно независимые величины, равномерно распределенные в интервале $0, 1$.

Примеры для применений. В некоторых процессах со столкновениями физическая частица расщепляется на две и ее масса m делится между ними. Различным процессам могут соответствовать различные законы разбиения, но чаще всего предполагается, что полученная каждой дочерней частицей часть исходной массы распределена равномерно в $0, 1$. Если одна из двух частиц выбрана наудачу и подвергается новому столкновению, то (в предположении отсутствия взаимодействия, т. е. независимости столкновений) массы двух частиц второго поколения задаются произведениями $mX_1 X_2$ и т. д. (см. задачу 21). С небольшими терминологическими изменениями эта модель применима также к разбиениям зерен минералов или галек и т. п. Вместо масс можно рассматривать *потери энергии* при столкновениях, и описание несколько упрощается, если заниматься изменениями энергии *одной и той же* частицы при последовательных столкновениях. В качестве последнего примера рассмотрим изменения в *интенсивности света*, проходящего через вещество. Пример 10, а) показывает, что когда световой луч проходит через сферу радиуса R «в случайном направлении», расстояние, пробегаемое лучом в сфере, распределено равномерно между 0 и $2R$. При наличии равномерного поглощения такое прохождение будет уменьшать

интенсивность случайного луча на величину, которая распределена равномерно в интервале $\overline{0, a}$ (где $a < 1$ зависит от степени поглощения). Масштабный параметр не влияет серьезно на нашу модель, и видно, что n независимых прохождений будут уменьшать интенсивность света на величину вида Z_n . ►

Для того чтобы найти распределение Z_n , мы можем действовать двумя способами.

(i) *Редукция к показательным распределениям.* Так как суммы обычно предпочтительнее произведений, мы перейдем к логарифмам, полагая $Y_k = -\log X_k$. Величины Y_k взаимно независимы, и при $t > 0$

$$P\{Y_k \geq t\} = P\{X_k \leq e^{-t}\} = e^{-t}. \quad (8.1)$$

Функция распределения G_n суммы $S_n = Y_1 + \dots + Y_n$ n независимых показательно распределенных величин была получена в формуле (3.5), и тогда функция распределения величины $Z_n = e^{-S_n}$ задается выражением $1 - G_n(\log t^{-1})$, где $0 < t < 1$. Плотность этого распределения равна $t^{-1} g_n(\log t^{-1})$ или

$$f_n(t) = \frac{1}{(n-1)!} \left(\log \frac{1}{t} \right)^{n-1}, \quad 0 < t < 1. \quad (8.2)$$

Наша задача решается в явном виде. Этот метод показывает преимущества вывода с помощью соответствующего преобразования, однако успех зависит от случайного совпадения нашей задачи и задачи, предварительно решенной.

(ii) *Рекуррентная процедура* имеет то преимущество, что она применима к родственным задачам. Пусть $F_n(t) = P\{Z_n \leq t\}$ и $0 < t < 1$. По определению $F_1(t) = t$. Допустим, что F_{n-1} известно, и рассмотрим $Z_n = Z_{n-1} X_n$ как произведение двух независимых величин. При условии $X_n = x$ событие $\{Z_n \leq t\}$ происходит тогда и только тогда, когда $Z_{n-1} \leq t/x$, и имеет вероятность $F_{n-1}(t/x)$. Суммируя по всем возможным x , мы получаем при $0 \leq t \leq 1$

$$F_n(t) = \int_0^1 F_{n-1}\left(\frac{t}{x}\right) dx = \int_t^1 F_{n-1}\left(\frac{t}{x}\right) dx + t. \quad (8.3)$$

Эта формула в принципе позволяет нам вычислить последовательно $F_1, F_2, \dots$. На практике удобнее иметь дело с соответствующими плотностями f_n . По предположению f_1 существует. Предположим по индукции существование f_{n-1} . Вспоминая, что $f_{n-1}(s) = 0$ при $s > 1$, мы получаем после дифференцирования (8.3)

$$f_n(t) = \int_t^1 f_{n-1}\left(\frac{t}{x}\right) \frac{dx}{x}, \quad 0 < t < 1, \quad (8.4)$$

и элементарные вычисления показывают, что f_n действительно задается формулой (8.2).

§ 9. СВЕРТКИ И ТЕОРЕМЫ О ПОКРЫТИИ

Результаты этого параграфа в какой-то мере заняты и сами по себе, и в связи с некоторыми очевидными приложениями. Кроме того, они несколько неожиданно возникают в совсем, казалось бы, не связанных разделах, таких, как, например, критерии значимости в гармоническом анализе [пример 3, е) гл. III], пуассоновские процессы [гл. XIV, 2, а)] и случайные сдвиги [пример 10, д)]. Поэтому не удивительно, что все формулы, так же как и их варианты, выводились несколько раз различными способами. Способ, используемый в последующем, выделяется простотой и применимостью к близким задачам.

Пусть $a > 0$ фиксировано, и пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины, равномерно распределенные на $\overline{0, a}$. Пусть $S_n = X_1 + \dots + X_n$. Наша первая задача заключается в нахождении распределения U_n суммы S_n и ее плотности $u_n = U'_n$.

По определению $u_1(x) = a^{-1}$ при $0 < x < a$ и $u_1(x) = 0$ в других точках (прямоугольная плотность). Для $n > 1$ плотности u_n определяются формулой свертки (2.13), которая в данном случае имеет вид

$$u_{n+1}(x) = \frac{1}{a} \int_0^a u_n(x-y) dy = \frac{1}{a} [U_n(x) - U_n(x-a)]. \quad (9.1)$$

Легко видеть, что

$$u_2(x) = \begin{cases} xa^{-2}, & 0 \leq x \leq a, \\ (2a-x)a^{-2}, & a \leq x \leq 2a, \end{cases} \quad (9.2)$$

и, конечно, $u_2(x) = 0$ для всех других x . График u_2 представляет собой равнобедренный треугольник с основанием $\overline{0, 2a}$, и поэтому u_2 называется *треугольной плотностью*. Аналогично, u_3 сосредоточена на $\overline{0, 3a}$ и определяется тремя квадратными полиномами соответственно в интервалах $\overline{0, a}$, $\overline{a, 2a}$ и $\overline{2a, 3a}$. Чтобы найти общую формулу, мы введем следующее

Обозначение. Мы записываем положительную часть действительного числа x в виде

$$x_+ = \frac{x + |x|}{2}. \quad (9.3)$$

В последующем мы используем символ x_+^n для $(x_+)^n$, т. е. для функции, которая равна 0 при $x \leq 0$ и равна x^n при $x \geq 0$. Заметим, что $(x-a)_+$ есть нуль при $x < a$ и линейная функция, когда $x > a$. С помощью этого обозначения равномерное распределение можно записать в форме

$$U_1(x) = (x_+ - (x-a)_+) a^{-1}. \quad (9.4)$$

Теорема 1. Пусть S_n — сумма n независимых случайных величин,

распределенных равномерно на $\overline{0, a}$. Пусть $U_n(x) = P\{S_n \leq x\}$, и пусть $u_n = U'_n$ — плотность этого распределения. Тогда при $n = 1, 2, \dots$ и для всех $x \geq 0$

$$U_n(x) = \frac{1}{a^n n!} \sum_{v=0}^n (-1)^v \binom{n}{v} (x - va)_+^n; \quad (9.5)$$

$$u_{n+1}(x) = \frac{1}{a^{n+1} n!} \sum_{v=0}^{n+1} (-1)^v \binom{n+1}{v} (x - va)_+^n. \quad (9.6)$$

(Эти формулы остаются верными для $x < 0$ и $n = 0$ при условии, что x_+^0 равно 0, если $x < 0$, и 1, если $x > 0$.)

Заметим, что если точка x расположена между $(k-1)a$ и ka , то только k членов в формулах отличны от 0. Для удобства практических вычислений можно, игнорируя пределы суммирования в (9.5), (9.6), считать, что v меняется от $-\infty$ до ∞ . Это возможно в силу того, что по соглашению биномиальные коэффициенты в (9.5) равны 0 при $v < 0$ и $v > n$ (см. 1; II, 8).

Доказательство. При $n = 1$ справедливость (9.5) очевидна, так как (9.5) сводится к (9.4). Докажем оба утверждения теоремы вместе методом индукции. Допустим, что формула (9.5) верна при некотором $n \geq 1$. Подставляя (9.5) в (9.1), получаем u_{n+1} в виде разности двух известных сумм. Заменяя индекс суммирования v во второй сумме на $v-1$, получаем

$$u_{n+1}(x) = \frac{1}{a^{n+1} n} \sum (-1)^v \left[\binom{n}{v} + \binom{n}{v-1} \right] (x - va)_+^n,$$

что совпадает с правой частью (9.6). Интегрирование доказывает справедливость (9.5) для $n+1$, и этим завершается доказательство. ►

(Другое доказательство, использующее переход к пределу в дискретной модели, содержится в 1; гл. XI, 7, задача 20.)

Если $a = 2b$, то случайные величины $X_k - b$ равномерно распределены на симметричном (относительно нуля) интервале $[-b, b]$ и поэтому сумма n таких величин имеет то же распределение, что и величина $S_n - nb$. Последнее распределение задается функцией $U_n(x + nb)$. Поэтому наша теорема может быть сформулирована в следующей эквивалентной форме.

Теорема 1а. Плотность суммы n независимых случайных величин, равномерно распределенных на $[-b, b]$, задается равенством

$$u_n(x + nb) = \frac{1}{(2b)^n (n-1)!} \sum_{v=0}^n (-1)^v \binom{n}{v} (x + (n-2v)b)_+^{n-1}. \quad (9.7)$$

Мы переходим теперь к теореме, которая допускает две эквивалентные формулировки, одинаково полезные для многих частных задач, возникающих в приложениях. Отыскиваемая в этих теоремах вероятность по счастливому совпадению очень просто выражается

в терминах плотности u_n . Мы докажем это аналитическим способом, который имеет широкую применимость. Доказательство, основанное на геометрических соображениях, см. в задаче 23.

Теорема 2. На окружности длины t заданы $n \geq 2$ дуг длины a , центры которых выбраны независимо и наудачу. Вероятность $\varphi_n(t)$ того, что эти n дуг покрывают всю окружность, равна

$$\varphi_n(t) = a^n (n-1)! u_n(t) \frac{1}{t^{n-1}}, \quad (9.8)$$

или, что то же самое,

$$\varphi_n(t) = \sum_{v=0}^n (-1)^v \binom{n}{v} \left(1 - v \frac{a}{t}\right)_+^{n-1}. \quad (9.9)$$

Прежде чем доказывать это, придадим теореме форму, в которой она будет использована позже. Выберем один из n центров в качестве начала и развернем окружность в интервал длины t . Оставшиеся $n-1$ центров случайно распределяются в интервале $\overline{0, t}$, и тогда теорема 2 выражает то же самое, что и

Теорема 3. Пусть интервал $\overline{0, t}$ разбит на n подынтервалов посредством независимого и случайного выбора $n-1$ точек деления $X_1, \dots, X_{n-1}$. Вероятность $\varphi_n(t)$ того, что ни один из этих интервалов не превосходит a , равна (9.9).

Отметим, что $\varphi_n(t)$, рассматриваемая при фиксированном t как функция a , представляет собой функцию распределения максимальной из длин n интервалов, на которые разбит интервал $\overline{0, t}$. В связи с этим см. задачи 22—27.

Доказательство. Достаточно доказать теорему 3. Мы докажем рекуррентную формулу

$$\varphi_n(t) = (n-1) \int_0^a \varphi_{n-1}(t-x) \left(\frac{t-x}{t}\right)^{n-2} \frac{dx}{t}. \quad (9.10)$$

Ее справедливость прямо следует из определения φ_n как $(n-1)$ -кратного интеграла, но предпочтительнее толковать (9.10) вероятностно следующим образом. Наименьшее значение среди $X_1, \dots, X_{n-1}$ должно быть меньше чем a , причем существует $n-1$ возможностей выбрать это значение. При условии, что $X_1 = x$, вероятность того, что точка X_1 самая левая, равна $\left[\frac{t-x}{t}\right]^{n-2}$; тогда $X_2, \dots, X_{n-1}$ распределены равномерно в интервале x, t и условная вероятность того, что они удовлетворяют условиям теоремы, равна $\varphi_{n-1}(t-x)$. Суммируя все возможности, мы получаем (9.10)¹⁾.

¹⁾ Читатели, которых беспокоит вопрос вычисления условных вероятностей по плотностям, могут заменить условие $X_j = x$ условием $x-h < X_j < x$, которое имеет положительную вероятность, а затем перейти к пределу при $h \rightarrow 0$.

Если предположить, что φ_n из (9.8) удовлетворяет равенству (9.10), то для u_n , входящих в (9.8), равенство (9.10) сведется к

$$u_n(t) = a^{-1} \int_0^a u_{n-1}(t-x) dx, \quad (9.11)$$

что есть в точности рекуррентная формула (9.1), которая служит для определения u_n . Поэтому достаточно доказать теорему для случая $n=2$. Но и так ясно, что $\varphi_2(t)=1$ при $0 < t < a$ и $\varphi_2(t) = (2a-t)/t$ при $a < t < 2a$ в полном согласии с (9.8). ►

§ 10. СЛУЧАЙНЫЕ НАПРАВЛЕНИЯ

Выбор случайного направления на плоскости $\mathcal{R}^2$ —это то же самое, что выбор наудачу точки на окружности. Если желательно задать направление углом, который оно составляет с правой частью x -оси, то на окружности следует ввести координату θ , $0 \leq \theta < 2\pi$. Для случайных направлений в пространстве $\mathcal{R}^3$ в качестве выборочного пространства служит единичная сфера; каждая область имеет вероятность, равную ее площади, деленной на 4π . Выбор случайного направления в $\mathcal{R}^3$ эквивалентен выбору наудачу точки на этой единичной сфере. Так как это включает пару случайных величин (долготу и широту), то следовало бы отложить обсуждение трехмерного случая до гл. III, но в данном контексте он появляется более естественно.

Предложения. (i) Обозначим через L длину проекции единичного вектора со случайным направлением в $\mathcal{R}^3$ на фиксированную прямую, скажем x -ось. Тогда величина L распределена равномерно на $[0, 1]$ и $E(L) = 1/2$.

(ii) Пусть U —длина проекции того же вектора на фиксированную плоскость, скажем (x, y) -плоскость. Тогда U имеет плотность $t/\sqrt{1-t^2}$ при $0 < t < 1$, и $E(U) = \pi/4$.

Важно то, что две проекции имеют различные распределения. Тот факт, что первое распределение равномерно, не есть свойство случайности, а есть свойство, присущее пространству трех измерений. Аналогом теоремы (i) в $\mathcal{R}^2$ служит

Предложение. (iii) Пусть L —длина проекции случайного единичного вектора в $\mathcal{R}^2$ на x -ось. Тогда L имеет плотность $2/(\pi\sqrt{1-x^2})$ и $E(L) = 2/\pi$.

Доказательства. (iii) Если θ —угол между нашим случайным направлением и y -осью, то $L = |\sin \theta|$ и, следовательно, при $0 < x < 1$ в силу симметрии

$$P\{L \leq x\} = P\{0 < \theta < \arcsin x\} = \frac{2}{\pi} \arcsin x. \quad (10.1)$$

Предложение (iii) теперь доказывается дифференцированием.

(i), (ii) Вспомним элементарную теорему о том, что площадь

сферического пояса между двумя параллельными плоскостями пропорциональна высоте пояса. При $0 < t < 1$ событие $\{L \leq t\}$ представляется поясом $|x_1| \leq t$ высоты $2t$, тогда как событие $\{U \leq t\}$ соответствует поясам $|x_1| \geq \sqrt{1-t^2}$ общей высоты $2 - 2\sqrt{1-t^2}$. Этим обе функции распределения определяются с точностью до числовых множителей. Их значения просто находятся из условия, что обе функции распределения равны 1 при $t=1$. ►

Примеры. а) *Прохождение через сферу.* Пусть Σ — сфера радиуса r и N — точка на ней. Линия, проведенная через N в случайном направлении, пересекает Σ в точке P . Тогда длина сегмента NP является случайной величиной, равномерно распределенной между 0 и $2r$.

Чтобы увидеть это, рассмотрим ось NS сферы и треугольник NPS , который имеет прямой угол в точке P и угол Θ в точке N . Длина NP равна тогда $2r \cos \Theta$. Но $\cos \Theta$ является в то же время проекцией единичного вектора прямой NP на диаметр NS , и поэтому величина $\cos \Theta$ равномерно распределена на $\overline{0, 1}$.

В физике эта модель употребляется для описания прохождения света через «случайно распределенные сферы». Возникающее при этом *поглощение света* использовалось в качестве одного из примеров процессов случайных разбиений в предыдущем параграфе (см. задачу 28).

б) *Круговые объекты под микроскопом.* Через микроскоп наблюдается проекция клетки на (x_1, x_2) -плоскость, а не ее действительная форма. В некоторых биологических экспериментах клетки имеют линзообразную форму и могут рассматриваться как круговые диски. Один лишь горизонтальный диаметр проектируется в свою натуральную величину, а весь диск проектируется в эллипс, малая ось которого является проекцией наиболее круто наклоненного радиуса. Обычно предполагается, что ориентация диска случайна; это означает, что направление его нормали выбрано наудачу. В этом случае проекция единичной нормали на x_3 -ось распределена равномерно в $\overline{0, 1}$. Но угол между этой нормалью и x_3 -осью равен углу между наиболее круто наклоненным радиусом и (x_1, x_2) -плоскостью. Следовательно, отношение малой и большой осей распределено равномерно в $\overline{0, 1}$. Иногда обработка экспериментов основывалась на ошибочном мнении, что угол между наиболее круто наклоненным радиусом и (x_1, x_2) -плоскостью должен быть распределен равномерно.

в) *Почему две скрипки звучат вдвое громче, чем одна?* (Вопрос серьезен, так как громкость пропорциональна квадрату амплитуды колебаний.) Поступающие волны могут быть изображены случайными *единичными* векторами, и эффект наложения звука двух скрипок соответствует сложению двух независимых случайных

векторов. По теореме косинусов квадрат длины результирующего вектора равен $2 + 2 \cos \Theta$. Здесь Θ — угол между двумя случайными векторами, и, следовательно, $\cos \Theta$ распределен равномерно в $[-1, 1]$ и имеет нулевое математическое ожидание. Математическое ожидание квадрата длины суммарного вектора поэтому действительно равно 2.

На плоскости $\cos \Theta$ уже не является равномерно распределенной случайной величиной, но по соображениям симметрии по-прежнему имеет нулевое математическое ожидание. Таким образом, наш результат справедлив при любом числе измерений. См. также пример V, 4, д). ►

Под *случайным вектором* в $\mathcal{R}^3$ понимается вектор, проведенный в случайном направлении, длина которого L является случайной величиной, не зависящей от его направления. Вероятностные свойства случайного вектора полностью определяются свойствами его проекции на x -ось, и, используя последнюю, можно избежать анализа в трех измерениях. Для этого важно знать связь между функцией распределения V истинной длины L и распределением F длины L_x проекции на x -ось. Имеем $L_x = XL$, где X — длина проекции *единичного* вектора с данным направлением. Соответственно величина X равномерно распределена на $[0, 1]$ и не зависит от L . При условии $X = x$ событие $\{L_x \leq t\}$ происходит тогда и только тогда, когда $L \leq t/x$, и поэтому¹⁾

$$F(t) = \int_0^1 V\left(\frac{t}{x}\right) dx, \quad t > 0. \quad (10.2)$$

Соответствующие *плотности* мы получаем дифференцированием

$$f(t) = \int_0^1 v\left(\frac{t}{x}\right) \frac{dx}{x} = \int_t^\infty v(y) \frac{dy}{y}. \quad (10.3)$$

Повторное дифференцирование приводит к

$$v(t) = -tf'(t), \quad t > 0. \quad (10.4)$$

Мы нашли таким образом *аналитическую связь между плотностью v длины случайного вектора в $\mathcal{R}^3$ и плотностью f длины его проекции на фиксированное направление*. Соотношение (10.3) используется для отыскания f , когда v известно, а соотношение (10.4) — в противоположном случае. (Асимметрия между двумя формулами обусловлена тем фактом, что направление не является независимым от длины проекции.)

Примеры. г) *Распределение Максвелла для скоростей*. Рассмотрим случайные векторы в пространстве, проекции которых на

¹⁾ Это рассуждение повторяет доказательство формулы (8.3).

x -ось имеют нормальную плотность с нулевым математическим ожиданием и единичной дисперсией. Длина проекции равна ее абсолютному значению, и поэтому

$$f(t) = 2n(t) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}t^2}, \quad t > 0. \quad (10.5)$$

Тогда из (10.4) выводим

$$v(t) = \sqrt{\frac{2}{\pi}} t^2 e^{-\frac{1}{2}t^2}, \quad t > 0. \quad (10.6)$$

Это есть плотность Максвелла для скоростей в статистической механике. Обычный вывод объединяет предыдущее рассуждение с доказательством того, что плотность f должна иметь вид (10.5). (Другой вывод см. в гл. III, 4.)

д) *Случайные сдвиги в $\mathcal{R}^3$ (Рэлей)*. Рассмотрим n единичных векторов, направления которых выбираются независимо друг от друга и наудачу. Мы разыскиваем распределение длины L_n их суммы. Вместо непосредственного изучения этой суммы мы рассмотрим ее проекцию на x -ось. Эта проекция, очевидно, равна сумме n независимых случайных величин, распределенных равномерно на $[-1, 1]$. Плотность этой суммы дается формулой (9.7) при $b=1$. Длина этой проекции равна ее абсолютному значению, и, используя (10.4), мы получаем плотность длины L_n ¹⁾:

$$v_n(x) = \frac{-x}{2^{n-1}(n-2)!} \sum_{v=0}^n (-1)^v \binom{n}{v} (x-2v)_+^{n-2}, \quad x > 0. \quad (10.7)$$

Эта задача возникает в физике и химии (векторы изображают, например, плоские волны или молекулярные связи). Сведение к одному измерению, по-видимому, делает эту известную задачу тривиальной.

Тот же метод применим к случайным векторам произвольной длины, и таким образом (10.4) позволяет нам свести задачи случайного блуждания в $\mathcal{R}^3$ к более простым задачам в $\mathcal{R}^1$. Даже если трудно получить точное решение, ценную информацию дает центральная предельная теорема [см. пример 4,6] гл. VIII]. ►

Случайные векторы в $\mathcal{R}^3$ определяются тем же способом. Аналогом (10.2) служит соотношение, связывающее распределение V истинной длины и распределение F длины проекции:

$$F(x) = \frac{2}{\pi} \int_0^{\pi/2} V\left(\frac{x}{\sin \theta}\right) d\theta. \quad (10.8)$$

¹⁾ Обычно ссылаются на статью С. Чандрасекара, который вычислил v_1, v_4, v_6 и преобразование Фурье для v_n . Так как он использовал полярные координаты, его $W_n(x)$ нужно умножить на $4\pi x^2$, чтобы получить нашу плотность v_n . Статья Чандрасекара перепечатана в книге Уэкса (1954).

Однако обратное соотношение (10.4) не имеет простого аналога, и, чтобы выразить V через F , мы должны опираться на относительно глубокую теорию интегрального уравнения Абеля¹⁾. Мы констатируем здесь без доказательства, что если F имеет непрерывную плотность f , то

$$1 - V(x) = x \int_0^{\pi/2} f\left(\frac{x}{\sin \theta}\right) \frac{d\theta}{\sin^2 \theta}. \quad (10.9)$$

(См. задачи 29, 30.)

е) *Двойные орбиты.* При наблюдении спектрально-двойных орбит астрономы могут измерить только проекции векторов на плоскость, перпендикулярную к линии зрения. Эллипс в пространстве проектируется в эллипс в этой плоскости. Большая ось исходного эллипса лежит в плоскости, определяемой лучом зрения и проекцией этой оси, и поэтому разумно предположить, что угол между большой осью и ее проекцией равномерно распределен. Распределение проекции определяется (в принципе) измерениями. Распределение исходной большой оси дается тогда решением (10.9) интегрального уравнения Абеля. ►

§ 11. ИСПОЛЬЗОВАНИЕ МЕРЫ ЛЕБЕГА

Если множество A в $\overline{0, 1}$ представляет собой объединение конечного множества непересекающихся интервалов $I_1, I_2, \dots$ с длинами $\lambda_1, \lambda_2, \dots$, то равномерное распределение приписывает ему вероятность

$$P\{A\} = \lambda_1 + \lambda_2 + \dots \quad (11.1)$$

Следующие примеры покажут, что некоторые простые, но важные задачи приводят к объединениям бесконечного числа непересекающихся интервалов. Определение (11.1) по-прежнему применимо и отождествляет $P\{A\}$ с лебеговой мерой A . Это согласуется с нашей программой отождествления вероятностей с интегралом от плотности $f(x) = 1$, за исключением того, что мы используем интеграл Лебега, а не интеграл Римана (который может не существовать). Из теории Лебега нам требуется только тот факт, что если A представляет собой объединение (возможно, пересекающихся) интервалов $I_1, I_2, \dots$, то мера $P\{A\}$ существует и не превосходит суммы $\lambda_1 + \lambda_2 + \dots$ их длин. Для непересекающихся интервалов выполняется равенство (11.1). Использование меры Лебега соответствует свободной интуиции и упрощает дело, так как обосновы-

¹⁾ Преобразование в интегральное уравнение Абеля достигается посредством замены переменных

$$F_1(x) = F\left(\frac{1}{\sqrt{x}}\right), \quad V_1(x) = V\left(\frac{1}{\sqrt{x}}\right) \text{ и } x \sin^2 \theta = y.$$

Тогда (10.8) принимает вид

$$F_1(t) = \frac{1}{\pi} \int_0^t \frac{V_1(y)}{\sqrt{y(t-y)}} dy.$$

ваются все формальные переходы к пределу. Множество N называется *нулевым множеством*, если оно содержится в множествах произвольно малой меры, т. е. если для каждого ε существует множество $A \supset N$, такое, что $P\{A\} < \varepsilon$. В этом случае $P\{N\} = 0$.

В дальнейшем X обозначает случайную величину, распределенную равномерно в $0, \overline{1}$.

Примеры. а) *Какова вероятность того, что величина X рациональна?* Последовательность $\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \dots$ содержит все рациональные числа интервала $0, \overline{1}$ (расположенные в порядке возрастания знаменателей). Выберем $\varepsilon < \frac{1}{2}$ и обозначим через J_k интервал длины ε^{k+1} с центром в k -й точке последовательности. Сумма длин интервалов J_k равна $\varepsilon^2 + \varepsilon^3 + \dots < \varepsilon$, и их объединение покрывает рациональные числа. Поэтому по нашему определению *множество всех рациональных чисел имеет вероятность нуль*, и, следовательно, X иррационально с вероятностью единица.

Здесь уместно спросить: зачем такие множества рассматриваются в теории вероятностей? Один из ответов состоит в том, что их исключением ничего нельзя достичь и что использование теории Лебега действительно упрощает дело, не требуя новых методов. Второй ответ может быть более убедителен для начинающих и для нематематиков; нижеследующие варианты приводят к задачам, несомненно, вероятностной природы.

б) *С какой вероятностью цифра 7 встречается в десятичном разложении величины X ?* В десятичном разложении каждой точки x из открытого интервала между 0,7 и 0,8 цифра 7 появляется на первом месте. Для каждого n существуют 9^{n-1} интервалов длины 10^{-n} , содержащих только такие числа, что цифра 7 появляется на n -м месте, но не раньше (при $n=2$ их концами служат 0,07 и 0,08, затем 0,17 и 0,18 и т. д.). Эти интервалы не пересекаются, и их общая длина равна $\frac{1}{10} \left(1 + \frac{9}{10} + \left(\frac{9}{10} \right)^2 + \dots \right) = 1$. Таким образом, наше событие имеет *вероятность 1*.

Заметим, что некоторые числа имеют два представления, например, $0,7 = 0,6999\dots$. Чтобы сделать обсуждаемый нами вопрос определенным, мы должны точно установить, должна или может цифра 7 появиться в представлении, но наше рассуждение не зависит от этого различия. Причина в том, что только рациональные числа могут иметь два представления, а множество всех рациональных чисел имеет вероятность нуль.

в) *Бросание монеты и случайный выбор.* Посмотрим теперь, как «случайный выбор точки X между 0 и 1» может быть описан в терминах дискретных случайных величин. Обозначим через $X_k(x)$ k -й десятичный знак x (чтобы избежать двусмысленностей, мы будем использовать, когда это возможно, обрывающиеся разложе-

ния). Случайная величина X_k принимает значения $0, 1, \dots, 9$, каждое с вероятностью $\frac{1}{10}$, и величины X_k взаимно независимы. Кроме того, мы имеем тождество

$$X = \sum_{k=1}^{\infty} \frac{1}{10^k} X_k. \quad (11.2)$$

Эта формула сводит случайный выбор точки X к последовательным выборам ее десятичных знаков.

В последующих рассуждениях мы перейдем от десятичного разложения к двоичному, т. е. мы заменим основание 10 на основание 2. Вместо (11.2) мы теперь имеем

$$X = \sum_{k=1}^{\infty} \frac{1}{2^k} X_k, \quad (11.3)$$

где X_k — взаимно независимые случайные величины, принимающие значения 0 и 1 с вероятностью $\frac{1}{2}$. Эти величины определены на интервале $\overline{0, 1}$, для которого вероятность отождествляется с мерой Лебега (длиной). Эта постановка задачи напоминает игру с бросанием монеты из первого тома, в которой выборочное пространство состоит из бесконечных последовательностей гербов и решетоков или нулей и единиц. В этом выборочном пространстве можно теперь заново интерпретировать (11.3). В этом пространстве X_k суть координатные величины, а X есть случайная величина, ими определяемая; ее функция распределения, конечно, равномерна. Заметим, что второе выборочное пространство содержит две различные выборочные точки 011111... и 1000000..., хотя соответствующие двоичные разложения изображают одну и ту же точку $\frac{1}{2}$. Однако понятие нулевой вероятности дает нам возможность отождествить два выборочных пространства. Другими словами, при пренебрежении событием нулевой вероятности случайный выбор точки X между 0 и 1 может быть осуществлен последовательностью бросаний монеты и, наоборот, результат бесконечных бросаний монеты может быть изображен точкой x из интервала $\overline{0, 1}$. Каждая случайная величина, связанная с игрой в монету, может быть представлена функцией на $\overline{0, 1}$ и т. д. Этот удобный и наглядный прием использовался в теории вероятностей издавна, но он связан с пренебрежением событиями, имеющими нулевую вероятность.

г) *Распределения канторовского типа.* Если мы будем рассматривать в (11.3) сумму слагаемых с четными номерами, или, что то же самое, рассматривать случайную величину

$$Y = 3 \sum_{v=1}^{\infty} \frac{1}{4^v} X_v, \quad (11.4)$$

то мы приходим к распределению с неожиданными свойствами. (Множитель 3 введен для упрощения рассуждений. Сумма слагаемых с нечетными номерами распределена так же, как $\frac{2}{3} Y$.) Функция распределения $F(x) = P\{Y \leq x\}$ служит примером так называемых сингулярных распределений.

При подсчете мы рассматриваем Y как выигрыш игрока, который получает величину $3 \cdot 4^{-k}$, когда k -е бросание симметричной монеты дает в результате решетку. Полный выигрыш лежит между 0 и $3(4^{-1} + 4^{-2} + \dots) = 1$. Если первое испытание приводит к 1, то выигрыш $\geq \frac{1}{4}$, тогда как в противоположном случае $Y \leq 3(4^{-2} + 4^{-3} + \dots) = \frac{1}{4}$. Таким образом, неравенство $\frac{1}{4} < Y < \frac{3}{4}$ не может быть осуществлено ни при каких обстоятельствах, и поэтому $F(x) = \frac{1}{4}$ в этом интервале длины $\frac{1}{4}$. Отсюда следует, что F не может иметь скачка, превышающего $\frac{1}{4}$.

Отметим теперь, что с точностью до множителя $\frac{1}{4}$ испытания с номерами 2, 3, ... представляют собой точную копию всей последовательности, и поэтому график $F(x)$ в интервале $0, \frac{1}{4}$ отличается от всего графика только преобразованием подобия

$$F(x) = \frac{1}{2} F(4x), \quad 0 < x < \frac{1}{4}. \quad (11.5)$$

Отсюда следует, что $F(x) = \frac{1}{4}$ всюду в интервале длины $\frac{1}{8}$ с центром в точке $\frac{1}{8}$. По соображениям симметрии $F(x) = \frac{3}{4}$ всюду в интервале длины $\frac{1}{8}$ с центром в $\frac{7}{8}$. Мы нашли сейчас три интервала общей длины $\frac{1}{2} + \frac{1}{8} = \frac{5}{8}$, на каждом из которых F принимает постоянное значение, а именно $\frac{1}{4}$, $\frac{1}{2}$ или $\frac{3}{4}$. Следовательно, F не может иметь скачка, превышающего $\frac{1}{4}$. Остаются четыре интервала длиной $\frac{1}{16}$ каждый, и в каждом из них график F отличается от всего графика только преобразованием подобия. Каждый из четырех интервалов поэтому содержит подынтервал половины его длины, на котором F принимает постоянное значение (а именно $\frac{1}{8}$, $\frac{3}{8}$, $\frac{5}{8}$, $\frac{7}{8}$ соответственно). Продолжая в том же духе, мы найдем за n шагов $1 + 2 + 2^2 + \dots + 2^{n-1}$ интервалов общей длины $2^{-1} + 2^{-2} + 2^{-3} + \dots + 2^{-n} = 1 - 2^{-n}$, в каждом из которых F принимает постоянное значение.

Таким образом, F представляет собой непрерывную функцию, возрастающую от $F(0) = 0$ до $F(1) = 1$ так, что сумма длин интервалов постоянно равна 1. Грубо говоря, все возрастание F происходит на множестве меры 0. Мы имеем здесь непрерывную функцию распределения F без плотности f . ►

§ 12. ЭМПИРИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ

«Эмпирическая функция распределения» F_n для n точек $a_1, \dots, a_n$ на прямой есть ступенчатая функция со скачками $1/n$ в точках $a_1, \dots, a_n$. Другими словами, $nF_n(x)$ равно числу точек

a_n в $-\infty, x$, и F_n —функция распределения. Если даны n случайных величин $X_1, \dots, X_n$, то их значения в некоторой точке выборочного пространства образуют строку из n чисел и ее эмпирическая функция распределения называется эмпирическим распределением выборки. При каждом x значение $F_n(x)$ эмпирического распределения выборки определяет новую случайную величину, а эмпирическое распределение $(X_1, \dots, X_n)$ —это целое семейство случайных величин, зависящих от параметра x (формально мы имеем дело со случайным процессом, где x —временной параметр). Мы не будем пытаться развить здесь теорию эмпирических распределений, но это понятие можно использовать для иллюстрации возникновения сложных случайных величин в простых приложениях. Сверх того мы в новом свете увидим равномерное распределение.

Пусть $X_1, \dots, X_n$ обозначают взаимно независимые случайные величины с одинаковым непрерывным распределением F . Вероятность того, что любые две величины принимают одно и то же значение, равна нулю, и поэтому мы можем ограничить наше внимание выборками из n различных значений. При фиксированном x число величин X_k , таких, что $X_k \leq x$, имеет биномиальное распределение с вероятностью «успеха» $p = F(x)$, и поэтому случайная величина $F_n(x)$ имеет биномиальное распределение с возможными значениями $0, 1/n, \dots, 1$. Следовательно, при большом n и фиксированном x значение $F_n(x)$ с большой вероятностью будет довольно близким к $F(x)$, и центральная предельная теорема информирует нас о возможных отклонениях. Более интересен (зависящий от случая) график F_n в целом и то, насколько он близок к F . Мерой этой близости может быть максимум расхождения, т. е.

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|. \quad (12.1)$$

Эта новая случайная величина представляет большой интерес для статистиков в силу следующего свойства. Распределение вероятностей случайной величины D_n не зависит от F (конечно, при условии, что F непрерывна).

Для доказательства этого свойства достаточно проверить, что распределение D_n остается неизменным, когда F заменяется равномерным распределением. Мы сначала покажем, что величины $Y_k = F(X_k)$ распределены равномерно в $[0, 1]$. Для этой цели мы ограничим t интервалом $[0, 1]$, и в этом интервале мы определим v как функцию, обратную F . Событие $\{F(X_k) \leq t\}$ равносильно тогда событию $\{X_k \leq v(t)\}$, которое имеет вероятность $F(v(t)) = t$. Таким образом, $P\{Y_k \leq t\} = t$, как и утверждалось.

Величины $Y_1, \dots, Y_n$ взаимно независимы, и мы обозначим их эмпирическое распределение через G_n . Только что приведенные рассуждения показывают также, что при фиксированном t случайная величина $G_n(t)$ тождественна с $F_n(v(t))$. Так как $t = F(v(t))$,

то отсюда следует, что в каждой точке выборочного пространства $\mathcal{A}^n$

$$\sup |G_n(t) - t| = \sup |F_n(v(t)) - F(v(t))| = D_n.$$

Этим доказательство завершается.

Тот факт, что распределение D_n не зависит от исходного распределения F , дает статистикам возможность построить критерии и процедуры оценивания, применимые в ситуации, когда исходное распределение неизвестно. В этой связи другие величины, связанные с D_n , находят даже большее практическое использование.

Пусть $X_1, \dots, X_n, X_1^*, \dots, X_n^*$ — $2n$ взаимно независимых случайных величин с одинаковым непрерывным распределением F . Обозначим эмпирические распределения $(X_1, \dots, X_n)$ и $(X_1^*, \dots, X_n^*)$ через F_n и F_n^* соответственно. Положим

$$D_{n,n} = \sup_x |F_n(x) - F_n^*(x)|. \quad (12.2)$$

Это есть *максимальное расхождение между двумя эмпирическими распределениями*. Оно обладает тем же свойством, что и D_n , а именно не зависит от распределения F . По этой причине $D_{n,n}$ используется при построении статистического критерия проверки «гипотезы о том, что $X_1, \dots, X_n$ и $X_1^*, \dots, X_n^*$ являются случайными выборками из одной и той же совокупности».

Распределение $D_{n,n}$ было найдено с помощью громоздких вычислений и исследований, но в 1951 г. Гнеденко и Корольюк показали, что вся проблема сводится к задаче случайного блуждания с хорошо известным решением. Их доказательство отличается изяществом, и мы приведем его для иллюстрации возможностей простых комбинаторных методов.

Теорема. Вероятность $P\{D_{n,n} < r/n\}$ равна вероятности того, что в симметричном случайном блуждании траектория длины $2n$, начинающаяся и заканчивающаяся в нуле, не достигнет точек $\pm r$.

Доказательство. Достаточно рассмотреть целые r . Расположим $2n$ величин $X_1, \dots, X_n^*$ в порядке возрастания величины и положим $\varepsilon_k = 1$ или $\varepsilon_k = -1$ в зависимости от того, какая величина занимает k -е место, X_j или X_j^* . Окончательная запись содержит n «+1» и n «-1», и все $\binom{2n}{n}$ расположений равновероятны.

Поэтому возникающие при этом строки $(\varepsilon_1, \dots, \varepsilon_{2n})$ находятся во взаимно однозначном соответствии с траекториями длины $2n$, начинающимися и заканчивающимися в начале. Если теперь $\varepsilon_1 + \dots + \varepsilon_j = k$, то первые j мест заняты $(j+k)/2$ неотмеченными и $(j-k)/2$ отмеченными величинами, и поэтому существует чка x , такая, что $F_n(x) = (j+k)/2n$ и $F_n^*(x) = (j-k)/2n$. Но тогда

$|F_n(x) - F_n^*(x)| = |k|/n$ и, следовательно, $D_{n,n} \geq |k|/n$. То же рассуждение, проведенное в обратном порядке, завершает доказательство.

Точное выражение для рассматриваемой вероятности содержится в 1, гл. XIV, (9.1). В самом деле, выражение

$$\binom{2n}{n} P \left\{ D_{n,n} \leq \frac{r}{n} \right\} = \omega_{r,n}$$

есть вероятность того, что частица, выходящая из начала отсчета, вернется в момент $2n$ в начало, не касаясь $\pm r$. Последнее условие можно реализовать, если установить в точках $\pm r$ поглощающие экраны, и поэтому $\omega_{r,n}$ есть вероятность возвращения в начало координат в момент $2n$, когда точки $\pm r$ являются поглощающими экранами. [В 1; гл. XIV, (9.1) рассматривается интервал $0, a$, а не $-r, r$. Наша вероятность $\omega_{r,n}$ тождественна $u_{r,2n}(r)$.]

В 1; гл. XIV было показано, что предельный переход приводит от случайных блужданий к диффузионным процессам, и, таким образом, легко видеть, что распределение $\sqrt{n} D_{n,n}$ стремится к пределу. Фактически этот предел был найден Смирновым еще в 1939 г., а аналогичный предел для $\sqrt{n} D_n$ — Колмогоровым в 1933 г. Эти вычисления очень сложны и не объясняют связи с диффузионными процессами в противоположность подходу Гнеденко — Королука. С другой стороны, они стимулировали плодотворную работу по сходности стохастических процессов (Биллингсли, Донскер, Прохоров, Скороход и другие).

Можно упомянуть, что теоремы Смирнова в равной степени применимы и к расхождениям $D_{m,n}$ между эмпирическими распределениями выборок, имеющих различные объемы m и n . Подход, связанный со случайным блужданием, применим к этой проблеме, но он во многом теряет свою изящность и простоту (Гнеденко, Рвачева). Статистиками было исследовано множество вариантов $D_{m,n}$ (см. задачу 36).

§ 13. ЗАДАЧИ

Во всех задачах предполагается, что заданные величины *взаимно независимы*.

1. Пусть X и Y имеют плотности ae^{-ax} , сосредоточенные на $0, \infty$. Найдите плотности величин

(i) X^2	(ii) $3+2X$
(iii) $X-Y$	(iv) $ X-Y $
(v) Наименьшей из величин X и Y^2	(vi) Наибольшей из величин X и Y^2

2. Решите ту же задачу, если плотности X и Y равны $1/2$ в интервале $-1, 1$ и 0 вне его.

3. Найдите плотности для $X+Y$ и $X-Y$, если X имеет плотность ae^{-ax} ($a > 0$), а плотность Y равна h^{-1} при $0 < x < h$.

4. Найдите вероятность того, что $\lambda^2 - 2a\lambda + b$ имеет комплексные корни, если коэффициенты a и b являются одинаково распределенными случайными величинами, плотность которых

- (i) равномерна, т. е. равна h^{-1} при $0 < x < h$;
 (ii) показательна, т. е. равна ae^{-ax} при $x > 0$.

5. Найдите функции распределения величин $\frac{X+Y}{X}$ и $\frac{X+Y}{Z}$, если величины X , Y и Z имеют одинаковое показательное распределение.

6. Выведите формулу свертки (3.6) для показательного распределения

прямым переходом к пределу из формулы свертки для «отрицательного биномиального» распределения из 1; гл. VI, (8.1).

7. Для пуассоновского процесса § 4 обозначим через Z время между моментом t и последним предшествующим прибытием или 0 («возраст» текущего времени между прибытиями). Найдите распределение Z и покажите, что оно стремится к показательному распределению при $t \rightarrow \infty$.

8. В примере 5, а) покажите, что вероятность появления первого рекордного значения на n -м месте и при этом не превосходящего x равна

$$\frac{1}{n(n+1)} (1 - e^{-\alpha x})^{n+1}.$$

Докажите, что распределение вероятностей первого рекордного значения есть $1 - (1 + \alpha x) e^{-\alpha x}$.

[Вообще, если величины X_j положительны и подчинены произвольному непрерывному распределению F , то первая вероятность равна $[n(n+1)]^{-1} \times \times F^{n+1}(x)$, а распределение первого рекорда равно $F - (1-F) \log(1-F)^{-1}$.]

9. *Нисходящие серии.* Случайная величина N определена как единственный индекс, такой, что $X_1 \geq X_2 \geq \dots \geq X_N < X_{N+1}$. Если X_j имеют одинаковое распределение с непрерывной функцией распределения, то докажите, что $P(N=n) = (n-1)/n!$ и $E(N) = e$. *Указание.* Используйте метод примера 5, а), относящийся к рекордным значениям.

10. *Образование колонн на транспорте¹⁾.* Автомобили отправляются друг за другом из начала координат и движутся с различными, но постоянными скоростями по бесконечной дороге, на которой запрещены остановки. Если автомобиль догоняет другой, медленнее едущий автомобиль, то ему самому приходится тащиться с той же скоростью. Таким образом будет сформирована колонна. Предельный размер колонны будет зависеть от скоростей автомобилей, но не от времен между последовательными отправлениями из начальной точки.

Рассмотрим скорости автомобилей как независимые случайные величины с одинаковым непрерывным распределением. Выберем наудачу любой автомобиль, например первый из отправляющихся после данного момента. Используя комбинаторные методы примера 5, а), покажите, что

а) вероятность того, что данный автомобиль не будет тащиться за каким-либо другим автомобилем, стремится к $1/2$;

б) вероятность того, что данный автомобиль ведет за собой колонну общим размером n (т. е. равно $n-1$ автомобилей тащатся за ним), стремится к $1/(n+1)(n+2)$;

в) вероятность того, что данный автомобиль является последним в колонне размера n (т. е. тащится за $n-1$ автомобилями), стремится к тому же пределу, что и предыдущая вероятность.

11. *Обобщение²⁾ примера 5, а) о рекордном значении.* Вместо того чтобы рассматривать одно предварительное наблюдение X_0 , мы начинаем с выборки $(X_1, \dots, X_m)$ с порядковыми статистиками $(X_{(1)}, \dots, X_{(m)})$. (Общее для всех величин распределение F не играет роли, коль скоро оно непрерывно.)

а) Пусть N — первый индекс n , такой, что $X_{m+n} \geq X_{(m)}$. Покажите, что $P\{N > n\} = m/(m+n)$. [В примере 5, а) мы имели $m=1$.]

б) Пусть N — первый индекс n , такой, что $X_{m+n} \geq X_{(m-r+1)}$; покажите, что

$$P\{N > n\} = \frac{\binom{m}{r}}{\binom{m+n}{r}}.$$

¹⁾ Newell G. F., Opns Res. 7 (1959), 589—598.

²⁾ Wilks S. S., J. Austr. Math. Soc., 1 (1959), 106—112.

При $r \geq 2$ мы имеем $E(N) = < \infty$ и

$$P\{N \leq mx\} \rightarrow 1 - \frac{1}{(1+x)^r}, \quad m \rightarrow \infty.$$

в) Пусть N — первый индекс, такой, что X_{m+n} находится вне интервала между $X_{(1)}$ и $X_{(m)}$. Тогда

$$P\{N > n\} = \frac{m(m-1)}{(m+n)(m+n-1)} \quad \text{и} \quad E(N) < \infty.$$

12. (Свертки показательных распределений.) Пусть X_j при $j=0, \dots, n$ имеет плотность $\lambda_j e^{-\lambda_j x}$ для $x > 0$. Здесь $\lambda_j \neq \lambda_k$, если только j не равно k . Положим

$$\psi_{k,n} = [(\lambda_0 - \varphi_k) \dots (\lambda_{k-1} - \lambda_k) (\lambda_{k+1} - \lambda_k) \dots (\lambda_n - \lambda_k)]^{-1}.$$

Покажите, что $X_0 + \dots + X_n$ имеет плотность

$$P_n(t) = \lambda_0 \dots \lambda_{n-1} [\psi_{0,n} e^{-\lambda_0 t} + \dots + \psi_{n,n} e^{-\lambda_n t}]. \quad (*)$$

Указание. Используйте индукцию, симметрию и (2.14). Вычислений можно избежать.

13. Продолжение. Если Y_j имеет плотность $j e^{-jx}$, то плотность суммы $Y_1 + \dots + Y_n$ равна

$$f_n(x) = n \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} e^{-kx}, \quad x > 0.$$

Используя утверждение примера 6,6), выведите, что f_{n-1} есть плотность размаха $X_{(n)} - X_{(1)}$ выборки $X_1, \dots, X_n$, где все X_j имеют одинаковую плотность e^{-x} .

14. Процессы чистого размножения. В процессе чистого размножения из 1; гл. XVII, 3 система проходит через последовательность состояний $E_0 \rightarrow E_1 \rightarrow \dots$, причем время пребывания в состоянии E_k имеет плотность $\lambda_k e^{-\lambda_k x}$. Таким образом, $S_n = X_0 + \dots + X_n$ есть момент перехода $E_n \rightarrow E_{n+1}$. Обозначьте через $P_n(t)$ вероятность E_n в момент t . Покажите, что $P_n(t) = P\{S_n > t\} - P\{S_{n-1} > t\}$ и, следовательно, $P_n(t)$ определяется формулой (*) задачи 12. Дифференциальные уравнения процесса: $P'_0(t) = -\lambda_0 P_0(t)$,

$$P'_n(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \quad n \geq 1,$$

можно вывести (а) из (1) и (б) из свойств сумм S_n .

Указание. По соображениям симметрии достаточно рассмотреть коэффициент при $e^{-\lambda_0 t}$.

15. В примере 6,а), где рассматриваются параллельные очереди, мы скажем, что система находится в состоянии k , если k окошек свободны. Покажите, что здесь применима модель процесса чистого размножения (задача 14) с $\lambda_k = (n-k)\alpha$. Докажите, что

$$P_k(t) = \binom{n}{k} (1 - e^{-\alpha t})^k e^{-(n-k)\alpha t}.$$

Выведите отсюда распределение $X_{(k)}$.

16. Рассмотрите две независимые очереди из m и $n > m$ лиц соответственно, предполагая, что времена обслуживания имеют одно и то же показательное распределение. Покажите, что вероятность более длинной очереди окончиться первой равна вероятности получить n гербов раньше, чем m решето, при бросаниях симметричной монеты. Найдите ту же вероятность,

рассматривая отношение X/Y двух величин с гамма-распределениями G_m и G_n , заданными формулой (3.5).

17. *Пример статистического оценивания.* Предполагается, что продолжительности существования электрических ламп имеют показательное распределение с неизвестным математическим ожиданием α^{-1} . Чтобы оценить α , берется выборка из n ламп и наблюдаются продолжительности существования

$$X_{(1)} < X_{(2)} < \dots < X_{(r)}$$

первых r портящихся ламп. «Наилучшая несмещенная оценка» α^{-1} дается линейной комбинацией $U = \lambda_1 X_{(1)} + \dots + \lambda_r X_{(r)}$, такой, что $E(U) = \alpha^{-1}$ и $\text{Var}(U)$ имеет наименьшее возможное значение. Покажите, что

$$U = (X_{(1)} + \dots + X_{(r)}) \frac{1}{r} + X_{(r)} (n-r) \frac{1}{r},$$

и тогда $\text{Var}(U) = \frac{1}{r} \alpha^{-2}$.

Указание. Проведите вычисления в терминах *независимых* величин $X_{(k)} - X_{(k-1)}$ (см. пример 6, б)).

18. Пусть величины $X_1, \dots, X_n$ распределены равномерно на $[0, 1]$. Покажите, что размах $X_{(n)} - X_{(1)}$ имеет плотность $n(n-1)x^{n-2}(1-x)$ и математическое ожидание $(n-1)/(n+1)$. Какова вероятность, что все n точек лежат внутри некоторого интервала длины t ?

19. Ответьте на вопросы примера 6, б), когда три времени ожидания равномерно распределены в интервале $[0, 1]$. (Задача включает «скучные» вычисления, однако это может оказаться полезным упражнением по части технических приемов.)

20. На окружности независимо и наудачу выбраны четыре точки. Найдите вероятность того, что хорды X_1X_2 и X_3X_4 пересекаются а) без вычислений, используя лишь соображения симметрии; б) из определения через интеграл.

21. В процессе случайного разбиения из § 8 обозначьте через $X_{11}, X_{12}, X_{21}, X_{22}$ массы четырех осколков второго поколения, индекс 1 относится к наименьшей, а 2 — к наибольшей части. Найдите плотности и математические ожидания этих величин.

Замечание. Следующие несколько задач содержат новые теоремы, касающиеся случайных разбиений интервала на части (см. пример 7, б)). Предполагается, что случайные величины $X_1, \dots, X_n$ независимы и равномерно распределены на $[0, t]$. Эти случайные величины индуцируют разбиение интервала $[0, t]$ на $n+1$ подынтервалов. Длины этих интервалов, взятые в надлежащем порядке, мы обозначим $L_1, \dots, L_{n+1}$. В терминах порядковых статистик

$$L_1 = X_{(1)}, L_2 = X_{(2)} - X_{(1)}, \dots, L_{n+1} = t - X_{(n)}.$$

22. Обозначим через $p_n(t)$ вероятность того, что длины всех $n+1$ подынтервалов разбиения больше чем h (иными словами, $p_n(t) = P\{\min L_k > h\}$, что представляет собой хвост функции распределения самого короткого из интервалов). Докажите рекуррентное соотношение

$$p_n(t) = \frac{n}{t^n} \int_0^{t-h} x^{n-1} p_{n-1}(x) dx. \quad (*)$$

Выведите отсюда, что $p_n(t) = t^{-n} (t - (n+1)h)_+^n$.

23. Используя рекуррентное соотношение, аналогичное (*), докажите без вычислений, что для произвольных $x_1 \geq 0, \dots, x_{n+1} \geq 0$

$$P\{L_1 > x_1, \dots, L_{n+1} > x_{n+1}\} = t^{-n} (t - x_1 - \dots - x_{n+1})_+^n. \quad (**)$$

(Этот изящный результат был получен Б. де Финетти¹⁾ из геометрических соображений. Он включает множество интересных частных случаев. Если $x_j = h$ при всех j , то мы получаем предыдущую задачу. Пример 7, б) соответствует частному случаю, когда ровно одна из величин x_j отлична от нуля. Теорема о покрытиях (теорема 3 § 9) следует из (**) и формулы I; IV, (1.5) для появления по крайней мере одного из $(n+1)$ событий.)

24. Пусть $q_n(t)$ — вероятность того, что все взаимные расстояния между X_k превышают h . (Отличие от задачи 22 в том, что здесь не накладываются ограничения на концевые интервалы L_1 и L_{n+1} .) Найдите соотношение, аналогичное (*), и выведите отсюда $q_n(t)$.

25. Продолжение. Не используя решения предыдущих задач, найдите, что $p_n(t) = (t-2h)^n t^{-n} q_n(t-2h)$.

26. Сформулируйте аналог задачи 24 для окружности и покажите, что задача 23 дает ее решение.

27. Равнобедренный треугольник образован единичным вектором в x -направлении и единичным вектором в случайном направлении. Найдите распределение длины третьей стороны (i) в $\mathcal{R}^3$ и (ii) в $\mathcal{R}^2$.

28. Единичный круг (шар) с центром в 0 имеет на положительной x -оси северный полюс. Из полюса выходит луч, и его угол с x -осью распределен равномерно на $[-1/2\pi, 1/2\pi]$. Найдите распределение длины хорды внутри круга (шара).

Замечание. В $\mathcal{R}^3$ луч имеет случайное направление, и мы имеем дело с аналогом примера 10, а). В $\mathcal{R}^3$ же задача другая.

29. Отношение средних длин случайного вектора и его проекции на x -ось равно 2 в $\mathcal{R}^3$ и $\pi/2$ в $\mathcal{R}^2$.

Указание. Используйте (10.2) и (10.8).

30. Длина случайного вектора распределена равномерно на $[0, 1]$. Найдите плотность длины его проекции на x -ось а) в $\mathcal{R}^3$ и б) в $\mathcal{R}^2$. **Указание.** Используйте (10.4) и (10.9).

31. Найдите функцию распределения угла между x -осью и случайно выбранным направлением в $\mathcal{R}^4$.

32. Найдите в $\mathcal{R}^4$ аналог соотношения (10.2) между распределением длины случайного вектора и распределением длины его проекции на x -ось. Проверьте результат задачи 31, ограничившись единичным вектором.

33. **Предельная теорема для порядковых статистик.** а) Пусть $X_1, \dots, X_n$ распределены равномерно в $[0, 1]$. Докажите, что при фиксированном k и $n \rightarrow \infty$

$$P \left\{ X_{(k)} \leq \frac{x}{n} \right\} \rightarrow G_{k-1}(x), \quad x > 0,$$

где G_k есть гамма-распределение (3.5) [см. пример 7, д)].

б) Если X_k имеют произвольную непрерывную функцию распределения F , то тот же предел существует и для $P \{X_{(k)} \leq \Phi(x/n)\}$, где Φ есть функция, обратная к F (Смирнов).

34. **Предельная теорема для выборочной медианы.** Порядковая статистика $X_{(n)}$ выборки $(X_1, \dots, X_{2n-1})$ называется **выборочной медианой**. Если X_j независимы и равномерно распределены на $[0, 1]$, то покажите, что

$$P \left\{ X_{(n)} - \frac{1}{2} < t/\sqrt{8n} \right\} \rightarrow \mathcal{N}(t),$$

где $\mathcal{N}$ обозначает функцию распределения стандартного нормального распределения.

35. **Продолжение.** Пусть случайные величины X_j имеют одинаковое распределение F с непрерывной плотностью f . Пусть m — медиана распределе-

¹⁾ Giornale Ist. Ital. degli Attuari, 27 (1964), 151—173.

ния F , т. е. $F(m) = 1/2$. Покажите, что

$$P\{X_{(n)} < x\} = (2n-1) \binom{2n-2}{n-1} \int_{-\infty}^x F^{n-1}(y) [1-F(y)]^{n-1} f(y) dy,$$

откуда в силу предыдущей задачи

$$P\left\{X_{(n)} - m < \frac{t}{f(m)\sqrt{8n}}\right\} \rightarrow \mathfrak{N}(t).$$

36. Докажите следующий вариант теоремы Гнеденко—Королука в § 12:

$$P\left\{\sup_x [F_n(x) - F_n^*(x)] \geq \frac{r}{n}\right\} = \frac{\binom{2n}{n-r}}{\binom{2n}{n}},$$

где $r=1, 2, \dots, n$. [В отличие от первоначальной формулировки знаки абсолютных величин слева опущены, и поэтому в соответствующем случайном блуждании появляется только один поглощающий барьер в точке r .]

37. Образование показательно распределенных случайных величин из равномерно распределенных величин¹⁾. Пусть $X_1, X_2, \dots$ — независимые величины с равномерным распределением в $\overline{0, 1}$. Определим случайную величину N как такой индекс, что $X_1 \geq X_2 \geq \dots \geq X_{N-1} < X_N$ (см. задачу 9). Докажите, что

$$P\{X_1 \leq x, N=n\} = \frac{x^{n-1}}{(n-1)!} - \frac{x^n}{n!},$$

откуда $P\{X_1 \leq x, N \text{ четно}\} = 1 - e^{-x}$.

Определим Y следующим образом: «испытанием» служит последовательность $X_1, \dots, X_N$; мы говорим о «неудаче», если N нечетно. Мы повторяем независимые испытания до появления первого «успеха». Пусть Y равно числу неудач плюс первая величина в успешном испытании. Докажите, что

$$P\{Y < x\} = 1 - e^{-x}.$$

¹⁾ Von Neumann J., National Bureau of Standards, Appl. Math. Series, 12 (1951), 36—38.

Основная задача этой главы состоит в том, чтобы перечислить для облегчения ссылок те плотности, которые чаще всего будут появляться в дальнейшем. Рандомизация, описанная во второй части, полезна для многих целей. В качестве иллюстрации дается вывод формул для некоторых распределений, связанных с функциями Бесселя и встречающихся в различных переменных. Мы увидим, что этот простой вероятностный прием заменяет сложные вычисления и трудоемкий анализ.

§ 1. ОБОЗНАЧЕНИЯ И ОПРЕДЕЛЕНИЯ

Мы говорим, что плотность f и соответствующее ей распределение F *сосредоточены* (или *сконцентрированы*) на интервале $I = \overline{a, b^1}$, если $f(x) = 0$ для всех x , не принадлежащих I . Тогда $F(x) = 0$ для $x < a$ и $F(x) = 1$ для $x > b$. Говорят, что два распределения F и G , а также их плотности f и g принадлежат *одному и тому же типу*, если они связаны соотношениями

$$G(x) = F(ax + b), \quad g(x) = af(ax + b), \quad (1.1)$$

где $a > 0$. Мы будем часто называть b *центрирующим* параметром и a — *масштабным* параметром. Эти термины становятся хорошо понятными после следующего разъяснения: если F служит функцией распределения случайной величины X , то G есть функция распределения величины

$$Y = \frac{X - b}{a}. \quad (1.2)$$

Во многих ситуациях в действительности важен только тип распределения.

Математическое ожидание m и *дисперсия* σ^2 , соответствующие

¹⁾ В соответствии с обычной терминологией наименьший замкнутый интервал I с таким свойством следовало бы называть *носителем* f . Новый термин введен в связи с тем, что мы будем употреблять его в более общем смысле: так, распределение может быть сосредоточено на множестве всех целых или всех рациональных чисел.

плотности f (или F), определяются равенствами

$$m = \int_{-\infty}^{+\infty} xf(x) dx, \quad \sigma^2 = \int_{-\infty}^{+\infty} (x-m)^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - m^2, \quad (1.3)$$

если только интегралы сходятся абсолютно. Из формулы (1.2) следует, что плотности g в этом случае соответствует математическое ожидание $(m-b)/a$ и дисперсия σ^2/a^2 . Отсюда ясно, что для каждого типа существует самое большее одна плотность с нулевым математическим ожиданием и единичной дисперсией.

Вспомним [см. гл. I, (2.12)], что *свертка* $f = f_1 * f_2$ двух плотностей f_1 и f_2 есть плотность вероятности, определяемая формулой

$$f(x) = \int_{-\infty}^{+\infty} f_1(x-y) f_2(y) dy. \quad (1.4)$$

Если f_1 и f_2 сосредоточены на $0, \infty$, эта формула сводится к

$$f(x) = \int_0^x f_1(x-y) f_2(y) dy, \quad x > 0. \quad (1.5)$$

Функция f представляет собой плотность суммы двух независимых случайных величин с плотностями f_1 и f_2 . Заметим, что для $g_i(x) = f_i(x+b_i)$ свертка $g = g_1 * g_2$ дается равенством $g(x) = f(x+b_1+b_2)$, как это видно из (1.2).

Мы заканчиваем этот параграф введением обозначений для плотности и функции распределения *нормального распределения*:

$$n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \mathfrak{N}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy. \quad (1.6)$$

Давно знакомая нам нормальная плотность с математическим ожиданием m и дисперсией σ^2 определяется как

$$\frac{1}{\sigma} n\left(\frac{x-m}{\sigma}\right), \quad \sigma > 0.$$

В центральной предельной теореме неявно содержится существенный факт, что *семейство нормальных плотностей замкнуто относительно операции свертки*; другими словами, свертка двух нормальных плотностей с математическими ожиданиями m_1, m_2 и дисперсиями σ_1^2, σ_2^2 есть нормальная плотность с математическим ожиданием $m_1 + m_2$ и дисперсией $\sigma_1^2 + \sigma_2^2$. Ввиду ранее сказанного достаточно доказать это для $m_1 = m_2 = 0$. Утверждается, что

$$\frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{x^2}{2\sigma^2}\right] = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left[-\frac{(x-y)^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}\right] dy. \quad (1.7)$$

Справедливость этого утверждения становится очевидной, если произвести замену переменных $z = y \frac{\sigma}{\sigma_1 \sigma_2} - x \frac{\sigma_2}{\sigma \sigma_1}$, где x фиксировано. (См. задачу 1.)

§ 2. ГАММА-РАСПРЕДЕЛЕНИЯ

Гамма-функция Γ определяется как

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx, \quad t > 0. \quad (2.1)$$

[См. 1; гл. II, (12.22).] Гамма-функция интерполирует факториалы в том смысле, что

$$\Gamma(n+1) = n! \quad \text{для } n = 0, 1, \dots$$

Интегрирование по частям показывает, что $\Gamma(t) = (t-1) \Gamma(t-1)$ при всех $t > 0$ (задача 2).

Гамма-плотности, сосредоточенные на $\overline{0, \infty}$, определяются формулой

$$f_{\alpha, \nu}(x) = \frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x}, \quad x > 0. \quad (2.2)$$

Здесь $\alpha > 0$ — тривиальный масштабный параметр, но $\nu > 0$ — существенный параметр. В частном случае $f_{\alpha, 1}$ мы имеем дело с показательной плотностью, а плотности $f_{\alpha, n}$ ($n = 1, 2, \dots$) совпадают с плотностями g_n из гл. I, (3.4). Простые вычисления показывают, что математическое ожидание $f_{\alpha, \nu}$ равно ν/α , а дисперсия равна ν/α^2 .

Семейство гамма-плотностей замкнуто относительно операции свертки

$$f_{\alpha, \mu} * f_{\alpha, \nu} = f_{\alpha, \mu+\nu}, \quad \mu > 0, \nu > 0. \quad (2.3)$$

Это важное свойство обобщает теорему из гл. I, 3 и будет постоянно использоваться; доказательство чрезвычайно просто. Согласно (1.5), левая часть равна

$$\frac{\alpha^{\mu+\nu}}{\Gamma(\mu) \Gamma(\nu)} e^{-\alpha x} \int_0^x (x-y)^{\mu-1} y^{\nu-1} dy. \quad (2.4)$$

После подстановки $y = xt$ это выражение будет отличаться от $f_{\alpha, \mu+\nu}$ только численным множителем, а он равен единице, так как $f_{\alpha, \mu+\nu}$ и (2.4) являются плотностями вероятности.

Последний интеграл при $x=1$ — это так называемая бета-функция $B(\mu, \nu)$, и как побочный результат доказательства мы получаем, что

$$B(\mu, \nu) = \int_0^1 (1-y)^{\mu-1} y^{\nu-1} dy = \frac{\Gamma(\mu) \Gamma(\nu)}{\Gamma(\mu+\nu)} \quad (2.5)$$

для всех $\mu > 0$, $\nu > 0$. [С целыми μ и ν эта формула используется в 1; гл. VI, (10.8) и (10.9). См. также задачу 3 этой главы.]

Что касается *графика* $f_{1,\nu}$, то он, очевидно, монотонен, если $\nu \leq 1$, и неограничен вблизи нуля, когда $\nu < 1$. При $\nu > 1$ график $f_{1,\nu}$ колоколообразен, $f_{1,\nu}$ достигает при $x = \nu - 1$ максимума, равного $(\nu - 1)^{\nu-1} e^{-(\nu-1)} / \Gamma(\nu)$, что близко к $[2\pi(\nu - 1)]^{-1/2}$ [формула Стирлинга, 1; гл. II, 12, задача 12]. Из центральной предельной теоремы следует, что

$$\frac{\sqrt{\nu}}{\alpha} f_{1,\nu} \left(\frac{\nu}{\alpha} + x \frac{\sqrt{\nu}}{\alpha} \right) \rightarrow \pi(x), \quad \nu \rightarrow \infty. \quad (2.6)$$

§ 3*. РАСПРЕДЕЛЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ, СВЯЗАННЫЕ С ГАММА-РАСПРЕДЕЛЕНИЕМ

Гамма-плотности играют значительную, хотя не всегда явную роль в математической статистике. Прежде всего они фигурируют как «тип III» в классической (теперь до некоторой степени устаревшей) системе плотностей, введенной К. Пирсоном (1894 г.). Более часто они встречаются благодаря тому, что для случайной величины X с нормальной плотностью π квадрат X^2 имеет плотность $x^{-1/2} \pi(x^{1/2}) = f_{1/2, 1/2}(x)$. Ввиду свойства свертки (2.3) отсюда следует, что

Если $X_1, \dots, X_n$ — взаимно независимые нормально распределенные величины с математическим ожиданием 0 и дисперсией σ^2 , то величина $X_1^2 + \dots + X_n^2$ имеет плотность $f_{1/2\sigma^2, n/2}$.

Для статистиков величина $\chi^2 = X_1^2 + \dots + X_n^2$ совпадает с «выборочной дисперсией для нормальной совокупности», и ее распределение постоянно используется. В этой связи $f_{1/2, 1/2n}$ по традиции (восходящей к К. Пирсону) называется *хи-квадрат плотностью с n степенями свободы*.

В статистической механике величина $X_1^2 + X_2^2 + X_3^2$ появляется как квадрат скорости частиц. Сама скорость поэтому имеет плотность

$$v(x) = 2x f_{1/2, 3/2}(x^2).$$

Это не что иное, как *плотность Максвелла*, найденная другим способом в гл. I, (10.6). (См. также пример в гл. III, 4.)

В теории очередей гамма-распределение иногда называют *распределением Эрланга*.

Некоторые случайные величины (или «статистики»), важные для статистиков, имеют форму $T = X/Y$, где X и Y — независимые случайные величины, $Y > 0$. Обозначим их функции распределения через F и G , а их плотности через f и g соответственно. Так как величина Y предполагается положительной, то g сосре-

*! Этот параграф затрагивает специальные вопросы и не используется в дальнейшем.

доточено на $\overline{0, \infty}$, и поэтому

$$P\{T \leq t\} = P\{X \leq tY\} = \int_0^{\infty} F(ty) g(y) dy. \quad (3.1)$$

Дифференцируя, находим, что отношение $T = X/Y$ обладает плотностью

$$w(t) = \int_0^{\infty} f(ty) yg(y) dy. \quad (3.2)$$

Примеры. а) Если X и Y имеют плотности $f_{1/2, m/2}$ и $f_{1/2, n/2}$, то X/Y имеет плотность

$$w(t) = \frac{\Gamma\left(\frac{1}{2}(m+n)\right) t^{\frac{1}{2}m-1}}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)(1+t)^{\frac{1}{2}(m+n)}}, \quad t > 0. \quad (3.3)$$

Действительно, интеграл (3.2) равен

$$\frac{t^{\frac{1}{2}m-1}}{2^{\frac{1}{2}(m+n)}\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \int_0^{\infty} y^{\frac{1}{2}(m+n)-1} e^{-\frac{1}{2}(1+t)y} dy, \quad (3.4)$$

а подстановкой $\frac{1}{2}(1+t)y = s$ он сводится к (3.3).

В дисперсионном анализе рассматривают частный случай $X = X_1^2 + \dots + X_m^2$ и $Y = Y_1^2 + \dots + Y_n^2$, где $X_1, \dots, X_m, Y_1, \dots, Y_n$ — взаимно независимые нормально распределенные величины с плотностью n . Случайная величина $F = \frac{n}{m} \frac{X}{Y}$ называется статистикой Снедекора, а ее плотность $\frac{m}{n} w\left(\frac{m}{n} x\right)$ — плотностью Снедекора или F -плотностью. Величина $Z = \frac{1}{2} \log F$ есть Z -статистика Фишера, а ее плотность в свою очередь — Z -плотность Фишера. Обе статистики являются, конечно, вариантами записи одной и той же величины.

б) T -плотность Стьюдента. Пусть $X, Y_1, \dots, Y_n$ независимы и одинаково распределены с нормальной плотностью n . Случайная величина

$$T = \frac{X \sqrt{n}}{\sqrt{Y_1^2 + \dots + Y_n^2}} \quad (3.5)$$

называется в математической статистике T -статистикой Стьюдента. Покажем, что распределение случайной величины T задается

плотностью

$$w(t) = \frac{C_n}{(1+t^2/n)^{\frac{1}{2}(n+1)}}, \text{ где } C_n = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{1}{2}n\right)}. \quad (3.6)$$

В самом деле, числитель в (3.5) имеет нормальную плотность с нулевым математическим ожиданием и дисперсией n , тогда как плотность знаменателя равна $2xf_{1/2, n/2}(x^2)$. Таким образом, используя формулу (3.2), получаем плотность вида

$$\frac{1}{\sqrt{\pi n} 2^{(n-1)/2} \Gamma(n/2)} \int_0^\infty e^{-\frac{1}{2}(1+t^2/n)y^2} y^n dy. \quad (3.7)$$

Подстановка $s = \frac{1}{2}(1+t^2/n)y^2$ сводит интеграл (3.7) к гамма-функции, и мы получаем (3.6).

§ 4. НЕКОТОРЫЕ РАСПРОСТРАНЕННЫЕ ПЛОТНОСТИ

а) *Двусторонняя показательная плотность* определяется выражением $\frac{1}{2}ae^{-\alpha|x|}$, где α — масштабный параметр. Ей соответствует нулевое математическое ожидание и дисперсия $2\alpha^{-2}$. Эта плотность представляет собой *свертку показательной плотности* $ae^{-\alpha x}$ ($x > 0$) со своим зеркальным отображением $ae^{\alpha x}$ ($x < 0$). Иными словами, двусторонняя показательная плотность является плотностью величины $X_1 - X_2$, где X_1 и X_2 независимы и обладают одной и той же показательной плотностью $ae^{-\alpha x}$ ($x > 0$). Во французской литературе это распределение обычно называют «вторым законом распределения Лапласа» (первым считается нормальное распределение).

б) *Равномерная (или прямоугольная) плотность* ρ_a и *треугольная плотность* τ_a , сосредоточенные на $-a, a$, определяются соответственно выражениями

$$\rho_a(x) = \frac{1}{2a}, \quad \tau_a(x) = \frac{1}{a} \left(1 - \frac{|x|}{a}\right), \quad |x| < a. \quad (4.1)$$

Легко видеть, что $\rho_a * \rho_a = \tau_{2a}$, или словами сумма двух равномерно распределенных на $-a, a$ величин имеет треугольную плотность распределения на $-2a, 2a$. [Повторные свертки $\rho_a * \dots * \rho_a$ были описаны в гл. I, (9.7).]

в) *Бета-плотности на $0, 1$* определяются как

$$\beta_{\mu, \nu}(x) = \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\Gamma(\nu)} (1-x)^{\mu-1} x^{\nu-1}, \quad 0 < x < 1, \quad (4.2)$$

где $\mu > 0$ и $\nu > 0$ — свободные параметры. То, что (4.2) действительно определяет плотность вероятности, следует из (2.5). Из той

же самой формулы видно, что $\beta_{\mu, \nu}$ имеет математическое ожидание $\nu/(\mu + \nu)$ и дисперсию $\mu\nu/[(\mu + \nu)^2(\mu + \nu + 1)]$. Если $\mu < 1$, $\nu < 1$, то график $\beta_{\mu, \nu}$ является U-образным, уходящим в бесконечность при x , стремящемся к 0 или 1. При $\mu > 1$, $\nu > 1$ график колоколообразен. При $\mu = \nu = 1$ получаем как частный случай равномерную плотность.

Простой вариант бета-плотности определяется формулой

$$\frac{1}{(1+t)^2} \beta_{\mu, \nu} \left(\frac{1}{1+t} \right) = \frac{\Gamma(\mu + \nu)}{\Gamma(\mu) \Gamma(\nu)} \frac{t^{\mu-1}}{(1+t)^{\mu+\nu}}, \quad 0 < t < \infty. \quad (4.3)$$

Если случайная величина X имеет плотность (4.2), то величина $Y = X^{-1} - 1$ имеет плотность (4.3).

В системе распределений Пирсона плотности (4.2) и (4.3) именуются плотностями типа I и VI соответственно. Плотность Снедекора (3.3) есть не что иное, как частный случай плотности (4.3). Плотности (4.3) иногда называются *плотностями Парето* (в честь экономиста Парето). Считалось (несколько наивно с современной статистической точки зрения), что распределения дохода должны иметь хвост с плотностью $\sim Ax^{-2}$ при $x \rightarrow \infty$, а (4.3) удовлетворяет этому требованию.

г) Так называемая «арксинус-плотность»

$$\frac{1}{\pi \sqrt{x(1-x)}}, \quad 0 < x < 1, \quad (4.4)$$

есть фактически то же самое, что и бета-плотность $\beta_{1/2, 1/2}$, однако эта плотность заслуживает специального упоминания вследствие ее частого появления в теории флуктуаций. (Мы ввели ее в I, гл. III, 4, в связи с неожиданным поведением «времени пребывания».) Вводящее в заблуждение название, к сожалению, общепотребительно; фактически функция распределения равна $2\pi^{-1} \arcsin \sqrt{x}$. (Бета-плотности с $\mu + \nu = 1$ иногда называются «обобщенными арксинус-плотностями».)

д) *Плотность Коши* с центром в начале координат определяется как

$$\gamma_t(x) = \frac{1}{\pi} \cdot \frac{t}{t^2 + x^2}, \quad -\infty < x < \infty, \quad (4.5)$$

где $t > 0$ — масштабный параметр. Соответствующая функция распределения равна $\frac{1}{\pi} + \pi^{-1} \operatorname{arctg}(x/t)$. График γ_t напоминает график нормальной плотности, но приближается к оси x столь медленно, что *математического ожидания не существует*.

Важность плотностей Коши обусловлена формулой свертки

$$\gamma_s * \gamma_t = \gamma_{s+t}. \quad (4.6)$$

Этим утверждается, что *семейство плотностей Коши (4.5) замкнуто относительно свертки*. Формула (4.6) может быть доказана

элементарным (но скучным) способом с помощью обычного разложения подынтегральной функции на элементарные дроби. Более простое доказательство опирается на анализ Фурье.

Формула свертки (4.6) имеет весьма интересное следствие, а именно: для независимых случайных величин $X_1, \dots, X_n$, имеющих одинаковую плотность распределения (4.5), *среднее значение $(X_1 + \dots + X_n)/n$ распределено с той же самой плотностью, что и X_j .*

Пример. Рассмотрим лабораторный опыт, заключающийся в том, что вертикальное зеркало проецирует горизонтальный световой луч на стену. Зеркало может свободно вращаться относительно вертикальной оси, проходящей через A . Мы предполагаем, что направление отраженного луча выбирается «случайно», т. е. угол ϕ между этим лучом и перпендикуляром AO к стене распределен равномерно между $-\frac{1}{2}\pi$ и $\frac{1}{2}\pi$. Световой луч пересекает стену в точке, которая лежит от точки O на расстоянии

$$X = t \cdot \operatorname{tg} \phi$$

(t равно расстоянию AO центра зеркала A от стены). Теперь очевидно, что случайная величина X имеет плотность (4.5)¹⁾. Если эксперимент повторен n раз, то среднее $(X_1 + \dots + X_n)/n$ имеет ту же самую плотность распределения и, таким образом, средние не накапливаются вокруг нуля, как можно было бы предположить по аналогии с законом больших чисел. ►

Плотность распределения Коши обладает любопытным свойством: если X имеет плотность γ_t , то $2X$ имеет плотность $\gamma_{2t} = \gamma_t * \gamma_t$. Таким образом, величина $2X = X + X$ является суммой двух независимых величин, однако ее плотность задается формулой свертки. Более того, если U и V — две независимые величины с одинаковой плотностью γ_t и $X = aU + bV$, $Y = cU + dV$, то величина $X + Y$ имеет плотность $\gamma_{(a+b+c+d)t}$, что представляет собой свертку плотностей $\gamma_{(a+b)t}$ величины X и $\gamma_{(c+d)t}$ величины Y ; однако X и Y не являются независимыми. (Сложный пример см. в задаче 1 в гл. III, 9.)

[Плотность Коши соответствует частному случаю $n=1$ семейства (3.5) T -плотностей Стюдента. Иными словами, если X и Y — независимые случайные величины с нормальной плотностью π , то величина $X/|Y|$ имеет плотность распределения Коши (4.5) *с $t=1$* . Некоторые другие родственные плотности см. в задачах 5, 6.]

Свойство свертки гамма-плотностей, выраженное формулой (2.3), выглядит точным подобием свойства (4.6), однако существует важное различие, состоящее в том, что для гамма-плотностей параметр v является существенным, в то время как формула (4.6) содержит

¹⁾ Простая переформулировка этого эксперимента приводит к физической интерпретации формулы свертки (4.6). Если рассуждение показывает, что если единственный источник света находится в начале отсчета, то γ_t представляет распределение интенсивности света вдоль линии $y=t$ на (x, y) -плоскости. Тогда (4.6) выражает принцип Гюйгенса, согласно которому интенсивность света вдоль направления $y=s+t$ такая же, как если бы источник был распределен вдоль линии $y=t$ с плотностью γ_t . (Я обязан этим замечанием Дж. В. Уолшу.)

только масштабный параметр. Тип распределений Коши является *устойчивым типом*. Эта устойчивость относительно операций свертки свойственна плотностям Коши и нормальным плотностям; различие заключается в том, что масштабные параметры определяются согласно правилам $t = t_1 + t_2$ и $\sigma^2 = \sigma_1^2 + \sigma_2^2$ соответственно. Имеются другие *устойчивые плотности* с подобными свойствами, и после систематизации терминологии мы будем называть нормальные плотности и плотности Коши «симметричными, устойчивыми с показателями 2 и 1» (см. гл. VI, 1).

е) *Одностороннее устойчивое распределение с показателем $1/2$* . Если $\mathfrak{N}$ — нормальное распределение, заданное формулой (1.6), то равенство

$$F_a(x) = 2 \left[1 - \mathfrak{N} \left(\frac{a}{\sqrt{x}} \right) \right], \quad x > 0, \quad (4.7)$$

определяет функцию распределения с плотностью

$$f_a(x) = \frac{a}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{x^3}} e^{-\frac{1}{2} a^2/x}, \quad x > 0. \quad (4.8)$$

Очевидно, что математическое ожидание не существует. Это распределение было найдено в 1; гл. III, (7,7), и снова в 1; гл. X, 1, как предел распределения времен возвращения. Отсюда легко вывести *правило свертки*

$$f_\alpha * f_\beta = f_\gamma, \quad \text{где } \gamma = \alpha + \beta. \quad (4.9)$$

(Возможна проверка элементарным, но несколько громоздким способом интегрирования. Доказательство, основанное на анализе Фурье, проще.) Если $X_1, \dots, X_n$ — независимые случайные величины с распределением (4.7), то из равенства (4.9) следует, что величина $(X_1 + \dots + X_n)n^{-\alpha}$ имеет то же самое распределение, и поэтому средние $(X_1 + \dots + X_n)n^{-1}$, вероятно, должны быть порядка n ; вместо того чтобы сходиться, они безгранично возрастают (см. задачи 7 и 8).

ж) *Распределения вида $e^{-x^{-\alpha}}$ ($x > 0, \alpha > 0$)* появляются в связи с порядковыми статистиками (см. задачу 8). Вместе с разнотипностью $1 - e^{-x^{-\alpha}}$ они появляются (несколько таинственно) под именем распределений *Вейбула* в статистической теории надежности.

з) *Логистическая функция распределения*

$$F(t) = \frac{1}{1 + e^{-\alpha t - \beta}}, \quad \alpha > 0, \quad (4.10)$$

может служить предостережением. Существует невероятно большая литература, где делаются попытки доказать трансцендентный «закон логистического развития». Предполагалось возможным представить практически все процессы развития (измеренные в соответствующих единицах) функцией типа (4.10) с t , изображающим время. Весьма длинные таблицы с хи-квадрат критериями подтверждали это положение для человеческих популяций, бактериальных колоний, развития железных дорог и т. д. Было обнаружено, что как высота,

так и вес растений и животных подчиняются логическому закону, хотя из теоретических соображений ясно, что эти две величины не могут подчиниться одному и тому же распределению. Лабораторные эксперименты на бактериях показали, что даже систематические нарушения не могут привести к другим результатам. Теория популяций была основана на логистических экстраполяциях (даже если они оказывались очевидным образом ненадежными). Единственное затруднение «логистической» теории заключается в том, что не только логистическое распределение, но также нормальное, Коши и другие распределения могут быть подогаданы под тот же самый материал в том же или лучшим согласии¹⁾. В этой конкуренции логистическое распределение не играет никакой выдающейся роли: самые противоречивые теоретические модели могут быть подтверждены на том же материале наблюдений.

Теории этого рода недолговечны, так как они не открывают новые пути, а новые подтверждения одних и тех же старых вещей очень скоро становятся надоедливыми. Однако наивное рассуждение само по себе не было заменено здравым смыслом, и поэтому может быть полезно иметь очевидное доказательство того, как могут вводить в заблуждение взятые сами по себе критерии согласия.

§ 5. РАНДОМИЗАЦИЯ И СМЕСИ

Пусть F — функция распределения, зависящая от параметра θ , и u — некоторая плотность вероятности. Тогда

$$W(x) = \int_{-\infty}^{+\infty} F(x, \theta) u(\theta) d\theta \quad (5.1)$$

есть монотонная функция x , возрастающая от 0 до 1, и, следовательно, функция распределения. Если F имеет непрерывную плотность f , то и W имеет плотность w , равную

$$w(x) = \int_{-\infty}^{+\infty} f(x, \theta) u(\theta) d\theta. \quad (5.2)$$

Вместо интегрирования относительно плотности u мы можем суммировать по отношению к дискретному распределению вероятностей: если $\theta_1, \theta_2, \dots$ выбраны произвольно и если $p_k \geq 0$, $\sum p_k = 1$, то равенство

$$w(x) = \sum_k f(x, \theta_k) p_k \quad (5.3)$$

определяет новую плотность вероятности. Процесса может быть описан вероятностно как *рандомизация*; параметр θ рассматривается как случайная величина, и новое распределение вероятностей определяется в (x, θ) -плоскости, которая служит выборочным пространством. Плотности вида (5.3) называются *смесями*, и это выражение здесь используется вообще для распределений и плотностей вида (5.1) и (5.2).

¹⁾ Feller W., On the logistic law of growth and its empirical verifications in biology, Acta Biotheoretica, 5 (1940), 51—66.

Мы не предполагаем на этом этапе развивать общую теорию. Наша цель скорее в том, чтобы проиллюстрировать на нескольких примерах возможности этого метода и его вероятностное содержание. Примеры служат также подготовительным материалом к понятию условных вероятностей. Следующий параграф содержит примеры дискретных распределений, полученных посредством рандомизации непрерывного параметра. Наконец, § 7 иллюстрирует построение непрерывных процессов на основе случайных блужданий; попутно мы получим распределения, встречающиеся во многих приложениях и требующие при другом подходе громоздких вычислений.

Примеры. а) *Отношения.* Если X — случайная величина с плотностью f , то при фиксированном $y > 0$ величина X/y имеет плотность $f(xy)y$. Рассматривая параметр y как случайную величину с плотностью g , мы получаем новую плотность

$$\omega(x) = \int_{-\infty}^{+\infty} f(xy)yg(y)dy. \quad (5.4)$$

Это то же самое, что и формула (3.2), которая была основой для обсуждений в § 3.

На вероятностном языке рандомизация знаменателя y в X/y означает рассмотрение случайной величины X/Y , и мы можем просто повторить вывод формулы (3.2) для плотности величины X/Y . В этом частном случае терминология есть дело вкуса.

б) *Суммы случайного числа слагаемых.* Пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины с одинаковой плотностью f . Сумма $S_n = X_1 + \dots + X_n$ имеет плотность f^{n*} , а именно n -кратную свертку f с собой (см. гл. I, 2). Число слагаемых n является параметром, который мы теперь рандомизируем с помощью распределения вероятностей $P\{N=n\} = p_n$. Плотность результирующей суммы S_N со случайным числом слагаемых N равна

$$\omega = \sum_1^{\infty} p_n f^{n*}. \quad (5.5)$$

Возьмем, например, в качестве $\{p_n\}$ геометрическое распределение $p_n = q r^{n-1}$, а в качестве f — показательную плотность. Тогда $f^{n*} = g_n$ дается формулой (2.2) и

$$\omega(x) = q a e^{-ax} \sum_{n=1}^{\infty} p^{n-1} \frac{(ax)^{n-1}}{(n-1)!} = q a e^{-ax}. \quad (5.6)$$

в) *Применение к теории очередей.* Рассмотрим один обслуживающий прибор с показательным распределением времени обслуживания (плотность $f(t) = \mu e^{-\mu t}$) и предположим, что поступающая нагрузка — это нагрузка пуассоновского типа, т. е. что интервалы

времени между вызовами независимы и имеют плотность распределения $\lambda e^{-\lambda t}$, $\lambda < \mu$. Данная модель описана в 1; гл. XVII, 7, б). Поступающие вызовы становятся в (возможно, пустую) «очередь» и обслуживаются в порядке поступления без перерывов.

Рассмотрим вызов, к моменту поступления которого в системе обслуживания уже имеется очередь из $n \geq 0$ других вызовов. Полное время, которое этот вызов проводит в системе обслуживания, равно сумме времен обслуживания n вызовов, стоящих в очереди, и времени обслуживания самого этого вызова. Полное время есть случайная величина с плотностью $f^{(n+1)}$ *. Мы видели в 1; XVII, (7.10), что в стационарном состоянии вероятность наличия в очереди ровно n вызовов равна $q r^n$, где $r = \lambda/\mu$. Если имеет место стационарное состояние, то *полное время, которое вызов проведет в системе обслуживания, есть случайная величина с плотностью распределения*

$$\sum_{n=0}^{\infty} q r^n f^{(n+1)} * (t) = q \mu e^{-\mu t} \sum_{n=0}^{\infty} (r \mu t)^n / n! = (\mu - \lambda) e^{-(\mu - \lambda)t}.$$

Поэтому $E(T) = 1/(\mu - \lambda)$. (См. также задачу 10.)

г) *Очереди на автобус*. Предполагается, что автобус появляется каждый час (в час, в два часа... и т. д.), но может запаздывать. Мы рассмотрим последовательные запаздывания X_k как независимые случайные величины с одним и тем же распределением F и плотностью f . Для простоты мы предполагаем, что $0 \leq X_k \leq 1$. Обозначим через T_x время ожидания для человека, прибывшего в момент $x < 1$ после полудня. Вероятность, что автобус, назначенный на полдень, уже ушел, равна $F(x)$. Ясно видно, что

$$P\{T_x \leq t\} =$$

$$\begin{cases} F(t+x) - F(x) & \text{при } 0 < t < 1-x, \\ 1 - F(x) + F(x)F(t+x-1) & \text{при } 1-x < t < 2-x, \end{cases} \quad (5.7)$$

и, конечно, $P\{T_x \leq t\} = 1$ для всех больших t . Соответствующая плотность равна

$$\begin{cases} f(t+x) & \text{при } 0 < t < 1-x, \\ F(x)f(t+x-1) & \text{при } 1-x < t < 2-x. \end{cases} \quad (5.8)$$

Здесь момент x прибытия является свободным параметром, и его естественно рандомизировать. Например, для человека, прибывающего «наугад», момент прибытия представляет собой случайную величину, распределенную равномерно на $[0, 1]$. Среднее время ожидания равно в данном случае $1/2 + \sigma^2$, где σ^2 есть дисперсия времени запаздывания. Другими словами, *среднее время ожидания является наименьшим тогда, когда автобусы идут точно по расписанию*, и возрастает вместе с дисперсией запаздывания. (См. задачи 11, 12.) ▶

§ 6. ДИСКРЕТНЫЕ РАСПРЕДЕЛЕНИЯ

Этот параграф посвящен беглому обзору некоторых результатов применения рандомизации к биномиальному и пуассоновскому распределениям.

Число S_n успехов в испытаниях Бернулли подчиняется распределению, зависящему от вероятности успеха p . Рассматривая p как случайную величину с плотностью u , мы приходим к новому распределению

$$P\{S_n = k\} = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} u(p) dp, \quad k=0, \dots, n. \quad (6.1)$$

Пример. а) Если $u(p)=1$, то интегрированием по частям можно показать, что распределение, заданное формулой (6.1), не зависит от k и сводится к дискретному *равномерному распределению вероятностей* $P\{S_n = k\} = (n+1)^{-1}$. Более разъясняющей является аргументация Байеса. Рассмотрим $n+1$ независимых величин $X_0, \dots, X_n$, распределенных равномерно между 0 и 1. Интеграл в выражении (6.1) (при $u=1$) равен вероятности того, что ровно k величин из $X_1, \dots, X_n$ будет меньше X_0 , или, другими словами, что при перечислении точек $X_0, \dots, X_n$ в порядке возрастания величина X_0 появляется на $(k+1)$ -м месте. Но ввиду симметрии все положения равновероятны, и поэтому интеграл равен $(n+1)^{-1}$. ►

На языке азартных игр (6.1) соответствует ситуации, когда несимметричная монета выбирается посредством случайного механизма, после чего с этой монетой (неизвестной структуры) производятся испытания. Игроку испытания не кажутся независимыми; действительно, если наблюдается длинная последовательность гербов, то правдоподобно, что для нашей монеты p близко к 1, и поэтому безопасно делать ставку на дальнейшее появление гербов. Два формальных примера иллюстрируют оценивание и предсказание в задачах такого рода.

Примеры. б) Дано, что n испытаний закончились k успехами (гипотеза H). Какова вероятность того, что $p < \alpha$? По определению условных вероятностей

$$P\{A|H\} = \frac{P\{AH\}}{P\{H\}} = \frac{\int_0^\alpha p^k (1-p)^{n-k} u(p) dp}{\int_0^1 p^k (1-p)^{n-k} u(p) dp}. \quad (6.2)$$

Этот способ оценивания использовался Байесом [при $u(p)=1$]. В рамках нашей модели (т. е. если мы действительно занимаемся смешанной популяцией монет с известной плотностью u) не воз-

никнет возражений против этого способа. Опасность в том, что этот способ постоянно использовался без оснований применительно к суждениям о «вероятностях причин», когда о рандомизации не было и речи; эта точка зрения широко обсуждалась в примере 2, д) в 1; гл. V, в связи с вычислением пресловутой вероятности того, что завтра солнце взойдет.

в) Вариант предыдущей задачи можно сформулировать следующим образом. Дано, что n испытаний закончились k успехами; какова вероятность того, что следующие m испытаний закончатся j успехами? То же самое рассуждение приводит к ответу

$$\frac{\binom{m}{j} \int_0^1 p^{j+k} (1-p)^{m+n-j-k} u(p) dp}{\int_0^1 p^k (1-p)^n u(p) dp}. \quad (6.3)$$

(См. задачу 13.)

Обращаясь к *пуассоновскому распределению*, мы будем интерпретировать его как распределение числа «прибытий» за интервал времени длительностью t . Математическое ожидание числа прибытий равно αt . Мы проиллюстрируем две существенно различные процедуры рандомизации.

Примеры. г) *Рандомизированное время.* Если длительность интервала времени является случайной величиной с плотностью u , то вероятность p_k ровно k прибытий равна

$$p_k = \int_0^{\infty} e^{-\alpha t} \frac{(\alpha t)^k}{k!} u(t) dt. \quad (6.4)$$

Например, если интервал времени распределен показательно, то вероятность $k=0, 1, \dots$ новых прибытий равна

$$p_k = \int_0^{\infty} e^{-(\alpha+\beta)t} \frac{(\alpha t)^k}{k!} \beta dt = \frac{\beta}{\alpha+\beta} \cdot \left(\frac{\alpha}{\alpha+\beta} \right)^k, \quad (6.5)$$

что представляет собой геометрическое распределение.

д) *Расслоение.* Допустим, что имеется несколько независимых источников для случайных прибытий, каждый источник имеет пуассоновский выход, но параметры различны. Например, *аварии* в какой-либо установке в продолжение фиксированного промежутка времени t могут быть представлены пуассоновскими величинами, но параметр будет меняться от установки к установке. Аналогично *телефонные вызовы* от отдельных абонентов могут быть пуассоновскими с математическим ожиданием числа вызовов, меняющимся от абонента к абоненту. В таких процессах параметр α появ-

ляется как случайная величина с плотностью u , а вероятность ровно n вызовов за время t равна

$$P_n(t) = \int_0^{\infty} e^{-\alpha t} \frac{(\alpha t)^n}{n!} u(\alpha) d\alpha. \quad (6.6)$$

В частном случае гамма-плотности $u = f_{\beta, \nu+1}$ мы получим

$$P_n(t) = \binom{n+\nu}{n} \left(\frac{\beta}{\beta+t} \right)^{\nu+1} \left(\frac{t}{\beta+t} \right)^n, \quad (6.7)$$

что является предельной формой *распределения Пуассона*, данного в 1; гл. V, 8, задача 24, и 1; гл. XVII, (10.2) (если $\beta = a^{-1}$, $\nu = a^{-1} - 1$). ▶

Замечание о ложном заражении. Курьезная и поучительная история приписывается распределению (6.7) и его двойственной природе.

Урновая модель Пуассона и процесс Пуассона, которые приводят к (6.7), являются моделями истинного заражения, где каждый случай эффективно увеличивает вероятность будущих случаев. Эта модель пользовалась большой популярностью. Для множества явлений эмпирически проверялось согласие с (6.7). При этом хороший результат воспринимался как *признак истинного заражения*.

По случайному совпадению то же самое распределение (6.7) было получено еще ранее (1920) М. Гринвудом и Дж. Юлом с тем замыслом, что хорошее согласие с ним *опровергает наличие заражения*. Их вывод приблизительно соответствует нашей модели расслоения, для которой начальным является предположение о том, что в основе лежит пуассоновский процесс, т. е. что полностью отсутствует последствие. Таким образом, мы имеем курьезный факт, что хорошее согласие наблюдений с одним и тем же распределением может быть истолковано двумя различными способами, диаметрально противоположными как по своей природе, так и по своим практическим последствиям. Это послужит предостережением против слишком поспешной интерпретации статистических данных.

Объяснение этому лежит в явлении *ложного заражения*, описанного в 1; гл. V, 2, г) и выше в связи с (6.1). В данной ситуации, получив за интервал времени длины s m прибытий, можно оценить вероятность n прибытий за следующий интервал времени t по формуле, аналогичной (6.3). Результат будет зависеть от m , и эта зависимость обусловлена способом выбора, а не природой данных; информация, относящаяся к прошлому, дает нам возможность делать лучшие предсказания относительно будущего поведения нашей выборки, но это не следует смешивать с будущим всей популяции.

§ 7. БЕССЕЛЕВЫ ФУНКЦИИ И СЛУЧАЙНЫЕ БЛУЖДЕНИЯ

Удивительно много явных решений задач теории диффузии, теории очередей и других приложений содержат бесселевы функции. Обычно далеко не очевидно, что решения представляют собой распределения вероятностей, а аналитическая теория, которая требуется для получения их преобразований Лапласа и других характеристик, довольно сложна. К счастью, рассматриваемые распределения (и многие другие) могут быть получены простой процедурой рандомизации. На этом пути многие соотно-

шения теряют свой случайный характер, и можно избежать весьма трудного анализа.

Под *бесселевой функцией порядка* $\rho > -1$ мы будем понимать функцию I_ρ :

$$I_\rho(x) = \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(k + \rho + 1)} \left(\frac{x}{2}\right)^{2k + \rho}, \quad (7.1)$$

определенную для всех вещественных x^1 .

Мы приступаем к описанию трех процедур, ведущих к трем различным типам распределений, содержащих бесселевы функции.

а) Рандомизированные гамма-плотности

Пусть мы имеем при фиксированном $\rho > -1$ гамма-плотность $f_{1, \rho+k+1}$, заданную формулой (2.2). Рассматривая параметр k как целочисленную случайную величину, подчиненную пуассоновскому распределению, мы получаем в соответствии с (5.3) новую плотность

$$\omega_\rho(x) = e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} f_{1, \rho+k+1}(x) = e^{-t-x} \sum_{k=0}^{\infty} \frac{t^k x^{\rho+k}}{k! \Gamma(\rho+k+1)}. \quad (7.2)$$

Сравнив выражения (7.1) и (7.2), мы можем записать

$$\omega_\rho(x) = e^{-t-x} \sqrt{\left(\frac{x}{t}\right)^\rho} I_\rho(2\sqrt{tx}), \quad x > 0. \quad (7.3)$$

Если $\rho > -1$, то ω_ρ является *плотностью вероятности*, сосредоточенной на $0, \infty$ (при $\rho = -1$ правая часть не интегрируема по x). Заметим, что t не масштабный параметр, так что эти плотности принадлежат различным типам.

Между прочим, из этого построения и формулы свертки (2.3) для гамма-плотностей следует, что

$$\omega_\rho * f_{1, \nu} = \omega_{\rho+\nu}. \quad (7.4)$$

б) Рандомизированные случайные блуждания

При анализе случайных блужданий обычно считают, что последовательные переходы совершаются в моменты времени $1, 2, \dots$. Должно быть, однако, ясно, что это соглашение просто придает гармоничность описанию и что модель полностью независима от

¹⁾ Согласно обычному словопотреблению, I_ρ — это «модифицированная» бесселева функция или бесселева функция «с минимальным аргументом». «Обыкновенная» бесселева функция, всегда обозначаемая J_ρ , определяется добавлением множителя $(-1)^k$ в правую часть (7.1). Наше использование названия бесселевой функции должно рассматриваться как сокращение, а не как нововведение.

времени. Настоящие случайные процессы с непрерывным временем получаются из обычных случайных блужданий в предположении, что интервалы времени между последовательными скачками являются независимыми случайными величинами с одной и той же плотностью e^{-t} . Другими словами, моменты скачков регулируются пуассоновским процессом, а скачки сами по себе являются случайными величинами, принимающими значения $+1$ и -1 с вероятностями p и q , независимо друг от друга и от пуассоновского процесса.

Каждому распределению, связанному со случайным блужданием, соответствует некоторое распределение для процесса с непрерывным временем, которое формально получается рандомизацией числа скачков. Чтобы увидеть процедуру в деталях, рассмотрим состояние процесса в заданный момент t . В исходном случайном блуждании n -й шаг приводит к состоянию $r \geq 0$ тогда и только тогда, когда среди первых n скачков $\frac{1}{2}(n+r)$ положительны и $\frac{1}{2}(n-r)$ отрицательны. Это возможно, если только $n-r=2v$ чётно. В этом случае вероятность состояния r точно после n -го скачка равна

$$\binom{n}{\frac{1}{2}(n+r)} p^{\frac{1}{2}(n+r)} q^{\frac{1}{2}(n-r)} = \binom{r+2v}{r+v} p^{r+v} q^v. \quad (7.5)$$

Для нашего пуассоновского процесса вероятность того, что вплоть до момента t произойдет ровно $n=2v+r$ скачков, равна $e^{-t} t^n / n!$, и поэтому для нашего процесса с непрерывным временем вероятность состояния $r \geq 0$ в момент t равна

$$e^{-t} \sum_{v=0}^{\infty} \frac{t^{r+2v}}{(r+2v)!} \binom{r+2v}{r+v} p^{r+v} q^v = \sqrt{\left(\frac{p}{q}\right)^r} e^{-t} I_r(2\sqrt{pq}t), \quad (7.6)$$

и мы приходим к двум заключениям.

(i) Если мы определим $I_{-r} = I_r$ при $r = 1, 2, 3, \dots$, тогда при фиксированных $t > 0$, p , q выражение

$$a_r(t) = \sqrt{\left(\frac{p}{q}\right)^r} e^{-t} I_r(2\sqrt{pq}t), \quad r = 0, \pm 1, \pm 2, \dots \quad (7.7)$$

представляет распределение вероятностей (т. е. $a_r \geq 0$, $\sum a_r = 1$).

(ii) В нашем случайном блуждании с непрерывным временем $a_r(t)$ равно вероятности состояния r в момент t .

Две известные формулы для бesselевых функций являются непосредственными следствиями этого результата. Во-первых, если изменить обозначения, полагая $2\sqrt{pq}t = x$ и $p/q = u^2$, то тождество $\sum a_r(t) = 1$ превращается в

$$e^{\frac{1}{2}x(u+u^{-1})} = \sum_{-\infty}^{+\infty} u^r I_r(x). \quad (7.8)$$

Это так называемая производящая функция для бесселевых функций (формула Шлемилля, которая иногда служит в качестве определения для I_r).

Во-вторых, из природы нашего процесса ясно, что вероятности $a_r(t)$ должны удовлетворять уравнению Колмогорова—Чепмена

$$a_r(t+\tau) = \sum_{k=-\infty}^{\infty} a_k(t) a_{r-k}(\tau), \quad (7.9)$$

которое отражает тот факт, что в момент t частица должна быть в некотором состоянии k и что переход из k в r эквивалентен переходу из 0 в $r-k$. Мы возвратимся к этому соотношению в гл. XV11, 3. [Его легко проверить непосредственно, используя представление (7.6) и аналогичную формулу для вероятностей в случайном блуждании.] Уравнение Колмогорова—Чепмена (7.9) эквивалентно формуле

$$I_r(t+\tau) = \sum_{k=-\infty}^{\infty} I_k(t) I_{r-k}(\tau), \quad (7.10)$$

которая известна как *тождество К. Неймана*.

в) Первые прохождения

Для простоты сосредоточим наше внимание на симметричных случайных блужданиях $p=q=1/2$. Согласно 1, гл. III, (7.5), вероятность того, что *первое прохождение* через точку $r > 0$ произойдет во время скачка с номером $2n-r$, равна

$$\frac{r}{2n-r} \binom{2n-r}{n} 2^{-2n+r}, \quad n \geq r. \quad (7.11)$$

Случайное блуждание возвратно. Поэтому первое прохождение происходит с вероятностью единица, т. е. при фиксированном r величины (7.11) дают в сумме единицу. В нашем процессе момент k -го скачка имеет гамма-плотность $f_{1,k}$, заданную формулой (2.2). Отсюда следует, что момент первого прохождения через $r > 0$ распределен с плотностью

$$\begin{aligned} \sum_n \frac{r}{2n-r} \binom{2n-r}{n} 2^{-2n+r} f_{1,2n-r}(t) = \\ = e^{-t} \sum \frac{t^{2n-r-1}}{(2n-r-1)!} \left(\frac{r}{2n-r} \right) \cdot \frac{(2n-r)!}{n!(n-r)!} 2^{-2n+r} = e^{-t} \frac{r}{t} I_r(t). \end{aligned} \quad (7.12)$$

Таким образом, (i) при фиксированном $r=1, 2, \dots$ выражение

$$v_r(t) = e^{-t} \frac{r}{t} I_r(t) \quad (7.13)$$

определяет плотность вероятности, сосредоточенную на $0, \infty$.

(ii) Момент первого прохождения через $r > 0$ имеет плотность v_r . (См. задачу 15.)

Этот результат позволяет сделать другие интересные заключения. Первое прохождение через $r+\rho$ в момент t предполагает предварительное первое прохождение через r в некоторый момент

$s < t$. Вследствие независимости скачков в интервалах времени $0, s$ и s, t и свойства отсутствия последействия у показательного времени ожидания мы должны иметь

$$v_r * v_p = v_{r+p}. \quad (7.14)$$

[Проверка этого соотношения по формуле (7.12) будет легкой, если использовать соответствующее свойство свертки для вероятностей (7.11).]

Фактически утверждение (i) и соотношение (7.14) справедливы для всех положительных значений параметров r и p ¹⁾.

§ 8. РАСПРЕДЕЛЕНИЯ НА ОКРУЖНОСТИ

Полуоткрытый интервал $0, 1$ можно рассматривать как изображение окружности единичной длины, но предпочтительнее обернуть всю прямую вокруг окружности. Окружность тогда приобретает ориентацию, а длина дуги изменяется от $-\infty$ до ∞ , но $x, x \pm 1, x \pm 2, \dots$ интерпретируются как одна и та же точка. Сложение переходит по модулю 1 (так же как сложение углов по модулю 2π). Плотность распределения вероятностей на окружности есть периодическая функция $\varphi \geq 0$ такая, что

$$\int_0^1 \varphi(x) dx = 1. \quad (8.1)$$

Пример. а) *Задача Бюффона об игле* (1777). Традиционная формулировка такова. Плоскость разбивается на параллельные полосы единичной ширины. Произвольным образом бросается на плоскость игла единичной длины. Какова вероятность, что игла ляжет на плоскости так, что заденет две полосы? Чтобы поставить задачу формально, рассмотрим сначала центр иглы. Его положение определяется двумя координатами, но y мы пренебрегаем, а x приводим по модулю 1. Таким образом, «центр иглы» превращается в случайную величину X , равномерно распределенную на окружности. Направление иглы может быть описано углом (измеренным по часовой стрелке) между иглой и x -осью. Поворот на угол π возвращает иглу в то же самое положение, и поэтому угол определяется с точностью до умножения на π . Обозначим рассматриваемый угол через $Z\pi$. В задаче Бюффона подразумевается, что X и Z — независимые равномерно распределенные на окружности единичной длины величины²⁾.

Если мы выбираем значения X между 0 и 1, а значения Z

¹⁾ Feller W., Infinitely divisible distributions and Bessel functions associated with random walks. J. Soc. Indust. Appl. Math. (1966), 864—875.

²⁾ Выборочное пространство пары (X, Z) есть тор.

между $-\frac{1}{2}$ и $\frac{1}{2}$, то игла пересекает границу тогда и только тогда, когда $\frac{1}{2} \cos Z\pi > X$ или $\frac{1}{2} \cos Z\pi > 1 - X$. При данном значении z , $-\frac{1}{2} < z < \frac{1}{2}$, вероятность того, что $X < \frac{1}{2} \cos z\pi$, равна вероятности того, что $1 - X < \frac{1}{2} \cos z\pi$, а именно равна $\frac{1}{2} \cos z\pi$. Поэтому искомая вероятность равна

$$\int_{-1/2}^{1/2} \cos z\pi \cdot dz = \frac{2}{\pi}. \quad (8.2) \blacktriangleright$$

Случайная величина X на прямой может быть преобразована по модулю 1 в величину ${}^0X$ на окружности. Случайными величинами такого рода являются *ошибки округления* при численных подсчетах. Если X имеет плотность f , то плотность величины ${}^0X$ равна ¹⁾

$$\varphi(x) = \sum_{-\infty}^{+\infty} f(x+n). \quad (8.3)$$

Любая плотность на прямой, таким образом, индуцирует плотность вероятности на окружности. [В гл. XIX, 5, мы увидим, что та же самая функция φ допускает совершенно иное представление в виде рядов Фурье. Частный случай нормальных плотностей см. в примере 5, д) гл. XIX.]

Примеры. б) *Задача Пуанкаре о рулетке.* Рассмотрим число поворотов колеса рулетки как случайную величину X с плотностью f , сосредоточенной на положительной полуоси. Наблюдаемый результат, а именно точка ${}^0X$, против которой колесо останавливается, является величиной X , приведенной по модулю 1. Ее плотность дается формулой (8.3).

Естественно ожидать, что «при обычных обстоятельствах» плотность ${}^0X$ должна быть приблизительно равномерной. В 1912 г. А. Пуанкаре придал этой расплывчатой мысли форму предельной теоремы. Мы не будем повторять здесь его анализ, так как подобный результат легко следует из (8.3). Подразумевается, конечно, что данная плотность, по существу, «расплывается» на длинный интервал, так что ее *максимум мал*. Предположим для простоты, что f возрастает вплоть до точки a , где она достигает своего максимума $m = f(a)$, и что f убывает при $x > a$. Тогда

¹⁾ Читателям, которых беспокоит вопрос сходимости, следует рассматривать только плотности f , сосредоточенные на конечном интервале. Если f монотонна при достаточно больших x и $-x$, но, очевидно, имеет место равномерная сходимость. Если не накладывать на f никаких ограничений, то ряд может расходиться в некоторых точках, то φ всегда представляет собой плотность, так как частные суммы в (8.3) образуют *монотонную* последовательность функций, интегралы которых стремятся к 1 (см. гл. IV, 2).

для плотности φ случайной величины ^0X верно равенство

$$\varphi(x) - 1 = \sum_n f(x+n) - \int_{-\infty}^{+\infty} f(s) ds. \quad (8.4)$$

При фиксированном x обозначим через x_k единственную точку вида $x+n$, такую, что $a+k \leq x_k < a+k+1$. Тогда формула (8.4) может быть переписана в виде

$$\varphi(x) - 1 = \sum_{k=-\infty}^{\infty} \int_{x_k}^{x_{k+1}} [f(x_k) - f(s)] ds. \quad (8.5)$$

При $k < -1$ подынтегральная функция ≤ 0 , и поэтому

$$\varphi(x) - 1 \leq f(a) + \sum_{k=0}^{\infty} [f(x_k) - f(x_{k+1})] \leq 2f(a) = 2m.$$

Аналогичные аргументы доказывают, что $\varphi(x) - 1 \geq -m$. Таким образом, $|\varphi(x) - 1| < 2m$, и, следовательно, φ действительно приближенно постоянна.

Условия монотонности были наложены на плотность только ради простоты изложения и могут быть ослаблены многими способами. [Хорошие достаточные условия могут быть получены применением формулы суммирования Пуассона, гл. XIX, (5.2).]

в) *Распределение первых значащих цифр*. Известный ученый, прикладной математик, имел громадный успех в заключаемых им пари на то, что число, выбранное наугад в *Farmer's Almanac* или *Census Report* или в подобном компендиуме, будет иметь первую значащую цифру меньше 5. Наивно предполагается, что все девять цифр равновероятны; в этом случае вероятность того, что цифра не превосходит 4, будет равна 4/9. Практически¹⁾ она близка к 0,7.

Рассмотрим дискретное распределение вероятностей, приписывающее цифре k вероятность $p_k = \text{Log}(k+1) - \text{Log} k$ (где Log обозначает логарифм по основанию 10 и $k=1, \dots, 9$). Эти вероятности равны приближенно

$$\begin{array}{lll} p_1 = 0,3010 & p_2 = 0,1761 & p_3 = 0,1249 \\ p_4 = 0,0969 & p_5 = 0,0792 & p_6 = 0,0669 \\ p_7 = 0,0580 & p_8 = 0,0512 & p_9 = 0,0458 \end{array}$$

Распределение $\{p_k\}$ заметно отличается от равномерного распределения с весами $\frac{1}{9} = 0,111\dots$.

¹⁾ Эмпирические данные см. в статье: Benford F., The law of anomalous numbers, Proc. Amer. Philos. Soc., 78 (1938), 551—572.

Мы теперь покажем (следуя Р. С. Пинкхэму), что распределение $\{p_k\}$ правдоподобно как эмпирическое распределение первой значащей цифры числа, выбранного наугад из большой совокупности физических или наблюдаемых данных. Действительно, такое число может быть рассмотрено как случайная величина $Y > 0$ с некоторым неизвестным распределением. Первая значащая цифра Y равна k тогда и только тогда, когда $10^k \leq Y < 10^{k+1}$ при некотором n . Для случайной величины $X = \text{Log } Y$ это означает, что

$$n + \text{Log } k \leq X < n + \text{Log } (k+1). \quad (8.6)$$

Если размах Y очень велик, то распределение случайной величины ^0X , которая есть X , приведенная по модулю, является приближенно равномерным. Тогда вероятность неравенства (8.6) близка к $\text{Log } (k+1) - \text{Log } k = p_k$. ►

Формула свертки (1.5) и умозаключения, приводящие к ней, остаются справедливыми в случае, когда сложение производится по модулю 1. Таким образом, *свертка двух плотностей на окружности длины 1 есть плотность, определяемая формулой*

$$w(x) = \int_0^1 f_1(x-y) f_2(y) dy. \quad (8.7)$$

Если X_1 и X_2 — независимые величины с плотностями f_1 и f_2 , то $X_1 + X_2$ имеет плотность w . Так как эти плотности — периодические функции, *свертка равномерной плотности с любой другой плотностью является равномерной*. (См. задачу 16.)

§ 9. ЗАДАЧИ

1. Покажите, что нормальное приближение для биномиального распределения, выведенное в I, гл. VI, позволяет обосновать формулу свертки (1.7) для нормальных плотностей.

2. Используя подстановку $x = \frac{1}{2} y^2$, докажите, что

$$\Gamma\left(\frac{1}{2}\right) = \pi^{1/2}.$$

3. *Формула удвоения Лежандра*. Из формулы (2.5) при $\mu = \nu$ вывести, что

$$\Gamma(2\nu) = \frac{1}{\sqrt{\pi}} 2^{2\nu-1} \Gamma(\nu) \Gamma\left(\nu + \frac{1}{2}\right).$$

Указание. Используйте подстановку $4(y-y^2) = z$ в интервале $0 < y < 1/2$.

4. Пусть $g(x) = \frac{1}{2} e^{-1/x^2}$. Найдите свертки $g * g$ и $g * g * g$, а также g^{4*} .

5. Пусть X и Y — одинаково распределенные независимые случайные величины с плотностью Коши $y_1(x)$, определенной по (4.5). Докажите, что произведение XY имеет плотность $2\pi^{-2} \frac{\ln|x|}{x^2-1}$.

Указание. Для вычисления не требуется ничего, кроме соотношения

$$\frac{a-1}{(1+s)(a+s)} = \frac{1}{1+s} - \frac{1}{a+s}.$$

6. Пусть

$$f(x) = \frac{2}{\pi} \cdot \frac{1}{e^x + e^{-x}}.$$

Докажите тогда, что

$$f * f(x) = \frac{4}{\pi^2} \frac{x}{e^x - e^{-x}},$$

двумя способами: а) рассматривая величины $\ln |X|$ и $\ln |Y|$ в обстановке предыдущей задачи; б) непосредственно при помощи подстановки $e^{2y} = t$ и разложения на элементарные дроби. (См. задачу 8 в гл. XV, 9.)

7. Если X имеет нормальную плотность n , то очевидно, что X^{-2} имеет устойчивую плотность (4.8). Используя это, покажите, что если X и Y — независимые нормально распределенные величины с нулевым математическим ожиданием и дисперсиями σ_1^2 и σ_2^2 , то величина $Z = XY/\sqrt{X^2 + Y^2}$ нормально распределена с дисперсией σ_3^2 , такой, что $1/\sigma_3 = 1/\sigma_1 + 1/\sigma_2$ (Л. Шепп).

8. Пусть величины $X_1, \dots, X_n$ независимы и $X_{(n)}$ — наибольшая среди них. Покажите, что если X_i имеют

а) плотность Коши (4.5), то

$$P\{n^{-1}X_{(n)} \leq x\} \rightarrow e^{-1/(\pi x)}, \quad x > 0;$$

б) устойчивую плотность (4.8), то

$$P\{n^{-2}X_{(n)} \leq x\} \rightarrow e^{-\alpha\sqrt{2/(\pi x)}}, \quad x > 0.$$

9. Пусть величины X и Y независимы с плотностями f и g , сосредоточенными на $0, \infty$. Если $E(X) < \infty$, то отношение X/Y имеет конечное математическое ожидание тогда и только тогда, когда

$$\int_0^1 \frac{1}{y} g(y) dy < \infty.$$

10. В примере 5, в) найдите плотность распределения времени ожидания следующей разгрузки:

а) если в момент 0 система обслуживания свободна;

б) в условиях стационарного состояния.

11. В условиях примера 5, г) покажите, что

$$E(T_x) = F(x)(\mu + 1 - x) + \int_0^{1-x} tf(t+x) dt,$$

где μ есть математическое ожидание F . Используя это, проверьте утверждение относительно $E(T)$, когда x распределено равномерно.

12. В примере 5, г) найдите распределение времени ожидания, когда $f(t) = 1$ для $0 < t < 1$.

13. В примере 6, в) предположим, что u есть бета-плотность, определенная в (4.2). Выразите условную вероятность (6.3) через биномиальные коэффициенты.

14. Пусть X и Y независимы и подчинены распределению Пуассона с од-

ним и тем же параметром: $P\{X=n\}=e^{-t}t^n/n!$. Покажите, что

$$P\{X-Y=r\}=e^{-2t}I_{|r|}(2t), \quad r=0, \pm 1, \pm 2, \dots$$

[См. задачу 9 гл. V, 11.]

15. Результаты § 7, в) остаются справедливыми для несимметричных случайных блужданий при условии, что вероятность первого прохождения через $r > 0$ равна единице, т. е. при условии $p \geq q$. Покажите, что единственное изменение в (7.11) состоит тогда в том, что 2^{-2n+r} заменяется на $p^n q^{n-r}$, а вывод таков: при $p \geq q$ и $r=1, 2, \dots$

$$\sqrt{\left(\frac{p}{q}\right)^r} e^{-t} \frac{r}{t} I_r(2\sqrt{pq}t)$$

определяет плотность вероятности, сосредоточенную на $\overline{0, \infty}$.

16. Пусть X и Y — независимые случайные величины, а ${}^0X$ и ${}^0Y$ те же величины, приведенные по модулю 1. Покажите, что величина $X+Y$ приводится по модулю 1 к ${}^0X+{}^0Y$. Непосредственным вычислением проверьте соответствующую формулу для сверток.

По понятным причинам многомерные распределения встречаются реже, чем одномерные распределения, и материал этой главы почти не будет играть роли в последующих главах. С другой стороны, глава содержит важный материал, например известную характеризацию нормального распределения и методы, применяемые в теории случайных процессов. Истинная природа этих методов становится более понятной, если отделить их от усложненных задач, с которыми они иногда связываются.

§ 1. ПЛОТНОСТИ

В ходе дальнейшего изложения будет очевидно, что число измерений несущественно, и лишь ради типографских удобств мы будем иметь дело с декартовой плоскостью $\mathcal{R}^2$. Мы связываем с плоскостью фиксированную систему координат с координатными величинами X_1, X_2 (более удобное однобуквенное обозначение будет введено в § 5).

Неотрицательная интегрируемая функция f , заданная в $\mathcal{R}^2$ и такая, что ее интеграл равен единице, называется *плотностью вероятности* или просто плотностью. (Все плотности, встречающиеся в этой главе, кусочно-непрерывны, и поэтому их интегрируемость не требует разъяснений.) Плотность f приписывает области Ω вероятность

$$P(\Omega) = \iint_{\Omega} f(x_1, x_2) dx_1 dx_2 \quad (1.1)$$

при условии, конечно, что область Ω достаточно регулярна для того, чтобы интеграл существовал. Все такие вероятности однозначно определяются через вероятности, соответствующие прямоугольникам, параллельным осям, т. е.

$$P\{a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2\} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_1 dx_2 \quad (1.2)$$

при всех комбинациях $a_i < b_i$. Полагая $a_1 = a_2 = -\infty$, мы получим функцию распределения F , а именно

$$F(x_1, x_2) = P\{X_1 \leq x_1, X_2 \leq x_2\}. \quad (1.3)$$

Очевидно, что $F(b_1, x_2) - F(a_1, x_2)$ представляет собой вероятность, соответствующую полубесконечной полосе шириной $b_1 - a_1$, и так как прямоугольник в (1.2) выражается разностью двух таких полос, то вероятность (1.2) равна так называемой смешанной разности

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Отсюда следует, что по функции распределения F однозначно определяются все вероятности (1.1). Несмотря на формальную аналогию с прямой линией, понятие функции распределения на плоскости гораздо менее полезно, поэтому лучше сосредоточиться на задании вероятностей (1.1) в терминах плотности. Это задание отличается от совместного распределения вероятностей двух дискретных случайных величин (1, гл. IX, 1) в двух отношениях. Во-первых, суммирование заменено интегрированием, и, во-вторых, вероятности теперь приписываются лишь «достаточно регулярным областям», тогда как в дискретных выборочных пространствах всем множествам соответствовали вероятности. Так как в настоящей главе рассматриваются простые примеры, в которых это различие едва уловимо, понятия и термины дискретной теории переносятся на новый случай в понятной форме. Так же как и в предыдущих главах, мы используем поэтому вероятностный язык без всякой попытки привлечения общей теории (которая будет изложена в гл. V).

Из (1.3) очевидно, что¹⁾

$$P\{X_1 \leq x_1\} = F(x_1, \infty). \quad (1.4)$$

Таким образом, $F_1(x) = F(x, \infty)$ определяет функцию распределения величины X_1 , а ее плотность f_1 равна

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy. \quad (1.5)$$

Если желательно подчеркнуть связь между X_1 и парой (X_1, X_2) , мы говорим об F_1 как о *маргинальном* (частном) распределении²⁾ и об f_1 как о маргинальной (частной) плотности.

Математическое ожидание μ_1 и *дисперсия* σ_1^2 величины X_1 , если они существуют, даются следующими формулами:

$$\mu_1 = E(X_1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 f(x_1, x_2) dx_1 dx_2, \quad (1.6)$$

¹⁾ Здесь и в дальнейшем $U(\infty) = \lim_{x \rightarrow \infty} U(x)$ при $x \rightarrow \infty$, и использование символа $U(\infty)$ подразумевает существование предела.

²⁾ Другой принятый термин: *проекция на ось*.

и

$$\sigma_1^2 = \text{Var}(\mathbf{X}_1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 - \mu_1)^2 f(x_1, x_2) dx_1 dx_2. \quad (1.7)$$

По симметрии эти определения применимы также и к $\mathbf{X}_2$. Наконец, ковариация $\mathbf{X}_1$ и $\mathbf{X}_2$ равна

$$\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2) dx_1 dx_2. \quad (1.8)$$

Нормированные величины $\mathbf{X}_i \sigma_i^{-1}$ безразмерны, и их ковариация, а именно $\rho = \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) \sigma_1^{-1} \sigma_2^{-1}$, есть коэффициент корреляции $\mathbf{X}_1$ и $\mathbf{X}_2$ (см. 1, гл. IX, 8).

Случайная величина $\mathbf{U}$ — это функция координатных величин $\mathbf{X}_1$ и $\mathbf{X}_2$; здесь мы снова рассматриваем только такие функции, что вероятности $\mathbf{P}\{\mathbf{U} \leq t\}$ могут быть выражены интегралами вида (1.1). Таким образом, каждая случайная величина будет иметь единственную функцию распределения, каждая пара будет иметь совместное распределение и т. д.

Во многих ситуациях выгодно переименовать координатные величины, т. е. приписать величинам $\mathbf{Y}_1, \mathbf{Y}_2$ роль, ранее предназначенную для $\mathbf{X}_1, \mathbf{X}_2$. В простейшем случае величины $\mathbf{Y}_j$ определяются линейным преобразованием

$$\mathbf{X}_1 = a_{11}\mathbf{Y}_1 + a_{12}\mathbf{Y}_2, \quad \mathbf{X}_2 = a_{21}\mathbf{Y}_1 + a_{22}\mathbf{Y}_2 \quad (1.9)$$

с детерминантом $\Delta = a_{11}a_{22} - a_{12}a_{21} > 0$. Обычно преобразование вида (1.9) можно описать или как отображение одной плоскости на другую, или как изменение координат в той же самой плоскости. Производя замену переменных (1.9) в интеграле (1.1), мы получаем

$$\mathbf{P}\{\Omega\} = \int_{\Omega_*} \int f(a_{11}y_1 + a_{12}y_2, a_{21}y_1 + a_{22}y_2) \cdot \Delta dy_1 dy_2, \quad (1.10)$$

где область Ω_* содержит все точки (y_1, y_2) , образ которых (x_1, x_2) принадлежит Ω . Так как события $(\mathbf{X}_1, \mathbf{X}_2) \in \Omega$ и $(\mathbf{Y}_1, \mathbf{Y}_2) \in \Omega_*$ идентичны, ясно, что совместная плотность $(\mathbf{Y}_1, \mathbf{Y}_2)$ равна

$$g(y_1, y_2) = f(a_{11}y_1 + a_{12}y_2, a_{21}y_1 + a_{22}y_2) \cdot \Delta. \quad (1.11)$$

Все это в равной степени справедливо и при большем числе измерений.

Подобные аргументы применимы к более общим преобразованиям с той лишь разницей, что детерминант Δ заменяется якобианом. Мы будем использовать явно только переход к полярным координатам

$$\mathbf{X}_1 = R \cos \Theta, \quad \mathbf{X}_2 = R \sin \Theta, \quad (1.12)$$

где пара случайных величин (R, Θ) ограничена условиями $R \geq 0$, $-\pi < \Theta \leq \pi$. Плотность пары (R, Θ) равна

$$g(r, \theta) = f(r \cos \theta, r \sin \theta) r. \quad (1.13)$$

В пространстве трех измерений используются географические долгота φ и широта θ ($-\pi < \varphi \leq \pi$ и $-\frac{1}{2}\pi \leq \theta \leq \frac{1}{2}\pi$). Тогда координатные величины выражаются в полярной системе следующим образом:

$$X_1 = R \cos \Phi \cos \Theta, \quad X_2 = R \sin \Phi \cos \Theta, \quad X_3 = R \sin \Theta. \quad (1.14)$$

Их совместная плотность записывается как

$$g(r, \varphi, \theta) = f(r \cos \varphi \cos \theta, r \sin \varphi \cos \theta, r \sin \theta) r^2 \cos \theta. \quad (1.15)$$

При преобразовании (1.14) «плоскости» $\Theta = -\frac{1}{2}\pi$ и $\Theta = \frac{1}{2}\pi$ соответствуют полуосям оси x_3 , но эта вырожденность не играет роли, так как эти полуоси имеют нулевую вероятность. Аналогичное замечание можно сделать и в отношении начала координат для полярной системы на плоскости.

Примеры. а) *Независимые случайные величины.* В предыдущих главах мы рассматривали независимые случайные величины X_1 и X_2 с плотностями f_1 и f_2 . Это равносильно определению двумерной плотности равенством $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, где f_i — маргинальные плотности.

б) «Случайный выбор». Пусть Γ — ограниченная область; предположим для простоты, что Γ — выпуклая область. Обозначим площадь Γ через γ . Пусть теперь функция f равна γ^{-1} внутри Γ и 0 вне Γ . Тогда f является плотностью, и вероятность, соответствующая любой области $\Omega \subset \Gamma$, равна отношению площадей Ω и Γ . По очевидной аналогии с одномерным случаем мы будем говорить, что пара (X_1, X_2) *распределена равномерно в Γ* . Маргинальная плотность X_1 в точке с абсциссой x_1 равна ширине области Γ в этой точке в очевидном смысле этого слова. (См. задачу 1.)

в) *Равномерное распределение на сфере.* Единичная сфера Σ в пространстве трех измерений может быть описана при помощи географических долготы φ и широты θ равенствами

$$x_1 = \cos \varphi \cos \theta, \quad x_2 = \sin \varphi \cos \theta, \quad x_3 = \sin \theta. \quad (1.16)$$

Каждой паре (φ, θ) , такой, что $-\pi < \varphi \leq \pi$, $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$, соответствует здесь ровно одна точка на сфере, и каждую точку Σ , исключая полюса, можно получить таким путем. Исключительность полюсов не должна нас беспокоить, так как им соответствует нулевая вероятность. Область Ω на сфере определяется своим образом на (φ, θ) -плоскости, и, следовательно, площадь Ω равна интегралу от $\cos \theta d\varphi d\theta$ по этому образу [см. (1.15)]. Для

описания мысленного эксперимента, заключающегося в «случайном выборе точки на Σ », мы должны положить $4\pi P(\Omega) = \text{площадь } \Omega$. Это эквивалентно определению в (φ, θ) -плоскости плотности

$$g(\varphi, \theta) = \begin{cases} \frac{1}{4\pi} \cos \theta & \text{при } -\pi < \varphi \leq \pi, \quad |\theta| < \frac{1}{2}\pi, \\ 0 & \text{при других } \varphi \text{ и } \theta. \end{cases} \quad (1.17)$$

При таком определении координатные величины независимы и догота распределена равномерно на $-\pi, \pi$.

Идея перехода от сферы Σ к (φ, θ) -плоскости хорошо знакома нам по географическим картам и полезна для теории вероятностей. Заметим, однако, что координатные величины в значительной степени произвольны и их математические ожидания и дисперсии ничего не значат для первоначального мысленного эксперимента.

г) *Двумерная нормальная плотность.* Многомерные нормальные плотности будут введены более систематично в § 6. Оправдание преждевременному появлению двумерного случая в том, что тем самым обеспечивается более легкий подход к нему в дальнейшем. По очевидной аналогии с нормальной плотностью π из II, (1.6), можно записать плотность вида $ce^{-q(x_1, x_2)}$, где $q(x_1, x_2) = a_1 x_1^2 + 2bx_1 x_2 + a_2 x_2^2$. Нетрудно видеть, что функция e^{-q} будет интегрируемой тогда и только тогда, когда $a_1 a_2 - b^2 > 0$ и $a_1, a_2 > 0$. Для целей теории вероятностей предпочтительнее выразить коэффициенты a_i и b в терминах дисперсий и *определить двумерную нормальную плотность, центрированную в начале координат как*

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{x_1^2}{\sigma_1^2} - 2\rho \frac{x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2} \right) \right], \quad (1.18)$$

где $\sigma_1 > 0, \sigma_2 > 0$ и $-1 < \rho < 1$. Интегрирование по x_2 легко выполняется с помощью подстановки $t = \frac{x_2}{\sigma_2} - \rho \frac{x_1}{\sigma_1}$ (дополнение до полного квадрата). Видно, что f действительно представляет плотность в $\mathbb{R}^2$. Кроме того, становится очевидным то, что *маргинальные* распределения X_1 и X_2 снова нормальны¹⁾ и что $E(X_i) = 0, \text{Var}(X_i) = \sigma_i^2, \text{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$. Иными словами, ρ есть коэффициент *корреляции* X_1 и X_2 . Заменяя x_i на $x_i - c_i$

¹⁾ Вопреки широко распространенному мнению *существуют не являющиеся нормальными двумерные* плотности с нормальными маргинальными плотностями (два типа описаны в задачах 2, 3, еще два — в задачах 5 и 7 из гл. V, 12). Для того чтобы иметь дело с нормальными плотностями, статистики иногда вводят новые координатные величины $Y_1 = g_1(X_1), Y_2 = g_2(X_2)$, которые распределены нормально. При этом, увы, совместное распределение (Y_1, Y_2) не является нормальным.

в (1.18), мы приходим к нормальной плотности, центрированной в точке (c_1, c_2) .

Важно то, что линейное преобразование (1.9) переводит нормальное распределение в другое нормальное распределение. Это очевидно из определения и формулы (1.11). [Продолжение см. в примере 2, а).]

д) Симметричное распределение Коши в $\mathcal{R}^2$. Положим

$$u(x_1, x_2) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{(1+x_1^2+x_2^2)^3}}. \quad (1.19)$$

Чтобы удостовериться, что это есть плотность, заметим¹⁾, что

$$\int_{-\infty}^{+\infty} u(x_1, y) dy = \frac{1}{2\pi} \cdot \frac{1}{1+x_1^2} \cdot \frac{y}{\sqrt{1+x_1^2+y^2}} \Big|_{-\infty}^{+\infty} = \frac{1}{\pi} \cdot \frac{1}{1+x_1^2}. \quad (1.20)$$

Отсюда следует, что u является плотностью и что маргинальная плотность X_1 не что иное, как плотность Коши γ_1 из гл. II, (4.5). Очевидно, что величина X_1 не имеет математического ожидания.

При переходе к полярным координатам [как в (1.12)] видим, что плотность R не зависит от θ , и поэтому величины R и θ стохастически независимы. Следуя терминологии гл. I, 10, мы можем тогда сказать, что имеющая симметричное распределение Коши пара (X_1, X_2) изображает вектор со случайно выбранным направлением длины R , плотность которой равна $r\sqrt{(1+r^2)^{-3}}$, откуда $P\{R \leq r\} = 1 - \sqrt{(1+r^2)^{-1}}$. [Продолжение в примере 2, в).]

е) Симметричное распределение Коши в $\mathcal{R}^3$. Положим

$$v(x_1, x_2, x_3) = \frac{1}{\pi^2} \cdot \frac{1}{(1+x_1^2+x_2^2+x_3^2)^2}. \quad (1.21)$$

Легко видеть²⁾, что маргинальная плотность (X_1, X_2) является симметричной плотностью Коши u из (1.19). Маргинальная плотность X_1 есть поэтому плотность Коши γ_1 . (Продолжение в задаче 5.)

Хотя это и не используется явно в последующем, но мы должны отметить, что свертки можно определить так же, как и в одномерном случае. Рассмотрим две пары (X_1, X_2) и (Y_1, Y_2) с совместными плотностями f и g соответственно. Утверждение о том, что две пары независимы, означает, что мы берем четырехмерное пространство с четырьмя координатными величинами X_1, X_2, Y_1, Y_2 в качестве выборочного пространства и определяем в нем плотность произведением $f(x_1, x_2)g(y_1, y_2)$. Так же как

¹⁾ Подстановка $y = \sqrt{1+x_1^2} \operatorname{tg} t$ облегчает вычисление.

²⁾ Используя подстановку $z = \sqrt{1+x_1^2+x_2^2} \operatorname{tg} t$.

в $\mathcal{R}^1$, тогда легко видеть, что совместная плотность v суммы $(X_1 + Y_1, X_2 + Y_2)$ задается формулой свертки

$$v(z_1, z_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(z_1 - x_1, z_2 - x_2) g(x_1, x_2) dx_1 dx_2, \quad (1.22)$$

которая является очевидным аналогом формулы гл. I, (2.12). (См. задачи 15—17.)

§ 2. УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ

Предположим, что пара (X_1, X_2) имеет непрерывную плотность f и что маргинальная плотность f_1 величины X_1 строго положительна. Рассмотрим условную вероятность события $X_2 \leq \eta$ при условии, что $\xi < X_1 \leq \xi + h$, а именно

$$P\{X_2 \leq \eta \mid \xi < X_1 \leq \xi + h\} = \frac{\int_{\xi}^{\xi+h} dx \int_{-\infty}^{\eta} f(x, y) dy}{\int_{\xi}^{\xi+h} f_1(x) dx}. \quad (2.1)$$

Деля числитель и знаменатель на h , устанавливаем, что при $h \rightarrow 0$ правая часть стремится к

$$U_{\xi}(\eta) = \frac{1}{f_1(\xi)} \int_{-\infty}^{\eta} f(\xi, y) dy. \quad (2.2)$$

При фиксированном ξ это функция распределения по η с плотностью

$$u_{\xi}(\eta) = \frac{1}{f_1(\xi)} f(\xi, \eta). \quad (2.3)$$

Мы называем u_{ξ} *условной плотностью величины X_2 при условии, что $X_1 = \xi$* . *Условное математическое ожидание X_2 при условии, что $X_1 = \xi$, определяется как*

$$E(X_2 \mid X_1 = \xi) = \frac{1}{f_1(\xi)} \int_{-\infty}^{+\infty} y f(\xi, y) dy \quad (2.4)$$

в предположении, что интеграл сходится абсолютно. Если рассматривать ξ как переменную величину, то правая часть есть функция от ξ . В частности, отождествляя ξ с координатной величиной X_1 , получим случайную величину, называемую *регрессией X_2 на X_1* и обозначаемую $E(X_2 \mid X_1)$. Присутствие X_2 не должно затемнять тот факт, что эта случайная величина является функцией одной переменной X_1 [ее значения даются формулой (2.4)].

До сих пор мы предполагали, что $f_1(\xi) > 0$ для всех ξ . Выра-

жение (2.4) не имеет смысла при $f_1(\xi) = 0$, но множество таких точек имеет вероятность нуль, и мы условимся интерпретировать (2.4) как нуль во всех точках, где f_1 обращается в нуль. Тогда $E(X_2|X_1)$ определено всякий раз, когда плотность непрерывна. (В гл. V, 9—11, условные вероятности будут введены для произвольных распределений.)

Излишне говорить, что регрессия $E(X_2|X_1)$ величины X_1 на X_2 определяется подобным же образом. Кроме того, *условная дисперсия* $\text{Var}(X_2|X_1)$ определяется по очевидной аналогии с (2.4).

Эти определения переносятся на случаи большего числа измерений, за тем исключением, что плотность в $\mathcal{R}^3$ порождает три двумерные и три одномерные условные плотности. (См. задачу 6.)

Примеры. а) *Нормальная плотность.* Для плотности (1.18), очевидно, имеем

$$u_{\frac{1}{2}}(y) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}} \exp \left[-\frac{\left(y - \rho \frac{\sigma_2}{\sigma_1} \xi\right)^2}{2(1-\rho^2)\sigma_2^2} \right], \quad (2.5)$$

что представляет собой нормальную плотность с математическим ожиданием $\rho \frac{\sigma_2}{\sigma_1} \xi$ и дисперсией $(1-\rho^2)\sigma_2^2$. Таким образом,

$$E(X_2|X_1) = \rho \frac{\sigma_2}{\sigma_1} X_1, \quad \text{Var}(X_2|X_1) = (1-\rho^2)\sigma_2^2. \quad (2.6)$$

Одним из приятных свойств нормального распределения является то, что регрессии — *линейные функции*.

Возможно, самым ранним применением этих соотношений мы обязаны Гальтону, и одним из его примеров можно проиллюстрировать их *эмпирическое значение*. Предположим, что X_1 и X_2 характеризуют рост (измеренный по отклонению от среднего в дюймах) соответственно отцов и сыновей в некоторой популяции. Рост случайно выбранного сына есть тогда нормальная случайная величина с математическим ожиданием 0 и с дисперсией σ_2^2 . Однако в подпопуляции сыновей, отцы которых имеют фиксированный рост ξ , рост сыновей представляется нормальной величиной с математическим ожиданием $\rho \frac{\sigma_2}{\sigma_1} \xi$ и дисперсией $\sigma_2^2(1-\rho^2) < \sigma_2^2$. Таким образом, регрессия X_2 на X_1 показывает, как много статистической информации относительно X_2 содержится в наблюдении над X_1 .

б) Пусть случайные величины X_1 и X_2 независимы и равномерно распределены на $[0, 1]$. Обозначим через $X_{(1)}$ и $X_{(2)}$ соответственно наименьшую и наибольшую среди этих величин. Пара $(X_{(1)}, X_{(2)})$ имеет плотность, которая тождественно равна 2 внутри треугольной области $0 \leq x_1 \leq x_2 \leq 1$ и равна 0 во всех других точках плоскости. Интегрируя по переменной x_2 , выводим, что маргинальная плотность $X_{(1)}$ есть $2(1-x_1)$. Плотность условного

распределения $X_{(1)}$ при условии, что $X_{(1)} = x_1$, равна, следовательно, постоянной $1/(1-x_1)$ внутри интервала $x_1, 1$ и равна нулю вне этого интервала. Иными словами, при условии, что $X_{(1)}$ принимает значение x_1 , случайная величина $X_{(2)}$ равномерно распределена на $x_1, 1$.

в) *Распределение Коши в $\mathbb{R}^2$* . Для двумерной плотности (1.19) маргинальная плотность X_1 задается формулой (1.20), и поэтому условная плотность X_2 при данном X_1 равна

$$u_{\xi}(y) = \frac{1}{2} \cdot \frac{1 + \xi^2}{\sqrt{(1 + \xi^2 + y^2)^3}}. \quad (2.7)$$

Заметим, что u_{ξ} отличается от плотности $u_0(y)$ только масштабным множителем $\sqrt{1 + \xi^2}$, и поэтому все плотности u_{ξ} принадлежат одному и тому же типу. Условных математических ожиданий в этом примере не существует. (См. задачу 6.) ►

В терминах условных плотностей (2.3) функция распределения величины X_2 принимает вид

$$P\{X_2 \leq y\} = \int_{-\infty}^y \int_{-\infty}^{+\infty} u_{\xi}(\eta) \cdot f_1(\xi) d\xi d\eta. \quad (2.8)$$

Иными словами, распределение X_2 получается посредством *рандомизации* параметра ξ в условных плотностях u_{ξ} , и *поэтому каждое⁴⁾ распределение может быть представлено в виде смеси*. Несмотря на эту теоретическую универсальность, имеется большое различие в положении «ударения». В некоторых ситуациях, таких, как в примере а), *начальным* является двумерное распределение для (X_1, X_2) , а затем выводятся условные распределения, тогда как при истинной рандомизации условные вероятности u_x являются исходным понятием и плотность $f(x, y)$ фактически *определяется* как $u_x(y) f_1(x)$. (Эта процедура определения вероятностей в терминах условных вероятностей объяснялась элементарным путем в 1; гл. V, 2.)

§ 3. ВОЗВРАЩЕНИЕ К ПОКАЗАТЕЛЬНОМУ И РАВНОМЕРНОМУ РАСПРЕДЕЛЕНИЯМ

Задача этого параграфа в том, чтобы дать примеры-иллюстрации к предыдущим параграфам и в то же время дополнить теорию первой главы.

Примеры. а) *Характеристическое свойство показательного распределения*. Пусть X_1 и X_2 — две независимые случайные величины с плотностями f_1 и f_2 . Обозначим плотность их суммы $S =$

⁴⁾ Мы рассматривали до сих пор непрерывные плотности. Общий случай будет разобран в гл. V, 9. Понятие рандомизации обсуждалось в гл. II, 5.

$= X_1 + X_2$ через g . Пары (X_1, S) и (X_1, X_2) связаны линейным преобразованием $X_1 = X_1$, $X_2 = S - X_1$ с детерминантом 1, и совместная плотность пары (X_1, S) по формуле (1.11) равна $f_1(x) \times f_2(s-x)$. Интегрируя по всем x , мы получаем маргинальную плотность g суммы S . Условная плотность u_s величины X_1 при условии, что $S=s$, удовлетворяет соотношению

$$u_s(x) = \frac{f_1(x) f_2(s-x)}{g(s)}. \quad (3.1)$$

Для частного случая показательных плотностей $f_1(x) = f_2(x) = \alpha e^{-\alpha x}$ ($x > 0$) мы получаем $u_s(x) = s^{-1}$ при $0 < x < s$. Иными словами, при условии, что $X_1 + X_2 = s$, случайная величина X_2 распределена равномерно на интервале $\overline{0, s}$. Образно говоря, знание того, что $S=s$, не дает нам никакой информации о возможном положении случайной точки X_1 внутри интервала $\overline{0, s}$. Этот результат согласуется с представлением о полной случайности, свойственной показательному распределению. (Более сильный вариант содержится в примере г). См. также задачу 12.)

б) Случайные разбиения интервала. Пусть $X_1, \dots, X_n$ — n точек выбранных независимо и наудачу в (одномерном) интервале $\overline{0, 1}$. Как и прежде, мы обозначим через $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ случайные точки $X_1, \dots, X_n$, расположенные в возрастающем порядке. Эти точки делят интервал $\overline{0, 1}$ на $n+1$ подынтервалов, которые мы обозначим $I_1, I_2, \dots, I_{n+1}$, пронумеровав их слева направо, так что $X_{(j)}$ является правым концом интервала I_j . В первую очередь мы вычислим совместную плотность $(X_{(1)}, \dots, X_{(n)})$.

Выборочное пространство, соответствующее $(X_1, \dots, X_n)$, представляет n -мерный гиперкуб $\Gamma: 0 < x_k < 1$, а вероятности равны $(n$ -мерному) объему. Естественное выборочное пространство, где где $X_{(k)}$ — координатные величины, есть подмножество $\Omega \subset \Gamma$, состоящее из всех точек, таких, что $0 < x_1 \leq \dots \leq x_n < 1$. Объем Ω равен $1/n!$. Очевидно, что гиперкуб Γ содержит $n!$ конгруэнтных «копий» множества Ω , и в каждой из них упорядоченная строка $(X_{(1)}, \dots, X_{(n)})$ совпадает с фиксированной перестановкой величин $X_1, \dots, X_n$. (Внутри Γ , в частности, $X_{(k)} = X_k$.) Вероятность того, что $X_j = X_k$ для некоторой пары $j \neq k$, равна нулю, а только это событие вызывает перекрытия между различными «копиями». Отсюда следует, что для любого подмножества $A \subset \Omega$ вероятность того, что $(X_{(1)}, \dots, X_{(n)})$ принадлежит A , равна вероятности того, что $(X_1, \dots, X_n)$ принадлежит одной из $n!$ «копий» A , а эта вероятность в свою очередь равна умноженному на $n!$ объему A . Таким образом, вероятность $P\{(X_{(1)}, \dots, X_{(n)}) \in A\}$ равна отношению объемов A и Ω , а это означает, что строка $(X_{(1)}, \dots, X_{(n)})$ распределена равномерно на мно-

жесткое Ω . Совместная плотность величин этой строки равна $n!$ внутри Ω и нулю вне Ω .

Из совместной плотности $(X_{(1)}, \dots, X_{(n)})$, предполагая x_k фиксированным и интегрируя по всем оставшимся переменным, можно вычислить плотность $X_{(k)}$. Легко видеть, что этот результат согласуется с формулой для плотности, вычисленной другими методами в гл. I, (7.3).

Этот пример был рассмотрен детально в качестве упражнения в обращении с многомерными плотностями.

в) *Распределение длин.* В обстановке предыдущего примера обозначим длину k -го интервала I_k через U_k . Тогда

$$U_1 = X_{(1)}, \quad U_k = X_{(k)} - X_{(k-1)} \quad k=2, 3, \dots, n. \quad (3.2)$$

Это не что иное, как линейное преобразование типа (1.9) с детерминантом 1. Множество Ω точек $0 < x_1 \leq \dots \leq x_n < 1$ отображается в множество Ω^* таких точек, что $u_j \geq 0$, $u_1 + \dots + u_n < 1$, и, следовательно, строка $(U_1, \dots, U_n)$ *распределена равномерно в этой области*. Этот результат сильнее ранее установленного факта, что величины U_k одинаково распределены [пример 7, б) из гл. I и задача 13 из гл. I].

г) *Еще раз о случайности показательного распределения.* Пусть $X_1, \dots, X_{n+1}$ независимы и одинаково распределены с плотностью $\alpha e^{-\alpha x}$ при $x > 0$. Положим $S_j = X_1 + \dots + X_j$. Тогда строка $(S_1, S_2, \dots, S_{n+1})$ получается из $(X_1, \dots, X_{n+1})$ линейным преобразованием типа (1.9) с детерминантом 1. Обозначим через Ω «контант», состоящий из точек $x_j > 0$ ($j=1, \dots, n+1$). Плотность $(X_1, \dots, X_{n+1})$ сосредоточена на Ω и равна

$$\alpha^{n+1} e^{-\alpha(x_1 + \dots + x_{n+1})},$$

если $x_j > 0$. Величины $S_1, \dots, S_{n+1}$ отображают Ω на область Ω^* , определяемую неравенствами $0 < s_1 \leq s_2 \leq \dots \leq s_{n+1} < \infty$, и [см. (1.11)] внутри Ω^* плотность $(S_1, \dots, S_{n+1})$ равна $\alpha^{n+1} e^{-\alpha s_{n+1}}$. Маргинальная плотность S_{n+1} есть не что иное, как гамма-плотность $\alpha^{n+1} s^n e^{-\alpha s} / n!$, и, следовательно, *условная плотность строки $(S_1, \dots, S_n)$ при условии, что $S_{n+1} = s$, равна $n! s^{-n}$ при $0 < s_1 < \dots < s_n < s$ (и нулю в других случаях)*. Иными словами, при условии, что $S_{n+1} = s$, величины $(S_1, \dots, S_n)$ равномерно распределены в области их возможных значений. Сравнивая это с примером б), мы можем сказать, что *при условии $S_{n+1} = s$ величины $(S_1, \dots, S_n)$ изображают n точек, выбранных независимо и наудачу в интервале $0, s$ и занумерованных в их естественном порядке слева направо*.

д) *Другое распределение, связанное с показательным.* Имея в виду предыдущее неожиданное применение, мы дадим еще один пример преобразования. Пусть снова $X_1, \dots, X_n$ — независимые величины, имеющие одинаковое показательное распределение и $S_n = X_1 + \dots + X_n$. Определим величины $U_1, \dots, U_n$

следующим образом:

$$U_k = \frac{X_k}{S_n} \text{ при } k=1, \dots, n-1, U_n = S_n, \quad (3.3)$$

или, что равносильно,

$$\begin{aligned} X_k &= U_k U_n \text{ при } k=1, \dots, n-1, \\ X_n &= U_n (1 - U_1 - \dots - U_{n-1}). \end{aligned} \quad (3.4)$$

Якобиан преобразования (3.4) равен U_n^{n-1} . Совместная плотность $(X_1, \dots, X_n)$ сосредоточена в области Ω , определяемой неравенствами $x_k > 0$, и внутри ее эта плотность равна $\alpha^n e^{-\alpha(x_1 + \dots + x_n)}$. Отсюда следует, что совместная плотность $(U_1, \dots, U_n)$ равна $\alpha^n u_n^{n-1} e^{-\alpha u_n}$ в области Ω^* , определенной неравенствами

$$u_1 + \dots + u_{n-1} < 1, \quad u_k > 0, \quad k=1, \dots, n,$$

и равна нулю вне Ω^* . Интегрирование относительно u_n показывает, что совместная плотность $(U_1, \dots, U_{n-1})$ равна $(n-1)!$ в Ω^* и нулю в других случаях. Сравнивая с примером в), мы видим, что $(U_1, \dots, U_{n-1})$ имеет такое же распределение, как если бы величина U_k была длиной k -го интервала при случайном разбиении интервала $[0, 1]$ набором $n-1$ точек.

е) *Критерий значимости в анализе периодограмм и теорема о покрытии.* Практически любая непрерывная функция времени t может быть приближена тригонометрическим полиномом. Если эта функция представляет собой выборочную функцию случайного процесса, то коэффициенты становятся случайными величинами и аппроксимирующий полином может быть записан в виде

$$\sum_{v=1}^n (X_v \cos \omega_v t + Y_v \sin \omega_v t) \equiv \sum_{v=1}^n R_v \cos(\omega_v t - \varphi_v), \quad (3.5)$$

где $R_v^2 = X_v^2 + Y_v^2$ и $\operatorname{tg} \varphi_v = Y_v/X_v$. Обратно, разумные предположения относительно случайных величин X_v, Y_v приводят к случайному процессу с выборочной функцией, задаваемой (3.5). В свое время было модным вводить модели такого рода и открывать «скрытую периодичность» для солнечных пятен, цен на пшеницу, поэтического творчества и т. д. Подобные скрытые периодичности отыскивали с такой же легкостью, как ведьм в средневековые. Но даже сильная вера должна быть подкреплена статистическим критерием. Грубо говоря, метод заключается в следующем. Тригонометрический полином вида (3.5) с как-то выбранными частотами $\omega_1, \dots, \omega_n$ сравнивается с некоторыми результатами наблюдений. Допустим, что при этом обнаруживается особенно большая амплитуда R_v . Желательно доказать, что это не может быть случайным и, следовательно, что ω_v — истинный период. Для проверки этого предположения выясним, правдоподобно ли совместить большое наблюдаемое значение R_v с гипотезой, что все n

компонент играют одинаковую роль. В соответствии с этим предполагается, что коэффициенты $X_1, \dots, Y_n$ взаимно независимы и имеют одинаковое нормальное распределение с нулевым математическим ожиданием и дисперсией σ^2 . В этом случае (см. II, 3) величины R_j^2 взаимно независимы и имеют одно и то же показательное распределение с математическим ожиданием $2\sigma^2$. Если наблюдаемое значение R_j^2 «значимо» отклонилось от ожидаемого, то мы приходим к заключению, что гипотеза об одинаковой роли компонент была несостоятельна и R_j^2 отражает «скрытую периодичность».

Ошибочность этих соображений была раскрыта Р. А. Фишером (1929). Он отметил, что максимальный из результатов n независимых наблюдений не подчиняется тому же самому распределению вероятностей, которое имеет каждый из них в отдельности. Ошибка в такой статистической трактовке худшего случая, как будто он выбирался случайно, до сих пор распространена в медицинской статистике, но причиной обсуждения этого вопроса здесь является неожиданная и занятая связь критерия значимости Фишера с теоремами о покрытии.

Так как важны только отношения компонент, то мы нормируем коэффициенты, полагая

$$V_j = \frac{R_j^2}{R_1^2 + \dots + R_n^2}, \quad j = 1, \dots, n. \quad (3.6)$$

Поскольку R_j^2 имеют одно и то же показательное распределение, мы можем использовать предыдущий пример с $X_j = R_j^2$. Тогда $V_1 = U_1, \dots, V_{n-1} = U_{n-1}$, но $V_n = 1 - U_1 - \dots - U_{n-1}$. Соответственно строка $(V_1, \dots, V_n)$ распределена так, как если бы V_j были длинами n интервалов, на которые разбивается интервал $0, 1$ при случайном расположении в нем $n-1$ точек. Вероятность того, что все V_j будут меньше a , задается формулой гл. I, (9.9) теоремы о покрытии. Этот результат иллюстрирует факт неожиданных отношений между на первый взгляд не связанными задачами¹⁾. ►

§ 4*). ХАРАКТЕРИЗАЦИЯ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Рассмотрим невырожденное линейное преобразование координатных величин

$$Y_1 = a_{11}X_1 + a_{12}X_2, \quad Y_2 = a_{21}X_1 + a_{22}X_2 \quad (4.1)$$

¹⁾ Фишер получил распределение максимальной из величин V_j в 1929 г. без знания теоремы о покрытии, а в 1940 г. объяснил связь с теоремой о покрытии, после того как Стивенс доказал последнюю. [См. Fisher, Contributions to Mathematical Statistics, John Wiley, N. Y. (1950), статьи 16 и 37.] Дрюгитт выводит, использующий анализ Фурье, см.: Grenander U., Rosenblatt M., (1957).

*) Этот параграф затрагивает специальную тему и не используется в дальнейшем.

и допустим (не ограничивая общности), что детерминант $\Delta = 1$. Если X_1 и X_2 — независимые нормально распределенные величины с дисперсиями σ_1^2 и σ_2^2 , то распределение пары (Y_1, Y_2) нормально с ковариацией $a_{11}a_{21}\sigma_1^2 + a_{12}a_{22}\sigma_2^2$ [см. пример 1, г)]. В этом случае существуют нетривиальные наборы коэффициентов a_{jk} , такие, что Y_1 и Y_2 независимы. Следующая теорема показывает, что это свойство одномерного нормального распределения *не разделяется никаким другим распределением*. Мы докажем это только для распределений с непрерывными плотностями, в случае которых наше утверждение сводится к лемме относительно функционального уравнения (4.3). Более общий случай сводится (с использованием характеристических функций) к *тому же самому* уравнению. Поэтому наше доказательство фактически устанавливает теорему в ее наибольшей общности (см. гл. XV, 8). Элементарное рассмотрение плотностей лучше раскрывает основную суть теоремы.

Преобразование (4.1) имеет смысл лишь в том случае, когда коэффициенты a_{jk} не обращаются в нуль. В самом деле, положим, например, $a_{11} = 0$. Без ограничения общности мы можем выбрать масштабные параметры так, чтобы $a_{12} = 1$. Тогда $Y_1 = X_2$ и, заглядывая вперед, из (4.4) получаем, что в этом случае Y_2 должно иметь ту же плотность распределения, что и X_1 . Иными словами, такое преобразование равнозначно простому переименованию случайных величин и потому не представляет никакого интереса.

Теорема. *Допустим, что X_1 и X_2 независимы и что величины Y_1 и Y_2 также не зависят одна от другой. Если ни один из коэффициентов a_{jk} не равен нулю, то все четыре величины являются нормальными.*

Наиболее интересный частный случай (4.1) доставляется *вращениями*, т. е. преобразованиями вида

$$\begin{aligned} Y_1 &= X_1 \cos \omega + X_2 \sin \omega, \\ Y_2 &= -X_1 \sin \omega + X_2 \cos \omega, \end{aligned} \quad (4.2)$$

где ω не кратно $\frac{1}{2}\pi$. Применив теорему к подобным вращениям, мы получаем

Следствие. *Если X_1 и X_2 — независимые случайные величины и существует вращение (4.2), такое, что Y_1 и Y_2 также независимы, то X_1 и X_2 имеют нормальные распределения с одинаковой дисперсией. В этом случае Y_1 и Y_2 независимы при каждом ω .*

Пример. *Распределение Максвелла для скоростей.* При изучении распределений скоростей молекул в $\mathcal{R}^3$ Максвелл предполагал, что в любой декартовой системе координат три компоненты скорости являются независимыми случайными величинами с нулевым математическим ожиданием. Примененное к вращениям, оставляющим одну ось фиксированной, наше следствие непосредст-

венно показывает, что три компонента нормально распределены с одинаковой дисперсией. Как мы видели в гл. II,3, это влечет распределение Максвелла для скоростей. ►

Эта теорема имеет длинную историю, восходящую к исследованиям Максвелла. Чисто вероятностное изучение было предпринято М. Качем (1940) и С. Бернштейном (1941), который доказал наше следствие в предположении конечности дисперсий. Многие авторы вносили улучшения и исследовали варианты, иногда при помощи более сильных методов. Кульминации это развитие достигло в результате, доказанном В. П. Скутзовичем¹⁾.

Перейдем к доказательству для случая непрерывных плотностей. Обозначим плотности X_j через u_j , а плотности Y_j через f_j . Положим для краткости

$$y_1 = a_{11}x_1 + a_{12}x_2, \quad y_2 = a_{21}x_1 + a_{22}x_2. \quad (4.3)$$

По условию теоремы имеем

$$f_1(y_1)f_2(y_2) = u_1(x_1)u_2(x_2). \quad (4.4)$$

Покажем, что из соотношения (4.4) следует, что

$$f_j(y) = \pm e^{\varphi_j(y)}, \quad u_j(x) = \pm e^{\omega_j(x)}, \quad (4.5)$$

где $\varphi_j(y)$ и $\omega_j(x)$ — полиномы не выше второй степени. Единственными распределениями с плотностью такого вида являются нормальные распределения. Поэтому для распределений с непрерывными плотностями теорема содержится в следующей лемме.

Лемма. *Предположим, что четыре непрерывные функции f_j и u_j , $j=1, 2$, связаны функциональным уравнением (4.4) и что ни один из коэффициентов a_{jk} не равен нулю. Тогда эти функции имеют вид (4.5), где показателями являются полиномы, степень которых не выше 2.*

(Предполагается, конечно, что ни одна из функций не равна тождественно нулю.)

Доказательство. Вначале отметим, что ни одна из наших функций не имеет нулей. Действительно, в противном случае должна существовать на (x_1, x_2) -плоскости область Ω , внутри которой обе части равенства (4.4) не имеют нулей и на границе которой они обращаются в нуль. Приравнявая обе части равенства (4.4) нулю, выводим, с одной стороны, что граница состоит из отрезков прямых, параллельных осям, с другой стороны, из отрезков, параллельных прямым $y_j = \text{const}$. Это противоречие показывает, что такая граница не существует.

Мы можем таким образом считать наши функции строго поло-

¹⁾ Изв. АН СССР, т. 18 (1954), стр. 185—200. Теорема. Пусть $X_1, \dots, X_n$ взаимно независимы, $Y_1 = \sum a_i X_i$ и $Y_2 = \sum b_i X_i$, где ни один коэффициент не равен нулю. Если Y_1 и Y_2 независимы, то величины X_i распределены нормально,

жительными. Переходя к логарифмам, перепишем (4.4) в виде

$$\varphi_1(y_1) + \varphi_2(y_2) = -\omega_1(x_1) + \omega_2(x_2). \quad (4.6)$$

Определим при любых фиксированных h_1 и h_2 смешанный разностный оператор Δ равенством

$$\Delta v(x_1, x_2) = v(x_1 + h_1, x_2 + h_2) - v(x_1 + h_1, x_2 - h_2) - \\ - v(x_1 - h_1, x_2 + h_2) + v(x_1 - h_1, x_2 - h_2). \quad (4.7)$$

Так как каждая функция ω_j зависит только от одной переменной x_j , то отсюда следует, что $\Delta\omega_j = 0$. Имеем

$$\Delta\varphi_1(y_1) = \varphi_1(y_1 + t_1) - \varphi_1(y_1 + t_2) - \varphi_1(y_1 - t_2) + \varphi_1(y_1 - t_1), \quad (4.8)$$

где для краткости введены

$$t_1 = a_{11}h_1 + a_{12}h_2, \quad t_2 = a_{11}h_1 - a_{12}h_2. \quad (4.9)$$

Таким образом, имеем $\Delta\varphi_1 + \Delta\varphi_2 = 0$, где φ_j зависит только от переменной y_j . При фиксированном y_2 очевидно, что $\Delta\varphi_1(y_1)$ есть постоянная, зависящая только от h_1 и h_2 . Выберем теперь h_1 и h_2 так, что $t_1 = t$ и $t_2 = 0$, где t произвольно, но фиксированно. Тогда соотношение $\Delta\varphi_1 = \text{const}$ примет вид

$$\varphi_1(y_1 + t) + \varphi_1(y_1 - t) - 2\varphi_1(y_1) = \lambda(t). \quad (4.10)$$

Вблизи точки y_1 , где функция φ_1 принимает минимальное значение, левая часть (4.10) не меньше нуля. Следовательно, такая точка y_1 существует только в том случае, если $\lambda(t) \geq 0$ при всех t из некоторой окрестности нуля. Но в таком случае φ_1 не может иметь максимальное значение. Далее, непрерывная функция, обращающаяся в нуль в трех точках, принимает как максимальное, так и минимальное значение. Мы выводим, что если непрерывное решение соотношения (4.10) равно нулю в трех различных точках, то оно равно нулю тождественно.

Каждый квадратный полином $q(y_1) = \alpha y_1^2 + \beta y_1 + \gamma$ удовлетворяет уравнению вида (4.10) (с другой правой частью), и, следовательно, то же справедливо для разности $\varphi_1(y_1) - q(y_1)$. Но q можно выбрать так, что эта разность будет равна нулю в трех точках и тогда $\varphi_1(y_1)$ совпадает с q . Те же рассуждения применимы к φ_2 , и этим доказывается утверждение относительно f_1 и f_2 . Так как случайные величины X_j и Y_j играют одинаковую роль, то те же аргументы применимы к плотностям u_j . ►

§ 5. МАТРИЧНЫЕ ОБОЗНАЧЕНИЯ. КОВАРИАЦИОННАЯ МАТРИЦА

Обозначения, использованные в § 1, громоздки и становятся еще более громоздкими в случае большего числа измерений. Изящества и экономии можно достичь при использовании матричной системы обозначений.

Чтобы облегчить ссылки, мы кратко изложим немногие факты из теории матриц и матричных обозначений, нужные в дальнейшем. Основное правило таково: сначала строки, затем столбцы. Таким образом, $(\alpha \times \beta)$ -матрица A имеет α строк и β столбцов; ее элементы обозначаются a_{jk} , при этом первый индекс указывает строку. Если B есть $(\beta \times \gamma)$ -матрица с элементами b_{jk} , то произведение AB представляет собой $(\alpha \times \gamma)$ -матрицу с элементами $a_{j1}b_{1k} + a_{j2}b_{2k} + \dots + a_{j\beta}b_{\beta k}$. Если число столбцов матрицы A не совпадает с числом строк матрицы B , то произведение не определено. Выполняется ассоциативный закон $(AB)C = A(BC)$, хотя, вообще говоря, $AB \neq BA$. Транспонированная матрица A^T есть $(\beta \times \alpha)$ -матрица с элементами $a_{jk}^T = a_{kj}$. Очевидно, что $(AB)^T = B^T A^T$.

$(1 \times \alpha)$ -матрица с единственной строкой называется *вектор-строкой*, а матрица с единственным столбцом — *вектор-столбцом*¹⁾. Вектор-строка $r = (r_1, \dots, r_\alpha)$ легко изображается при печати, а вектор-столбец лучше определять результатом его транспонирования $c^T = (c_1, \dots, c_\alpha)$. Заметим, что cr есть $(\alpha \times \alpha)$ -матрица (типа «таблицы умножения»), в то время как rc есть (1×1) -матрица, или скаляр. В случае $\alpha = 2$

$$cr = \begin{pmatrix} c_1 r_1 & c_1 r_2 \\ c_2 r_1 & c_2 r_2 \end{pmatrix}, \quad rc = (r_1 c_1 + r_2 c_2).$$

Нулевой вектор имеет нулевые компоненты.

Матрицы с одинаковым числом строк и столбцов называются *квадратными* матрицами. Для каждой квадратной матрицы A определяется ее *детерминант* (определитель), который будет обозначаться через $|A|$. Для наших целей достаточно знать, что детерминанты обладают свойством мультипликативности: если A и B — две квадратные матрицы и $C = AB$, то $|C| = |A| \cdot |B|$. Матрица A и транспонированная матрица A^T имеют один и тот же детерминант.

Под *единичной матрицей* подразумевается квадратная матрица с единицами по главной диагонали и с нулями на всех остальных местах. Если I — единичная матрица с r строками и r столбцами и A — любая $(r \times r)$ -матрица, то очевидно, что $IA = A = AI$. Матрицей, *обратной* для A , называется матрица A^{-1} , такая, что $AA^{-1} = A^{-1}A = I$. [Только квадратные матрицы могут иметь обратные. Обратная матрица единственна: так, если B — любая обратная матрица для A , то $AB = I$ и по ассоциативному закону $A^{-1} = (A^{-1}A)B = B$.] Квадратная матрица, не имеющая обратной, называется *вырожденной*. Из мультипликативного свойства детерминантов следует, что матрица, имеющая нулевой детерминант, вырождена. Справедливо также и обратное: если $|A| \neq 0$, то матрица A невырождена. Иными словами, матрица A является вырожденной тогда и только тогда, когда существует ненулевая вектор-строка x , такая, что $xA = 0$.

Квадратная матрица A *симметрична*, если $a_{jk} = a_{kj}$, т. е. если $A^T = A$. *Квадратичная форма*, соответствующая симметричной $(r \times r)$ -матрице A , определяется как

$$xAx^T = \sum_{j,k=1}^r a_{jk}x_jx_k,$$

где $x_1, \dots, x_r$ являются переменными. Матрица называется *положительно определенной*, если $xAx^T > 0$ для всех ненулевых векторов x . Из последнего критерия следует, что положительно определенная матрица невырождена.

Вращения в $\mathcal{R}^\alpha$. Для полноты мы кратко коснемся одного геометрического приложения матричного анализа, хотя это и не будет использовано в дальнейшем.

¹⁾ На самом деле это — злоупотребление языком. В конкретном случае x_1 может означать количество фунтов, а x_2 — число коров; тогда (x_1, x_2) не является «вектором» в строгом смысле.

Скалярным (внутренним) произведением двух векторов-строк $x = (x_1, \dots, x_\alpha)$ и $y = (y_1, \dots, y_\alpha)$ называется величина

$$xy^T = yx^T = \sum_{j=1}^{\alpha} x_j y_j.$$

Длина L вектора x определяется из равенства $L^2 = xx^T$. Для векторов x и y единичной длины угол δ между ними определяется величиной $\cos \delta = xy^T$.

Любая $(\alpha \times \alpha)$ -матрица A индуцирует преобразование, переводящее x в $\xi = xA$; для транспонированной матрицы x преобразуется в $\xi^T = A^T x^T$. Матрица A называется *ортогональной*, если индуцированное ею преобразование сохраняет длины и углы, иначе говоря, если любые два вектора-строки имеют то же самое скалярное произведение, что и их образы при преобразовании. Таким образом, матрица A ортогональна тогда и только тогда, когда для любой пары векторов-строк x, y

$$xAA^T y^T = xy^T.$$

Отсюда следует, что AA^T есть единичная матрица I (в этом можно убедиться, выбирая в качестве x и y векторы с $\alpha - 1$ компонентами, равными нулю). Таким образом, мы пришли к заключению, что A ортогональна тогда и только тогда, когда $AA^T = I$. Поскольку A и A^T имеют один и тот же детерминант, то этот детерминант равен $+1$ или -1 . Ортогональная матрица с детерминантом 1 называется *матрицей вращения*, а индуцируемое ею преобразование — *вращением*.

Отныне мы будем обозначать точку n -мерного пространства $\mathcal{R}^n$ одной буквой, интерпретируя ее как *вектор-строку*. Таким образом, $x = (x_1, \dots, x_r)$ и $f(x) = f(x_1, \dots, x_r)$ и т. д. Неравенства должны интерпретироваться по координатам: $x < y$ тогда и только тогда, когда $x_k < y_k$ для $k = 1, \dots, r$; аналогично и для других неравенств. На плоскости $\mathcal{R}^2$ соотношение $x < y$ может быть прочитано как « x лежит юго-западнее y ». Новая особенность этого обозначения в том, что две точки не обязаны состоять в одном из двух соотношений: либо $x \leq y$, либо $y < x$, т. е. в многомерном случае знак $<$ определяет только частичное упорядочение.

Мы вводим обозначение $X = (X_1, \dots, X_r)$ для вектора-строки из координатных величин и будем использовать это обозначение для случайных величин вообще (в основном для нормально распределенных величин).

Если величины $X_1, \dots, X_r$ имеют математические ожидания $E(X_j)$, то мы обозначим через $E(X)$ *вектор-строку с компонентами* $E(X_j)$. Вектор $X - E(X)$ имеет нулевое математическое ожидание. Вообще, если M — матрица, элементы которой M_{jk} являются случайными величинами, мы обозначаем через $E(M)$ матрицу элементов $E(M_{jk})$ в предположении, что она существует.

Определение. Если $E(X) = 0$, то ковариационная матрица $\text{Var}(X)$ для X есть симметричная $(r \times r)$ -матрица с элементами $E(X_j X_k)$ (при условии, что все они существуют). Иными словами,

$$\text{Var}(X) = E(X^T X). \quad (5.1)$$

В случае произвольного X мы определим $\text{Var} X$ как $\text{Var}(X - E(X))$.

Использование векторов-строк неизбежно влечет за собой запись *линейного преобразования* $\mathcal{R}^r$ в $\mathcal{R}^m$ в виде

$$Y = XA, \quad (5.2)$$

т. е.

$$y_k = \sum_{j=1}^r a_{jk} x_j, \quad k=1, \dots, m, \quad (5.3)$$

где A — $(r \times m)$ -матрица. Очевидно, что $E(Y) = E(X)A$ всякий раз, когда $E(X)$ существует. Для того, чтобы найти дисперсии, мы положим, не ограничивая общности, что $E(X) = 0$. Тогда $E(Y) = 0$ и

$$E(Y^T Y) = E(A^T X^T X A) = A^T E(X^T X) A. \quad (5.4)$$

Мы получаем, таким образом, важный результат:

$$\text{Var}(Y) = A^T \text{Var}(X) A. \quad (5.5)$$

Особый интерес представляет частный случай $m=1$, когда

$$Y = a_1 X_1 + \dots + a_r X_r, \quad (5.6)$$

есть обыкновенная случайная величина. Здесь $\text{Var}(Y)$ —(скалярная) квадратичная форма:

$$\text{Var}(Y) = \sum_{j,k=1}^r E(X_j X_k) a_j a_k. \quad (5.7)$$

Линейная форма (5.6) равна нулю с вероятностью единица, если $\text{Var}(Y) = 0$, и в этом случае любая область за пределами гиперплоскости $\sum a_k x_k = 0$ имеет вероятность нуль. Распределение вероятностей сосредоточено тогда на $(r-1)$ -мерном многообразии и вырождено, если его рассматривать в r измерениях. Мы доказали, что ковариационная матрица любого невырожденного распределения вероятностей положительно определена. Обратно, каждая такая матрица может служить ковариационной матрицей нормальной плотности (см. теорему 4 следующего параграфа).

§ 6. НОРМАЛЬНЫЕ ПЛОТНОСТИ И РАСПРЕДЕЛЕНИЯ

Всюду в этом параграфе Q обозначает симметричную $(r \times r)$ -матрицу, а $q(x)$ —соответствующую ей квадратичную форму

$$q(x) = \sum_{j,k=1}^r q_{jk} x_j x_k = x Q x^T, \quad (6.1)$$

где $x = (x_1, \dots, x_r)$ есть вектор-строка. Плотности в $\mathcal{R}^r$, определенные как показательные функции с квадратичной формой в показателе, являются естественным обобщением нормальной плотности на прямой, и мы начнем поэтому с того, что введем следующее

Определение. Плотность φ в пространстве r измерений называется нормальной¹⁾ (с центром в начале координат), если она представляется в виде

$$\varphi(x) = \gamma^{-1} \cdot e^{-\frac{1}{2} q(x)}, \quad (6.2)$$

где γ — некоторая постоянная. Нормальная плотность с центром в точке $a = (a_1, a_2, \dots, a_r)$ определяется как $\varphi(x - a)$.

Частный случай двух измерений обсуждался в примерах 1, г) и 2, а).

Возьмем в качестве выборочного пространство $\mathcal{R}$ с вероятностным распределением (6.2) и обозначим через $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_r)$ вектор-строку, образованную координатными случайными величинами. Соответствующую ковариационную матрицу обозначим через M :

$$M = \text{Var}(\mathbf{X}) = E(\mathbf{X}^T \mathbf{X}). \quad (6.3)$$

Наша задача состоит в исследовании матриц Q и M , а также связи между ними.

Вначале заметим, что диагональные элементы Q не могут быть нулями. В самом деле, если бы мы имели $q_{rr} = 0$, то при фиксированных значениях $x_1, \dots, x_{r-1}$ плотность (6.2) имела бы вид $\gamma^{-1} e^{-ax_r + b}$ и интеграл от плотности по x_r расходился бы. Введем теперь подстановку $y = xA$, определенную равенствами

$$y_1 = x_1, \dots, y_{r-1} = x_{r-1}, \quad y_r = q_{1r}x_1 + \dots + q_{rr}x_r. \quad (6.4)$$

Проверка показывает, что $q(x) = y_r^2 / q_{rr}$ есть квадратичная форма от величин $x_1, \dots, x_{r-1}$, но не от x_r . Таким образом,

$$q(x) = \frac{1}{q_{rr}} y_r^2 + \bar{q}(y), \quad (6.5)$$

где $\bar{q}(y)$ — квадратичная форма от $y_1, \dots, y_{r-1}$. Отсюда следует, что вектор $\mathbf{Y} = \mathbf{X}A$ имеет нормальную плотность, которая является произведением двух нормальных плотностей для случайных величин $\mathbf{Y}_r$ и $(\mathbf{Y}_1, \dots, \mathbf{Y}_{r-1})$ соответственно. Первый полученный результат содержится в простой, но важной теореме.

Теорема 1. Все маргинальные плотности нормальной плотности снова нормальны.

Более неожиданный результат содержит

Теорема 2. Существует матрица C с положительным детерминантом, такая, что $\mathbf{Z} = \mathbf{X}C$ есть вектор-строка, чьи компоненты $\mathbf{Z}_j$ суть взаимно независимые нормально распределенные случайные величины.

Матрица C не единственна; в действительности теорема может

¹⁾ «Вырожденные» нормальные распределения будут введены в конце этого параграфа.

быть усилена до утверждения, что в качестве C можно выбрать матрицу вращения (см. задачу 19).

Доказательство. Используем индукцию. При $r=2$ утверждение теоремы выводится непосредственно из представления (6.5). Если теорема справедлива в $r-1$ измерениях, то случайные величины $Y_1, \dots, Y_{r-1}$ являются линейными комбинациями независимых нормальных величин $Z_1, \dots, Z_{r-1}$, тогда как случайная величина Y_r сама распределена нормально и независимо от остальных величин. Поскольку $X=YA^{-1}$, отсюда следует также, что величины X_j являются линейными комбинациями $Z_1, \dots, Z_{r-1}$ и Y_r . Детерминант матрицы A равен q_{rr} и, как следует из (6.5), положителен. Детерминант преобразования $X \rightarrow Z$ равен произведению детерминанта A и детерминанта преобразования $Y \rightarrow Z$ и, следовательно, положителен.

Теорема 3. Матрицы Q и M взаимно обратны и

$$\gamma^2 = (2\pi)^r \cdot |M|, \quad (6.6)$$

где $|M| = |Q|^{-1}$ есть детерминант M .

Доказательство. В обозначениях предыдущей теоремы положим

$$D = E(Z^T Z) = C^T M C. \quad (6.7)$$

Диагональные элементы этой матрицы равны $E(Z_j^2) = \sigma_j^2$, а остальные элементы равны нулю. Плотность Z равна произведению нормальных плотностей $\pi(x\sigma_j^{-1})\sigma_j^{-1}$ и, следовательно, порождается матрицей D^{-1} с диагональными элементами σ_j^{-2} . Далее, плотность величины Z получается из плотности (6.2) величины X путем подстановки $x = zC^{-1}$ и умножения на детерминант $|C^{-1}|$. Таким образом

$$zD^{-1}z^T = xQx^T \quad (6.8)$$

и

$$(2\pi)^r |D| = \gamma^2 \cdot |C|^2. \quad (6.9)$$

Из (6.8) видно, что

$$Q = CD^{-1}C^T, \quad (6.10)$$

и отсюда с учетом (6.7) следует, что $Q = M^{-1}$. Из (6.7) вытекает также, что $|D| = |M| \cdot |C|^2$ и, следовательно, (6.9) равносильно (6.6). ►

Из теоремы следует, в частности, что разложение для матрицы M соответствует аналогичному разложению для Q , и поэтому имеет место

Следствие. Если (X_1, X_2) имеет нормальное распределение, то X_1 и X_2 независимы в том и только в том случае, когда $\text{Cov}(X_1, X_2) = 0$, т. е. когда X_1 и X_2 некоррелированы.

Вообще, если $(X_1, \dots, X_r)$ имеет нормальную плотность, то

$(X_1, \dots, X_n)$ и $(X_{n+1}, \dots, X_r)$ независимы тогда и только тогда, когда $\text{Cov}(X_j, X_k) = 0$ при $j \leq n, k > n$.

Предостережение. Для справедливости следствия существенно то, что совместная плотность пары (X_1, X_2) нормальна. Следствие неприменимо, если известно только то, что маргинальные плотности X_1 и X_2 являются нормальными. В последнем случае плотность (X_1, X_2) не обязана быть нормальной и в действительности даже не обязана существовать. Этот факт часто понимается неправильно (см. задачи 2, 3).

Теорема 4. Матрица M представляет собой ковариационную матрицу нормальной плотности тогда и только тогда, когда она положительно определена.

Поскольку плотность порождается матрицей $Q = M^{-1}$, можно сформулировать эквивалентное утверждение: матрица Q порождает нормальную плотность (6.2) тогда и только тогда, когда она положительно определена.

Доказательство. Мы видели в конце § 5, что всякая ковариационная матрица нормальной плотности положительно определена. Обратное тривиально в случае $r=1$. К большим r мы перейдем по индукции. Предположим, что матрица Q положительно определенная. При $x_1=0, \dots, x_{r-1}=0$ мы получаем $q(x) = q_{rr}x_r^2$ и, следовательно, $q_{rr} > 0$. При этом предположении, как мы видели, q можно привести к виду (6.5). Выбирая x_r так, что $y_r = 0$, мы видим, что положительная определенность Q влечет $\bar{q}(x) > 0$ при всех наборах $x_1, \dots, x_{r-1}$. Поэтому по предположению индукции $\bar{q}$ соответствует нормальной плотности в пространстве $r-1$ измерений. Из (6.5) теперь очевидным образом следует, что q соответствует нормальной плотности в пространстве r измерений, и этим завершается доказательство. ►

Мы закончим эту общую теорию интерпретацией (6.5) в терминах *условных плотностей*, которая приводит к общим формулировкам теории регрессии, разобранный для двумерного случая в примере 2, а).

Положим для краткости $a_k = -q_{kr}/q_{rr}$, так, что

$$y_r = q_{rr}(x_r - a_1x_1 - \dots - a_{r-1}x_{r-1}). \quad (6.11)$$

Для вероятностной интерпретации коэффициентов a_k мы вспомним, что Y_r по построению не зависит от $X_1, \dots, X_{r-1}$. Иными словами, a_k — такие числа, что

$$T = X_r - a_1X_1 - \dots - a_{r-1}X_{r-1} \quad (6.12)$$

не зависит от $(X_1, \dots, X_{r-1})$, и это свойство однозначно характеризует коэффициенты a_k .

Чтобы найти условную плотность X_r для данных $X_1 = x_1, \dots, X_{r-1} = x_{r-1}$, мы должны разделить плотность $(X_1, \dots, X_r)$ на

маргинальную плотность $(X_1, \dots, X_{r-1})$. Ввиду (6.5) мы получим показательную функцию с показателем $-\frac{1}{2} y_r^2 / q_{rr}$. Отсюда следует, что условная плотность X_r при данных $X_1 = x_1, \dots, X_{r-1} = x_{r-1}$ есть нормальная плотность с математическим ожиданием $a_1 x_1 + \dots + a_{r-1} x_{r-1}$ и дисперсией $1/q_{rr}$. Соответственно

$$E(X_r | X_1, \dots, X_{r-1}) = a_1 X_1 + \dots + a_{r-1} X_{r-1}. \quad (6.13)$$

Таким образом, мы доказали следующее обобщение двумерной задачи регрессии, выражение которой было дано в (2.6).

Теорема 5. Если $(X_1, \dots, X_r)$ имеет нормальную плотность, то условная плотность X_r при данных $(X_1, \dots, X_{r-1})$ снова нормальна. Кроме того, условное математическое ожидание (6.13) есть единственная линейная функция от $X_1, \dots, X_{r-1}$, делающая величину T не зависящей от $(X_1, \dots, X_{r-1})$. Условная дисперсия равна $\text{Var}(T) = q_{rr}^{-1}$.

Пример. Выборочное среднее значение и дисперсия. В статистике случайные величины

$$\hat{X} = \frac{1}{r} (X_1 + \dots + X_r), \quad \hat{\sigma}^2 = \frac{1}{r} \sum_{k=1}^r (X_k - \hat{X})^2 \quad (6.14)$$

называются выборочным средним значением и выборочной дисперсией для $X = (X_1, \dots, X_r)$. Весьма любопытен тот факт, что если величины $X_1, \dots, X_r$ независимы и нормальны с $E(X_k) = 0$, $E(X_k^2) = \sigma^2$, то случайные величины $\hat{X}$ и $\hat{\sigma}^2$ являются независимыми¹⁾.

Доказательство демонстрирует применимость предыдущих результатов. Мы положим $Y_k = X_k - \hat{X}$ при $k = 1, \dots, r-1$, но $Y_r = \hat{X}$. Преобразование X в $Y = (Y_1, \dots, Y_r)$ линейно и невырождено. Поэтому Y имеет нормальную плотность. Далее $E(Y_k Y_r) = 0$ при $k = 1, \dots, r-1$, и поэтому Y_r не зависит от $(Y_1, \dots, Y_{r-1})$. Однако

$$r\hat{\sigma}^2 = Y_1^2 + \dots + Y_{r-1}^2 + (Y_1 + \dots + Y_{r-1})^2 \quad (6.15)$$

зависит только от $Y_1, \dots, Y_{r-1}$, и, таким образом, $\hat{\sigma}^2$ действительно не зависит от $Y_r = \hat{X}$. ►

Нормальные распределения общего вида

Из предыдущего следует, что если $X = (X_1, \dots, X_r)$ имеет нормальную плотность, то каждая ненулевая линейная комбинация $Y_i = a_1 X_1 + \dots + a_r X_r$ также имеет нормальную плотность.

¹⁾ То, что этот факт характеризует нормальное распределение в $\mathcal{M}^1$, было показано Гири и Лукачем.

То же самое верно для каждой пары (Y_1, Y_2) при условии, что не выполняется линейное соотношение типа $c_1 Y_1 + c_2 Y_2 = 0$. В этом исключительном случае распределение вероятностей пары (Y_1, Y_2) сосредоточено на прямой с уравнением $c_1 y_1 + c_2 y_2 = 0$, и, следовательно, оно вырождено, если его рассматривать как *двумерное распределение*. Для многих целей желательно сохранить термин нормального распределения также и для вырожденных распределений, сосредоточенных на многообразиях низшей размерности, скажем на какой-либо прямой. Простейшее общее определение выглядит следующим образом: *распределение $Y = (Y_1, \dots, Y_p)$ является нормальным, если существует вектор $X = (X_1, \dots, X_r)$ с нормальной r -мерной плотностью, такой, что $Y = a + XA$, где A — (постоянная), $(r \times p)$ -матрица и $a = (a_1, \dots, a_p)$. Если $p > r$, то распределение Y вырождено в p измерениях. При $p \leq r$ это распределение невырождено тогда и только тогда, когда p форм, определяющих Y_k , линейно независимы.*

§ 7*). СТАЦИОНАРНЫЕ НОРМАЛЬНЫЕ ПРОЦЕССЫ

Цель этого параграфа состоит отчасти в том, чтобы дать примеры нормальных распределений, и отчасти в том, чтобы установить некоторые соотношения, имеющие важное применение в теории дискретных случайных процессов и временных рядов. Они носят аналитический характер и легко отделимы от более глубокого стохастического анализа. Фактически мы будем иметь дело только с конечномерными нормальными плотностями, или, что равносильно, с их ковариационными матрицами. Ссылки на случайные величины существенны для вероятностной интуиции и как подготовка к применениям, но на данном этапе мы касаемся только их совместных распределений; случайные величины сами используются единственно как удобный способ описания всех маргинальных плотностей посредством указания соответствующих совокупностей $(X_{\alpha_1}, \dots, X_{\alpha_k})$. Кроме того, ссылка на бесконечную последовательность $\{X_k\}$ влечет за собой лишь то, что число элементов в $(X_1, \dots, X_n)$ может быть взято сколько угодно большим.

Мы будем фактически рассматривать бесконечную в обе стороны последовательность $\{\dots, X_{-2}, X_{-1}, \dots\}$. Под этим мы просто подразумеваем, что для каждой конечной совокупности $(X_{n_1}, \dots, X_{n_r})$ задана нормальная плотность с очевидными условиями согласованности. Последовательность называется *стационарной*, если эти распределения инвариантны относительно сдвигов во времени, т. е. если все строки вида $(X_{n_1+v}, \dots, X_{n_r+v})$ с фиксированными $(n_1, \dots, n_r)$ имеют одно и то же распределе-

* Не используется в последующем. В частности, § 8 может быть прочитан независимо (см. также гл. XIX, А).

ние, не зависящее от n . При $r=1$ отсюда следует, что математические ожидания и дисперсии постоянны и, следовательно, можно предположить, не ограничивая общности, что $E(X_n)=0$. Совместные распределения полностью определяются через ковариации $r_{jk}=E(X_j X_k)$, а свойство стационарности требует, чтобы r_{jk} зависели только от разности $|k-j|$. Поэтому мы положим $r_{j, j+n}=r_n$. Таким образом,

$$r_n = E(X_k X_{k+n}) = E(X_{k-n}, X_k), \quad (7.1)$$

откуда $r_n = r_{-n}$. Мы будем иметь дело только с последовательностями чисел r_n , которые могут служить ковариациями случайного процесса.

Всюду в этом параграфе $\{Z_n\}$ изображает бесконечную в обе стороны последовательность *взаимно независимых нормально распределенных случайных величин*, нормированных так, что

$$E(Z_n)=0, \quad E(Z_n^2)=1. \quad (7.2)$$

Будут описаны три метода построения стационарных последовательностей в терминах данной последовательности $\{Z_n\}$. Эти методы постоянно используются в анализе временных рядов и могут служить упражнением в стандартных рассуждениях.

Пример. а) *Обобщенные процессы скользящего среднего.* При произвольных постоянных $b_0, b_1, \dots, b_N$ положим

$$X_n = b_0 Z_n + b_1 Z_{n-1} + \dots + b_N Z_{n-N}. \quad (7.3)$$

В частном случае равных коэффициентов $b_k = 1/(N+1)$ случайная величина X_n является средним арифметическим того типа, которое применяется в анализе временных рядов для «сглаживания данных» (т. е. для устранения локальных иррегулярностей). В общем случае (7.3) представляет собой линейный оператор, переводящий стационарную последовательность $\{Z_n\}$ в новую стационарную последовательность $\{X_n\}$. Такие операции модно называть «фильтрами». Последовательность $\{X_n\}$ имеет ковариации

$$r_k = r_{-k} = E(X_n X_{n+k}) = \sum_v b_v b_{v+k} \quad (k \geq 0), \quad (7.4)$$

где сумма справа содержит ряды, имеющие лишь конечное число членов.

Так как $2|b_v b_{v+k}| \leq b_v^2 + b_{v+k}^2$, то выражение (7.4) имеет смысл также и для бесконечных последовательностей, для которых $\sum b_v^2 < \infty$. Легко видеть, что предел последовательности ковариационных матриц есть снова ковариационная матрица. Полагая $N \rightarrow \infty$, мы заключаем, что для любой последовательности $b_0, b_1, b_2, \dots$, такой, что $\sum b_n^2 < \infty$, числа r_k , определенные в (7.4), могут служить ковариациями стационарного процесса $\{X_n\}$.

Формально мы получаем для нового процесса

$$X_n = \sum_{k=0}^{\infty} b_k Z_{n-k}. \quad (7.5)$$

Можно без затруднений показать, что любой стационарный процесс с ковариациями (7.4) представим в такой форме, но выражение (7.5) включает бесконечно много членов, и мы не можем сделать это в настоящий момент (см. гл. XIX, 8).

б) *Процесс авторегрессии.* С тех пор как родился анализ временных рядов, предлагались многие теоретические модели для объяснения эмпирических явлений, например, таких, как поведение экономических временных рядов, солнечных пятен и наблюдаемые (или воображаемые) периодичности. Наиболее распространенная модель предполагает, что величины X_n исследуемого процесса связаны с нашей последовательностью Z_n независимых нормальных величин (7.2) посредством уравнения авторегрессии вида

$$a_0 X_n + a_1 X_{n-1} + \dots + a_N X_{n-N} = Z_n. \quad (7.6)$$

Эта модель основывается на эмпирическом предположении, что значение величины X_n в момент n (цена, содержание или интенсивность) зависит от ее прошлого развития и наложенного на него «случайного возмущения» Z_n , которое не связано с прошлым. Как часто бывает, предположение о *линейной* зависимости упрощает (или делает возможным) теоретический анализ. Более общие модели получаются, если положить $N \rightarrow \infty$ или выбрать в качестве Z_n величины другого стационарного процесса.

Если $a_0 \neq 0$, то можно произвольным образом выбрать $(X_0, \dots, X_{N-1})$ и затем последовательно вычислить $X_N, X_{N+1}, \dots$ и $X_{-1}, X_{-2}, \dots$. В этом смысле (7.6) определяет некоторый процесс, но нас интересует, существует ли *стационарное* решение.

Для ответа на этот вопрос мы перепишем (7.6) в форме, не включающей элементов, непосредственно предшествующих X_n . Рассмотрим (7.6), заменяя n последовательно на $n-1, n-2, \dots, n-v$. Умножим эти равенства на $b_1, b_2, \dots, b_v$ соответственно и прибавим к (7.6). Величины $X_{n-1}, \dots, X_{n-v}$ не появятся в новом уравнении в том и только в том случае, если b_j таковы, что

$$a_0 b_1 + a_1 b_0 = 0, \dots, a_v b_v + a_1 b_{v-1} + \dots + a_v b_0 = 0, \quad (7.7)$$

где $b_0 = 1$. В конце концов приходим к выражению вида

$$a_0 X_n = b_0 Z_n + b_1 Z_{n-1} + \dots + b_v Z_{n-v} + Y_{n,v}, \quad (7.8)$$

где $Y_{n,v}$ представляет собой линейную комбинацию $X_{n-v-1}, \dots, X_{n-N-v}$ (с коэффициентами, которые нас не интересуют). В (7.8) мы представим величину X_n как сумму случайных возму-

щений в моменты $n, n-1, \dots, n-v$ и величины $Y_{n,v}$, изображающей влияние «прошлого» до момента $n-v$. При $v \rightarrow \infty$ это время становится «бесконечно далеким прошлым», и в большинстве ситуаций оно не будет иметь влияния. При переходе к пределу мы (по крайней мере временно) предположим именно этот случай, т. е. мы разыскиваем процесс, удовлетворяющий предельному соотношению вида

$$a_0 X_n = \sum_{k=0}^{\infty} b_k Z_{n-k}. \quad (7.9)$$

(Грубо говоря, мы предполагаем, что остаточные величины $Y_{n,v}$ стремятся к нулю. Другие возможные пределы изучены в следующем примере.)

Процессы вида (7.9) рассматривались в примере а), и мы видели, что всякий раз, когда $\sum b_k^2 < \infty$, существует стационарное решение (если ряд расходится, то даже выражение для ковариаций теряет смысл). Для решения уравнений (7.7) относительно b_k мы используем формальные производящие функции

$$A(s) = \sum a_k s^k, \quad B(s) = \sum b_k s^k. \quad (7.10)$$

Уравнения (7.7) выполняются тогда и только тогда, когда $A(s)B(s) = a_0 b_0$. Так как A — полином, то B — рациональная функция. Поэтому мы можем использовать теорию разложения на простые дроби, развитую в 1, гл. XI, 4. Если полином $A(s)$ имеет различные корни $s_1, \dots, s_N$, мы получаем

$$B(s) = \frac{A_1}{s_1 - s} + \dots + \frac{A_N}{s_N - s} \quad (7.11)$$

и, следовательно,

$$b_n = A_1 s_1^{-n-1} + \dots + A_N s_N^{-n-1}. \quad (7.12)$$

Очевидно, что $\sum b_n^2 < \infty$ тогда и только тогда, когда все корни удовлетворяют неравенству $|s_j| > 1$. Легко проверить, что это остается справедливым также в случае кратных корней. Мы таким образом сейчас показали, что стационарное решение авторегрессионной модели (7.6) существует всегда, когда все корни полинома $A(s)$ лежат вне единичного круга. Ковариации нашего процесса задаются формулой (7.4), и «бесконечно далекое прошлое» не играет в процессе никакой роли.

Полученное решение $\{X_n\}$ уравнения авторегрессии (7.6) единственно. В самом деле, разность двух решений должна удовлетворять однородному уравнению (7.13), и теперь мы покажем, что условие $|s_j| > 1$ исключает появление имеющего вероятностный смысл решения этого уравнения.

в) *Вырожденные процессы.* Обратимся к стационарным последовательностям $\{Y_n\}$, удовлетворяющим стохастическому раз-

ностному уравнению

$$a_0 Y_n + a_1 Y_{n-1} + \dots + a_N Y_{n-N} = 0. \quad (7.13)$$

Эти процессы интересны своей противоположностью процессам авторегрессии, определяемым уравнением (7.6). Типичными являются примеры

$$Y_n = \lambda (Z_1 \cos n\omega + Z_{-1} \sin n\omega), \quad (7.14)$$

$$Y_n = \alpha_1 Z_1 + (-1)^n \alpha_2 Z_{-1}, \quad (7.15)$$

где коэффициенты и величина ω постоянны, а Z_1 и Z_{-1} являются независимыми нормальными случайными величинами, нормированными условиями (7.2). Эти процессы удовлетворяют уравнению (7.13), первый с $a_0 = a_2 = 1$ и $a_1 = -2 \cos \omega$, второй с $a_0 = -a_2 = 1$ и $a_1 = 0$. Эти процессы вырождены в том смысле, что весь процесс полностью определяется двумя наблюдениями, скажем Y_{k-1} и Y_k . Эти два наблюдения можно взять как угодно далеко в прошлом, и в этом смысле процесс полностью определяется своим «бесконечно далеким прошлым». Аналогичные рассуждения применимы к любому процессу, удовлетворяющему разностному уравнению вида (7.13), и, следовательно, эти процессы составляют противоположность примеру б), где «бесконечно далекое прошлое» совсем не имеет влияния. ►

Эти примеры объясняют тот широкий интерес, который проявляется по отношению к стохастическому разностному уравнению (7.13). Прежде чем перейти к соответствующей теории, заметим, что любой процесс $\{Y_n\}$, удовлетворяющий (7.13), удовлетворяет также различным разностным уравнениям более высокого порядка, например

$$a_0 Y_n + (a_1 - a_0) Y_{n-1} + \dots + (a_N - a_{N-1}) Y_{n-N} - a_N Y_{n-N-1}.$$

Для того чтобы придать задаче смысл, мы должны предположить, что (7.13) представляет собой разностное уравнение *наимизшего порядка*, которому удовлетворяет $\{Y_n\}$. Это равносильно тому, что строка $(Y_1, \dots, Y_n)$ имеет невырожденное нормальное распределение в N измерениях. Отсюда следует, что $a_0 \neq 0$ и $a_N \neq 0$.

Теперь можно показать, что теория стационарных решений разностного уравнения (7.13) тесно связана с «характеристическим уравнением»

$$a_0 \xi^N + a_1 \xi^{N-1} + \dots + a_N = 0. \quad (7.16)$$

Каждому квадратному множителю полинома, стоящего слева, здесь соответствует стохастическое разностное уравнение второго порядка и вследствие этого процесс вида (7.14) или (7.15). В соответствии с разложением на множители характеристического полинома мы представим, таким образом, общее решение уравнения (7.13) как сумму компонент вида (7.14) и (7.15).

Как и раньше, мы предполагаем $E(Y_n) = 0$. Вся теория строится на следующей лемме.

Лемма 1. *Стационарная последовательность с $E(Y_n Y_{n+k}) = r_k$ удовлетворяет стохастическому разностному уравнению (7.13) в том и только том случае, когда*

$$a_0 r_n + a_1 r_{n-1} + \dots + a_N r_{n-N} = 0. \quad (7.17)$$

Доказательство. Умножив (7.13) на Y_0 и взяв математические ожидания, приходим к (7.17). Возводя в квадрат левую часть равенства (7.13) и вновь вычисляя математические ожидания, получаем в результате $\sum a_j (\sum a_k r_{k-j})$, и поэтому из (7.17) следует, что левая часть в (7.13) имеет нулевую дисперсию. Это доказывает лемму. ►

Мы приступаем к выводу канонической формы для r_n . Эти величины, конечно, действительны, но так как в формулу входят корни характеристического уравнения (7.16), мы должны прибегнуть к временному использованию комплексных чисел.

Лемма 2. *Если $\{Y_n\}$ удовлетворяет уравнению (7.13), но не удовлетворяет никакому разностному уравнению низшего порядка, то характеристическое уравнение (7.16) обладает N различными корнями $\xi_1, \dots, \xi_N$, по модулю равными единице. В этом случае*

$$r_n = c_1 \xi_1^n + \dots + c_N \xi_N^n, \quad (7.18)$$

где $c_j > 0$ при $j = 1, \dots, N$.

Доказательство. Предположим вначале, что характеристическое уравнение (7.16) имеет N различных корней $\xi_1, \dots, \xi_N$. Найдем решение уравнения (7.17) методом частных решений, который применялся для подобных целей в первом томе. Проверка устанавливает, что (7.18) изображает формальное решение (7.17), зависящее от N свободных параметров $c_1, \dots, c_N$. Теперь r_n полностью определяются N значениями $r_1, \dots, r_N$. Поэтому для того, чтобы показать, что *каждое* решение (7.17) имеет вид (7.18), достаточно показать, что c_j могут быть выбраны так, что соотношения (7.18) приводят к предписанным значениям $r_1, \dots, r_N$. Это означает, что c_j должны удовлетворять N линейным уравнениям с матрицей A , элементами которой служат $a_{jk} = \xi_k^j$ ($j, k = 1, \dots, N$). Детерминант A не равен нулю¹⁾, и, следовательно, искомое решение существует.

¹⁾ Этот детерминант обычно называется в честь Вандермонда. Чтобы показать, что детерминант не равен нулю, заменим ξ некоторой свободной переменной x . Проверка показывает, что детерминант имеет при этом вид $xP(x)$, где P есть полином степени $N-1$. Далее $P(x) = 0$ при $x = \xi_2, \dots, \xi_N$, потому что при этих значениях x два столбца детерминанта становятся одинаковыми. Поэтому детерминант не обращается в нуль ни при каком другом значении x , и в частности при $x = \xi_1$.

Таким образом, мы установили, что (в случае различных корней) r_n действительно имеет вид (7.18). Теперь мы покажем, что фактически в (7.18) присутствуют только корни, равные по модулю единице. Мы знаем, что $a_N \neq 0$ и поэтому 0 не может быть корнем характеристического уравнения. Далее отметим, что ковариации r_n ограничены общим значением r_0 дисперсии Y_n . Однако если ξ_j не равны по модулю единице, то $|\xi_j|^n \rightarrow \infty$ при $n \rightarrow \infty$ или при $n \rightarrow -\infty$. Отсюда следует, что при любом j либо $|\xi_j| = 1$, либо же $c_j = 0$.

Предположим теперь, что ξ_1 и ξ_2 представляют собой пару сопряженных корней и что $c_1 \neq 0$. Тогда ξ_1 равен по модулю единице и, следовательно, $\xi_2 = \xi_1^{-1}$. В силу симметрии $r_n = r_{-n}$, поэтому $c_2 = c_1$. С другой стороны, $\xi_1^n + \xi_2^n$ действительно, и поэтому значение c_1 должно быть действительным. Таким образом комплексные корни присутствуют в (7.18) в сопряженных парах с действительными коэффициентами, и если некоторый коэффициент c_j равен нулю, то r_n будет удовлетворять разностному уравнению порядка меньшего, чем N . Таким образом, все корни равны по модулю единице, все c_j действительны и $c_j \neq 0$.

Для того чтобы убедиться в том, что c_j положительны, рассмотрим ковариационную матрицу R для случайных величин $(Y_1, \dots, Y_N)$. Элементами матрицы R являются r_{j-k} , и легко вывести из (7.18), что

$$R = AC\bar{A}^T, \quad (7.19)$$

где C — диагональная матрица с элементами c_j , матрица A введена выше, а матрица $\bar{A}$ сопряжена с A (т. е. получается из A заменой ξ_j на $\bar{\xi}_j^{-1}$). Далее, R является действительной и положительно определенной матрицей и поэтому для любой комплексной N -мерной ненулевой вектор-строки $x = u + iv$ имеет место соотношение

$$xR\bar{x}^T = uRu^T + vRv^T > 0. \quad (7.20)$$

Полагая $y = xA$, получаем вместо (7.20)

$$y\bar{C}y^T = \sum_{j=1}^N c_j |y_j|^2 > 0. \quad (7.21)$$

Поскольку детерминант A не равен нулю, последнее неравенство выполняется для произвольного y и, таким образом, $c_j > 0$, как и утверждалось.

Для завершения доказательства нам осталось показать, что характеристическое уравнение не имеет кратных корней. Предположим, что $\xi_1 = \xi_2$, а другие корни различны. Мы снова получим представление в форме (7.18) с той разницей, что член $c_1 \xi_1^n$ заменится членом $c_1 n \xi_1^n$. Ограниченность r_n снова с необходимостью приводит к тому, что $c_1 = 0$. Следовательно, в случае одного

двойного корня мы получим представление вида (7.18) с числом отличных от нуля членов, меньшим N , а, как мы уже видели, это невозможно. Те же аргументы применимы и в более общей ситуации и показывают, что кратные корни невозможны.

Сформулируем теперь окончательный результат для случая, когда N — нечетное целое число. Изменения, необходимые для перехода к четному N , должны быть очевидны.

Теорема. *Предположим, что стационарная последовательность $\{Y_n\}$ удовлетворяет разностному уравнению (7.13) с $N = 2\nu + 1$, но не удовлетворяет никакому разностному уравнению низшего порядка. Тогда характеристическое уравнение (7.16) обладает ν парами комплексных корней $\xi_j = \cos \omega_j \pm i \sin \omega_j$ (где ω_j вещественны) и одним вещественным корнем $\omega_0 = \pm 1$. Последовательность $\{Y_n\}$ имеет вид*

$$Y_n = \lambda_0 Z_0 \omega_0^n + \sum_{j=1}^{\nu} \lambda_j [Z_j \cos n\omega_j + Z_{-j} \sin n\omega_j], \quad (7.22)$$

где Z_j — взаимно независимые нормальные величины с нулевыми математическими ожиданиями и единичными дисперсиями, а λ_j — постоянные. Для этой последовательности

$$r_n = \lambda_0^2 \omega_0^n + \sum_{j=1}^{\nu} \lambda_j^2 \cos n\omega_j. \quad (7.23)$$

Обратно, выберем произвольные действительные $\lambda_\nu \neq 0$ и $\omega_0 = \pm 1$. Пусть $\omega_1, \dots, \omega_\nu$ — различные действительные числа, такие, что $0 < \omega_j < \pi$. Тогда (7.22) определяет стационарный процесс с ковариациями (7.23), удовлетворяющий разностному уравнению порядка $2\nu + 1$ (но не удовлетворяющий никакому разностному уравнению низшего порядка).

Доказательство. Пусть λ_j и ω_j и нормальные случайные величины Z_j удовлетворяют условиям теоремы. Определим величины Y_n формулой (7.22). Простые вычисления показывают, что ковариации r_n величин $\{Y_n\}$ задаются равенством (7.23). Существует алгебраическое уравнение с действительными коэффициентами вида (7.16) с корнями, описанными в теореме. Тогда ковариации r_j удовлетворяют разностному уравнению (7.17) и в силу леммы 1 отсюда следует, что Y_n удовлетворяет стохастическому разностному уравнению (7.13). По построению это есть разностное уравнение наименьшего порядка, которому удовлетворяет Y_n .

Обратно, пусть $\{Y_n\}$ обозначает решение данного разностного уравнения (7.13). Ковариации r_n величин $\{Y_n\}$ определяют числа λ_j и ω_j , присутствующие в (7.22). Рассмотрим эти уравнения при $n = 0, 1, \dots, 2\nu$ как линейное преобразование произвольной строки из нормальных величин $(Z_{-\nu}, \dots, Z_\nu)$ в $(Y_0, \dots, Y_n)$. Это преоб-

разование невырождено, и поэтому ковариационные матрицы для $(Z_{-v}, \dots, Z_v)$ и $(Y_0, \dots, Y_N)$ определяют одна другую единственным образом. Мы только что показали, что если ковариационная матрица величин Z_j сводима к единичной матрице, то величины Y_k имеют своими ковариациями r_n . Поэтому обратное также справедливо, и тогда существуют нормальные величины Z_j , удовлетворяющие условиям теоремы и такие, что (7.22) выполняется при $n=0, \dots, N$. Однако обе части каждого из этих уравнений представляют собой решения стохастического разностного уравнения (7.13), и поскольку они совпадают при $n=0, \dots, N$, то они с необходимостью тождественны.

§ 8. МАРКОВСКИЕ НОРМАЛЬНЫЕ ПЛОТНОСТИ

Мы перейдем к обсуждению одного класса нормальных плотностей, встречающихся в марковских процессах. Не ограничивая общности, мы рассмотрим только плотности, *центрированные в начале координат*. Тогда $E(X_k) = 0$, и мы используем обычные сокращения

$$E(X_k^2) = \sigma_k^2, \quad E(X_j X_k) = \sigma_j \sigma_k \rho_{jk}. \quad (8.1)$$

Величины ρ_{jk} суть *коэффициенты корреляции*, и $\rho_{kk} = 1$.

Определение. *Нормальная r -мерная плотность $(X_1, \dots, X_r)$ является марковской, если при $k \leq r$ условная плотность X_k при данных $X_1, \dots, X_{k-1}$ тождественна с условной плотностью X_k при данной X_{k-1} .*

Грубо говоря, если мы знаем X_{k-1} («настоящее»), то дополнительное знание «прошлого» $X_1, \dots, X_{k-2}$ не приносит никакой полезной информации относительно «будущего», т. е. относительно любой из величин X_j с $j \geq k$.

Как обычно в подобных случаях, мы применяем термин «марковский» и к последовательности $(X_1, \dots, X_r)$, и к ее плотности.

Теорема 1. *Для того чтобы последовательность $(X_1, \dots, X_r)$ была марковской, каждое из следующих двух условий является необходимым и достаточным:*

(i) для $k \leq r$

$$E(X_k | X_1, \dots, X_{k-1}) = E(X_k | X_{k-1}); \quad (8.2)$$

(ii) для $j \leq v < k \leq r$

$$\rho_{jk} = \rho_{jv} \rho_{vk}. \quad (8.3)$$

Для выполнения (8.3) достаточно, чтобы

$$\rho_{jk} = \rho_{j, k-1} \rho_{k-1, k}, \quad j \leq k. \quad (8.4)$$

Доказательство. Совпадение плотностей влечет равенство математических ожиданий, и поэтому необходимость (8.2) тривиальна.

С другой стороны, если (8.2) справедливо, то теорема 5 из § 6 показывает, что условная плотность X_k при данных $X_1, \dots, X_{k-1}$ зависит только от X_{k-1} , но не от предшествующих величин. Условная плотность X_k при данном X_{k-1} получается интегрированием относительно величин $X_1, \dots, X_{k-2}$, и, следовательно, две условные плотности совпадают. Таким образом, (8.2) необходимо и достаточно.

Обращаясь снова к теореме 5 из § 6, заключаем, что величина

$$T = X_k - E(X_k | X_{k-1}) \quad (8.5)$$

совпадает с величиной

$$T = X_k - \frac{\sigma_k}{\sigma_{k-1}} \rho_{k-1,k} X_{k-1}, \quad (8.6)$$

поскольку это есть единственная случайная величина вида $X_k - cX_{k-1}$, не коррелированная с X_{k-1} . В силу той же теоремы равенство (8.2) выполняется тогда и только тогда, когда величина T не коррелирована также с $X_1, \dots, X_{k-2}$, т. е. тогда и только тогда, когда выполняется (8.4). Таким образом, (8.4) необходимо и достаточно. Поскольку (8.4) есть частный случай (8.3), то последнее условие является достаточным. Оно также и необходимо, так как повторное применение (8.4) показывает, что при $j < v < k \leq r$

$$\frac{\rho_{jk}}{\rho_{vk}} = \frac{\rho_{j,k-1}}{\rho_{v,k-1}} = \frac{\rho_{j,k-2}}{\rho_{v,k-2}} = \dots = \frac{\rho_{jv}}{\rho_{vv}} = \rho_{jv}. \quad (8.7)$$

Поэтому из (8.4) следует (8.3). ►

Следствие. Если последовательность $(X_1, \dots, X_r)$ марковская, то каждая подпоследовательность $(X_{\alpha_1}, \dots, X_{\alpha_v})$ с $\alpha_1 < \alpha_2 < \dots < \alpha_v \leq r$ является марковской.

Это очевидно, так как (8.3) автоматически распространяется на все подпоследовательности. ►

Примеры. а) *Независимые приращения.* Говорят, что последовательность (конечная или бесконечная) $\{X_k\}$ нормальных случайных величин с $E(X_k) = 0$ образует процесс с независимыми приращениями, если при $j < k$ приращение $X_k - X_j$ не зависит от $(X_1, \dots, X_j)$. Отсюда следует, в частности, что $E(X_j(X_k - X_j)) = 0$ или

$$\rho_{jk} = \frac{\sigma_j}{\sigma_k}, \quad j < k. \quad (8.8)$$

Сравнивая это с (8.3), видим, что нормальный процесс с независимыми приращениями автоматически оказывается марковским. Его структура чрезвычайно проста: X_k есть сумма k взаимно независимых нормальных величин

$$X_1, X_2 - X_1, \dots, X_k - X_{k-1}.$$

б) *Модели авторегрессии.* Рассмотрим нормальную марковскую последовательность $X_1, X_2, \dots$ с $E(X_k) = 0$. Существует единственная постоянная a_k , такая, что $X_k - a_k X_{k-1}$ не зависит от X_{k-1} и, следовательно, от $X_1, \dots, X_{k-1}$. Положим

$$\lambda_k^2 = \text{Var}(X_k - a_k X_{k-1})$$

и, следовательно,

$$\begin{aligned} X_1 &= \lambda_1 Z_1, \\ X_k &= a_k X_{k-1} + \lambda_k Z_k \quad k = 2, 3, \dots \end{aligned} \quad (8.9)$$

Легко видеть, что определенные таким образом величины Z_k независимы и

$$E(Z_k) = 0, \quad E(Z_k^2) = 1. \quad (8.10)$$

Верно и обратное. Если Z_k нормальны и удовлетворяют (8.10), то (8.9) определяет последовательность $\{X_n\}$, и сама по себе структура соотношений (8.9) показывает, что последовательность $\{X_n\}$ марковская. В качестве упражнения мы проверим это утверждение вычислением. Умножим (8.9) на X_j и возьмем математические ожидания. Так как Z_k не зависит от $X_1, \dots, X_{k-1}$, мы получаем при $j < k$

$$a_k = \frac{\sigma_k}{\sigma_{k-1}} \frac{\rho_{j,k}}{\rho_{j,k-1}}. \quad (8.11)$$

Теперь (8.4) есть простое следствие (8.11), а мы знаем, что выполнение (8.4) влечет за собой марковский характер X_k . Таким образом, $(X_1, \dots, X_r)$ является марковской тогда и только тогда, когда выполняются соотношения вида (8.9), где Z_j — независимые нормальные величины, удовлетворяющие (8.10). [Это есть частный случай примера 7, б).] ►

До сих пор мы рассматривали только конечные последовательности $(X_1, \dots, X_r)$, но число r не играет никакой роли, и мы можем также говорить о бесконечных последовательностях $\{X_n\}$. Это не затрагивает пространств бесконечных последовательностей или какую-либо новую теорию, это просто указание на то, что распределение для $(X_1, \dots, X_r)$ определяется при всех r . Аналогично мы говорим о марковском семействе $\{X(t)\}$, когда любая конечная совокупность $X_1 = X(t_1), \dots, X_r = X(t_r)$ является марковской. Ее описание зависит от функций

$$E(X^2(t)) = \sigma^2(t), \quad E(X(s)X(t)) = \sigma(s)\sigma(t)\rho(s, t). \quad (8.12)$$

В силу критерия (8.3) очевидно, что семейство является марковским тогда и только тогда, когда при $s < t < \tau$

$$\rho(s, t)\rho(t, \tau) = \rho(s, \tau). \quad (8.13)$$

Терминология не должна затемнять то, что фактически мы будем иметь дело только с семействами конечномерных нормальных распределений с ковариациями, которые удовлетворяют (8.13).

Как обстоятельно объяснялось в начале § 7, последовательность $\{X_n\}$ называют *стационарной*, если для каждого фиксированного набора $\alpha_1, \dots, \alpha_n$ распределение $(X_{z_1+v}, \dots, X_{z_n+v})$ не зависит от v . Конечный отрезок такой последовательности может быть продолжен в обе стороны, и, следовательно, естественно рассматривать только бесконечные в обе стороны последовательности $\{\dots, X_{-2}, X_{-1}, X_0, X_1, \dots\}$. Эти понятия тривиальным образом переносятся на семейства $\{X(t)\}$.

Для стационарной последовательности $\{X_n\}$ дисперсия σ_n^2 не зависит от n и в марковском случае (8.3) влечет равенство $\rho_{jk} = \rho_{12}^{j-k-1}$. Таким образом, для *стационарной марковской последовательности*

$$E(X_j X_k) = \sigma^2 \rho^{|k-j|}, \quad (8.14)$$

где σ^2 и ρ — постоянные, $|\rho| < 1$. Обратно, последовательность с нормальными распределениями, удовлетворяющая (8.14), является марковской и стационарной.

В случае стационарного семейства $\{X(t)\}$ корреляция $\rho(s, t)$ зависит только от разности $|t-s|$ и (8.13) приобретает вид

$$\rho(t)\rho(\tau) = \rho(t+\tau) \text{ для } t, \tau > 0.$$

Очевидно, из равенства $\rho(\tau) = 0$ должно следовать $\rho(t) = 0$ при всех $t > \tau$ и также $\rho\left(\frac{1}{\tau}\right) = 0$. Поэтому ρ не может иметь нулей, за исключением случая $\rho(t) = 0$ при всех $t > 0$. Следовательно, $\rho(t) = e^{-\lambda t}$ (здесь используется результат из 1, гл. XVII, 6). Соответственно для *стационарного марковского семейства*

$$E(X(s)X(s+t)) = \sigma^2 e^{-\lambda t}, \quad t > 0, \quad (8.15)$$

за исключением случая, когда $X(s)$ и $X(t)$ не коррелированы при всех $s \neq t$.

Пример. в) Стационарные последовательности могут быть построены по схеме последнего примера. Вследствие (8.11) мы должны иметь

$$X_k = \rho X_{k-1} + \sigma \sqrt{1-\rho^2} Z_k. \quad (8.16)$$

При каждом k можно выразить X_k как линейную комбинацию $Z_k, Z_{k-1}, \dots, Z_{k-v}$ и X_{k-v-1} . Формальный переход к пределу приводит к представлению

$$X_k = \sigma \sqrt{1-\rho^2} \sum_{j=0}^{\infty} \rho^j Z_{k-j}, \quad (8.17)$$

т. е. к представлению $\{X_k\}$ через бесконечную в обе стороны последовательность независимых нормальных величин Z_j , нормиро-

ванных условием (8.10). Так как $|\rho| < 1$, то правдоподобно, что ряд сходится, однако формула по существу связана с пространством бесконечных последовательностей. [См. замечания, касающиеся формулы (7.5), частным случаем которой является (8.17).] ►

Полезно, быть может, обсудить связь теоремы 1 с непосредственным описанием марковских последовательностей в терминах плотностей. Обозначим через g_i плотность X_i и через $g_{ik}(x, y)$ — значение в точке y условной плотности X_k при условии $X_i = x$. (В теории случайных процессов g_{ik} называется переходной плотностью от X_i к X_k .) Для нормальных марковских последовательностей g_i представляет собой нормальную плотность с нулевым математическим ожиданием и дисперсией σ_i^2 . Что касается условных вероятностей, то в примере 2, а) было показано, что

$$g_{ik}(x, y) = \frac{1}{\sigma_k \sqrt{1 - \rho_{ik}^2}} \pi \left(\frac{y - \sigma_i^{-1} \rho_{ik} \sigma_k x}{\sigma_k \sqrt{1 - \rho_{ik}^2}} \right), \quad (8.18)$$

где π обозначает стандартную нормальную плотность. Мы не будем, однако, использовать этот результат и проведем анализ свойств g_{ik} независимым способом. Как обычно, мы интерпретируем индексы как временные параметры.

Совместная плотность (X_i, X_j) задается как $g_i(x) g_{ij}(x, y)$. Совместная плотность (X_i, X_j, X_k) равна произведению предыдущей плотности на условную плотность X_k при данных X_j и X_i , но ввиду марковского характера индекс i может быть опущен, если $i < j < k$, и плотность (X_i, X_j, X_k) становится равной

$$g_i(x) g_{ij}(x, y) g_{jk}(y, z). \quad (8.19)$$

В марковском случае плотность каждой строки $(X_{a_1}, \dots, X_{a_n})$ задается произведением вида (8.19), однако плотности g_{ik} не могут быть выбраны произвольно. В самом деле, интегрирование (8.19) относительно y дает маргинальную плотность для (X_i, X_k) . Поэтому мы получаем условие согласованности

$$g_{ik}(x, z) = \int_{-\infty}^{+\infty} g_{ij}(x, y) g_{jk}(y, z) dy \quad (8.20)$$

при всех $i < j < k$. Это есть частный случай *уравнения Колмогорова—Чепмена* для марковских процессов¹⁾. Говоря очень приблизительно, это уравнение означает, что переход из x в момент i в z в момент k происходит через все промежуточные состояния y , причем переход от y к z не зависит от прошлого. Очевидно, что при любой системе переходных вероятностей g_{ik} , удовлетворяющих уравнению Колмогорова—Чепмена, схема умножения (8.19)

¹⁾ Другие частные случаи встречались в 1, гл. XV, (13.3), и гл. XVII, (9.1). Заметим, что формула (8.19) представляет собой аналог определения (1.1) в 1, гл. XV, вероятностей для марковских цепей, за исключением того, что здесь суммирование заменено интегрированием и что были рассмотрены только стационарные переходные вероятности.

приводит к согласованной системе плотностей для $(X_1, X_2, \dots, X_r)$, и эта последовательность является марковской. Таким образом, мы пришли к следующему аналитическому дополнению теоремы 1.

Теорема 2. *Плотности семейства $\{g_{ik}\}$ тогда и только тогда могут служить переходными плотностями в нормальном марковском процессе, когда они удовлетворяют уравнению Колмогорова — Чепмена и когда $g_{ik}(x, y)$ при каждом фиксированном x является нормальной плотностью относительно y .*

Обе теоремы содержат необходимые и достаточные условия, и поэтому они в некотором смысле эквивалентны. Тем не менее они имеют различную природу. Вторая из них на самом деле не ограничивается нормальными процессами; примененная к семействам $\{X(t)\}$, она приводит к дифференциальным и интегральным уравнениям для переходных вероятностей и на этом пути способствует введению новых классов марковских процессов. С другой стороны, из теоремы 2 трудно угадать, что g_{ik} необходимо имеет вид (8.18) — результат, содержащийся в более специальной теореме 1.

Для последующих ссылок и сравнений мы приводим здесь два наиболее важных марковских семейства $\{X(t)\}$.

Примеры. г) *Броуновское движение или процесс Винера — Башелле.* Процесс определяется тем условием, что $X(0) = 0$ и что при $t > s$ величина $X(t) - X(s)$ не зависит от $X(s)$ и имеет дисперсию, зависящую только от $t - s$. Иными словами, процесс имеет независимые приращения [пример а)] и стационарные переходные вероятности [но он не стационарный, поскольку $X(0) = 0$]. Очевидно, что $E(X^2(t)) = \sigma^2 t$ и $E(X(s)X(t)) = \sigma^2 s$ при $s < t$. При $t > s$ переходные плотности, характеризующие переход из (t, x) в (τ, y) , нормальны с математическим ожиданием x и дисперсией $\sigma^2(\tau - t)$. Они зависят только от $(y - x)/(\tau - t)$, и уравнения Колмогорова — Чепмена сводятся к свертке.

д) *Процесс Орнштейна — Уленбека.* Под этим понимается наиболее общий нормальный стационарный марковский процесс с нулевыми математическими ожиданиями. Его ковариации определяются формулой (8.15). Иными словами, при $t > s$ переходные плотности, характеризующие переход из (t, x) в (τ, y) , нормальны с математическим ожиданием $e^{-\lambda(\tau-t)}x$ и дисперсией $\sigma^2(1 - e^{-2\lambda(\tau-t)})$. При $t \rightarrow \infty$ математическое ожидание стремится к нулю, а дисперсия к σ^2 . Этот процесс был рассмотрен Орнштейном и Уленбеком с совершенно другой точки зрения. Его связь с диффузией будет обсуждена в гл. X, 4. ►

§ 8. ЗАДАЧИ

1. Рассмотрим Ω — область на плоскости (площадью $\frac{1}{2}$), ограниченную четырехугольником с вершинами $(0, 0)$, $(1, 1)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 1)$ и треугольником с вершинами $(\frac{1}{2}, 0)$, $(1, 0)$, $(1, \frac{1}{2})$. (Единичный квадрат представляется

объединением Ω и области, симметричной Ω относительно биссектрисы.) Пусть пара (X, Y) распределена равномерно на Ω . Докажите, что частные распределения являются равномерными и что $X + Y$ имеет такую же плотность, как если бы X и Y были независимыми¹⁾.

Указание. Рисунок делает вычисления излишними.

2. Плотности с маргинальными нормальными плотностями. Пусть u — нечетная непрерывная функция на прямой, равная нулю вне интервала $-1, 1$. Если $|u| < (2\pi e)^{-1/2}$, то выражение

$$n(x)n(y) + u(x)u(y)$$

изображает двумерную плотность, которая не является нормальной, но маргинальные плотности которой нормальны (Э. Нелсон).

3. Второй пример. Пусть φ_1 и φ_2 — две двумерные нормальные плотности с единичными дисперсиями, но с различными коэффициентами корреляции. Смесь $\frac{1}{2}(\varphi_1 + \varphi_2)$ не является нормальной плотностью, но две ее маргинальные плотности совпадают с n .

Замечание. В последующем все случайные величины рассматриваются в $\mathcal{R}^1$. Векторные величины обозначаются парами (X_1, X_2) и т. д.

4. Пусть $X_1, \dots, X_n$ — независимые случайные величины с одинаковой плотностью f и функцией распределения F . Если X и Y являются соответственно наименьшей и наибольшей из них, то совместная плотность пары (X, Y) равна

$$n(n-1)f(x)f(y)[F(y)-F(x)]^{n-2}, \quad y > x.$$

5. Покажите, что симметричное распределение Коши в $\mathcal{R}^3$ [определенное в (1.21)] соответствует случайному вектору, длина которого распределена с плотностью $v(r) = 4\pi^{-1}r^2(1+r^2)^{-2}$ при $r > 0$.

Указание. Используйте полярные координаты и либо (1.15), либо же общее соотношение (10.4) гл. I для проекций.

6. Для распределения Коши (1.21) условная плотность X_3 при данных X_1, X_2 равна

$$v_{\xi_1, \xi_2}(z) = \frac{2}{\pi} \cdot \frac{\sqrt{(1+\xi_1^2+\xi_2^2)^3}}{(1+\xi_1^2+\xi_2^2+z^2)^2},$$

а двумерная условная плотность X_2, X_3 при условии, что $X_1 = \xi$, равна

$$v_{\xi}(y, z) = \frac{1}{\pi} \cdot \frac{1+\xi^2}{(1+\xi^2+y^2+z^2)^2},$$

7. Пусть $0 < a < 1$ и $f(x, y) = [(1+ax)(1+ay) - a]e^{-x-y-axy}$ при $x > 0, y > 0$ и $f(x, y) = 0$ в других случаях.

а) Докажите, что f есть плотность пары (X, Y) . Найдите маргинальные плотности и функцию распределения.

б) Найдите условную плотность $u_x(y)$ и $E(Y|X)$, $\text{Var}(Y|X)$.

8. Пусть f — плотность, сосредоточенная на $0, \infty$. Положим $u(x, y) = f(x+y)/(x+y)$ при $x > 0, y > 0$ и $u(x, y) = 0$ в других случаях. Докажите, что u есть плотность в $\mathcal{R}^2$, и найдите ее ковариационную матрицу.

9. Пусть X_1, X_2, X_3 взаимно независимы и равномерно распределены на $0, 1$. Пусть $X_{(1)}, X_{(2)}, X_{(3)}$ — соответствующие порядковые статистики. Найдите плотность пары

$$\left(\frac{X_{(1)}}{X_{(2)}}, \frac{X_{(2)}}{X_{(3)}} \right)$$

¹⁾ Иными словами, распределение суммы может быть задано сверткой, даже если величины зависимы. Этот интуитивно понятный пример предложен Г. Роббинсом, Другой парадокс такого же типа см. в гл. II, 4, д).

и покажите, что эти два отношения независимы. Обобщите на n измерений.

10. Пусть X_1, X_2, X_3 независимы и одинаково показательно распределены. Найдите плотность $(X_2 - X_1, X_3 - X_1)$.

11. Частица единичной массы расщепляется на два осколка с массами X и $1 - X$. Плотность f случайной величины X сосредоточена на $[0, 1]$, и по соображениям симметрии $f(x) = f(1 - x)$. Обозначим наименьший осколок через X_1 , а наибольший осколок через X_2 . Предположим, что эти два осколка независимо друг от друга расщепляются таким же образом и дают в результате четыре осколка с массами $X_{11}, X_{12}, X_{21}, X_{22}$. Найдите (а) плотность X_{11} , (б) совместную плотность X_{11} и X_{22} . Используйте (б) для проверки (а).

12. Пусть $X_1, X_2, \dots$ независимы и имеют одну и ту же нормальную плотность ϕ . Обозначим $S_k = X_1 + \dots + X_k$. При $m < n$ найдите совместную плотность (S_m, S_n) и условную плотность для S_m при условии, что $S_n = t$.

13. В предыдущей задаче найдите условную плотность $X_1^2 + \dots + X_n^2$ при данном значении $X_1^2 + \dots + X_n^2$.

14. Пусть (X, Y) имеет двумерную нормальную плотность, центрированную в начале координат, с $E(X^2) = E(Y^2) = 1$ и $E(XY) = \rho$. При переходе к полярным координатам (X, Y) превращается в (R, Φ) , где $R^2 = X^2 + Y^2$. Докажите, что Φ имеет плотность, равную

$$\frac{\sqrt{1-\rho^2}}{2\pi(1-2\rho\sin\varphi\cos\varphi)}, \quad 0 < \varphi < 2\pi,$$

и распределена равномерно тогда и только тогда, когда $\rho = 0$. Выведите, что

$$P\{XY > 0\} = \frac{1}{2} + \pi^{-1} \arcsin \rho \quad \text{и} \quad P\{XY < 0\} = \pi^{-1} \arccos \rho.$$

15. Пусть f — равномерная плотность в треугольнике с вершинами $(0, 0)$, $(0, 1)$, $(1, 0)$ и g — равномерная плотность в симметричном треугольнике в третьем квадранте. Найдите $f * f$ и $f * g$.

Предостережение. Требуется утомительное отдельное рассмотрение отдельных интервалов.

16. Пусть f — равномерное распределение в единичном круге. Найдите $f * f$ в полярных координатах.

17. Пусть u и v — плотности в $\mathcal{R}^2$ вида

$$u(x, y) = f(\sqrt{x^2 + y^2}), \quad v(x, y) = g(\sqrt{x^2 + y^2}).$$

Найдите $u * v$ в полярных координатах.

18. Пусть $X = (X_1, \dots, X_r)$ имеет r -мерную нормальную плотность. Существует такой единичный вектор $a = (a_1, \dots, a_r)$, что

$$\text{Var}(a_1 X_1 + \dots + a_r X_r) \geq \text{Var}(c_1 X_1 + \dots + c_r X_r)$$

для всех единичных векторов $c = (c_1, \dots, c_r)$. Если $a = (1, 0, \dots, 0)$ — такой вектор, то X_1 не зависит от остальных величин X_j .

19. Докажите теорему.

Теорема. Пусть дана нормальная плотность в $\mathcal{R}^r$. Тогда оси координат могут быть повернуты таким образом, что новые координатные величины будут взаимно независимыми нормальными величинами.

Иными словами, в теореме 2 из § 6 в качестве матрицы C может быть взята матрица вращения.

Указание. Пусть $Y = XC$ и матрица C выбрана так, что

$$Y_r = a_1 X_1 + \dots + a_r X_r,$$

где $a = (a_1, \dots, a_r)$ есть вектор из задачи 18. Остальное несложно.

20. Найдите общий нормальный стационарный процесс, удовлетворяющий

соотношениям:

а) $X_{n+2} + X_n = 0$;

б) $X_{n+2} - X_n = 0$;

в) $X_{n+2} - X_{n+1} + X_{n+1} - X_n = 0$.

21. *Сервостохастический процесс* (Г. Д. Миллс). Сервомеханизм подвергается случайным толчкам, но в любое время могут быть введены поправки. Таким образом, ошибка Y_n в момент n имеет (в подходящих единицах) вид $Y_{n+1} = Y_n + C_n + X_{n+1}$, где C_n — поправка, а X_n — независимые нормально распределенные величины с $E(X_n) = 0$, $E(X_n^2) = 1$. Величины C_n в принципе являются произвольными функциями прошлых наблюдений, т. е. Y_k и X_k при $k \leq n$. Желательно выбрать их так, чтобы минимизировать дисперсию $\text{Var}(Y_n)$, которая представляет собой меру того, сколь хорошо работает механизм, и дисперсию $\text{Var}(C_n)$, которая указывает, сколь тяжелы условия его работы.

а) Рассмотрите ковариационную функцию $\{Y_n\}$ и покажите, что $\text{Var}(Y_n) \geq 1$.

б) В предположении, что $\text{Var}(C_n) \rightarrow \alpha^2$, $\text{Var}(Y_n) \rightarrow \sigma^2$ (тенденция к стационарности), покажите, что $\sigma > \frac{1}{2}(\alpha + \alpha^{-1})$.

в) Рассмотрите, в частности, линейную схему $C_n = a - p(Y_n - b)$, $0 < p \leq 1$. Найдите ковариационную функцию и представление вида (7.8) для Y_n .

22. *Продолжение*. Если существует запаздывание во времени в поступлении информации или корректировке, то модель остается, по существу, той же самой, за исключением того, что C_n следует заменить на C_{n+N} . Обсудите эту ситуацию.

Как уже говорилось во введении, технический аппарат теории меры используется в этой книге лишь в небольшой степени, и для понимания большей ее части настоящая глава не нужна¹⁾. Тем не менее желательно дать краткий обзор важнейших понятий, образующих теоретический фундамент книги, а также привести основные теоремы для удобства ссылок. Лежащие в основе идеи и факты нетрудны, однако доказательства в теории меры включают в себя нагромождение технических деталей. Для начинающего и неспециалиста подступ затрудняется многогранностью теории меры, а также разнообразием ее применений. Существуют превосходно написанные вводные курсы, однако материал в них в силу необходимости дается в слишком общей форме, и, кроме того, в них излагаются различные аспекты теории меры, не являющиеся существенными для настоящей книги. Даваемый ниже обзор содержит главным образом лишь то, что нужно для дальнейшего; многие доказательства и технические детали опускаются²⁾. (Следует отметить, что простота теории обманчива: значительно более трудные задачи теории меры возникают в связи со случайными процессами, зависящими от непрерывного временного параметра. Изложение условных математических ожиданий откладывается до гл. V, 10, 11; теорема Радона—Никодима содержится в гл. V, 3.)

Формулы, относящиеся к евклидовым пространствам $\mathcal{M}$, не зависят от числа измерений, и x в них нужно понимать как сокращенную запись для $(x_1, \dots, x_r)$.

¹⁾ Это относится к читателям, знакомым с основами теории меры, а также к тем читателям, которые интересуются главным образом результатами и фактами. Для удобства последних определение интеграла повторяется в гл. V, 1. За пределами этого материала они могут полагаться на свою интуицию, поскольку, по существу, теория меры дает обоснование простым формальным действиям.

²⁾ Бэровские функции и интегрирование по Лебегу—Стилтьесу превосходно изложены в книге Макшейна и Боттса, E. J. McShane and T. A. Bots Real analysis, D. Van Nostrand, Princeton, 1959. Результаты общей теории меры широко используются в изложении Халмоша, см.: Халмош П. Р., Теория меры, ИЛ, М., 1953, и Бурбаки, см.: N. Bourbaki, Elements de mathematiques [Livre VI, chapitres 3—5], Herman, Paris, 1952, 1956. Изложение общей теории меры с целью применения в теории вероятностей содержится в книгах Дуба, Крикеберга, Лоэва, Невё, Анникьяна и Тортра. См. также Парасарати К, Введение в теорию вероятностей и теорию меры, —М., Мир, 1983,

§ 1. БЭРОВСКИЕ ФУНКЦИИ

Нам нужно указать класс множеств, для которых определены вероятности, а также класс функций, являющихся случайными величинами. Эти две задачи связаны между собой, и, более того, их изложение объединяется применением современных обозначений. Введем сначала эти обозначения и напомним определение сходимости в терминах монотонных пределов.

*Индикатором*¹⁾ множества A называется функция, равная единице во всех точках A и равная нулю во всех точках множества A' , дополнительного к A . Индикатор A обозначается 1_A , так что $1_A(x) = 1$ при $x \in A$ и $1_A(x) = 0$ в противном случае. Каждому множеству отвечает его индикатор, и каждая функция, принимающая лишь значения 0 и 1, является индикатором некоторого множества. Если f — произвольная функция, то произведение $1_A f$ есть функция, равная f на A и нулю в остальных точках.

Рассмотрим теперь пересечение $C = A \cap B$ двух множеств. Его индикатор 1_C равен нулю, когда либо 1_A , либо 1_B обращается в нуль, так что $1_C = \inf(1_A, 1_B)$, т. е. значение 1_C равно наименьшему из значений 1_A , 1_B . Этот параллелизм отражается в обозначении $f \cap g$, которое используется вместо $\inf(f, g)$ для функции, равной в каждой точке x наименьшему из значений $f(x)$ и $g(x)$. Аналогично $f \cup g = \sup(f, g)$ обозначает наибольшее из двух значений²⁾. Операторы $\cap$ и $\cup$ применимы к любому множеству функций. Обычно пользуются сокращенными обозначениями

$$f_1 \cap \dots \cap f_n = \bigcap_{k=1}^n f_k, \quad f_1 \cup \dots \cup f_n = \bigcup_{k=1}^n f_k. \quad (1.1)$$

По определению в каждой точке x эти функции равны соответственно минимуму и максимуму среди n значений $f_1(x), \dots, f_n(x)$. Если f_k есть индикатор множества A_k , то функции (1.1) являются индикаторами соответственно пересечения $A_1 \cap \dots \cap A_n$ и объединения $A_1 \cup \dots \cup A_n$.

Рассмотрим теперь *бесконечную* последовательность $\{f_n\}$. Функции, определяемые формулой (1.1), монотонны по n , так что пределы $\bigcap_{k=1}^{\infty} f_k$ и $\bigcup_{k=1}^{\infty} f_k$ вполне определены, хотя, возможно, и бесконечны. Для фиксированного j функция

$$w_j = \bigcap_{k=j}^{\infty} f_k \quad (1.2)$$

¹⁾ Этот термин был введен Лозовом. Более старый термин «характеристическая функция» неудачен в теории вероятностей, поскольку его принято относить к другому объекту.

²⁾ Многие авторы предпочитают символы $\wedge$ и $\vee$ для функций, оставляя символы $\cap$ и $\cup$ для множеств. В нашем изложении двойные обозначения не дают никаких преимуществ.

есть предел монотонной последовательности функций $f_j \cap \dots \cap f_{j+n}$; сама последовательность $\{\omega_j\}$ тоже монотонна, именно $\omega_n = \omega_1 \cup \dots$

$\dots \cup \omega_n$. В наших обозначениях $\omega_n \rightarrow \bigcup_{k=1}^{\infty} \omega_k$. По определению $\omega_n(x)$ есть нижняя грань (инфимум) числовой последовательности $f_n(x)$, $f_{n+1}(x)$, $\dots$. Следовательно, предел последовательности ω_n совпадает с нижним пределом $\liminf f_n$, так что

$$\liminf f_n = \bigcup_{i=1}^{\infty} \bigcap_{k=i}^{\infty} f_k. \quad (1.3)$$

При таком подходе нижний предел $\liminf$ получается двумя последовательными переходами к пределу в *монотонных* последовательностях. Аналогичное справедливо и для верхнего предела $\limsup f_n$, равенство (1.3) сохранится, если в левой его части $\liminf$ заменить на $\limsup$, а в правой части значки $\cap$ и $\cup$ поменять местами.

Все сказанное переносится на множества. Так, например, по определению $A = \lim A_n$ тогда и только тогда, когда $1_A = \lim 1_{A_n}$. Это означает, что последовательность множеств $\{A_n\}$ сходится к множеству A тогда и только тогда, когда каждая точка A принадлежит *всем* A_n , за исключением не более чем конечного числа, и каждая точка дополнения A' принадлежит не более чем конечному числу множеств A'_n .

Пример. а) Множество $\{A_n \text{ б. ч.}\}^1$. В качестве важного с вероятностной точки зрения примера предельных операций над множествами рассмотрим событие A , состоящее в «реализации бесконечного числа событий из данной последовательности событий $A_1, A_2, \dots$ ». [Частные случаи были рассмотрены в I; гл. VIII, 3 (леммы Бореля—Кантелли), и в I; гл. XIII (рекуррентные события).] На более формальном языке это означает, что точка x принадлежит множеству A тогда и только тогда, когда она принадлежит бесконечному числу множеств последовательности $\{A_k\}$. Поскольку индикаторы могут принимать лишь два значения 0 и 1, то данное определение равносильно следующему: $1_A = \limsup 1_{A_n}$. В обычных обозначениях это записывается как $A = \limsup A_n$, однако обозначение $\{A_n \text{ б. ч.}\}$ (читается A_n «бесконечно часто») более приятно и наглядно. Оно было введено Чжуном и теперь общепринято в вероятностной литературе. ►

¹ В подлиннике « $A_n \text{ i. o.}$ » (infinitely often). — *Прим. перев.*

² В принципе мы интересуемся лишь конечнозначными функциями, однако иногда удобно допустить также в качестве возможных значений $\pm \infty$. Например, простая теорема, утверждающая сходимость всякой монотонной последовательности, неверна для конечнозначных функций, а без нее многие формулировки становятся тяжеловесными. По этой причине мы следуем обычному соглашению, по которому все функции принимают значения из расширенной

Наша следующая задача — ограничить класс функций¹⁾ на $\mathcal{R}'$, с которыми мы предполагаем иметь дело. Понятие произвольной функции слишком общо, чтобы быть полезным для наших целей. Более подходящим является модернизированный вариант эйлера-вого определения функции. Если считать, что непрерывные функции заданы, то единственный эффективный способ конструирования новых функций состоит в предельных переходах. Оказывается, все наши потребности будут удовлетворены, если мы будем знать, как обращаться с функциями, являющимися пределами последовательностей $\{f_n\}$ непрерывных функций или пределами последовательностей, в которых каждая f_n является таким пределом и т. д. Иными словами, нам интересен класс $\mathfrak{B}$ функций, обладающий следующими свойствами: 1) каждая непрерывная функция принадлежит $\mathfrak{B}$ и 2) если функции $f_1, f_2, \dots$ принадлежат $\mathfrak{B}$ и предел $f(x) = \lim f_n(x)$ существует при всех x , то f тоже принадлежит $\mathfrak{B}$. Такой класс называется *замкнутым* относительно поточечных пределов. Ясно, что такие классы существуют, например класс *всех* функций. Пересечение всех таких классов является замкнутым семейством и образует, очевидно, наименьший такой класс. Разумно ограничить наши рассуждения этим наименьшим классом.

Наименьший замкнутый класс функций, содержащий все непрерывные функции, называется классом Бэра и будет обозначаться $\mathfrak{B}$. Функции из $\mathfrak{B}$ называются бэровскими функциями²⁾.

Понятие бэровской функции будет использоваться не только по отношению к функциям, заданным на всем пространстве, но и по отношению к функциям, определенным лишь на некотором подмножестве (например, $\sqrt{x}$ и $\log x$ в случае пространства $\mathcal{R}^1$).

Из данного определения ясно, что сумма и произведение двух бэровских функций тоже являются бэровскими функциями. Более того, если ψ — непрерывная функция от r переменных и $f_1, \dots, f_r$ — бэровские функции, то функция $\psi(f_1, \dots, f_r)$ тоже бэровская. Заменяя ψ на ψ_n и переходя к пределу, можно показать, что вообще *всякая бэровская функция от бэровских функций есть бэровская функция*. Фиксация значений одной или нескольких переменных приводит опять к бэровской функции и т. д. Короче говоря, никакая из обычных операций над бэровскими функциями не выходит за пределы класса $\mathfrak{B}$, и, следовательно, класс $\mathfrak{B}$ представляет собой естественный объект для изучения. Оказывается, что никакого упрощения не достигается, если ограничиться меньшими классами.

вещественной прямой, т. е. их значения суть числа или $\pm\infty$. На практике значения $\pm\infty$ не будут играть роли. Чтобы сумма и произведение двух функций были определены, вводят следующие соглашения: $\infty + \infty = \infty$, $\infty - \infty = 0$, $\infty \cdot \infty = \infty$, $0 \cdot \infty = 0$ и т. д.

²⁾ Это определение основано лишь на понятии непрерывности и не зависит от других свойств евклидовых пространств. Его, следовательно, можно перенести на любое топологическое пространство.

§ 2. ФУНКЦИИ ИНТЕРВАЛОВ И ИНТЕГРАЛЫ В $\mathcal{R}'$

Интервалом мы будем называть множество точек, удовлетворяющих двойному неравенству одного из следующих четырех типов (попутно вводятся очевидные обозначения):

$$\begin{array}{ll} \overline{a, b}: a < x < b & \overline{a, b}: a < x \leq b, \\ \overline{a, b}: a \leq x \leq b & \overline{a, b}: a \leq x < b. \end{array}$$

В одномерном случае выписанные неравенства исчерпывают все возможные интервалы, включая вырожденный интервал нулевой длины. В двумерном случае эти неравенства понимаются в поординатном смысле, и интервалы представляют собой (возможно, вырожденные) прямоугольники со сторонами, параллельными осям координат. При числе измерений, большем или равном двум, возможны и другие типы интервалов, однако мы их исключаем из рассмотрения. Допускается предельный случай, когда одна или более координат a или b равны $\pm\infty$. В частности, все пространство есть интервал $-\infty, \infty$.

Функция точек f сопоставляет значения $f(x)$ отдельным точкам. Функция множеств F сопоставляет значения отдельным множествам или областям пространства. Объем в $\mathcal{R}^3$, площадь в $\mathcal{R}^2$ или длина в $\mathcal{R}^1$ являются типичными примерами функций множеств. Помимо этих есть много других функций множеств, и среди них вероятности представляют собой наиболее интересный для нас частный случай. Мы будем интересоваться лишь функциями множеств, обладающими следующим свойством: если множество A разбито на две части A_1 и A_2 , то $F\{A\} = F\{A_1\} + F\{A_2\}$. Такие функции называются *аддитивными*¹⁾.

Как мы видели, вероятности $F\{I\}$ часто определены для всех интервалов r -мерного пространства $\mathcal{R}'$, и их желательно определить на более общих множествах. Та же самая задача возникает в элементарном анализе, где площадь первоначально определена только для прямоугольников, и желательно определить ее для общей области A . Простейший подход состоит в том, что сначала определяется интеграл от функции двух переменных и затем «площадь A » приравнивается к интегралу от индикатора 1_A (т. е. к функции, равной 1 на A и обращающейся в нуль вне A). Аналогично мы определим интеграл

$$E(u) = \int_{\mathcal{R}'} u(x) F\{dx\} \quad (2.1)$$

¹⁾ Вот несколько примеров аддитивных функций из практики: масса и количество тепла в области, цена земли, площадь под пшеницей и число жителей в географическом районе, ежегодная добыча угля, расстояние, которое пролетел пассажир, или число израсходованных килоплатт-часов в течение некоторого периода, число телефонных вызовов и т. д.

от функции точек и по отношению к функции интервалов F . Вероятность A определяется затем как $E(1_A)$. При построении интеграла (2.1) неважно, как интерпретируется функция F , и мы, по существу, дадим общее понятие интеграла Лебега—Стилтьеса. Приступим теперь к аккуратному выполнению намеченной программы.

Пусть F —функция, сопоставляющая каждому интервалу I конечное значение $F\{I\}$. Такая функция называется (конечно) *аддитивной*, если для каждого разбиения интервала I на конечное число непересекающихся интервалов $I_1, \dots, I_n$

$$F\{I\} = F\{I_1\} + \dots + F\{I_n\}. \quad (2.2)$$

Примеры. а) *Распределения в $\mathcal{R}^1$* . В первом томе мы рассматривали дискретные вероятностные распределения, сопоставляющие вероятности $p_1, p_2, \dots$ точкам $a_1, a_2, \dots$. Здесь $F\{I\}$ есть сумма масс p_n во всех точках a_n , содержащихся в I , и $E(u) = \sum u(a_n) p_n$.

б) Если G —какая-либо непрерывная монотонная функция, возрастающая от 0 на $-\infty$ до 1 на ∞ , то можно положить $F\{a, b\} = G(b) - G(a)$.

в) *Случайные векторы в $\mathcal{R}^2$* . Вектор единичной длины выходит из начала координат в случайном направлении. Вероятность того, что его конечная точка лежит в двумерном интервале I , пропорциональна длине дуги, по которой I пересекается с единичной окружностью. Определяемое здесь непрерывное распределение вероятностей не имеет плотности. Оно *сингулярно* в том смысле, что вся вероятность сосредоточена на окружности. Можно думать, что такие распределения искусственны и что в рассматриваемом случае скорее окружность, а не плоскость должна быть естественным выборочным пространством. Это возражение неосновательно, поскольку сумма двух таких независимых случайных векторов может принять любое значение между 0 и 2, и имеет положительную плотность в круге радиуса 2 (см. пример 4, д)). Следовательно, в некоторых задачах, связанных с единичными случайными векторами, плоскость является естественным выборочным пространством. Во всяком случае, нашей целью было только показать на простом примере, что может происходить в более сложных ситуациях.

г) В заключение приведем пример, иллюстрирующий одно обстоятельство, которое будет *исключено* в дальнейшем. В пространстве $\mathcal{R}^1$ положим $F\{I\} = 0$ для любого интервала $I = \overline{a, b}$ с $b < \infty$ и $F\{I\} = 1$, когда $I = \overline{a, \infty}$. Так определенная функция интервалов, будучи аддитивной, является патологической в том смысле, что она не удовлетворяет естественному требованию непрерывности, согласно которому $F\{a, b\}$ должно стремиться к $F\{a, \infty\}$ при $b \rightarrow \infty$.

Последний пример показывает, что желательно усилить требование (2.2) конечной аддитивности. Мы скажем, что *функция интервалов* F *счетно- или σ -аддитивна*, если для каждого разбиения интервала I на счетное число интервалов $I_1, I_2, \dots$

$$F\{I\} = \sum F\{I_k\}. \quad (2.3)$$

«Счетное число» обозначает либо конечное число, либо бесконечное счетное число. Термины « *вполне аддитивная*» и « *счетно-аддитивная*» означают одно и то же. Условие (2.3) явно нарушено в последнем примере.

В дальнейшем мы будем рассматривать только счетно-аддитивные функции множеств. Это ограничение, с одной стороны, позволяет успешно развить теорию, а с другой — может быть оправдано а priori на эвристической и прагматической основе. В самом деле, если $A_n = I_1 \cup \dots \cup I_n$ есть объединение первых n интервалов, то $A_n \rightarrow I$. Можно считать, что «при достаточно большом n A_n практически неотлично от I ». Если значения функции F могут быть найдены из экспериментов, то $F\{A_n\}$ должно быть «практически неотлично» от $F\{I\}$, т. е. $F\{A_n\}$ должно стремиться к $F\{I\}$ ². Счетная аддитивность (2.3) и есть точное выражение этого требования.

Поскольку нас интересуют в первую очередь вероятности, мы ограничимся рассмотрением лишь неотрицательных функций интервалов F , нормированных условием $F\{-\infty, \infty\} = 1$. Требование нормированности не накладывает никаких серьезных ограничений, когда $F\{-\infty, \infty\} < \infty$, однако исключает такие функции интервалов, как длина в $\mathcal{R}^1$ или площадь в $\mathcal{R}^2$. Чтобы использовать излагаемую ниже теорию в подобных случаях, достаточно разбить прямую или плоскость на единичные интервалы и рассматривать их по отдельности. Этот прием так очевиден и настолько хорошо известен, что не требует никаких дальнейших пояснений.

Функция на $\mathcal{R}$ называется *ступенчатой функцией*, если она принимает лишь конечное число значений, причем каждое значение на некотором интервале. Пусть ступенчатая функция u принимает значения $a_1, \dots, a_n$ на интервалах $I_1, \dots, I_n$ (т. е. с вероятностями $F\{I_1\}, \dots, F\{I_n\}$) соответственно. По аналогии с определением математического ожидания дискретных случайных величин положим

$$E(u) = a_1 F\{I_1\} + \dots + a_n F\{I_n\}. \quad (2.4)$$

[Конечно, разбиение пространства на интервалы, на которых функция u постоянна, не единственно, однако, как и в дискретном случае, легко видеть, что определение (2.4) не зависит от разбиения.] Математическое ожидание $E(u)$ обладает следующими свойствами:

¹) См. Колмогоров А.Н. Основные понятия теории вероятностей, изд. 2, М., 1975, стр. 27, — *Прим., перев.*

а) *Аддитивностью* относительно линейных комбинаций:

$$E(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 E(u_1) + \alpha_2 E(u_2). \quad (2.5)$$

б) *Положительностью*:

$$\text{если } u \geq 0, \text{ то } E(u) \geq 0. \quad (2.6)$$

в) *Нормированностью*: для функции, равной во всех точках 1,

$$E(1) = 1. \quad (2.7)$$

Последние два свойства эквивалентны *теореме о среднем значении*: если $\alpha \leq u \leq \beta$, то $\alpha \leq E(u) \leq \beta$, так что функция $E(u)$ представляет собой один из видов *среднего значения*¹⁾.

Теперь задача состоит в том, чтобы продолжить определение $E(u)$ на более широкий класс функций с сохранением свойств (а)–(в). В классическом интегрировании по Риману используется тот факт, что для каждой непрерывной функции u на $\overline{0, 1}$ существует последовательность ступенчатых функций u_n , для которой $u_n \rightarrow u$ равномерно на $\overline{0, 1}$. По определению полагают $E(u) = \lim E(u_n)$. Оказывается, что требование равномерной сходимости необязательно, и то же самое определение $E(u)$ можно использовать, когда сходимость $u_n \rightarrow u$ лишь поточечная. Таким способом $E(u)$ можно продолжить на *все ограниченные бэровские функции*, и это продолжение *единственно*. При распространении $E(u)$ на неограниченные функции неизбежно появляются расходящиеся интегралы, однако, по крайней мере для *положительных бэровских функций*, можно определить $E(u)$ либо как число, либо как символ ∞ (указывающий расходямость). Никаких неприятностей в связи с этим не возникает, поскольку в лебеговской теории интегрирования рассматривается лишь абсолютная интегрируемость. Грубо говоря, исходя из определения (2.4) математического ожидания для простых функций, можно доопределить $E(u)$ для произвольных бэровских функций, используя очевидные приближения и переходы к пределу. Так определенное число $E(u)$ есть инте-

¹⁾ Когда функция F задает вероятности, среднее $E(u)$ можно интерпретировать как ожидаемый выигрыш игрока, возможные выигрыши которого суть $a_1, a_2, \dots$. Чтобы интуитивно осознать смысл среднего в других ситуациях, рассмотрим три примера, в которых $u(x)$ является соответственно температурой во время x , числом телефонных разговоров во время x , расстоянием материальной точки от начала координат, а F представляет собой соответственно длительность временного интервала, цену временного интервала (цену разговора) и механическую массу. В каждом случае интегрирование производится лишь по конечному интервалу и $E(u)$ есть соответственно средняя температура, накопленный доход и статистический момент. Эти примеры показывают, что определенное нами интегрирование по отношению к произвольной функции множеств проще и нагляднее, чем интегрирование по Риману, в котором независимая переменная играет несколько ролей и интерпретация как «площадь под кривой» ничуть не помогает начинающему. Следует иметь в виду, что понятие среднего (математического ожидания) встречается не только в теории вероятностей.

грал Лебега—Стилтьеса от функции u по отношению к F . (Когда функция F фиксирована, более предпочтителен термин «математическое ожидание» — никаких двусмысленностей при этом не возникает.) Ниже приводится без доказательства¹⁾ основной факт лебеговской теории; его природа и значение будут проанализированы в следующих параграфах. [Конструктивное определение $E(u)$ дается в § 4.]

Основная теорема. Пусть F — счетно-аддитивная функция интервалов в $\mathcal{R}$ и $F\{-\infty, \infty\} = 1$. Существует единственный интеграл Лебега—Стилтьеса $E(u)$ на классе бэровских функций, обладающий следующими свойствами:

Если $u \geq 0$, то $E(u)$ равно либо неотрицательному числу, либо ∞ . В случае произвольного u $E(u)$ существует тогда и только тогда, когда либо $E(u^+)$, либо $E(u^-)$ конечно; при этом $E(u) = E(u^+) - E(u^-)$. Функция u называется интегрируемой, если $E(u)$ конечно. Далее

(i) если u — ступенчатая функция, то $E(u)$ дается формулой (2.4);
(ii) условия (2.5)—(2.7) выполняются для всех интегрируемых функций;

(iii) (принцип монотонной сходимости) если $u_1 \leq u_2 \leq \dots \rightarrow u$ и u_n интегрируемы, то $E(u_n) \rightarrow E(u)$.

Сделав замену переменных $v_n = u_{n+1} - u_n$, можно переформулировать принцип монотонной сходимости в терминах рядов:

Если v_n интегрируемы и $v_n \geq 0$, то

$$\sum E(v_n) = E(\sum v_n), \quad (2.8)$$

причем обе стороны равенства конечны или бесконечны одновременно. Отсюда, в частности, следует, что если $v \geq u \geq 0$ и $E(u) = \infty$, то также $E(v) = \infty$.

Что будет, если в (iii) условие монотонности отбросить? Ответ на этот вопрос содержится в следующей важной лемме, имеющей широкую область применений.

Лемма Фату. Если $u_n \geq 0$ и u_n интегрируемы, то

$$E(\liminf u_n) \leq \liminf E(u_n). \quad (2.9)$$

В частности, если $u_n \rightarrow u$, то $\liminf E(u_n) \geq E(u)$.

Доказательство. Положим $v_n = u_n \cap u_{n+1} \cap \dots$. Тогда $v_n \leq u_n$ и, следовательно,

$$E(v_n) \leq E(u_n).$$

Но (как мы видели в § 1) v_n монотонно стремится к $\liminf u_n$, так что $E(v_n)$ стремится к левой части в (2.9) и лемма доказана. [Отметим, что обе стороны неравенства (2.9) могут быть равны ∞ .] ►

¹⁾ Метод доказательства намечен в § 5. Как обычно u^+ и u^- обозначают положительную и отрицательную части u , т. е. $u^+ = u \cup 0$ и $-u^- = u \cap 0$. Очевидно, $u = u^+ - u^-$.

Приводимый ниже пример д) показывает, что условие положительности не может быть отброшено, однако его можно заменить на формально более слабое условие, состоящее в том, что существует интегрируемая функция U , для которой $u_n \geq U$ (достаточно заменить u_n на $u_n - U$). Заменяя u_n на $-u_n$, устанавливаем следующий результат: если $u_n < U$ и $E(U) < \infty$, то

$$\limsup E(u_n) \leq E(\limsup u_n). \quad (2.10)$$

Для сходящихся последовательностей левая часть неравенства (2.9) совпадает с правой частью неравенства (2.10), и мы получаем следующий важный

Принцип мажорированной сходимости. Пусть u_n интегрируемы и $u_n \rightarrow u$ поточечно. Если существует интегрируемая функция U , такая, что $|u_n| \leq U$ при всех n , то u интегрируема и $E(u_n) \rightarrow E(u)$.

Эта теорема относится к единственному месту в лебеговской теории интегрирования, где наивные формальные действия могут привести к неверному результату. Необходимость условия $|u_n| \leq U$ иллюстрируется нижеследующим примером.

Пример. д) Рассмотрим $0, 1$ в качестве исходного интервала и определим математическое ожидание с помощью одномерного интеграла (по отношению к длине). Положим $u_n(x) = (n+1)(n+2) \times x^n(1-x)$. Эти функции стремятся к нулю в каждой точке, но $1 = E(u_n) \rightarrow 1$. Заменяя u_n на $-u_n$, видим, что неравенство Фату (2.9) может не выполняться для неположительных функций. ►

Отметим без доказательства одно правило обычного интегрального исчисления, применимое и в более общей ситуации.

Теорема Фубини о повторных интегралах. Если $u \geq 0$ — бэровская функция и F, G — вероятностные распределения, то с очевидной интерпретацией в случае расходящихся интегралов имеет место равенство

$$\int_{-\infty}^{\infty} F\{dx\} \int_{-\infty}^{\infty} u(x, y) G\{dy\} = \int_{-\infty}^{\infty} G\{dy\} \int_{-\infty}^{\infty} u(x, y) F\{dx\}. \quad (2.11)$$

Здесь x и y можно понимать как точки в $\mathcal{R}^m$ и $\mathcal{R}^n$, и теорема содержит в себе утверждение, что оба внутренних интеграла суть бэровские функции. (Эта теорема применима к произведению произвольных пространств, и более общий ее вариант приводится в § 6.) ►

Теорема об аппроксимации в среднем. Для всякой интегрируемой функции u и любого $\varepsilon > 0$ можно найти ступенчатую функцию v , такую, что $E(|u-v|) < \varepsilon$.

Вместо ступенчатых функций можно использовать аппроксимацию посредством непрерывных функций или посредством функций, бесконечное число раз дифференцируемых и равных нулю вне некоторого конечного интервала. [Сравните с теоремой об аппроксимации из примера VIII, 3, а).]

Замечание об обозначениях. Обозначение $E(u)$ подчеркивает зависимость от u и хорошо применимо в тех случаях, когда функция интервалов F фиксирована. Если F меняется или нужно подчеркнуть зависимость от F , то предпочтительнее интегральное обозначение (2.1). То же самое относится к интегралам по подмножеству A , поскольку интеграл от u по подмножеству A есть (по определению) интеграл от произведения $1_A u$ по всему пространству

$$\int_A u(x) F\{dx\} = E(1_A u)$$

(при этом предполагается, конечно, что индикатор 1_A есть бэровская функция). Обе части выписанного равенства означают в точности одно и то же, причем выражение, стоящее в левой части, подчеркивает зависимость от F . Когда $A = \overline{a, b}$ есть интервал, иногда предпочитают обозначение $\int_a^b$, однако чтобы сделать его точным, необходимо указать, включаются ли в интервал его конечные точки. Это можно сделать, используя запись $a+$ или $a-$. ►

В соответствии с намеченной в начале этого параграфа программой вероятность множества A полагается теперь, по определению, равной $E(1_A)$ для всех множеств, для которых 1_A есть бэровская функция. Для множеств, не обладающих этим свойством, вероятности не определяются. Теперь мы обсудим следствия из данного определения в более общей ситуации произвольного выборочного пространства.

§ 3. σ -АЛГЕБРЫ. ИЗМЕРИМОСТЬ

В дискретном случае можно было определить вероятности для всех подмножеств выборочного пространства $\mathfrak{E}$, однако в общем случае это или невозможно, или нежелательно. В предыдущих главах мы рассматривали специальный случай евклидовых пространств $\mathcal{M}$ и начинали с того, что приписывали вероятности всем интервалам. В § 2 было показано, что такое определение вероятностей может быть естественным образом распространено на более широкий класс множеств $\mathfrak{A}$. Основные свойства этого класса таковы:

(i) Если множество A принадлежит $\mathfrak{A}$, то $\mathfrak{A}$ принадлежит также его дополнение $A' = \mathfrak{E} - A$.

(ii) Для произвольной счетной системы $\{A_n\}$ множеств из $\mathfrak{A}$ объединение $\bigcup A_n$ и пересечение $\bigcap A_n$ множеств A_n также принадлежит $\mathfrak{A}$.

Короче говоря, класс $\mathfrak{A}$ замкнут относительно образования счетных объединений и пересечений, а также дополнений. Как было показано в § 1, отсюда также следует, что верхний и нижний пределы любой последовательности $\{A_n\}$ множеств из $\mathfrak{A}$ снова принадлежат $\mathfrak{A}$. Иными словами, ни одна из известных операций, производимых над множествами из $\mathfrak{A}$, не приведет нас к множествам, лежащим за пределами $\mathfrak{A}$, и поэтому не возникает необходимости рассматривать какие-либо другие множества. Эта ситуация типична настолько, что вообще вероятности будут приписываться только классу множеств со свойствами (i) и (ii). Введем поэтому следующее определение, которое применимо к произвольным вероятностным пространствам.

Определение 1. σ -алгеброй¹⁾ называется семейство $\mathfrak{A}$ подмножеств данного множества $\mathfrak{E}$, обладающее свойствами (i) и (ii).

Для любого заданного семейства $\mathfrak{F}$ подмножеств $\mathfrak{E}$ наименьшая σ -алгебра, содержащая все множества из $\mathfrak{F}$, называется σ -алгеброй, порожденной $\mathfrak{F}$.

В частности, множества, порожденные интервалами из $\mathfrak{A}'$, называются борелевскими множествами в $\mathfrak{A}'$.

В том, что наименьшая σ -алгебра, содержащая $\mathfrak{F}$, существует, можно убедиться с помощью рассуждений, используемых при определении бэровских функций в § 1. Заметим, что любая σ -алгебра содержит пространство $\mathfrak{E}$, поскольку $\mathfrak{E}$ представимо в виде объединения любого множества A и его дополнения.

Примеры. Наименьшая σ -алгебра содержит все подмножества $\mathfrak{E}$. Эта алгебра была нам очень полезна в дискретных пространствах, однако в общем случае она слишком обширна, чтобы ее можно было использовать. Другим крайним случаем является тривиальная алгебра, состоящая только из всего пространства и из пустого множества. В качестве нетривиального примера рассмотрим множества на действительной прямой, обладающие тем свойством, что если $x \in A$, то все точки $x \pm 1$, $x \pm 2$, ... тоже принадлежат A (периодические множества). Очевидно, семейство таких множеств образует σ -алгебру. ►

¹⁾ Алгебра множеств определяется аналогичным образом, но с заменой слова «счетный» в (ii) на слово «конечный». σ -алгебру иногда называют «борелевской алгеброй», но это ведет к противоречию с последней частью данного определения. (Интервалы в этом определении можно заменить на открытые множества, и тогда это определение применимо к произвольным топологическим пространствам.)

Наш опыт подсказывает нам, что главный объект теории вероятностей — это случайные величины, т. е. некоторые функции, определенные на выборочном пространстве. Мы хотим связать с каждой случайной величиной ее функцию распределения, и для этого нам нужно, чтобы событие $\{X \leq t\}$ имело бы определенную вероятность. Это рассуждение приводит нас к следующему определению.

Определение 2. Пусть $\mathcal{A}$ — произвольная σ -алгебра подмножеств $\mathcal{S}$. Действительная функция u , определенная на $\mathcal{S}$, называется $\mathcal{A}$ -измеримой¹⁾, если для любого действительного t множество всех точек x , для которых $u(x) \leq t$, принадлежит $\mathcal{A}$.

Множество точек x , для которых $u(x) < t$, представляет собой объединение счетной совокупности множеств с тем свойством, что $u(x) \leq t - n^{-1}$, и поэтому оно принадлежит $\mathcal{A}$. Так как $\mathcal{A}$ замкнуто относительно взятия дополнения, то отсюда следует, что в определении 2 знак $\leq$ может быть заменен на $<$, $>$ или $\geq$.

Из определения следует, что $\mathcal{A}$ -измеримые функции образуют замкнутое семейство в том смысле, который указан в § 1.

Следующая простая лемма часто оказывается полезной.

Лемма 1. Функция u $\mathcal{A}$ -измерима тогда и только тогда, когда она является равномерным пределом последовательности простых функций, т. е. таких функций, которые принимают только счетное число значений, причем каждое на множестве из $\mathcal{A}$.

Доказательство. По определению любая простая функция $\mathcal{A}$ -измерима. Вследствие замкнутости класса $\mathcal{A}$ -измеримых функций любой предел последовательности простых функций снова является $\mathcal{A}$ -измеримой функцией.

Обратно, пусть u — $\mathcal{A}$ -измеримая функция. При фиксированном $\varepsilon > 0$ определим множество A_n как множество всех точек x , для которых $(n-1)\varepsilon < u(x) \leq n\varepsilon$. Здесь целое n меняется от $-\infty$ до ∞ . Множества A_n взаимно исключают друг друга, а их объединение составляет все пространство $\mathcal{S}$. На множестве A_n определены $\underline{\sigma}_\varepsilon(x) = (n-1)\varepsilon$ и $\bar{\sigma}_\varepsilon(x) = n\varepsilon$. Этим путем мы получим две функции $\underline{\sigma}_\varepsilon$ и $\bar{\sigma}_\varepsilon$, определенные на $\mathcal{S}$ и такие, что

$$\underline{\sigma}_\varepsilon \leq u \leq \bar{\sigma}_\varepsilon, \quad \bar{\sigma}_\varepsilon - \underline{\sigma}_\varepsilon = \varepsilon \quad (3.1)$$

во всех точках x . Очевидно, что u есть равномерный предел $\underline{\sigma}_\varepsilon$ и $\bar{\sigma}_\varepsilon$ при $\varepsilon \rightarrow 0$.

Лемма 2. В $\mathcal{A}'$ класс бэровских функций тождествен классу функций, измеримых относительно σ -алгебры $\mathcal{A}$ борелевских множеств.

¹⁾ Использование этого термина, по существу, преждевременно, поскольку мера еще не введена.

Доказательство. (а) Очевидно, что любая непрерывная функция измерима по Борелю. Далее, эти функции образуют замкнутый класс, тогда как бэровские функции образуют *наименьший* класс, содержащий все непрерывные функции. Таким образом, всякая бэровская функция измерима по Борелю.

(б) Предыдущая лемма показывает, что для обратного утверждения достаточно показать, что любая *простая* измеримая по Борелю функция является бэровской функцией. Это равносильно утверждению, что для всякого борелевского множества A индикатор 1_A является бэровской функцией. Таким образом, борелевские множества могут быть определены следующим образом: множество A является борелевским тогда и только тогда, когда его индикатор 1_A принадлежит *наименьшему* замкнутому классу, содержащему все индикаторы интервалов. Поскольку бэровские функции образуют замкнутый класс, содержащий все индикаторы интервалов¹⁾, то отсюда вытекает, что 1_A есть бэровская функция для любого борелевского множества A . ►

Применим этот результат в частном случае евклидова пространства $\mathcal{R}^r$. В § 2 мы начинали с вполне аддитивных функций интервалов и определяли $P\{A\} = E(1_A)$ для каждого множества A , индикатор 1_A которого есть бэровская функция. Из предыдущего видно, что *при такой процедуре вероятность $P\{A\}$ определена тогда и только тогда, когда A есть борелевское множество.*

Аппроксимация борелевских множеств интервалами. Ввиду последнего замечания вероятности в $\mathcal{R}^r$, как правило, определяются на σ -алгебре борелевских множеств. Поэтому важно то, что *любое борелевское множество A может быть аппроксимировано множеством B , состоящим из конечной совокупности интервалов.* Точнее, для любого $\varepsilon > 0$ существует множество C , такое, что $P\{C\} < \varepsilon$ и вне C множества A и B совпадают (т. е. точка из дополнения C^c либо принадлежит A и B одновременно, либо не принадлежит ни тому, ни другому множеству. В качестве C можно взять объединение $A - AB \cup B - AB$).

Доказательство. В силу теоремы об аппроксимации в среднем из § 2 существует ступенчатая функция $v \geq 0$, для которой $E(|1_A - v|) < \varepsilon/2$. Возьмем в качестве B множество таких точек x , в которых $v(x) > 1/2$. Поскольку v — ступенчатая функция, то B состоит из конечного числа интервалов. Легко проверить, что

$$E|1_A(x) - 1_B(x)| \leq 2E|1_A(x) - v(x)| < \varepsilon$$

при всех x . Но $|1_A - 1_B|$ есть индикатор множества C , состоящего из всех точек, которые принадлежат либо A , либо B , но не обоим этим множествам одновременно. Из последнего неравенства следует, что $P\{C\} < \varepsilon$, и на этом доказательство завершается.

¹⁾ Чтобы убедиться в этом для открытого интервала I , определим v как непрерывную функцию, равную нулю вне I и такую, что $0 < v(x) \leq 1$ при $x \in I$. Тогда $\sqrt[n]{v} \rightarrow 1_I$.

§ 4. ВЕРОЯТНОСТНЫЕ ПРОСТРАНСТВА. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Мы теперь в состоянии описать общую конструкцию, используемую в теории вероятностей. Каково бы ни было выборочное пространство $\mathfrak{E}$, вероятности будут приписываться только множествам соответствующей σ -алгебры $\mathfrak{A}$. Поэтому начнем со следующего определения.

Определение 1. Вероятностной мерой P на σ -алгебре $\mathfrak{A}$ подмножеств $\mathfrak{E}$ называется функция, наделяющая каждое множество $A \in \mathfrak{A}$ числом $P\{A\} \geq 0$, таким, что $P\{\mathfrak{E}\} = 1$, и для любого счетного семейства взаимно непересекающихся множеств A_n из $\mathfrak{A}$ выполняется равенство

$$P\{\cup A_n\} = \sum P\{A_n\}. \quad (4.1)$$

Это свойство называется *вполне аддитивностью*, и вероятностная мера может быть описана как вполне аддитивная неотрицательная функция множества на $\mathfrak{A}$, подвергнутая нормировке¹⁾ $P\{\mathfrak{E}\} = 1$.

В отдельных случаях можно выбрать подходящую σ -алгебру и построить на ней вероятностную меру. Эти процедуры меняются от случая к случаю, и невозможно отыскать общий метод. Иногда можно приспособить подход, используемый в § 2, для построения вероятностной меры на борелевских множествах пространства $\mathcal{M}$. Типичный пример дает последовательность независимых случайных величин (§ 6). Исходным пунктом любой вероятностной задачи служит выборочное пространство, в котором выбрана σ -алгебра с соответствующей вероятностной мерой. Это приводит к следующему определению.

Определение 2. Вероятностным пространством называется тройка $(\mathfrak{E}, \mathfrak{A}, P)$, состоящая из выборочного пространства $\mathfrak{E}$, σ -алгебры $\mathfrak{A}$ подмножеств $\mathfrak{E}$ и вероятностной меры P на $\mathfrak{A}$.

Конечно, не любое воображаемое вероятностное пространство представляет собой интересный объект, однако данное определение включает в себя все, что требуется для формального развития теории, следуя схеме первого тома, и было бы бесплодным обсуждать заранее типы вероятностных пространств, которые могут встретиться на практике.

Случайные величины суть функции, определенные на выбороч-

¹⁾ Условие $P\{\mathfrak{E}\} = 1$ вводится лишь для нормировки, и никаких существенных изменений не произойдет, если его заменить на $P\{\mathfrak{E}\} < \infty$. В этом случае мы получаем пространство с конечной мерой. В теории вероятностей в связи с различными обстоятельствами возникают ситуации, когда $P\{\mathfrak{E}\} < 1$. В этом случае P называют *несобственной вероятностной мерой*. Даже условие $P\{\mathfrak{E}\} < \infty$ может быть ослаблено: можно потребовать лишь, чтобы $\mathfrak{E}$ было объединением счетного числа подмножеств $\mathfrak{E}_n$, для которых $P\{\mathfrak{E}_n\} < \infty$. (Типичными примерами являются длина и площадь.) Такие меры называют *о-конечными*.

ном пространстве, но в теории вероятностей используются только те функции, для которых можно определить функцию распределения. Определение 2 в § 3 было введено и как раз применительно к этой ситуации и приводит к определению 3.

Определение 3. Случайной величиной X называется действительная функция, которая измерима относительно исходной σ -алгебры $\mathfrak{U}$. Функция F , определенная равенством $F(t) = P\{X \leq t\}$, называется функцией распределения случайной величины X .

Исключение из рассмотрения функций, не являющихся случайными величинами, возможно в силу того, что, как мы сейчас увидим, все обычные операции, такие, как вычисление сумм или других функций, переходы к пределу и т. д., могут быть произведены внутри класса случайных величин без выхода за его границы. Прежде чем выразить это более точно, заметим, что случайная величина X отображает выборочное пространство $\mathfrak{S}$ в действительную прямую $\mathfrak{R}^1$ таким образом, что множество из $\mathfrak{S}$, для кото-

рого $a < X \leq b$, отображается в интервал a, b с соответствующей вероятностью $F(b) - F(a)$. Следовательно, каждый интервал I из $\mathfrak{R}^1$ получает вероятность $F\{I\}$. Вместо интервала I можно взять произвольное борелевское множество Γ в $\mathfrak{R}^1$ и рассмотреть множество A таких точек из $\mathfrak{S}$, для которых значение X принадлежит Γ . Символически $A = \{X \in \Gamma\}$. Ясно, что система всех таких множеств образует σ -алгебру $\mathfrak{U}_1$, которая может совпадать с $\mathfrak{U}$, но обычно меньше. Будем называть $\mathfrak{U}_1$ σ -алгеброй, порожденной случайной величиной X . Ее можно описать как наименьшую σ -алгебру в $\mathfrak{S}$, относительно которой X измерима. Случайная величина X отображает всякое множество из $\mathfrak{U}_1$ в борелевское множество Γ на прямой $\mathfrak{R}^1$, и, следовательно, соотношение $F\{\Gamma\} = P\{A\}$ определяет единственную вероятностную меру на σ -алгебре борелевских множеств $\mathfrak{R}^1$. Для интервала $I = a, b$ имеем $F(I) = F(b) - F(a)$, и потому F совпадает с единственной вероятностной мерой в $\mathfrak{R}^1$, которая связана с функцией распределения F так, как описано в § 2.

Эти рассуждения показывают, что, пока мы имеем дело с какой-либо отдельной случайной величиной X , мы можем забыть о первоначальном выборочном пространстве и делать вид, что вероятностное пространство есть прямая $\mathfrak{R}^1$ с σ -алгеброй борелевских множеств на нем и с мерой, индуцированной посредством функции распределения F . Мы видим, что в $\mathfrak{R}^1$ класс борелевских функций совпадает с функциями, измеримыми по Борелю. Выбор $\mathfrak{R}^1$ в качестве выборочного пространства означает, что класс случайных величин совпадает с классом функций, измеримых по Борелю. По отношению к исходному выборочному пространству это означает, что семейство борелевских функций от случайной величины X совпадает с семейством всех функций, которые измеримы относительно

σ -алгебры $\mathfrak{A}_1$, порожденной X . Поскольку $\mathfrak{A}_1 \subset \mathfrak{A}$, то отсюда следует, что любая бэровская функция X снова является случайной величиной.

Эти доводы переносятся без изменения на конечные совокупности случайных величин. Таким образом, вектор $(X_1, \dots, X_r)$ отображает $\mathfrak{E}$ в $\mathfrak{A}^r$ так, что открытому интервалу в $\mathfrak{A}^r$ соответствует множество в $\mathfrak{E}$, для которого имеют место r соотношений вида $a_k < X_k < b_k$. Это множество $\mathfrak{A}$ -измеримо, так как оно является пересечением r таких множеств. Как и в случае одиночной случайной величины, можно определить σ -алгебру $\mathfrak{A}_1$, порожденную $X_1, \dots, X_r$, как наименьшую σ -алгебру множеств из $\mathfrak{E}$, относительно которой r случайных величин измеримы. Получаем основную теорему.

Теорема. *Любая бэровская функция от конечного числа случайных величин тоже есть случайная величина.*

Случайная величина U является бэровской функцией переменных $X_1, \dots, X_r$, если она измерима относительно σ -алгебры, порожденной $X_1, \dots, X_r$.

Примеры. а) На прямой $\mathfrak{A}^1$ с X как координатной случайной величиной функция X^2 порождает σ -алгебру борелевских множеств, которые являются симметричными относительно начала координат (в том смысле, что если $x \in A$, то и $-x \in A$).

б) Рассмотрим $\mathfrak{A}^3$ с координатными случайными величинами X_1, X_2, X_3 и σ -алгеброй борелевских множеств. Пара (X_1, X_2) порождает семейство всех цилиндрических множеств с образующими, параллельными третьей оси, и основаниями, которыми служат борелевские множества на (X_1, X_2) -плоскости. ►

Математические ожидания

В § 2 мы начинали с функции интервалов в $\mathfrak{A}^r$ и использовали ее для построения вероятностного пространства. Там мы считали, что удобно сначала определить математические ожидания (интегралы) и затем определить вероятности борелевского множества A как математическое ожидание $E(1_A)$ его индикатора. Если начинать с вероятностного пространства, то процедура должна быть обратной: вероятности заданы и нужно определить математические ожидания случайных величин в терминах данных вероятностей. К счастью, эта процедура крайне проста.

Как и в предыдущем параграфе, мы будем говорить, что случайная величина U является *простой*, если она принимает только счетное число значений $a_1, a_2, \dots$, каждое на множестве A_j , принадлежащем основной σ -алгебре $\mathfrak{A}$. К таким случайным величинам применима дискретная теория первого тома, и мы определим

математическое ожидание U равенством

$$E(U) = \sum a_k P\{A_k\} \quad (4.2)$$

при условии, что ряд сходится абсолютно, в противном случае мы скажем, что U не имеет математического ожидания.

Для любой случайной величины U и произвольного числа $\varepsilon > 0$ мы определили в (3.1) две простые случайные величины $\underline{\sigma}_\varepsilon$ и $\bar{\sigma}_\varepsilon$, такие, что $\bar{\sigma}_\varepsilon = \underline{\sigma}_\varepsilon + \varepsilon$ и $\underline{\sigma}_\varepsilon \leq U \leq \bar{\sigma}_\varepsilon$. При любом разумном определении $E(U)$

$$E(\underline{\sigma}_\varepsilon) \leq E(U) \leq E(\bar{\sigma}_\varepsilon), \quad (4.3)$$

всякий раз когда величины $\underline{\sigma}_\varepsilon$ и $\bar{\sigma}_\varepsilon$ имеют математические ожидания. Так как эти функции отличаются не более чем на ε , то либо это же справедливо для их математических ожиданий, либо эти математические ожидания не существуют. В последнем случае мы говорим, что U не имеет математического ожидания, тогда как в первом случае $E(U)$ определяется однозначно из (4.3) переходом к пределу при $\varepsilon \rightarrow 0$. Короче, поскольку любая случайная величина U представляется равномерным пределом последовательности простых случайных величин σ_n , то математическое ожидание U может быть определено как предел $E(\sigma_n)$. Например, в терминах $\bar{\sigma}_\varepsilon$ имеем

$$E(U) = \lim_{\varepsilon \rightarrow 0} \sum_{n=-\infty}^{\infty} n\varepsilon P\{(n-1)\varepsilon < U \leq n\varepsilon\} \quad (4.4)$$

при условии, что ряд сходится абсолютно (при некотором, а потому и при всех $\varepsilon > 0$). Далее, вероятности, входящие в (4.4), совпадают с вероятностями, которые приписываются интервалам $[(n-1)\varepsilon, n\varepsilon]$ с помощью функции распределения F случайной величины U . Отсюда следует, что при таком изменении обозначений наше определение $E(U)$ сводится к определению, данному в § 2:

$$E(U) = \int_{-\infty}^{+\infty} tF\{dt\}. \quad (4.5)$$

Таким образом, $E(U)$ можно корректно определить либо непосредственно в исходном вероятностном пространстве, либо в терминах функции распределения. (То же замечание было сделано в 1; гл. IX, для дискретных случайных величин.) По этой причине излишне говорить, что в произвольных вероятностных пространствах математические ожидания обладают основными свойствами математических ожиданий в $\mathcal{M}$, введенных в § 2.

§ 5. ТЕОРЕМА О ПРОДОЛЖЕНИИ

При построении вероятностных пространств вероятности обычно определены а priori на не очень богатом классе множеств, и требуется подходящим образом расширить их область определения. Например, при рассмотрении в первом томе бесконечных последовательностей испытаний и рекуррентных событий были заданы вероятности всех событий, зависящих от конечного числа испытаний, и эту область определения вероятностей требовалось расширить так, чтобы включить такие события, как разорение, возвращение и полное вырождение. Аналогично при построении в § 2 мер в $\mathcal{R}^r$ сначала наделялись вероятностями $F\{I\}$ интервалы, а затем область определения продолжалась до класса всех борелевских множеств. Такое продолжение возможно благодаря теореме, имеющей значительно более широкую область применения и на которой основываются многие конструкции вероятностных пространств. Схема рассуждения состоит в следующем.

Аддитивность F позволяет однозначно определить

$$F\{A\} = \sum F\{I_k\} \quad (5.1)$$

для любого множества A , являющегося объединением *конечного* числа *непересекающихся интервалов* I_k . Эти множества образуют алгебру $\mathcal{A}_0$ (т. е. объединения и пересечения конечного числа множеств из $\mathcal{A}_0$, а также дополнения множеств из $\mathcal{A}_0$ тоже принадлежат $\mathcal{A}_0$). Начиная с этого пункта, природа пространства $\mathcal{R}^r$ не играет никакой роли и можно рассматривать произвольную алгебру $\mathcal{A}_0$ множеств в произвольном пространстве $\mathcal{S}$. Всегда существует наименьшая алгебра $\mathcal{A}$ множеств, содержащая $\mathcal{A}_0$ и замкнутая относительно *счетных* объединений и пересечений. Иными словами, существует наименьшая σ -алгебра $\mathcal{A}$, содержащая $\mathcal{A}_0$ (см. определение 1 из § 3). При построении мер в $\mathcal{R}^r$ σ -алгебра $\mathcal{A}$ совпадала с σ -алгеброй всех борелевских множеств. Расширение области определения вероятностей $\mathcal{A}_0$ до $\mathcal{A}$ основывается на следующей общей теореме.

Теорема о продолжении. Пусть $\mathcal{A}_0$ — произвольная алгебра множеств в некотором пространстве $\mathcal{S}$. Пусть F — определенная на $\mathcal{A}_0$ функция множеств, такая, что $F\{A\} \geq 0$ при всех $A \in \mathcal{A}_0$, $F\{\mathcal{S}\} = 1$, и, кроме того, соотношение (5.1) выполняется для любого разбиения A на счетное число непересекающихся множеств $I_k \in \mathcal{A}_0$.

Тогда существует единственное продолжение F до счетно-аддитивной функции множеств (т. е. до вероятностной меры) на наименьшей σ -алгебре $\mathcal{A}$, содержащей $\mathcal{A}_0$.

В следующем параграфе будет приведен типичный пример применения этой теоремы. Теперь мы обратимся к более общему и гибкому варианту теоремы о продолжении, ближе примыкающему к рассмотрению § 2 и 3. Мы исходили из математического ожидания (2.4) для ступенчатых функций (т. е. функций, принимающих

лишь конечное число значений, причем каждое из значений — на некотором интервале). Область определения этого математического ожидания была затем продолжена с довольно бедного класса ступенчатых функций на более широкий класс, включающий все ограниченные бэровские функции. Указанное продолжение непосредственно приводит к интегралу Лебега—Стилтьеса, и мера множества A получается как математическое ожидание его индикатора 1_A . Наметим соответствующую абстрактную схему.

Рассмотрим вместо алгебры множеств $\mathfrak{A}_0$ класс функций $\mathfrak{B}_0$, замкнутый относительно линейных комбинаций и операций $\cap$ и $\cup$. Иными словами, предполагается, что если функции u_1 и u_2 принадлежат $\mathfrak{B}_0$, то функции

$$\alpha_1 u_1 + \alpha_2 u_2, \quad u_1 \cap u_2, \quad u_1 \cup u_2 \quad (5.2)$$

тоже принадлежат $\mathfrak{B}_0$ ¹⁾. Отсюда, в частности, следует, что всякая функция u из $\mathfrak{B}_0$ может быть записана как разность двух неотрицательных функций, именно $u = u^+ - u^-$, где $u^+ = u \cup 0$ и $u^- = (-u) \cap 0$. Под *линейным функционалом* E на $\mathfrak{B}_0$ понимается сопоставление значений $E(u)$ всем функциям из $\mathfrak{B}_0$ с выполнением условия

$$E(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 E(u_1) + \alpha_2 E(u_2). \quad (5.3)$$

Функционал E называется *положительным*, если $E(u) \geq 0$ при $u \geq 0$. *Нормой* E называется верхняя грань значений $E(|u|)$ по всем функциям $u \in \mathfrak{B}_0$, удовлетворяющим условию $|u| \leq 1$. Если функция, равная всюду 1, принадлежит $\mathfrak{B}_0$, то норма E равна $E(1)$. Наконец, скажем, что E *счетно-аддитивен* на $\mathfrak{B}_0$, если

$$E\left(\sum_1^\infty u_k\right) = \sum_1^\infty E(u_k) \quad (5.4)$$

всякий раз, когда $\sum u_k$ принадлежит $\mathfrak{B}_0$ ²⁾. Условие счетной аддитивности эквивалентно следующему: если $\{v_n\}$ — последовательность функций из $\mathfrak{B}_0$, сходящаяся монотонно к нулю, то³⁾

$$E(v_n) \rightarrow 0. \quad (5.5)$$

Для каждого заданного класса функций $\mathfrak{B}_0$ существует *наименьший класс $\mathfrak{B}$, содержащий $\mathfrak{B}_0$ и замкнутый относительно переходов к пределу в смысле поточечной сходимости*. [Класс $\mathfrak{B}$ автоматически

¹⁾ Наши предположения равносильны требованию, чтобы класс $\mathfrak{B}_0$ был линейной структурой.

²⁾ Здесь следует добавить требование $u_k \geq 0$, иначе последующее утверждение неверно. — *Прим. перев.*

³⁾ Для доказательства эквивалентности (5.4) и (5.5) достаточно рассмотреть случай $u_k \geq 0$, $v_k \geq 0$. Тогда (5.4) следует из (5.5), если положить $v_n = u_{n+1} + u_{n+2} + \dots$, и (5.5) вытекает из (5.4), если положить $u_k = v_k - v_{k+1}$ (при этом $\sum u_k = v_1$).

замкнут относительно операций (5.2).] Дадим теперь другую формулировку теоремы о продолжении¹⁾. Каждый положительный счетно-аддитивный линейный функционал нормы 1 на $\mathfrak{B}_0$ может быть единственным образом продолжен до положительного счетно-аддитивного линейного функционала нормы 1 на всех ограниченных (и многих неограниченных) функциях из $\mathfrak{B}$.

В качестве примера применения этой теоремы докажем следующий важный результат.

Теорема Ф. Рисса о представлении²⁾. Пусть E — положительный линейный функционал нормы 1 на классе всех непрерывных функций на $\mathcal{R}$, обращающихся в нуль на бесконечности³⁾. Существует мера P на σ -алгебре борелевских множеств, для которой $P\{\mathcal{R}\} = 1$ и такая, что $E(u)$ совпадает с интегралом от u по P .

Иными словами, определенные нами интегралы дают представление наиболее общих положительных линейных функционалов.

Доказательство. В основе доказательства лежит тот факт, что если последовательность $\{v_n\}$ обращающихся в нуль на бесконечности непрерывных функций монотонно сходится к нулю, то сходимость эта автоматически *равномерна*. Пусть $v_n \geq 0$, и обозначим $\|v_n\| = \max v_n(x)$. Тогда $E(v_n) \leq \|v_n\|$, и поэтому для E выполнено условие счетной аддитивности (5.5). В силу теоремы о продолжении E можно продолжить на все ограниченные бэровские функции. Положив $P\{A\} = E(1_A)$, получаем меру на σ -алгебре борелевских множеств. Как мы видели, по заданной мере $P\{A\}$ интеграл Лебега—Стилтьеса однозначно определяется с помощью двойного

¹⁾ Основная идея доказательства (восходящая к Лебегу) проста и остроумна. Нетрудно видеть, что если две последовательности $\{u_n\}$ и $\{u'_n\}$ функций из $\mathfrak{B}_0$ сходятся монотонно к одному и тому же пределу u , то последовательности $E(u_n)$ и $E(u'_n)$ тоже сходятся к одному и тому же пределу. Для таких монотонных пределов u можно, следовательно, определить $E(u) = \lim E(u_n)$. Рассмотрим теперь класс $\mathfrak{B}_1$ функций u , для которых при любом $\varepsilon > 0$ существуют две функции $\underline{u}$ и $\bar{u}$, либо принадлежащие $\mathfrak{B}_0$, либо являющиеся монотонными пределами последовательностей из $\mathfrak{B}_0$ и такие, что $\underline{u} < u < \bar{u}$, $E(\bar{u}) - E(\underline{u}) < \varepsilon$. Класс $\mathfrak{B}_1$ замкнут относительно переходов к пределам, и для функций из $\mathfrak{B}_1$ значение $E(u)$ определяется очевидным образом, поскольку должно быть $E(\underline{u}) \leq E(u) \leq E(\bar{u})$.

Замечательной особенностью этого рассуждения является следующее обстоятельство: обычно класс $\mathfrak{B}_1$ больше класса $\mathfrak{B}$ и простое доказательство оказывается возможным благодаря тому, что доказывается больше, чем требуется. (Относительно сравнения классов $\mathfrak{B}$ и $\mathfrak{B}_1$ см. § 7.)

²⁾ Эта теорема справедлива для произвольных локально компактных пространств. См. другое доказательство в V, 1.

³⁾ u обращается в нуль на бесконечности, если для любого $\varepsilon > 0$ существует шар (компактное множество), вне которого $|u(x)| < \varepsilon$.

неравенства (4.3). Следовательно, интеграл Лебега—Стилтьеса по P от обращаемой в нуль на бесконечности непрерывной функции u совпадает со значением на u заданного функционала $E(u)$. ►

§ 6. ПРОИЗВЕДЕНИЯ ПРОСТРАНСТВ. ПОСЛЕДОВАТЕЛЬНОСТИ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Понятие прямого произведения пространств (1; гл. V, 4) является весьма важным в теории вероятностей и используется всякий раз, когда речь идет о повторных испытаниях. Представление точки плоскости $\mathcal{R}^2$ посредством двух координат означает, что $\mathcal{R}^2$ рассматривается как прямое произведение двух своих осей. Обозначим две координатные переменные через X и Y . Рассматриваемые как функции на плоскости они представляют собой бэровские функции, и если на σ -алгебре борелевских множеств в $\mathcal{R}^2$ определена вероятностная мера P , то существуют две функции распределения $P\{X \leq x\}$ и $P\{Y \leq y\}$. Эти две функции распределения индуцируют вероятностные меры на обеих осях, называемые частными (маргинальными) распределениями (или проекциями). Здесь у нас в качестве первичного понятия появляется плоскость, однако часто более естественной является противоположная ситуация. Например, когда говорят о двух независимых случайных величинах с заданными распределениями, два частных распределения являются первичным понятием и вероятности на плоскости получаются из них с помощью «правила умножения». Общий случай ничуть не сложнее случая плоскости.

Рассмотрим теперь два произвольных вероятностных пространства. Иными словами, нам заданы два выборочных пространства $\mathcal{S}^{(1)}$ и $\mathcal{S}^{(2)}$, две σ -алгебры $\mathcal{U}^{(1)}$ и $\mathcal{U}^{(2)}$ подмножеств $\mathcal{S}^{(1)}$ и $\mathcal{S}^{(2)}$ соответственно и определенные на $\mathcal{U}^{(1)}$ и $\mathcal{U}^{(2)}$ вероятностные меры $P^{(1)}$ и $P^{(2)}$. Прямое произведение $(\mathcal{S}^{(1)}, \mathcal{S}^{(2)})$ есть множество всех упорядоченных пар $(x^{(1)}, x^{(2)})$, где $x^{(i)}$ — точка пространства $\mathcal{S}^{(i)}$. Рассмотрим в этом произведении множества, являющиеся «прямоугольниками», т. е. прямые произведения $(A^{(1)}, A^{(2)})$ множеств $A^{(i)} \in \mathcal{U}^{(i)}$. Множества такого вида мы наделяем вероятностями согласно правилу умножения:

$$P\{A^{(1)}, A^{(2)}\} = P^{(1)}\{A^{(1)}\} P^{(2)}\{A^{(2)}\}. \quad (6.1)$$

Далее множества, представляющие собой конечные объединения непересекающихся прямоугольников, образуют алгебру $\mathcal{U}_0$, на которой с помощью (6.1) единственным образом определяется счетно-аддитивная функция. Таким образом по теореме о продолжении на наименьшей σ -алгебре, содержащей все прямоугольники, существует единственная вероятностная мера P , такая, что вероятности прямоугольников задаются правилом умножения (6.1). Эта наименьшая, содержащая все прямоугольники σ -алгебра обозначается

$\mathfrak{U}^{(1)} \times \mathfrak{U}^{(2)}$, а определенная мера на ней называется произведением мер.

Конечно, на произведении пространств можно определить и другие вероятности, например, в терминах условных вероятностей. Рассматриваемая σ -алгебра множеств $\mathfrak{U}$ всегда будет не меньше $\mathfrak{U}^{(1)} \times \mathfrak{U}^{(2)}$, и в то же время за пределы $\mathfrak{U}^{(1)} \times \mathfrak{U}^{(2)}$ приходится выходить весьма редко. Ниже в рассуждении о случайных величинах предполагается, что $\mathfrak{U} = \mathfrak{U}^{(1)} \times \mathfrak{U}^{(2)}$.

Понятие случайной величины (измеримой функции) связано с соответствующей σ -алгеброй, и при рассмотрении произведений пространств мы должны различать случайные величины на произведении и на пространствах $\mathfrak{E}^{(1)}$ и $\mathfrak{E}^{(2)}$. К счастью, отношение между этими тремя классами случайных величин весьма простое. Пусть u и v — случайные величины на $\mathfrak{E}^{(1)}$ и $\mathfrak{E}^{(2)}$, и рассмотрим на произведении пространств функцию w , которая в точке $(x^{(1)}, x^{(2)})$ принимает значение

$$w(x^{(1)}, x^{(2)}) = u(x^{(1)})v(x^{(2)}). \quad (6.2)$$

Покажем, что класс случайных величин на произведении пространств $(\mathfrak{E}^{(1)}, \mathfrak{E}^{(2)})$ совпадает с минимальным классом конечнозначных функций, замкнутым относительно поточечных переходов к пределу и содержащим все линейные комбинации функций вида (6.2).

Ясно, во-первых, что каждый множитель в правой части (6.2) является случайной величиной и в том случае, когда он рассматривается как функция на произведении пространств. Следовательно, w есть случайная величина, и поэтому класс случайных величин на $(\mathfrak{E}^{(1)}, \mathfrak{E}^{(2)})$ не меньше минимального класса, о котором говорилось выше. С другой стороны, случайные величины образуют наименьший класс, замкнутый относительно переходов к пределу и содержащий все линейные комбинации индикаторов прямоугольников. Эти индикаторы имеют вид (6.2), и, следовательно, класс случайных величин не больше минимального класса, о котором говорилось выше.

Частный случай произведения двух пространств $\mathcal{H}^n$ и $\mathcal{H}^n$ с вероятностными мерами F и G неявно встречался в связи с теоремой Фубини (2.11) о повторных интегралах. Теперь мы можем утверждать справедливость более общей теоремы, относящейся не только к $\mathcal{H}$.

Теорема Фубини для произведений мер. *Интеграл от любой неотрицательной бэровской функции и по произведению мер равен повторным интегралам из формулы (2.11).*

(При этом имеется в виду, что интегралы могут расходиться. Теорема очевидна для простых функций, а в общем случае получается с помощью повторной аппроксимации.) Обобщение на случай произведения трех и большего числа пространств вполне очевидно и не требует специальных пояснений.

Обратимся теперь к проблеме бесконечных последовательностей

случайных величин, с которой мы уже встречались в первом томе в связи с бесконечными последовательностями испытаний Бернулли, случайными блужданиями, рекуррентными событиями и т. д., а также в гл. III в связи с нормальными случайными процессами. Проблемы не возникает, когда имеется бесконечное множество случайных величин, определенных на заданном вероятностном пространстве. Например, действительная прямая с нормальным распределением есть вероятностное пространство и $\{\sin nx\}$ есть бесконечная последовательность определенных на нем случайных величин. Нас интересует только та ситуация, когда вероятности должны быть определены в терминах заданных случайных величин. Более точно проблема состоит в следующем.

Пусть $\mathcal{R}^\infty$ — пространство, точками которого являются бесконечные последовательности действительных чисел $(x_1, x_2, \dots)$ (т. е. $\mathcal{R}^\infty$ есть счетное прямое произведение действительных прямых). Обозначим n -ю координатную величину через X_n (X_n есть функция на $\mathcal{R}^\infty$, принимающая в точке $x = (x_1, x_2, \dots)$ значение x_n). Предположим, что нам заданы вероятностные распределения для $X_1, (X_1, X_2), (X_1, X_2, X_3), \dots$ и требуется определить соответствующие вероятности в $\mathcal{R}^\infty$. Конечно, заданные вероятности должны быть взаимно согласованы в том смысле, что распределения $(X_1, \dots, X_n)$ являются частными распределениями для $(X_1, \dots, X_{n+1})$ и т. д.

Формализуем теперь интуитивное понятие «события, определяемого исходом конечного множества испытаний». Скажем, что множество A в $\mathcal{R}^\infty$ зависит лишь от первых r координат, если существует борелевское множество A_r в $\mathcal{R}^r$, такое, что $x = (x_1, x_2, \dots)$ принадлежит A тогда и только тогда, когда $(x_1, \dots, x_r)$ принадлежит A_r . В теории вероятностей часто встречается ситуация, когда вероятности таких множеств заданы, и задача состоит в расширении этой области определения. Приведем без доказательства принадлежащую А. Н. Колмогорову фундаментальную теорему (доказанную им в несколько большей общности), впервые появившуюся в его ставшем теперь классическим аксиоматическом обосновании теории вероятностей (1933). Эта работа предвосхитила и стимулировала развитие современной теории меры.

Теорема 1. *Согласованную систему вероятностных распределений величин $X_1, (X_1, X_2), (X_1, X_2, X_3), \dots$ можно единственным способом продолжить до вероятностной меры на наименьшей σ -алгебре $\mathfrak{A}$ множеств в $\mathcal{R}^\infty$, содержащей все множества, зависящие лишь от конечного числа координат¹⁾.*

Весьма важным является то обстоятельство, что все вероятности определяются последовательными переходами к пределам, исходя из конечномерных множеств. Каждое множество A из $\mathfrak{A}$

¹⁾ Эта теорема справедлива и в более общем случае произведения локально компактных пространств. Например, она справедлива, когда переменные X_n являются векторными переменными (т. е. значения их — точки в $\mathcal{M}^r$).

можно аппроксимировать конечномерными множествами в следующем смысле. Для любого $\varepsilon > 0$ существует n и зависящее лишь от первых n координат множество A_n , такое, что

$$P\{A - A \cap A_n\} < \varepsilon, \quad P\{A_n - A \cap A_n\} < \varepsilon. \quad (6.3)$$

Иными словами, множество тех точек, которые принадлежат либо A , либо A_n , но не принадлежат обоим множествам A и A_n , имеют вероятность, меньшую 2ε . Отсюда следует, что можно так выбрать множества A_n , чтобы

$$P\{A_n\} \rightarrow P\{A\}. \quad (6.4)$$

Теорема 1 позволяет говорить о *бесконечной последовательности взаимно независимых случайных величин с произвольными заданными распределениями*. Такие последовательности уже встречались в первом томе, однако нам приходилось определять интересующие нас вероятности с помощью специальных переходов к пределу, в то время как теорема 1 является общим результатом, охватывающим все случаи. Это обстоятельство хорошо иллюстрируется следующими двумя важными теоремами, первая из которых принадлежит А. Колмогорову (1933), а вторая — Э. Хьюиту и Л. Дж. Сэвиджу (1955). Обе они являются типичными вероятностными предложениями и играют центральную роль во многих рассуждениях.

Теорема 2. (Закон нуля или единицы для остаточных событий.) Предположим, что случайные величины X_k взаимно независимы и что при каждом n событие A не зависит¹⁾ от $X_1, \dots, X_n$ ²⁾. Тогда или $P\{A\} = 0$, или $P\{A\} = 1$.

Доказательство. Вообще говоря, случайные величины X_k могут быть определены на произвольном вероятностном пространстве, однако они отображают это пространство в произведение прямых $\mathcal{R}^\infty$, в котором они играют роль координатных переменных. Следовательно, не уменьшая общности, можно ограничиться рассмотрением ситуации, описанной в настоящем параграфе. В обозначениях, использованных в (6.3), множества A и A_n независимы, и поэтому из первого неравенства в (6.3) вытекает, что $P\{A\} - P\{A\}P\{A_n\} < \varepsilon$. Следовательно, $P\{A\} = P^2\{A\}$. ►

Пример. а) Ряд $\sum X_n$ сходится с вероятностью единица или нуль. Точно так же множество точек, в которых $\limsup X_n = \infty$, имеет вероятность либо единица, либо нуль. ►

Теорема 3. (Закон нуля или единицы для симметричных собы-

¹⁾ Более точно, A не зависит от любого события, определяемого в терминах $X_1, \dots, X_n$. Иными словами, индикатор A есть случайная величина, независимая от $X_1, \dots, X_n$.

²⁾ Предполагается, конечно, что A определяется последовательностью $X_1, X_2, \dots$. — Прим. перев.

тий.) Предположим, что случайные величины X_k взаимно независимы и одинаково распределены. Если множество A инвариантно относительно конечных перестановок координат¹⁾, то либо $P\{A\} = 0$, либо $P\{A\} = 1$.

Доказательство. Как и в доказательстве теоремы 2, мы рассматриваем X_k как координатные величины и под множествами A_n понимаем то же, что и в (6.3). Пусть B_n — множество, полученное из A_n обращением порядка первых $2n$ координат, в то время как остальные координаты остаются без изменения. В силу сделанных предположений соотношения (6.3) остаются справедливыми при замене A_n на B_n . Отсюда вытекает, что множество точек, принадлежащих либо A , либо $A_n \cap B_n$, но не принадлежащих обоим множествам A и $A_n \cap B_n$, имеет вероятность, меньшую 4ϵ . Следовательно,

$$P\{A_n \cap B_n\} \rightarrow P\{A\}. \quad (6.5)$$

Далее A_n зависит лишь от первых n координат, и поэтому B_n зависит лишь от координат с номерами $n+1, \dots, 2n$. Таким образом, A_n и B_n независимы, что вместе с (6.5) приводит к равенству $P(A) = P^2\{A\}$. ►

Пример. б) Положим $S_n = X_1 + \dots + X_n$, и пусть $A = \{S_n \in I \text{ б. ч.}\}$, где I — произвольный интервал на прямой. Событие A инвариантно относительно конечных перестановок. [По поводу обозначений см. пример 1, а).] ►

§ 7. НУЛЕВЫЕ МНОЖЕСТВА. ПОПОЛНЕНИЕ

Множеством вероятности нуль обычно можно пренебрегать, и две случайные величины, различающиеся лишь на таком нулевом множестве, «практически совпадают». Выражаясь более формально, их называют *эквивалентными*. Это означает, что если изменить определение случайной величины на нулевом множестве, то все вероятностные соотношения сохраняются без изменения и, следовательно, случайную величину можно считать неопределенной на нулевом множестве. Типичным примером является время осуществления первого рекуррентного события: с вероятностью единица оно есть некоторое число, а с вероятностью нуль — не определено (или считается равным ∞). Таким образом, мы чаще имеем дело с классами эквивалентных случайных величин, а не с самими случайными величинами. Тем не менее обычно проще выбрать подходящие представители классов и оперировать с ними, а не с классами.

С нулевыми множествами связано единственное место в теории

¹⁾ Более точно, предполагается, что если $a = (a_1, a_2, \dots)$ — точка из A и n_1, n_2 — два произвольных натуральных числа, то A содержит также точку, у которой n_1 -я координата есть a_{n_2} , n_2 -я координата равна a_{n_1} , а остальные координаты те же, что и у a . Это условие автоматически распространяется на перестановки, включающие k координат.

вероятностей, которое идет вразрез с нашей интуицией. Ситуация одна и та же во всех вероятностных пространствах и достаточно описать ее для случая вероятностной прямой. В рамках нашей конструкции вероятности определены лишь на борелевских множествах, но каждое борелевское множество, вообще говоря, содержит много неборелевских подмножеств. Следовательно, нулевое множество может содержать подмножества, на которых вероятности не определены, в то время как естественно ожидать, что каждое подмножество нулевого множества само должно быть нулевым множеством. Это несоответствие не является серьезным и легко может быть устранено. В самом деле, введем такой *постулат*: если $A \subset B$ и $P\{B\} = 0$, то $P\{A\} = 0$. При этом мы должны расширить σ -алгебру $\mathfrak{A}$ борелевских множеств до (по крайней мере) наименьшей σ -алгебры $\mathfrak{A}_1$, содержащей все множества из $\mathfrak{A}$ и все подмножества нулевых множеств. Можно дать прямое описание σ -алгебры $\mathfrak{A}_1$. Именно, множество A принадлежит $\mathfrak{A}_1$ тогда и только тогда, когда оно отличается от некоторого борелевского множества A^0 лишь на нулевое множество¹⁾. Область определения вероятности можно продолжить с $\mathfrak{A}$ на $\mathfrak{A}_1$, просто положив $P\{A\} = P\{A^0\}$. Без труда показывается, что это продолжение единственно и дает вполне аддитивную меру на $\mathfrak{A}_1$. Таким образом, мы построили вероятностное пространство, которое удовлетворяет введенному постулату и в котором сохранились вероятности борелевских множеств.

Описанная конструкция называется *лебеговым пополнением* (заданного вероятностного пространства). Использование пополнения естественно в задачах, связанных с единственным вероятностным распределением. По этой причине длину интервалов в $\mathcal{R}^1$ обычно продолжают до лебеговой меры, которая определена не только на борелевских множествах. Однако, когда рассматриваются семейства распределений (например, бесконечные последовательности испытаний Бернулли с различными вероятностями p), пополнение обычно приводит к неприятностям. Именно $\mathfrak{A}_1$ зависит от рассматриваемого распределения, и поэтому случайная величина по отношению к $\mathfrak{A}_1$ может не быть таковой, если перейти к другому распределению.

Пример. Пусть $a_1, a_2, \dots$ — последовательность точек в $\mathcal{R}^1$, имеющих вероятности $p_1, p_2, \dots$ соответственно, причем $\sum p_k = 1$. Дополнение к $\{a_j\}$ имеет вероятность нуль, и поэтому $\mathfrak{A}_1$ содержит все подмножества $\mathcal{R}^1$. Каждая ограниченная функция u является при этом случайной величиной: математическим ожиданием $\sum p_k u(a_k)$, однако, когда рассматриваемое распределение не дискретно, оперировать в «произвольными функциями» становится опасно. ►

¹⁾ Более точно, требуется, чтобы оба множества $A - A \cap A^0$ и $A^0 - A \cap A^0$ содержались в некотором нулевом множестве.

В этой главе изучаются вероятностные распределения в r -мерном пространстве $\mathcal{R}^r$. В идейном отношении понятие вероятностного распределения основывается на изложенной в предыдущей главе теории интегрирования, однако, по существу, никаких сложных размышлений для понимания настоящей главы не требуется, поскольку вводимые понятия и формулы интуитивно близки к понятиям и формулам, уже известным из первого тома и первых трех глав.

Новая черта рассматриваемой теории состоит в том, что (в противоположность случаю дискретного выборочного пространства) не *каждое* множество имеет вероятность и не *каждая* функция является случайной величиной. К счастью, эта теоретическая сложность не очень заметна на практике, поскольку можно исходить из интервалов и непрерывных функций и ограничиться соответственно рассмотрением множеств и функций, которые могут быть получены из них посредством элементарных операций и (возможно, бесконечно многих) предельных переходов. Этим выделяются классы борелевских множеств и бэровских функций. Читатели, которых интересуют больше факты, нежели логические связи, могут не беспокоиться по поводу точных определений (приводимых в гл. IV). Им разумнее полагаться на собственную интуицию и считать, что все рассматриваемые множества и функции являются «хорошими». Приводимые теоремы настолько просты¹⁾, что для их понимания достаточно знакомства с элементарным анализом. Изложение является строгим, если условиться, что слова «множество» и «функция» сокращенно означают «борелевское множество» и «бэровская функция» соответственно.

¹⁾ Следует понимать, что этой простоты нельзя достичь ни в какой теории, использующей лишь непрерывные функции или какой-либо другой класс «хороших» функций. Например, в гл. II, (8.3), мы определили плотность ϕ бесконечным рядом. Скучно и бессмысленно устанавливать условия, при которых ϕ является хорошей функцией, однако в простых случаях формула очевидна, и она всегда имеет смысл, если ограничиваться бэровскими функциями, что и можно считать точным смыслом неопределенного выражения «формула всегда справедлива». Между прочим, будут указаны несколько случаев, где ограничение бэровскими функциями нетривиально [примером такого рода является теория выпуклых функций, см. 8, б)],

При первоначальном чтении можно ограничиться § 1—4 и 9. Параграфы 5—8 содержат результаты и неравенства, собранные для удобства ссылок. В последних параграфах, посвященных теории условных распределений и математических ожиданий, изложение ведется более полно, чем это необходимо для настоящего тома. В дальнейшем результаты этой теории будут использованы лишь от случая к случаю при рассмотрении мартингалов в гл. VI, II, и гл. VII, 9.

§ 1. РАСПРЕДЕЛЕНИЯ И МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

Даже совсем безобидное использование термина «случайная величина» может содержать косвенную ссылку на весьма непростое вероятностное пространство или на сложный умозрительный эксперимент. Например, теоретическая модель может включать в себя положения и скорости 10^{28} частиц, в то время как нам интересны лишь температура и энергия. Эти две случайные величины отображают исходное выборочное пространство в плоскость $\mathcal{R}^2$, порождая в ней вероятностное распределение. По существу, мы имеем дело с двумерной задачей, и первоначальное выборочное пространство утрачивает ясные очертания и превращается в фон. Конечномерные евклидовы пространства $\mathcal{R}^r$ представляют собой, следовательно, наиболее важные выборочные пространства, и мы переходим к систематическому изучению соответствующих вероятностных распределений.

Рассмотрим сначала случай прямой $\mathcal{R}^1$. Интервалы, определяемые неравенствами $a < x < b$ и $a \leq x \leq b$, обозначим через $\overline{a, b}$ и $\overline{a, b}$ соответственно. (Мы не исключаем предельный случай замкнутого интервала, сводящегося к одной точке.) Полуоткрытые интервалы обозначаются $\overline{a, b}$ и $\overline{a, b}$. В одномерном случае все случайные величины являются функциями координатной переменной X (иными словами, функции, которая в точке x принимает значение x). Все вероятности можно, следовательно, выразить в терминах функции распределения

$$F(x) = P\{X \leq x\}, \quad -\infty < x < \infty. \quad (1.1)$$

В частности, интервал $I = \overline{a, b}$ имеет вероятность

$$P\{I\} = F(b) - F(a).$$

Гибкое стандартное обозначение $P\{\}$ становится неудобным, когда рассматриваются различные распределения. Применение новой буквы было бы неэкономичным, а использование обозначения $P_F\{\}$ для указания зависимости от F слишком неуклюже. Намного проще использовать одну и ту же букву F как для функции точки (1.1), так и для функции интервала, и в дальнейшем вместо

$P\{I\}$ мы будем писать $F\{I\}$. Иными словами, использование фигурных скобок $\{ \}$ будет означать, что аргумент в $F\{A\}$ есть интервал или множество и что F представляет собой функцию интервалов (или меру). Когда используются круглые скобки, аргумент в $F(a)$ есть точка. Соотношение между функцией точки $F(\cdot)$ и функцией интервала $F\{ \}$ выражается равенствами

$$F(x) = F\{\overline{-\infty, x}\}, \quad F\{a, b\} = F(b) - F(a). \quad (1.2)$$

В действительности понятие функции точки $F(x)$ излишне и служит лишь для удобства обозначения и записи. Основным и первичным является наделение вероятностями *интервалов*. Функция точки $F(\cdot)$ называется *функцией распределения* функции интервалов $F\{ \}$. Символы $F(\cdot)$ и $F\{ \}$ представляют собой различное выражение одного и того же, и никакой двусмысленности не появляется, когда говорят о «вероятностном распределении F ». Следует привыкнуть думать в терминах функций *интервалов* и мер и использовать функцию распределения только как обозначение¹⁾.

Определение. *Определенная на прямой функция точки F есть функция распределения, если она*

- (i) *неубывающая, т. е. $F(a) \leq F(b)$ при $a < b$,*
- (ii) *непрерывна справа²⁾, т. е. $F(a) = F(a+)$,*
- (iii) *удовлетворяет условиям $F(-\infty) = 0$, $F(\infty) < \infty$.*

Функция распределения F , для которой $F(\infty) = 1$, называется *вероятностной функцией распределения*. Функция распределения называется *дефектной* (или *несобственной*), если $F(\infty) < 1$.

Покажем, что каждая функция распределения задает (соответствующие ей) вероятности всех множеств на прямой. Первым шагом является наделение вероятностями интервалов. Поскольку F монотонна, предел слева $F(a-)$ существует в каждой точке a . Определим функцию интервалов $F\{I\}$, положив

$$\begin{aligned} F\{a, b\} &= F(b) - F(a), & F\{a, \overline{b}\} &= F(b-) - F(a), \\ F\{\overline{a}, b\} &= F(b) - F(a-), & F\{\overline{a}, \overline{b}\} &= F(b-) - F(a-). \end{aligned} \quad (1.3)$$

Для интервала $\overline{a, a}$, сводящегося к единственной точке a , имеем

$$F\{\overline{a, a}\} = F(a) - F(a-),$$

¹⁾ Для вводных курсов представляется разумной педантичная тщательность в использовании обозначений, однако автор надеется, что читатели *не будут* требовать такого рода «состоятельности» и найдут в себе смелость писать без различия $F\{I\}$ и $F(x)$. Это не повлечет никакой двусмысленности. В лучшей математической литературе стало привычным, к счастью, использовать один и тот же символ (в частности, 1 и $=$) на одной и той же странице в нескольких смыслах.

²⁾ Как обычно, $f(a+)$ обозначает предел (если он существует) значений $f(x)$, когда $x \rightarrow a$ при условии $x > a$. Под $f(\infty)$ понимается предел $f(x)$ при $x \rightarrow \infty$. Аналогично определяются $f(a-)$ и $f(-\infty)$. Эти обозначения распространяются и на многомерный случай.

что представляет собой скачок F в точке a . (Вскоре мы увидим, что F непрерывна «почти всюду».)

Чтобы показать, что определяемые формулами (1.3) значения вероятностей для интервалов удовлетворяют требованиям теории вероятностей, докажем следующую простую лемму (читатели могут принять ее как интуитивно очевидную).

Лемма 1. (Счетная аддитивность.) *Если интервал I представляет собой объединение счетного числа непересекающихся интервалов $I_1, \dots, I_n$, то*

$$F\{I\} = \sum F\{I_k\}. \quad (1.4)$$

Доказательство. В частном случае, когда $I = \overline{a, b}$ и $I_1 = \overline{a, a_1}$, $I_2 = \overline{a_1, a_2}, \dots, I_n = \overline{a_{n-1}, b}$, утверждение леммы тривиально. Произвольное конечное разбиение $I = \overline{a, b}$ получается из этого разбиения перераспределением конечных точек a_k подынтервалов, так что правило сложения (1.4) для конечных разбиений выполняется.

При рассмотрении случая бесконечного множества интервалов I_k , не ограничивая общности, можно допустить, что интервал I замкнут. В силу непрерывности справа функции распределения F для любого заданного $\varepsilon > 0$ можно найти содержащий I_k открытый интервал $I_k^\#$, такой, что $0 \leq F\{I_k^\#\} - F\{I_k\} \leq \varepsilon 2^{-k}$. Поскольку существует конечная совокупность интервалов $I_{k_1}^\#, \dots, I_{k_n}^\#$, покрывающая I , то

$$F\{I\} \leq F\{I_{k_1}^\#\} + \dots + F\{I_{k_n}^\#\} \leq F\{I_1\} + \dots + F\{I_n\} + \varepsilon, \quad (1.5)$$

откуда

$$F\{I\} \leq \sum F\{I_k\}. \quad (1.6)$$

Обратное неравенство тоже справедливо, так как для каждого n существует конечное разбиение I , содержащее в качестве подынтервалов интервалы $I_1, \dots, I_n$. Тем самым лемма доказана. ►

Как уже говорилось в гл. IV, 2, теперь можно определить

$$F\{A\} = \sum F\{A_k\} \quad (1.7)$$

для любого множества A , представляющего собой конечное или счетное объединение непересекающихся интервалов A_k . Интуитивно можно ожидать, что любое множество допускает аппроксимацию такими объединениями интервалов, и теория меры это подтверждает¹⁾. Используя естественные приближения и переходы к пре-

¹⁾ Следует иметь в виду принятое нами соглашение, согласно которому слова «множество» и «функция» сокращенно означают соответственно «борелевское множество» и «борелевская функция».

делу, можно продолжить определение F на все множества так, чтобы свойство счетной аддитивности (1.7) сохранялось. Это продолжение единственно, и получающаяся в результате функция множеств называется вероятностным распределением или мерой.

Замечание о терминологии. Термин «распределение» свободно используется в литературе в различных смыслах, и поэтому здесь уместно точно оговорить, что понимается под распределением в настоящей книге.

Вероятностное распределение, или *вероятностная мера*, есть функция, ставящая в соответствие множествам числа $F\{A\} \geq 0$ и удовлетворяющая условию счетной аддитивности (1.7) и условию нормировки $F\{-\infty, \infty\} = 1$. Отбрасывая условие нормировки, получаем определение более общей меры (распределения массы). Весьма важным примером такой меры является мера Лебега (или обычная длина).

Иногда приходится иметь дело с мерами, относительно которых масса всей прямой $\rho = F\{-\infty, \infty\}$ меньше 1 (достаточно вспомнить теорию рекуррентных событий из первого тома). Мету с таким свойством мы называем вероятностной *несобственной* или *дефектной* мерой с *дефектом*, равным $1 - \rho$. Иногда с целью стилистической ясности или желая подчеркнуть, что рассматривается не мера вообще и не дефектная вероятностная мера, а именно вероятностная мера, мы будем называть ее *собственной* вероятностной мерой, хотя прилагательное *собственная* и излишне.

Аргументом меры $m\{A\}$ является множество; при письме буква, обозначающая это множество, заключается в фигурные скобки. Каждой ограниченной мере m соответствует *функция распределения*, т. е. функция точки, определяемая равенством $m(x) = m\{-\infty, x\}$. Функция распределения обозначается той же буквой, что и соответствующая ей мера, только аргумент ее заключается в круглые скобки. Двойное использование одной и той же буквы не вызывает никаких недоразумений, и, более того, термин «распределение» может сокращенно обозначать как вероятностное распределение, так и его функцию распределения. ►

В 1, гл. IX, *случайная величина* была определена как вещественная функция на выборочном пространстве. Мы сохраняем здесь это определение. В случае когда выборочным пространством является прямая, каждая вещественная функция есть случайная величина. Координатная случайная величина X является основной, и все остальные случайные величины можно рассматривать как функции от нее. Функция распределения случайной величины u , по определению равная $P\{u(X) \leq x\}$, может быть выражена в терминах функции распределения координатной случайной

величины X . Например, функция распределения величины X^3 есть $F(\sqrt[3]{x})$.

Функция u называется *простой*, если она принимает лишь счетное множество значений $a_1, a_2, \dots$. Пусть u — простая функция и A_n — то множество, на котором u равна a_n . Скажем, что *математическое ожидание функции u существует и равно*

$$E(u) = \sum a_k F\{A_k\}, \quad (1.8)$$

если ряд в правой части (1.8) абсолютно сходится. В противном случае u называется *неинтегрируемой* по отношению к F . Таким образом, u имеет математическое ожидание тогда и только тогда, когда существует $E(|u|)$. Исходя из (1.8), можно следующим образом определить математическое ожидание произвольной ограниченной функции u . Возьмем $\varepsilon > 0$ и обозначим через A_n множество тех точек x , в которых $(n-1)\varepsilon < u(x) \leq n\varepsilon$. При любом разумном определении $E(u)$ должно быть

$$\sum (n-1)\varepsilon \cdot F\{A_n\} \leq E(u) \leq \sum n\varepsilon F\{A_n\}. \quad (1.9)$$

(Суммы в (1.9) представляют собой математические ожидания двух аппроксимирующих простых функций $\underline{\sigma}$ и $\bar{\sigma}$, таких, что $\underline{\sigma} \leq u \leq \bar{\sigma}$ и $\bar{\sigma} - \underline{\sigma} = \varepsilon$.) Поскольку функция $\bar{u}$, согласно предположению, ограничена, то суммы в (1.9) содержат лишь конечное число отличных от нуля членов. Разность этих сумм равна $\varepsilon \sum F\{A_n\} = \varepsilon$. При замене ε на $\varepsilon/2$ левая сумма, очевидно, увеличивается, а правая — уменьшается. Отсюда легко вывести, что при $\varepsilon \rightarrow 0$ обе суммы в (1.9) стремятся к одному и тому же пределу, который и полагается равным $E(u)$. Для неограниченной функции применимо то же самое рассуждение, если оба ряда в (1.9) абсолютно сходятся; в противном случае $E(u)$ не определяется.

Определенное таким простым способом математическое ожидание представляет собой *интеграл Лебега—Стилтьеса* от функции u по отношению к распределению F . Когда желают подчеркнуть зависимость математического ожидания от F , предпочтительно использовать интегральное обозначение

$$E(u) = \int_{-\infty}^{\infty} u(x) F\{dx\}. \quad (1.10)$$

За исключением редких случаев, мы будем иметь дело лишь с монотонными функциями u или с кусочно-непрерывными функциями, такими, что для них множества A_n сводятся к объединению конечного числа интервалов. При этом суммы в (1.9) представляют собой, с точностью до порядка суммирования, верхние и нижние суммы, используемые при элементарном определении обычных интегралов. Общий интеграл Лебега—Стилтьеса, обладая основными свойствами обычного интеграла, имеет дополнительное

преимущество—при обращении с ним меньше приходится заботиться о законности формальных операций и переходов к пределу. Мы будем использовать математические ожидания лишь в очень простых ситуациях, и никакой общей теории не потребуются для того, чтобы следить за отдельными этапами рассуждений. Читатель, интересующийся теоретическим фундаментом и основными фактами, может обратиться к гл. IV.

Примеры. а) Пусть F —дискретное распределение, наделяющее точки $a_1, a_2, \dots$ массами $p_1, p_2, \dots$. Ясно, что в этом случае математическое ожидание $E(u)$ функции u , если оно существует, равно $\sum u(a_k) p_k$, и условием существования $E(u)$ является абсолютная сходимость ряда $\sum u(a_k) p_k$. Это согласуется с определением, данным в 1, гл. IX.

б) Для распределений, задаваемых непрерывной плотностью, математическое ожидание u представляет собой интеграл

$$E(u) = \int_{-\infty}^{\infty} u(x) f(x) dx, \quad (1.11)$$

если только этот интеграл сходится абсолютно. По поводу общего понятия плотности см. § 3. ►

Обобщение вышеизложенного на многомерные пространства можно описать в нескольких словах. Точка в $\mathcal{R}^2$ представляет собой пару действительных чисел $x = (x_1, x_2)$. Неравенства следует понимать в покоординатном смысле¹⁾; так $a < b$ означает, что $a_1 < b_1$ и $a_2 < b_2$ (или, иными словами, « a лежит юго-западнее b »). Это отношение порядка является лишь частичным упорядочением, т. е. две точки a и b не обязаны находиться в одном из двух отношений $a < b$, $a \geq b$. *Интервалами* мы называем множества, определяемые четырьмя возможными типами двойных неравенств $a < x < b$ и т. д. Интервалы представляют собой прямоугольники со сторонами, параллельными осям координат, и могут также вырождаться в отрезки или точки.

Единственно новым является то обстоятельство, что если $\overline{a, c}$ —двумерный интервал и $\overline{a < b < c}$, то $\overline{a, c}$ не есть объединение $\overline{a, b}$ и $\overline{b, c}$. Каждой функции интервала $F\{I\}$ можно поставить в соответствие отвечающую ей функцию распределения, которая, как и в одномерном случае, определяется равенством $F(x) = F\{\overline{-\infty, x}\}$. Теперь, однако, в выражение $F\{\overline{a, b}\}$ в терминах функции распределения входят все четыре вершины интервала. В самом деле, из рассмотрения двух бесконечных полос, параллельных x_2 —оси и имеющих основания—стороны прямо-

¹⁾ Это соглашение было введено в гл. III, 5.

угольника $\overleftarrow{a, b}$, непосредственно следует, что $F\{\overleftarrow{a, b}\}$ выражается через так называемую «смешанную разность»

$$F\{\overleftarrow{a, b}\} = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \quad (1.12)$$

Для функции распределения $F(x) = F(x_1, x_2)$ правая часть (1.12) неотрицательна. Отсюда следует монотонность $F(x_1, x_2)$ по x_1 и x_2 , однако обратное неверно: монотонность F не обеспечивает положительности (1.12). (См. задачу 4.)

Ясно, что в многомерном случае функции распределения являются не очень подходящим аппаратом: если бы не соображение аналогии со случаем пространства $\mathcal{R}^1$, все рассмотрения, возможно, велись бы в терминах функций интервалов. Формально определение функции распределения в $\mathcal{R}^1$ распространяется на $\mathcal{R}^2$, если условие монотонности (1) заменить на условие неотрицательности смешанных разностей (1.12) при $a \leq b$. Так определенная функция распределения индуцирует функцию интервалов (соотношения подобны (1.3), в случае $\mathcal{R}^1$ нужно лишь обычные разности заменить на смешанные). Лемма 1 сохраняется вместе с доказательством ¹⁾.

Одно простое, но существенно важное свойство математических ожиданий принимается иногда как само собой разумеющееся. Любая функция $u(\mathbf{X}) = u(X_1, X_2)$ от координатных переменных является случайной величиной и как таковая имеет функцию распределения G . Математическое ожидание $E(u(\mathbf{X}))$ можно определить двумя способами — как интеграл от $u(x_1, x_2)$ по отношению к заданной вероятностной мере на плоскости, а также в терминах функции распределения величины u :

$$E(u) = \int_{-\infty}^{\infty} uG\{dy\}. \quad (1.13)$$

Оба определения эквивалентны в силу самого определения первого из этих интегралов с помощью аппроксимирующих сумм (гл. IV, (4.3) ²⁾). Существенным здесь является то обстоятельство, что математическое ожидание случайной величины Z (если оно существует) имеет внутренний смысл и не зависит от того, рассматривается ли Z как функция на исходном вероятностном пространстве $\mathcal{B}$ или на пространстве, полученном при соответствующем отображении $\mathcal{B}$; в частности, Z сама отображает $\mathcal{B}$ на прямую и превращается на ней в координатную величину.

¹⁾ В доказательстве используется тот факт, что при конечном разбиении одномерного интервала подынтервалы имеют естественный порядок слева направо. Столь же хорошее упорядочение имеет место при шахматных разбиениях $\overleftarrow{a, b}$ двумерного интервала $\overleftarrow{a, b}$, т. е. разбиениях на m подынтервалов,

получаемых разбиениями отдельно обеих сторон $\overleftarrow{a, b}$ и проведением прямых, параллельных осям, через все точки подразбиений. Доказательство конечной аддитивности для таких шахматных разбиений не меняется, и каждому разбиению соответствует его шахматное подразбиение. Переход от конечных разбиений к счетным одинаков при любом числе измерений.

²⁾ Специальный случай содержится в теореме 1, гл. IX, 2.1.

Начиная с этого места, не будет никакой разницы между случаями пространства $\mathcal{R}^1$ и пространства $\mathcal{R}^2$. В частности, определение математического ожидания не зависит от числа измерений.

Резюмируя, можно сказать, что *любая функция распределения индуцирует вероятностную меру на σ -алгебре борелевских множеств в $\mathcal{R}^r$ и тем самым определяет вероятностное пространство*. Иначе говоря, мы показали, что вероятностная конструкция дискретных выборочных пространств, а также пространств с распределениями, задаваемыми плотностями, распространяется без формальных изменений на общий случай. Тем самым оправдана вероятностная терминология, использовавшаяся в первых трех главах. Когда мы говорим об r случайных величинах $X_1, \dots, X_r$, то подразумеваем, что все они определены на одном и том же вероятностном пространстве, и поэтому существует их совместное вероятностное распределение. При этом мы можем, если пожелаем, рассматривать X_k как координатные величины в выборочном пространстве $\mathcal{R}^r$.

Вряд ли необходимо объяснять способ распространения таких понятий, как *маргинальное (частное) распределение* (см. гл. III, 1, и 1, гл. IX, 1) или *независимость случайных величин*. Основные утверждения о независимых случайных величинах переносятся с дискретного на общий случай без изменений.

(i) Независимость случайных величин X и Y с (одномерными) распределениями F и G означает, что функция совместного распределения (X, Y) задается произведением $F(x_1)G(x_2)$. Это может относиться к двум случайным величинам, определенным на заданном вероятностном пространстве, но может также пониматься и как сокращенное выражение того, что мы вводим плоскость с X и Y в качестве координатных переменных и *определяем* вероятности по правилу произведения. Сказанное в равной степени применимо к парам, тройкам и вообще к любому числу случайных величин.

(ii) Если совокупность m случайных величин $(X_1, \dots, X_m)$ не зависит от совокупности n случайных величин $(Y_1, \dots, Y_n)$, то случайные величины $u(X_1, \dots, X_m)$ и $v(Y_1, \dots, Y_n)$ независимы, каковы бы ни были функции u и v .

(iii) Для независимых X и Y справедливо равенство $E(XY) = E(X)E(Y)$, если только математические ожидания X и Y существуют (т. е. если соответствующие интегралы сходятся абсолютно).

Следующий простой результат часто приходится использовать.

Лемма 2. *Вероятностное распределение однозначно определяется значениями $E(u)$ для всех непрерывных функций u , обращающихся в нуль вне некоторого конечного интервала.*

Доказательство. Пусть I — конечный открытый интервал и v — непрерывная функция, которая положительна на I и равна нулю

вне I . Тогда $\sqrt[n]{v(x)} \rightarrow 1$ при всех $x \in I$ и, следовательно, $E(\sqrt[n]{v}) \rightarrow F\{I\}$. Таким образом, значения математических ожиданий непрерывных функций, обращающихся в нуль вне некоторого интервала, однозначно определяют значения $F\{I\}$ для всех открытых интервалов, которые в свою очередь однозначно определяют F . $\blacktriangleright$

Замечание 1¹⁾. Теорема Рисса о представлении. В последней лемме математические ожидания определялись в терминах заданного вероятностного распределения. Часто (например, в проблеме моментов, см. гл. VII, 3) нам приходится отправляться от заданного функционала, назначая значения $E(u)$ некоторым известным функциям u . Вопрос в том, существует ли вероятностное распределение F , для которого

$$E(u) = \int_{-\infty}^{+\infty} u(x) F\{dx\}. \quad (1.14)$$

Оказывается, что следующие явно необходимые условия являются также и достаточными.

Теорема. Допустим, что каждой непрерывной функции u обращаемой в нуль вне конечного интервала, соответствует число $E(u)$ со следующими свойствами:

(i) Функционал $E(u)$ является линейным, т. е. для всех линейных комбинаций выполняется

$$E(c_1 u_1 + c_2 u_2) = c_1 E(u_1) + c_2 E(u_2).$$

(ii) Функционал $E(u)$ положителен, т. е. $E(u) \geq 0$ при $u \geq 0$.

(iii) Функционал $E(u)$ имеет норму 1, т. е. если $0 \leq u \leq 1$, то $E(u) \leq 1$, но при всяком $\varepsilon > 0$ существует такая функция u , что $0 \leq u \leq 1$ и $E(u) > 1 - \varepsilon$. Тогда существует единственное вероятностное распределение F , для которого выполняется (1.14).

Доказательство. Для произвольных t и $h > 0$ обозначим через $z_{t,h}$ непрерывную функцию переменного x , которая равна 1 при $x \leq t$, равна 0 при $x \geq t+h$ и линейна на промежуточном интервале $t \leq x \leq t+h$. Эта функция не обращается в нуль на бесконечности, но мы можем определить $E(z_{t,h})$ посредством простой аппроксимации. Возьмем функцию u_n , $|u_n| \leq 1$, такую, что $E(u_n)$ определено и $u_n(x) = z_{t,h}(x)$ при $|x| < n$. При $m > n$ разность $u_m - u_n$ внутри интервала $-n, n$ равна тождественно нулю. Поэтому из условия, что E имеет норму 1, можно легко вывести, что $E(u_n - u_m) \rightarrow 0$. Отсюда следует, что $E(u_n)$ сходится к конечному пределу и этот предел очевидным образом не зависит от конкретного выбора аппроксимирующих функций u_n . Таким образом можно законным способом определить $E(z_{t,h}) = \lim E(u_n)$. Легко видеть, что даже в пределах расширенной области определения функционала E обладают тремя свойствами, постулируемыми теоремой.

Положим далее $F_h(t) = E(z_{t,h})$. При фиксированном h эта функция монотонно меняется от 0 до 1. Она непрерывна, так как при $0 < \delta < h$ разность $z_{t+\delta,h} - z_{t,h}$ имеет треугольный график с высотой δ/h и, следовательно, отношение приращений для F_h ограничено величиной $1/h$. При $h \rightarrow 0$ функции F_h монотонно убывают к пределу, который мы обозначим через F . Покажем, что F представляет собой вероятностное распределение. Очевидно, что F монотонна и $F(-\infty) = 0$. Кроме того, $F(t) \geq F_h(t-h)$, откуда следует, что $F(\infty) = 1$. Остается показать, что F непрерывна справа. При

¹⁾ Это замечание относится к вопросу, имеющему принципиальный интерес, и не будет использоваться в дальнейшем. Другой подход см. в гл. IV, 5.

заданных t и $\varepsilon > 0$ выберем значение h столь малым, чтобы $F(t) > F_h(t) - \varepsilon$. Вследствие непрерывности F_h мы имеем для достаточно малых δ

$$F(t) > F_h(t) - \varepsilon > F_h(t + \delta) - 2\varepsilon \geq F(t + \delta) - 2\varepsilon,$$

откуда следует непрерывность справа.

Пусть u — непрерывная функция, обращающаяся в нуль вне конечного интервала a, b . Выберем $a = a_0 < a_1 < \dots < a_n = b$ такими, что внутри каждого подынтервала a_{k-1}, a_k колебания значений функции u меньше, чем ε . Если h меньше длины наименьшего среди этих интервалов, то

$$u_h = \sum_{k=1}^n u(a_k) [z_{a_k, h} - z_{a_{k-1}, h}] \quad (1.15)$$

есть кусочно-линейная функция с вершинами в точках a_k и $a_k + h$. Поскольку $u(a_k) = u_h(a_k)$, то $|u - u_h| \leq 2\varepsilon$ и, следовательно, $|E(u) - E(u_h)| \leq 2\varepsilon$. Однако при $h \rightarrow 0$

$$E(u_h) \rightarrow \sum_{k=1}^n u(a_k) F(a_k - 1, a_k), \quad (1.16)$$

причем сумма справа в (1.16) отличается от интеграла в (1.14) меньше чем на ε . Таким образом, левая и правая части в (1.14) отличаются меньше чем на 3ε , и, следовательно, (1.14) справедливо. ►

Замечание II. О независимости и корреляции. Теория статистической корреляции восходит к тому времени, когда формализация теории была еще невозможна и понятие стохастической независимости по необходимости носило мистический характер. Уже тогда понимали, что независимость двух ограниченных случайных величин о нулевыми математическими ожиданиями влечет равенство $E(XY) = 0$, однако сначала думали, что это равенство должно быть и достаточным для независимости X и Y . После того как была обнаружена ложность этого заключения, долгое время искали условия, при которых обращение в нуль корреляции влечет стохастическую независимость. Как часто случается, история задачи и красота частных результатов затемнили тот факт, что современные методы позволяют дать чрезвычайно простое ее решение. Следующая теорема содержит различные результаты, доказанные ранее трудоемкими методами.

Теорема. Случайные величины X и Y независимы тогда и только тогда, когда

$$E(u(X)v(Y)) = E(u(X))E(v(Y)) \quad (1.17)$$

для всех непрерывных функций u и v , обращающихся в нуль вне конечного интервала.

Доказательство. Необходимость условия очевидна. Достаточность его будет доказана, если показать, что для всякой непрерывной функции $w(X, Y)$, равной нулю вне некоторого конечного интервала, математическое ожидание $E(w)$ совпадает с математическим ожиданием w относительно пары независимых случайных величин, распределенных как X и Y . Но, согласно (1.17), это верно для функций вида $w(X, Y) = u(X)v(Y)$. Так как каждую непрерывную функцию w , обращающуюся в нуль вне некоторого конечного интервала, можно равномерно приблизить¹⁾ линейными комбинациями вида $\sum c_k u_k(X) v_k(Y)$, то, переходя к пределу, убеждаемся в справедливости нашего утверждения.

¹⁾ См. задачу 10 в гл. VIII, 10.

§ 2. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Настоящий параграф посвящен в основном известным или очевидным понятиям и фактам, связанным с функциями распределения в $\mathcal{R}^1$.

Как и в дискретном случае, k -й момент случайной величины X по определению полагается равным $E(X^k)$. При этом предполагается, что интеграл

$$E(X^k) = \int_{-\infty}^{\infty} x^k F\{dx\} \quad (2.1)$$

сходится абсолютно. Таким образом, $E(X^k)$ существует тогда и только тогда, когда $E(|X|^k) < \infty$. Величина $E(|X|^k)$ называется *абсолютным моментом* X порядка k . Она определена и для нецелых $k > 0$. Так как при $0 < a < b$ имеет место неравенство $|x|^a \leq |x|^b + 1$, то из существования абсолютного момента порядка b вытекает существование абсолютных моментов всех порядков $a < b$.

Второй момент случайной величины $X - m$, где m — математическое ожидание X , называется *дисперсией* X и обозначается $\text{Var } X$:

$$\text{Var}(X) = E((X - m)^2) = E(X^2) - m^2. \quad (2.2)$$

В рассматриваемом общем случае дисперсия обладает теми же свойствами и играет ту же роль, что и в дискретном случае. В частности, если X и Y независимы и дисперсии $\text{Var } X$ и $\text{Var } Y$ существуют, то

$$\text{Var}(X + Y) = \text{Var } X + \text{Var } Y. \quad (2.3)$$

[Две случайные величины, удовлетворяющие равенству (2.3), называются *некоррелированными*. В I, гл. IX, 8, было показано, что существуют зависимые некоррелированные случайные величины.]

Вспомним, что мы часто заменяли случайную величину на «нормированную величину» $X^* = (X - m)/\sigma$, где $m = E(X)$ и $\sigma^2 = \text{Var}(X)$. Выражаясь физическим языком, можно сказать, что X^* есть запись X в «безразмерных единицах». Вообще переход от X к $(X - \beta)/\alpha$ при $\alpha > 0$ есть изменение начала отсчета и единицы измерения. Функция распределения новой случайной величины равна $F(\alpha x + \beta)$, и нередко мы имеем дело по существу со всем классом функций распределения этого вида, а не с отдельным представителем класса. Введем для удобства следующее

Определение 1. Два распределения F_1 и F_2 в $\mathcal{R}^1$ называются *распределениями одного и того же типа*¹⁾, если $F_2(x) = F_1(\alpha x + \beta)$,

¹⁾ Это понятие было введено Хинчиным, который использовал немецкий термин «Klasse» (класс), однако в английской литературе «класс функций» имеет установившееся значение.

$\alpha > 0$. Параметр α называют масштабным множителем, а параметр β — центрирующей постоянной (или параметром расположения).

Данное определение позволяет использовать выражения вида « F центрировано к нулевому математическому ожиданию» или «центрирование не меняет дисперсии».

Медианой распределения F называется число ξ , удовлетворяющее условиям $F(\xi) \geq 1/2$ и $F(\xi -) \leq 1/2$. Медиана может быть определена неоднозначно; если $F(x) = 1/2$ при всех x из некоторого интервала $\overline{a, b}$, то каждое x из этого интервала есть медиана. Распределение можно центрировать так, чтобы его медиана стала равной нулю.

За исключением медианы, все рассмотренные понятия переносятся на случай любого конечного числа измерений, или, иными словами, на векторные случайные величины вида $\mathbf{X} = (X_1, \dots, X_n)$. Соответствующие векторные обозначения были введены в гл. III, § 5, и не требуют никаких изменений. Математическое ожидание $\mathbf{X}$ представляет собой в этом случае вектор, а дисперсия $\mathbf{X}$ — матрицу.

Первое, на что обращается внимание при рассмотрении графика функции распределения, — это разрывы и интервалы постоянства. Нередко приходится оговаривать, что точка не принадлежит интервалу постоянства. Введем следующую удобную терминологию, относящуюся к любому конечному числу измерений.

Определение 2. Точка x является атомом, если она имеет положительную массу. Точка x , для которой $F\{I\} > 0$, каков бы ни был открытый интервал I , содержащий x , называется точкой роста.

Распределение F сосредоточено на множестве A , если дополнение A' имеет вероятность $F\{A'\} = 0$.

Распределение F называется атомическим, если оно сосредоточено на множестве своих атомов.

Пример. Расположим рациональные числа из отрезка $\overline{0, 1}$ в последовательность $r_1, r_2, \dots$ так, чтобы с ростом номера числа рос и его знаменатель. Определим распределение F , положив вероятность каждой точки r_k равной 2^{-k} . Распределение F атомическое, однако все точки замкнутого интервала $\overline{0, 1}$ являются точками роста F . 

В силу условия счетной аддитивности (1.7) сумма масс атомов не превосходит единицы, так что имеется не больше одного атома с массой $> 1/2$, не больше двух атомов с массой $> 1/3$ и т. д. Все атомы, следовательно, можно расположить в простую последовательность $a_1, a_2, \dots$ так, чтобы соответствующие им массы не возрастали: $p_1 \geq p_2 \geq \dots$. Иными словами, число атомов не более чем счетно.

Распределение, не имеющее атомов, называется *непрерывным*. Предположим, что распределение имеет атомы в массах $p_1, p_2, \dots$, и обозначим $p = \sum p_k$. Положим

$$F_a(x) = \frac{1}{p} \sum_{a_k \leq x} p_k, \quad (2.4)$$

где суммирование распространяется на все атомы из интервала $-\infty, x$. Функция $F_a(x)$, представляющая собой, очевидно, функцию распределения, называется *атомической компонентой* F . Если $p=1$, то распределение F является атомическим. При $p < 1$ положим $q=1-p$. Легко видеть, что $[F - pF_a]/q = F_c$ есть *непрерывная* функция распределения. Таким образом, имеем

$$F = pF_a + qF_c, \quad (2.5)$$

т. е. F является линейной комбинацией двух функций распределения F_a и F_c , из которых F_a — атомическая, а F_c — непрерывная. Если F есть атомическая функция распределения, то (2.5) по-прежнему имеет место с $p=1$, причем в качестве F_c можно взять произвольную непрерывную функцию распределения. При отсутствии атомов в (2.5) будет $p=0$. Нами доказана, таким образом,

Теорема Жордана о разложении. *Всякое вероятностное распределение является смесью атомического и непрерывного распределений вида (2.5) с $p \geq 0, q \geq 0, p+q=1$.*

Имеется один класс атомических распределений, из-за которых приходится иногда загромождать простые формулировки перечислением тривиальных исключений. Распределения этого класса лишь произвольным масштабным множителем отличаются от распределений, сосредоточенных в целочисленных точках, однако встречаются они весьма часто и поэтому заслуживают специального названия для удобства ссылок.

Определение 3. *Распределение F на $\mathcal{A}^1$ называется арифметическим¹⁾, если оно сосредоточено на множестве точек вида $0, \pm \lambda, \pm 2\lambda, \dots$. Наибольшее число λ , обладающее этим свойством, называется шагом F .*

§ 3. ПЛОТНОСТИ

Первые две главы были посвящены изучению вероятностных распределений в $\mathcal{A}^1$, для которых вероятность произвольного

¹⁾ Пожалуй, более общепринятым является термин «решетчатое распределение», однако его употребляют в разных смыслах: некоторые авторы под решетчатым распределением понимают распределение, сосредоточенное на множестве $a, a \pm \lambda, a \pm 2\lambda, \dots$, где a — произвольно. (Биномиальное распределение с атомами в точках ± 1 в нашей терминологии является арифметическим распределением с шагом 1, в то время как в упомянутой другой терминологии оно — решетчатое распределение с шагом 2.)

интервала (а, следовательно, и вообще произвольного множества) задается интегралом

$$F\{A\} = \int_A \varphi(x) dx. \quad (3.1)$$

Распределения, рассматривавшиеся в гл. III, имеют такой же вид с той лишь разницей, что интегрирование происходит по лебеговой мере (площади или объему) в $\mathcal{R}$. Если плотность φ в (3.1) сосредоточена на интервале $\overline{0, 1}$, то (3.1) можно переписать в виде

$$F\{A\} = \int_A \varphi(x) U\{dx\}, \quad (3.2)$$

где U обозначает равномерное распределение на $\overline{0, 1}$. Последняя формула имеет смысл для произвольного вероятностного распределения U , и, когда $F\{-\infty, \infty\} = 1$, она определяет новое вероятностное распределение F . В этом случае мы говорим, что φ есть плотность F по отношению к U .

В формуле (3.1) мера U бесконечна, в то время как в формуле (3.2) имеем $U\{-\infty, \infty\} = 1$. Это различие несущественно, поскольку интеграл в (3.1) может быть разбит на интегралы вида (3.2), распространенные на конечные интервалы. Хотя мы будем использовать (3.2) лишь в случаях, когда U есть либо вероятностное распределение, либо мера Лебега, как в (3.1), дадим, однако, следующее общее определение.

Определение. Распределение F называется абсолютно непрерывным по отношению к мере U , если для некоторой функции φ имеет место соотношение (3.2). При этом φ называется плотностью ¹⁾ F по отношению к U .

Наиболее важным частным случаем является, конечно, тот случай, когда имеет место (3.1), т. е. когда U есть мера Лебега. Функция φ в этом случае называется «обычной» плотностью.

Введем теперь следующую сокращенную запись:

$$F\{dx\} = \varphi(x) U\{dx\}. \quad (3.3)$$

Эта запись имеет лишь тот смысл, что для всех множеств выполняется равенство (3.2), и никакого специального смысла символу dx придавать не следует. Пользуясь этой сокращенной записью, мы будем записывать (3.1) в виде $F\{dx\} = \varphi(x) dx$, и если U имеет обычную плотность u , то (3.2) принимает вид $F\{dx\} = \varphi(x) u(x) dx$.

Примеры. а) Пусть U — вероятностное распределение на $\mathcal{R}^1$ со вторым моментом, равным m_2 . Тогда

$$F\{dx\} = \frac{1}{m_2} x^2 U\{dx\}$$

¹⁾ В теории меры функция φ называется производной Радома — Никодима распределения F по отношению к U .

есть новое вероятностное распределение. В частности, если U — равномерное распределение на $\overline{0, 1}$, то $F(x) = x^3$ при $0 < x < 1$, а если U имеет плотность e^{-x} ($x > 0$), то F есть гамма-распределение с обычной плотностью $\frac{1}{2} x^2 e^{-x}$.

б) Пусть U — атомическое распределение, наделяющее атомы $a_1, a_2, \dots$ массами $p_1, p_2, \dots$ (здесь $\sum p_k = 1$). Распределение F имеет плотность φ по отношению к U тогда и только тогда, когда оно атомическое и все его атомы являются также атомами распределения U . Если F имеет в a_j массу q_j , то плотность φ задается посредством равенства $\varphi(a_k) = q_k/p_k$. Значение φ в точках, отличных от точек a_j , не играет никакой роли, и лучше всего считать φ в этих точках неопределенной. ►

Теоретически подынтегральная функция φ в (3.2) определена неоднозначно. В самом деле, если N — такое множество, что $U\{N\} = 0$, то φ можно как угодно изменить на N и при этом (3.2) останется в силе. К этому, однако, и сводится вся неоднозначность в определении φ : *плотность определена однозначно с точностью до значений на нулевом множестве*¹⁾. На практике условиями непрерывности диктуется обычно однозначный выбор плотности, и поэтому, как правило, говорят о некоторой определенной плотности, хотя, конечно, более правильно говорить о некоторой из плотностей.

Из (3.3), очевидно, следует, что для любой ограниченной функции v ²⁾.

$$v(x) F\{dx\} = v(x) \varphi(x) U\{dx\}. \quad (3.4)$$

В частности, если все значения функции φ больше (или меньше) нуля на некоторое число, то, взяв $v = \varphi^{-1}$, получим формулу

¹⁾ В самом деле, пусть φ и φ_1 — две плотности F по отношению к U . Рассмотрим множество A всех точек x , для которых $\varphi(x) > \varphi_1(x) + \varepsilon$. Из равенств

$$F\{A\} = \int_A \varphi(x) U\{dx\} = \int_A \varphi_1(x) U\{dx\}$$

следует, что $U\{A\} = 0$, и поскольку это верно для любого $\varepsilon > 0$, то $\varphi(x) = \varphi_1(x)$ всюду, за исключением множества N , такого, что $U\{N\} = 0$.

²⁾ Читателям, которых смущает справедливость соотношения (3.4), следует обратить внимание на то, что для непрерывных плотностей (3.4) сводится к правилу подстановки для интегралов. Следующее доказательство (3.4) в общем случае использует стандартное рассуждение, применимое и в более общих ситуациях. Формула (3.4) тривиальна, когда v есть простая функция, принимающая лишь конечное число различных значений. Для любой ограниченной функции v существуют две простые функции с конечным числом значений $\underline{v}$ и $\bar{v}$, такие, что $\underline{v} < v \leq \bar{v}$ и $\bar{v} - \underline{v} < \varepsilon$, и поэтому из справедливости формулы (3.4) для всех простых функций вытекает ее справедливость в общем случае.

обращения для (3.2):

$$U\{dx\} = \frac{1}{\varphi(x)} F\{dx\}. \quad (3.5)$$

Полезный критерий абсолютной непрерывности содержится в одной из основных теорем теории меры, которую мы приведем здесь без доказательства.

Теорема Радона—Никодима¹⁾. *Распределение F абсолютно непрерывно по отношению к мере U тогда и только тогда, когда*

$$U\{A\} = 0 \text{ влечет } F\{A\} = 0. \quad (3.6)$$

Условие (3.5) можно перефразировать следующим образом:

U -нулевые множества являются также F -нулевыми множествами. Приведем одно важное следствие теоремы Радона—Никодима, которое, однако, нигде в дальнейшем непосредственно использоваться не будет.

Критерий. *Распределение F абсолютно непрерывно по отношению к мере U тогда и только тогда, когда каждому $\varepsilon > 0$ соответствует $\delta > 0$, такое, что для любой совокупности непесекающихся интервалов $I_1, \dots, I_n$*

$$\sum_1^n U\{I_k\} < \delta \text{ влечет } \sum_1^n F\{I_k\} < \varepsilon. \quad (3.7)$$

Весьма важным частным случаем выполнения (3.7) является тот случай, когда для всех интервалов

$$F\{I\} \leq aU\{I\}. \quad (3.8)$$

При этом (3.7) выполняется с $\delta = \varepsilon/a$, и, как легко видеть, плотность F по отношению к U не превосходит a .

3а. Сингулярные распределения²⁾

Условие (3.6) теоремы Радона—Никодима наводит на рассмотрение случая, прямо противоположного случаю абсолютно непрерывных распределений.

Определение. *Вероятностное распределение F сингулярно по отношению к U , если оно сосредоточено на множестве N , таком, что $U\{N\} = 0$.*

Особую роль здесь играет мера Лебега $U\{dx\} = dx$, и когда говорят, что распределение «сингулярно», и не делают при этом дополнительных пояснений, имеют в виду, что оно сингулярно по отношению к мере Лебега. Каждое атомическое распределение сингулярно по отношению к dx , но имеются и *непрерывные* распределения в $\mathcal{M}^1$, сингулярные относительно dx : таковым яв-

¹⁾ Эту теорему нередко называют теоремой Лебега—Никодима. Соотношение (3.6) может быть принято в качестве *определения* абсолютной непрерывности, и в этом случае теорема будет утверждать существование плотности.

²⁾ Хотя принципиально сингулярные распределения очень важны, они появляются в этой книге лишь между прочим.

ляется канторовское распределение, рассмотренное в примере 11, г) гл. I. Сингулярные распределения не поддаются аналитическому изучению, и их явное представление практически невозможно. Чтобы иметь возможность применять аналитические методы, приходится, следовательно, накладывать ограничения, которые обеспечивают либо абсолютную непрерывность, либо атомичность рассматриваемых распределений. Сингулярные распределения, однако, играют важную принципиальную роль, и многие статистические критерии основываются на их существовании. Это обстоятельство затемняется бытующим мнением, что «в практике» сингулярные распределения не встречаются.

Примеры. в) *Испытания Бернулли.* В примере 11, в) гл. I было показано, что выборочное пространство последовательностей $УУ...Н... (успехов «У» и неудач «Н»)$ можно отобразить на единичный интервал, пользуясь простым приемом замены символов «У» и «Н» на 1 и 0 соответственно. Выборочным пространством становится тогда единичный интервал, и результат бесконечной последовательности испытаний представляется случайной величиной $Y = \sum 2^{-k} X_k$, где X_k — независимые случайные величины, принимающие значения 1 и 0 с вероятностями p и q . Обозначим распределение Y через F_p . Если испытания симметричны, то $F_p = F_{1/2}$ есть *равномерное* распределение, и модель принимает привлекательный простой вид. По существу эквивалентность симметричных испытаний Бернулли со «случайным выбором точки из $[0, 1]$ » использовалась с самого возникновения теории вероятностей. Далее в силу закона больших чисел распределение F_p сосредоточено на множестве N_p тех точек, в двоичном разложении которых частота цифры 1 стремится к p . При $p \neq \alpha$ множество N_α имеет вероятность нуль и, следовательно, *распределения F_p сингулярны по отношению друг к другу*. В частности, при $p \neq 1/2$ распределение F_p сингулярно по отношению к равномерному распределению dx . Точного представления для F_p практически не существует, и поэтому описанная конструкция редко используется при $p \neq 1/2$. Два обстоятельства заслуживают особого внимания.

Во-первых, посмотрим, что было бы, если бы частный случай $p = 1/3$ представлял особый интерес или часто встречался в приложениях. Мы заменили бы тогда двоичное разложение чисел троичным и ввели бы новый масштаб так, чтобы $F_{1/3}$ представляло собой равномерное распределение. «Практически» мы по-прежнему имели бы дело лишь с абсолютно непрерывными распределениями, однако в основе этого лежит скорее выбор подхода, чем истинная природа вещей.

Во-вторых, симметричность или несимметричность монеты можно проверить статистическими методами и можно достичь практически определенного решения после конечного числа испы-

таний. Это оказывается возможным лишь благодаря тому, что события, являющиеся правдоподобными при гипотезе $p = 1/2$, весьма неправдоподобны при гипотезе $p = 1/3$. Небольшое размышление показывает, что возможность принять решение после конечного числа испытаний основана на сингулярности F_p при $p \neq 1/2$ по отношению к $F_{1/2}$. Таким образом, существование сингулярных распределений является существенным для статистической практики.

г) *Случайные направления.* В гл. I, 10, было введено понятие единичного вектора в $\mathbb{R}^3$ со случайным направлением. Распределение такого вектора сосредоточено на единичной окружности и, следовательно, является сингулярным по отношению к мере Лебега (площади) в плоскости. Можно было бы думать, что в этом случае в качестве выборочного пространства следует взять окружность, однако в некоторых практических задачах такой вариант выборочного пространства невозможен [см. пример 4, д)]. ►

Теорема Лебега о разложении. Каждое вероятностное распределение F является смесью вида

$$F = pF_s + qF_{ac}, \quad (3.9)$$

где $p \geq 0$, $q \geq 0$, $p + q = 1$, двух вероятностных распределений F_s и F_{ac} , из которых F_s сингулярно, а F_{ac} абсолютно непрерывно по отношению к заданной мере U .

Применяя к F_s жорданово разложение (2.5), получаем, что F можно представить в виде смеси трех вероятностных распределений, из которых первое — атомическое, второе — абсолютно непрерывное по отношению к $U \{dx\}$, а третье непрерывно, но сингулярно.

Доказательство. Множество N , для которого $U\{N\} = 0$, будем называть для краткости нулевым множеством. Пусть p есть верхняя грань значений $F\{N\}$ по всем нулевым множествам N . Для каждого n существует нулевое множество N_n , такое, что $F\{N_n\} > p - 1/n$. При этом $F\{A\} \leq 1/n$ для всякого нулевого множества A , лежащего в дополнении N_n' . Для множества $N = \bigcup N_n$ имеем $U\{N\} = 0$ и $F\{N\} = p$, так что никакое нулевое множество, лежащее в дополнении N' , не может иметь положительной вероятности.

Если $p = 1$, то F сингулярно, тогда как равенство $p = 0$ означает, что F абсолютно непрерывно. При $0 < p < 1$ утверждение теоремы выполняется для двух вероятностных распределений, определенных формулами

$$p \cdot F_s\{A\} = F\{AN\}, \quad q \cdot F_{ac}\{A\} = F\{AN'\}. \quad (3.10) \quad \blacktriangleright$$

§ 4. СВЕРТКИ

Трудно переоценить важность операции свертки во многих областях математики. Нам придется иметь дело со сверткой с двух точек зрения: и как с операцией, применяемой к распределе-

ниям, и как с операцией, применяемой к распределению и непрерывной функции.

Для определенности мы будем рассматривать распределения в $\mathcal{R}^1$, однако при использовании векторных обозначений § 1 все формулы остаются в силе и для распределений в пространстве любой размерности. Определение *свертки на окружности* следует схеме, описанной в гл. II, 8, и не требует дополнительных пояснений. (Более общие свертки можно определить на произвольных группах.)

Пусть F — вероятностное распределение и φ — ограниченная функция точек. (В рассматриваемых нами приложениях φ будет либо непрерывной функцией, либо функцией распределения.) Определим новую функцию u , полагая

$$u(x) = \int_{-\infty}^{\infty} \varphi(x-y) F\{dy\}. \quad (4.1)$$

Если F имеет плотность f (по отношению к dx), то

$$u(x) = \int_{-\infty}^{\infty} \varphi(x-y) f(y) dy. \quad (4.2)$$

Определение 1. Функция u в (4.1) называется *сверткой функции φ с вероятностным распределением F* . Для такой свертки будет использоваться обозначение $u = F \star \varphi$. Если F имеет плотность f , то мы будем писать $u = f \star \varphi$.

Отметим, что порядок компонент в свертке (4.1) существен: символ $\varphi \star F$, вообще говоря, смысла не имеет. С другой стороны, интеграл в формуле (4.2) имеет смысл для любых интегрируемых функций f и φ (даже если f принимает значения разных знаков) и символ $\star$ используется в этом обобщенном смысле. Ограниченность функции φ была введена конечно, ради простоты и вовсе не является необходимой.

Примеры. а) Если F — равномерное распределение на $\overline{0, a}$, то

$$u(x) = \frac{1}{a} \int_{x-a}^x \varphi(s) ds. \quad (4.3)$$

Ясно, что в этом случае функция u непрерывна; если предположить еще, что функция φ непрерывна, то u будет обладать непрерывной производной и т. д. Вообще говоря, u ведет себя лучше, чем φ , так что свертка служит *сглаживающим оператором*.

б) Формулы свертки для экспоненциальных и равномерных распределений [гл. 1, (3.6), и гл. I, (9.1)] являются частными случаями общей формулы. Примеры свертки распределений в $\mathcal{R}^2$ имеются в гл. III, (1.22), и в задачах 15–17 гл. III. ►

Теорема 1. Если функция φ ограничена и непрерывна, то свертка $u = F \star \varphi$ тоже ограничена и непрерывна. Если φ — вероятностное распределение, то u тоже является вероятностным распределением.

Доказательство. Если φ — ограниченная непрерывная функция, то, согласно принципу мажорированной сходимости, $u(x+h) \rightarrow u(x)$. По тем же соображениям из непрерывности справа функции φ следует непрерывность справа u . Наконец, поскольку φ монотонно меняется от 0 до 1, то же самое справедливо для u . ►

Следующая теорема дает вероятностную интерпретацию свертки $F \star \varphi$, когда φ есть функция распределения.

Теорема 2. Пусть X и Y — независимые случайные величины с распределениями F и G . Тогда

$$P\{X + Y \leq t\} = \int_{-\infty}^{\infty} G(t-x) F\{dx\}. \quad (4.4)$$

Доказательство ¹⁾. Возьмем $\varepsilon > 0$ и обозначим через I_n интервал $ne < x \leq (n+1)\varepsilon$, $n = 0, \pm 1, \dots$. Появление при некотором n события $\{X \in I_{n-1}, Y \leq t - ne\}$ влечет появление события $\{X + Y \leq t\}$. Следовательно, поскольку события $\{X \in I_{n-1}, Y \leq t - ne\}$ взаимно исключают друг друга и случайные величины X, Y независимы, имеем

$$P\{X + Y \leq t\} \geq \sum G(t - ne) F\{I_{n-1}\}. \quad (4.5)$$

Справа в этой формуле стоит интеграл от ступенчатой функции G_ε , которая на I_{n-1} равна $G(t - ne)$. Поскольку $G_\varepsilon(y) \geq G(t - \varepsilon - y)$, то

$$P\{X + Y \leq t\} \geq \int_{-\infty}^{\infty} G(t - \varepsilon - x) F\{dx\}. \quad (4.6)$$

Аналогичное рассуждение приводит к противоположному неравенству с ε , замененным на $-\varepsilon$. Устремляя ε к нулю, получаем (4.4). ►

Пример. в) Пусть распределения F и G сосредоточены на множестве целых неотрицательных чисел $0, 1, 2, \dots$, причем соответствующие точке k массы равны p_k и q_k . В этом случае интеграл в формуле (4.4) сводится к сумме $\sum G(t-k) p_k$. Функция, представляемая этой суммой, обращается в нуль при $t < 0$ и постоянна на каждом интервале $n-1 < t < n$. Скачок ее при

¹⁾ Формула (4.4) представляет собой частный случай теоремы Фубини [гл. IV, (2.11)]. Утверждение, обратное теореме 2, неверно: мы видели в гл. II, 4, д), и в задаче 1 из гл. III, 9, что в исключительных случаях формула (4.4) может иметь место и для пары зависимых случайных величин X и Y .

$t = n$ равен

$$\sum_{k=0}^n q_{n-k} p_k = q_n p_0 + q_{n-1} p_1 + \dots + q_0 p_n, \quad (4.7)$$

что согласуется с формулой свертки [1, гл. XI, (2.1)] для неотрицательных целочисленных случайных величин. ►

Обе предыдущие теоремы показывают, что свертка двух функций распределения $F \star G$ снова приводит к функции распределения U . Из коммутативности сложения $X + Y$ вытекает, что $F \star G = G \star F$. При совершенной системе обозначений можно было бы ввести новый символ для свертки функций распределения, однако это вряд ли было бы полезным¹⁾. Конечно, U следует рассматривать как функцию интервалов или меру: для любого интервала $I = \overline{a, b}$ имеем, очевидно,

$$U\{I\} = \int_{-\infty}^{\infty} G\{I - y\} F\{dy\}, \quad (4.8)$$

где, как обычно, $I - y$ обозначает интервал $\overline{a - y, b - y}$. (Эта формула автоматически переносится на произвольные множества.) В силу коммутативности F и G в формуле (4.8) можно поменять местами.

Рассмотрим теперь три распределения F_1, F_2, F_3 . Из ассоциативности сложения случайных величин следует, что $(F_1 \star F_2) \star F_3 = F_1 \star (F_2 \star F_3)$. Следовательно, при обозначении свертки трех распределений можно не пользоваться скобками и писать просто $F_1 \star F_2 \star F_3$. Резюмируем сказанное выше в следующих теоремах.

Теорема 3. *В классе распределений операция свертки коммутативна и ассоциативна.*

Теорема 4. *Если распределение G непрерывно (т. е. не имеет атомов), то свертка $U = F \star G$ тоже непрерывна. Если G имеет*

¹⁾ Иными словами, символ $A \star B$ используется, когда интегрирование происходит по мере A . Свертка $A \star B$ является функцией точки или мерой в соответствии с тем, является ли B функцией точки [как в (4.1)] или мерой [как в (4.8)]. Звездочка $\star$ используется для операции между двумя функциями, когда интегрирование происходит по лебеговой мере. Мы применяем этот тип свертки почти исключительно к вероятностным плотностям.

Можно дать более общее определение свертки двух функций, положив

$$f \star g(x) = \int_{-\infty}^{\infty} f(x-y) g(y) m(dy),$$

где m — произвольная мера. Суммы вида (4.7) представляют собой частный случай этой формулы, а именно когда m сосредоточена на множестве неотрицательных целых чисел и каждому такому числу сопоставлена единичная масса. В этом смысле наше теперешнее использование звездочки согласуется с ее использованием для сверток последовательностей в 1, гл. XI, 2.

обычную плотность φ , то U имеет обычную плотность u , задаваемую формулой (4.1).

Доказательство. Первое утверждение содержится в теореме 1. Если φ есть плотность распределения G , то, интегрируя (4.1) по интервалу I , справа получим правую часть (4.8), так что u действительно является плотностью распределения U , определяемого формулой (4.8). ►

Из теоремы, в частности, следует, что если F и G имеют плотности f и g , то свертка $F \star G$ имеет плотность $h = f \star g$:

$$h(x) = \int_{-\infty}^{+\infty} f(x-y) g(y) dy. \quad (4.9)$$

Вообще говоря, функция h обладает большей гладкостью, чем любая из функций f и g . (См. задачу 14.)

Суммы $S_n = X_1 + \dots + X_n$ взаимно независимых случайных величин с общей функцией распределения F встречаются очень часто, и поэтому для их распределения удобно ввести специальное обозначение. Распределение суммы S_n есть n -кратная свертка распределения F с собой и обозначается F^{n*} . Имеем, таким образом,

$$F^1 = F, \quad F^{(n+1)*} = F^{n*} \star F. \quad (4.10)$$

Сумму без слагаемых условно считают равной нулю. В соответствии с этим мы определим F^{0*} как атомическое распределение, сосредоточенное в нуле. Тогда (4.10) будет выполняться и при $n=0$.

Если F имеет плотность f , то F^{n*} имеет плотность $f \star f \star \dots \star f$ (n раз), которую мы обозначим f^{n*} . Эти обозначения согласуются с обозначениями, введенными в гл. 1, 2.

Замечание. Следующие примеры показывают, что свертка двух сингулярных распределений может иметь непрерывную плотность. Из этих примеров видно также, что можно эффективно вычислять свертки, не используя определяющую их формулу.

Примеры. г) *Равномерное распределение на $[0, 1]$ является сверткой двух сингулярных распределений канторовского типа.* В самом деле, пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины, принимающие значения 0 и 1 с вероятностью $1/2$. В примере 11, в) гл. 1 мы видели, что случайная величина $X = \sum 2^{-k} X_k$ имеет равномерное распределение. Обозначим сумму членов ряда с четными и нечетными номерами через U и V соответственно. Очевидно, U и V независимы и $X = U + V$. Следовательно, равномерное распределение есть свертка распределений величин U и V . Ясно далее, что величинами U и $2V$ имеют одинаковые распределения, а величина V лишь обозначением отличается от величины $1/3 U$ примера 11, г) гл. 1. Иными словами, распределения величин U и V лишь масштабными множителями отличаются от рассмотренного в этом примере канторовского распределения.

д) *Случайные величины в $\mathbb{R}^2$.* Распределение единичного вектора со случайным направлением (см. гл. 1, 10) сосредоточено на единичной окружности и, следовательно, сингулярно относительно лебеговой меры на плоскости.

Тем не менее длина L суммы двух таких независимых векторов имеет распределение с плотностью $\frac{2}{\pi} \frac{1}{\sqrt{4-r^2}}$, сосредоточенное на $\overline{0, 2}$. В самом деле, по закону косинусов $L = \sqrt{2-2\cos\omega} = \left| 2\sin\frac{1}{2}\omega \right|$, где ω — угол между двумя векторами. Так как $\frac{1}{2}\omega$ имеет равномерное распределение на $\overline{0, \pi}$, то

$$P\{L \leq r\} = P\left\{\left| 2\sin\frac{1}{2}\omega \right| \leq r\right\} = \frac{2}{\pi} \arcsin\frac{1}{2}r, \quad 0 < r < 2, \quad (4.11)$$

что и требовалось доказать. (См. задачу 12.) ►

4а. О точках роста

Здесь необходимо прервать изложение, чтобы зафиксировать некоторые элементарные факты относительно точек роста свертки $F \star G$. Первая лемма интуитивно понятна, вторая имеет технический характер. Этот материал будет использован лишь в теории восстановления, и потому косвенно в теории случайных блужданий.

Лемма 1. Если a и b — точки роста распределений F и G , то $a+b$ есть точка роста свертки $F \star G$. Если a и b — атомы F и G , то $a+b$ есть атом $F \star G$ и каждый атом $F \star G$ может быть так получен.

Доказательство. Для независимых X и Y имеем

$$P\{|X+Y-a-b| < \varepsilon\} \geq P\left\{|X-a| < \frac{1}{2}\varepsilon\right\} \cdot P\left\{|Y-b| < \frac{1}{2}\varepsilon\right\}.$$

Если a и b суть точки роста, то правая часть этого неравенства положительна при любом $\varepsilon > 0$, и поэтому $a+b$ тоже является точкой роста $F \star G$.

Обозначим F_a и G_a атомические компоненты F и G в жордановом разложении (2.5). Очевидно, атомическая компонента композиции $F \star G$ равна $F_a \star G_a$, и поэтому все атомы $F \star G$ имеют вид $a+b$, где a и b суть атомы F и G соответственно. ►

Внутренняя простота следующей леммы немало теряет из-за той специальной роли, которую играют в ней, с одной стороны, арифметические распределения, а с другой — распределения положительных случайных величин.

Лемма 2. Пусть F — распределение в $\mathcal{R}^1$ и Σ — множество, образованное точками роста распределений $F, F^2 \star, F^3 \star, \dots$.

а) Предположим, что F не сосредоточено на полуоси. Тогда если F не является арифметическим распределением, то Σ плотно в $-\infty, \infty$, а если F — арифметическое распределение с шагом λ , то $\Sigma = \{0, \pm\lambda, \pm 2\lambda, \dots\}$.

б) Пусть F сосредоточено на $\overline{0, \infty}$, но не сосредоточено в нуле. Если F не является арифметическим распределением, то Σ «асимптотически плотно на ∞ » в том смысле, что для любого заданного $\varepsilon > 0$ при всех достаточно больших x интервал $x, x+\varepsilon$ содержит точки из Σ . Если F — арифметическое распределение с шагом λ , то Σ содержит все точки вида $n\lambda$ при достаточно больших n .

Доказательство. Пусть $a, b, 0 < a < b$, — две точки множества Σ , и пусть $h = b - a$. Будем различать два случая:

(i) При каждом $\varepsilon > 0$ можно выбрать a, b такие, что $h < \varepsilon$.

(ii) Существует $\delta > 0$ такое, что $h \geq \delta$ при всех возможных выборах точек a, b .

Обозначим через I_n интервал $na < x \leq nb$. Если $n(b-a) > a$, то этот интервал содержит $na, (n+1)a$ в качестве подынтервала и, следовательно, любая точка $x > x_0 = a^2/(b-a)$ принадлежит по крайней мере одному из ин-

тервалов $I_1, I_2, \dots$. По лемме 1 $(n+1)$ точек $na+kh, k=0, \dots, n$, принадлежат Σ , и эти точки разбивают I_n на n подынтервалов длины h . Таким образом, всякая точка $x > x_0$ находится на расстоянии, не превосходящем $h/2$, от точки множества Σ .

В случае (i) отсюда вытекает, что множество Σ асимптотически плотно на $+\infty$. Если при этом F сосредоточено на $0, \infty$, то ничего не надо доказывать. В ином случае пусть $-c < 0$ — точка роста F . Для произвольного y и достаточно большого n интервал $nc+y < x < nc+y+\varepsilon$ содержит точку s множества Σ . Поскольку $s=nc$ снова принадлежит Σ , то отсюда вытекает, что любой интервал длины ε содержит какие-либо точки множества Σ и, таким образом, Σ всюду плотно.

В случае (ii) можно допустить, что a и b были выбраны так, что $h < 2\delta$. Тогда точками вида $na+kh$ исчерпываются все точки множества Σ , лежащие внутри I_n . Так как точка $(n+1)a$ находится среди этих точек, это означает, что все точки Σ внутри I_n являются кратными h . Пусть теперь c — произвольная (положительная или отрицательная) точка роста F . При достаточно большом n интервал I_n содержит точку вида $kh+c$, а так как она принадлежит Σ , то отсюда вытекает, что c кратно h . Таким образом, в случае (ii) распределение F является арифметическим. ►

Один частный случай этой теоремы представляет особый интерес. Каждое число $x > 0$ можно представить единственным образом в виде $x=m+\xi$, где m — целое и $0 \leq \xi < 1$. Число ξ в этом представлении называется дробной долей x . Рассмотрим теперь распределение F , сосредоточенное в двух точках -1 и $\alpha > 0$. Соответствующее множество Σ , включая в себя все точки вида $na-m^1$, содержит дробные доли всех чисел $\alpha, 2\alpha, \dots$. Если $\alpha=p/q$, где p и q — взаимно простые натуральные числа, то F — арифметическое распределение с шагом $1/q$. Мы приходим, таким образом, к следующему результату (в гл. VIII, 6, будет доказана теорема 3 о равномерном распределении, уточняющая этот результат).

Следствие. Для любого иррационального $\alpha > 0$ дробные доли чисел $\alpha, 2\alpha, 3\alpha, \dots$ образуют множество, плотное в $0, 1$.

§ 5. СИММЕТРИЗАЦИЯ

Пусть X — случайная величина с распределением F . Распределение случайной величины $-X$ мы будем обозначать ${}^{-}F$. В точках непрерывности имеем

$${}^{-}F(x) = 1 - F(-x), \quad (5.1)$$

и это соотношение однозначно определяет ${}^{-}F$. Распределение F называется *симметричным*, если ${}^{-}F=F$. [В случае когда существует плотность f , симметричность означает, что $f(-x)=f(x)$.]

Пусть X_1, X_2 — независимые случайные величины с одним и тем же распределением F . Случайная величина X_1-X_2 имеет симметричное распределение ${}^0F$, равное

$${}^0F = F \star {}^{-}F. \quad (5.2)$$

Из свойства симметрии ${}^0F(x) = 1 - {}^0F(-x)$ непосредственно сле-

¹⁾ Множество Σ здесь в точности совпадает с множеством точек $na-m$, $m, n=0, 1, \dots$. — Прим. перев.

дует, что

$${}^0F(x) = \int_{-\infty}^{\infty} F(x+y) F(dy). \quad (5.3)$$

О распределении ${}^0F$ говорят, что оно получено в результате симметризации F .

Примеры. а) Симметризация показательного распределения приводит к двустороннему показательному распределению [гл. II, 4, а)]; результатом симметризации равномерного на $\overline{0, 1}$ распределения является треугольное распределение τ_2 [см. гл. II, (4.1)].

б) Симметричное распределение, имеющее атомы массы $1/2$ в точках ± 1 , не является результатом симметризации никакого распределения.

в) Пусть F — атомическое распределение, наделяющее точки $0, 1, \dots$ массами $p_0, p_1, \dots$. Тогда симметризованное распределение ${}^0F$ тоже является атомическим, имеющим в точках $\pm n$ массу

$$q_n = \sum_{k=0}^{\infty} p_k p_{k+n}, \quad q_{-n} = q_n. \quad (5.4)$$

В частности, для пуассоновского распределения F при $n \geq 0$ имеем

$$q_n = e^{-2\alpha} \sum_{k=0}^{\infty} \frac{\alpha^{n+2k}}{k! (n+k)!} = e^{-2\alpha} I_n(2\alpha), \quad (5.5)$$

где I_n — функция Бесселя, определенная в гл. II, (7.1) (см. задачу 9). ▶

Применение симметризации позволяет значительно упростить многие рассуждения. В связи с этим весьма важным является то обстоятельство, что хвосты распределений F и ${}^0F$ имеют сравнимую величину. Более точно это выражено в нижеследующих неравенствах. Смысл этих неравенств представляется более ясным, когда они выражены в терминах случайных величин, а не в терминах самих функций распределения.

Лемма 1. Неравенства симметризации. Если X_1, X_2 — независимые и одинаково распределенные случайные величины, то при $t > 0$

$$\mathbf{P}\{|X_1 - X_2| > t\} \leq 2\mathbf{P}\{|X_1| > \frac{1}{2}t\}. \quad (5.6)$$

Если $a \geq 0$ таково, что $\mathbf{P}\{X_1 \leq a\} \geq p$ и $\mathbf{P}\{X_1 \geq -a\} \geq p$, то

$$\mathbf{P}\{|X_1 - X_2| > t\} \geq p\mathbf{P}\{|X_1| > t + a\}. \quad (5.7)$$

В частности, когда медиана X , равна нулю, имеем

$$P\{|X_1 - X_2| > t\} \geq \frac{1}{2} P\{|X_1| > t\}. \quad (5.8)$$

Доказательство. Неравенство (5.6) является следствием того очевидного обстоятельства, что событие $|X_1 - X_2| > t$ осуществляется лишь тогда, когда осуществляется хотя бы одно из событий $|X_1| > \frac{1}{2}t$, $|X_2| > \frac{1}{2}t$. Для доказательства (5.7) достаточно заметить, что каждое из непересекающихся событий $X_1 > t + a$, $X_2 \leq a$ и $X_1 < -t - a$, $X_2 \geq -a$ влечет событие $|X_1 - X_2| \geq t$. ►

Симметризация часто используется при оценке сумм независимых случайных величин. В этой связи особенно полезно следующее неравенство.

Лемма 2. Если $X_1, \dots, X_n$ — независимые случайные величины с симметричными распределениями, то сумма $S_n = X_1 + \dots + X_n$ тоже имеет симметричное распределение и

$$P\{|X_1 + \dots + X_n| > t\} \geq \frac{1}{2} P\{\max |X_j| > t\}. \quad (5.9)$$

Если X_j имеют одинаковое распределение F , то

$$P\{|X_1 + \dots + X_n| > t\} \geq \frac{1}{2} (1 - e^{-n[1 - F(t) + F(-t)]}). \quad (5.10)$$

Доказательство. Пусть M — случайная величина, равная первой из величин среди $X_1, \dots, X_n$, имеющих наибольшее значение по абсолютной величине. Положим $T = S_n - M$. Двумерная случайная величина (M, T) имеет симметричное распределение в том смысле, что все четыре случайные величины $(\pm M, \pm T)$ имеют одинаковое распределение. Имеем, очевидно,

$$P\{M > t\} \leq P\{M > t, T \geq 0\} + P\{M > t, T \leq 0\}. \quad (5.11)$$

Слагаемые справа в (5.11) равны по величине, и поэтому

$$P\{S_n > t\} = P\{M + T > t\} \geq P\{M > t, T \geq 0\} \geq \frac{1}{2} P\{M > t\}, \quad (5.12)$$

что равносильно (5.9). ►

Чтобы доказать (5.10), заметим, что во всех точках непрерывности

$$P\{\max |X_j| \leq t\} = (F(t) - F(-t))^n \leq e^{-n[1 - F(t) + F(-t)]}, \quad (5.13)$$

в силу того, что $1 - x < e^{-x}$ при $0 < x < 1$. Отсюда следует (5.10). ►

§ 6. ИНТЕГРИРОВАНИЕ ПО ЧАСТЯМ. СУЩЕСТВОВАНИЕ МОМЕНТОВ

Известную формулу интегрирования по частям можно применять также к интегралам по распределениям в $\mathcal{M}^1$. Если функция u ограничена, имеет непрерывную производную u' , то

$$\int_a^{b+} u(x) F\{dx\} = u(b) F(b) - u(a) F(a) - \int_a^b u'(x) F(x) dx. \quad (6.1)$$

Доказательство. Простым преобразованием (6.1) сводится к виду

$$\int_a^{b+} [u(b) - u(a)] F\{dx\} - \int_a^b u'(x) [F'(x) - F(a)] dx = 0. \quad (6.2)$$

Предположим, что $|u'| \leq M$, и рассмотрим разбиение $\overline{a, b}$ на конгруэнтные интервалы I_k длины h . Легко убедиться в том, что вклад I_k в левую часть (6.2) по абсолютной величине меньше чем $2MhF\{I_k\}$. Суммируя по всем k , находим, что левая часть в (6.2) по модулю меньше $2Mh$, что в свою очередь может быть сделано таким малым, как мы пожелаем. Таким образом, левая часть (6.2) в самом деле равна нулю.

В качестве приложения выведем одну часто используемую формулу [обобщающую 1; гл. XI, (1.8)].

Лемма 1. Для любого $\alpha > 0$

$$\int_0^\infty x^\alpha F\{dx\} = \alpha \int_0^\infty x^{\alpha-1} [1 - F(x)] dx, \quad (6.3)$$

причем из сходимости одного из этих интегралов следует сходимость другого.

Доказательство. Поскольку интегрирование в (6.3) происходит по бесконечному интервалу, мы не можем применять (6.1) непосредственно, однако после тривиальных преобразований при $b < \infty$ из (6.1) следует, что

$$\int_0^{b+} x^\alpha F\{dx\} = -b^\alpha [1 - F(b)] + \alpha \int_0^b x^{\alpha-1} [1 - F(x)] dx. \quad (6.4)$$

Предположим сначала, что интеграл слева сходится при $b \rightarrow \infty$. Вклад в предельный интеграл по бесконечному промежутку от интегрирования по $\overline{b, \infty}$ не меньше $b^\alpha [1 - F(b)]$, так что первое слагаемое в правой части (6.4) стремится к нулю. Переходя теперь в (6.4) к пределу при $b \rightarrow \infty$, получаем (6.3). С другой стороны, в (6.4) интеграл слева не превосходит интеграла справа, и, следовательно, из сходимости правого интеграла вытекает сходимость левого, что, как мы видели выше, приводит к (6.3). ►

Для левой полуоси имеет место соотношение, аналогичное (6.3). Комбинируя обе формулы, получаем следующий результат.

Лемма 2. *Распределение F обладает абсолютным моментом порядка $\alpha > 0$ тогда и только тогда, когда функция $|x|^{\alpha-1} \times \times [1-F(x) + F(-x)]$ интегрируема на $\overline{0, \infty}$.*

В качестве приложения покажем, что имеет место

Лемма 3. *Пусть X, Y — независимые случайные величины и $S = X + Y$. Математическое ожидание $E(|S|^\alpha)$ существует тогда и только тогда, когда существуют $E(|X|^\alpha)$ и $E(|Y|^\alpha)$.*

Доказательство. Поскольку случайные величины X и $X - S$ обладают абсолютными моментами одних и тех же порядков, то, не ограничивая общности, можно предположить, что медианы величин X и Y равны нулю. Но тогда $P\{|S| > t\} \geq \frac{1}{2} P\{|X| > t\}$, и поэтому в силу предыдущей леммы $E(|S|^\alpha) < \infty$ влечет за собой $E(|X|^\alpha) < \infty$. Тем самым доказана необходимость. Достаточность следует из неравенства $|S|^\alpha \leq 2^\alpha (|X|^\alpha + |Y|^\alpha)$, справедливость которого вытекает из того очевидного факта, что во всех точках $|S|$ не превосходит наибольшего из чисел $2|X|$ и $2|Y|$.

§ 7. НЕРАВЕНСТВО ЧЕБЫШЕВА

Неравенство Чебышева является одним из наиболее часто используемых инструментов теории вероятностей. Как само неравенство, так и его доказательство не отличаются от их дискретного варианта (1; гл. IX, 6) и повторяются здесь главным образом для удобства ссылок. Интересные приложения неравенства будут даны в гл. VII, 1.

Неравенство Чебышева. *Если существует $E(X^2)$, то*

$$P\{|X| \geq t\} \leq \frac{1}{t^2} E(X^2), \quad t > 0. \quad (7.1)$$

В частности, если $E(X) = m$ и $\text{Var}(X) = \sigma^2$, то

$$P\{|X - m| \geq t\} \leq \frac{1}{t^2} \sigma^2 \quad (7.2)$$

Доказательство. Обозначим распределение величины X через F . Тогда

$$E(X^2) \geq \int_{|x| \geq t} x^2 F\{dx\} \geq t^2 \int_{|x| \geq t} F\{dx\}.$$

Поделив эти неравенства на t^2 , получаем (7.1). ▶

Полезность неравенства Чебышева определяется не его точностью (она невелика), а простотой и тем, что оно особенно удобно в применении к суммам случайных величин. Среди многих возможных обобщений неравенства Чебышева ни

одно не обладает этими ценными его качествами. (Большинство из обобщений весьма просто и их лучше выводить по мере надобности. Например, полезная комбинация неравенства Чебышева и урезания распределений описана в гл. VII, 7.)

Приведем один довольно общий метод получения нетривиальных неравенств. Если $u \geq 0$ на всей прямой и $u(x) > a > 0$ при всех x из некоторого интервала I , то

$$F\{I\} \leq a^{-1} E(u(X)). \quad (7.3)$$

С другой стороны, если $u \leq 0$ вне I и $u \leq 1$ на I , то мы получаем обратное неравенство $F\{I\} \geq E(u(X))$. Беря в качестве u многочлены, приходим к неравенствам, зависящим лишь от моментов F .

Примеры. а) Пусть $u(x) = (x+c)^2$, где $c > 0$. Тогда $u(x) \geq 0$ при всех x и $u(x) \geq (t+c)^2$ при $x \geq t > 0$. Следовательно,

$$P\{X > t\} \leq \frac{1}{(t+c)^2} E((X+c)^2). \quad (7.4)$$

При $E(X) = 0$ и $E(X^2) = \sigma^2$ правая часть в (7.4) достигает минимума, когда $c = \sigma^2/t$, и, следовательно,

$$P\{X > t\} \leq \frac{\sigma^2}{\sigma^2 + t^2}, \quad t > 0. \quad (7.5)$$

Это важное неравенство было обнаружено независимо многими авторами.

б) Пусть X — положительная случайная величина (т. е. $F(0) = 0$) и $E(X) = 1$, $E(X^2) = b$. Многочлен $u(x) = h^{-2}(x-a)(a+2h-x)$ положителен лишь для $a < x < a+2h$ и $u(x) \leq 1$ при всех x . Если $0 < a < 1$, то, как легко видеть, $E(u(X)) \geq [2h(1-a) - b]h^{-2}$. Беря $h = b(1-a)^{-1}$ и используя замечание, предшествовавшее примерам, получаем

$$P\{X > a\} \geq (1-a)^2 b^{-1}. \quad (7.6)$$

в) Пусть $E(X^2) = 1$ и $E(X^4) = M$. Применяя последнее неравенство к X^2 , находим, что при $0 < t < 1$

$$P\{|X| > t\} \geq (1-t^2)^2 M^{-1}. \quad (7.7)$$

По поводу обобщения неравенства Чебышева, принадлежащего Колмогорову, (см. пример 8, д).

§ 8. ДАЛЬНЕЙШИЕ НЕРАВЕНСТВА. ВЫПУКЛЫЕ ФУНКЦИИ

Собранные в этом параграфе неравенства широко используются в математике и ни в коей мере не являются специфическими для теории вероятностей. Наиболее общее из них — неравенство Шварца. Другие неравенства приведены по причине их применений в теории случайных процессов и математической статистике. (Настоящий параграф предназначен более для ссылок, чем для чтения.)

а) Неравенство Шварца

В вероятностной интерпретации это неравенство означает, что для всяких двух случайных величин φ и ψ , определенных на одном и том же выборочном пространстве,

$$(E(\varphi\psi))^2 \leq E(\varphi^2) E(\psi^2), \quad (8.1)$$

если только входящие в неравенство математические ожидания существуют. Равенство в (8.1) осуществляется только тогда, когда некоторая линейная комбинация $a\varphi + b\psi$ величин φ и ψ равна нулю с вероятностью единица. Более общим образом, если F — произвольная мера на множестве A ,

$$\left(\int_A \varphi(x) \psi(x) F\{dx\} \right)^2 \leq \int_A \varphi^2(x) F\{dx\} \int_A \psi^2(x) F\{dx\}, \quad (8.2)$$

каковы бы ни были функции φ и ψ , для которых интегралы справа существуют. Беря в качестве F атомическую меру, наделяющую единичными массами целочисленные точки, получаем *неравенство Шварца для сумм*

$$(\sum \varphi_i \psi_i)^2 \leq \sum \varphi_i^2 \sum \psi_i^2. \quad (8.3)$$

Принимая во внимание важность неравенства (8.1), мы дадим два его доказательства, намечающие различные обобщения. Аналогичные доказательства можно дать для (8.2) и (8.3).

Первое доказательство. Можно предположить, что $E(\psi^2) > 0$. Тогда математическое ожидание

$$E(\varphi + t\psi)^2 = E(\varphi^2) + 2tE(\varphi\psi) + t^2E(\psi^2) \quad (8.4)$$

является квадратным многочленом по t . Этот многочлен, будучи неотрицательным, имеет либо два комплексных корня, либо двойной действительный корень λ . Из формулы для решения квадратного уравнения следует, что в первом случае (8.1) выполняется как строгое неравенство. Во втором случае (8.1) является равенством и $E(\varphi + \lambda\psi)^2 = 0$, так что $\varphi + \lambda\psi = 0$, за исключением множества вероятности нуль.

Второе доказательство. Поскольку при замене функций φ и ψ в (8.1) на их кратные $a\varphi$ и $b\psi$, где a, b — некоторые числа, получается равносильное неравенство, то достаточно доказать (8.1) в случае, когда $E(\varphi^2) = E(\psi^2) = 1$. Для получения (8.1) в этом случае достаточно взять математические ожидания от обеих частей неравенства $2|\varphi\psi| \leq \varphi^2 + \psi^2$. ►

б) Выпуклые функции. Неравенство Иенсена

Пусть u — функция, определенная на открытом интервале I , и $P = (\xi, u(\xi))$ — точка ее графика. Проходящая через P прямая L называется *опорной прямой функции u в точке ξ* , если график u целиком лежит над L или на L . (Это условие исключает вертикальные прямые.) На аналитическом языке это означает, что при всех x из I

$$u(x) \geq u(\xi) + \lambda(x - \xi), \quad (8.5)$$

где λ — тангенс угла наклона L . Функция u называется *выпуклой*

в интервале I , если в каждой точке x из I существует ее опорная прямая. (Функция u называется *вогнутой*, если выпукла функция $-u$.)

Покажем сначала, что из этого определения вытекают различные свойства выпуклых функций, которые интуитивно ассоциируются нами с выпуклостью (исходя из примера выпуклых ломаных линий).

Пусть X — случайная величина с распределением F , сосредоточенным на I , и с математическим ожиданием $E(X)$. Положив $\xi = E(X)$ и взяв математическое ожидание от обеих частей неравенства (8.5), получим

$$E(u) \geq u(E(X)), \quad (8.6)$$

если только математическое ожидание слева существует. Неравенство (8.6) известно как *неравенство Иенсена*.

Наиболее важным является тот случай, когда F сосредоточено в двух точках x_1 и x_2 и имеет в них массы $1-t$ и t . При этом (8.6) принимает вид

$$(1-t)u(x_1) + tu(x_2) \geq u((1-t)x_1 + tx_2). \quad (8.7)$$

Это неравенство допускает простую геометрическую интерпретацию, которую мы сформулируем в виде следующей теоремы.

Теорема 1. *Функция u выпукла тогда и только тогда, когда все хорды ее графика лежат не ниже самого графика.*

Доказательство. (i) *Необходимость.* Пусть u — выпуклая функция, и рассмотрим хорду под произвольным интервалом x_1, x_2 . Когда t меняется от 0 до 1 точка $(1-t)x_1 + tx_2$ пробегает интервал x_1, x_2 , а левая часть (8.7) представляет собой ординату хорды в точке $(1-t)x_1 + tx_2$. Таким образом, (8.7) означает, что точки хорды лежат не ниже графика u .

(ii) *Достаточность.* Пусть u обладает указанным в формулировке теоремы свойством, и рассмотрим треугольник, образованный тремя точками P_1, P_2, P_3 на графике u с абсциссами $x_1 < x_2 < x_3$. Точка P_2 лежит ниже хорды P_1P_3 , и из трех сторон треугольника хорда P_1P_2 имеет наименьший наклон, а хорда P_2P_3 — наибольший. Следовательно, вне интервала x_2, x_3 график u лежит над прямой P_2P_3 . Будем считать теперь, что x_3 — переменная точка и пусть $x_3 \rightarrow x_2 + 0$. Наклон хорды P_2P_3 при этом монотонно убывает, оставаясь ограниченным снизу наклоном хорды P_1P_2 . Таким образом, прямые P_2P_3 стремятся к предельной прямой L , проходящей через P_2 . Поскольку вне x_2, x_3 график u лежит над прямой P_2P_3 , то весь график u лежит не ниже L . Следовательно, L является опорной прямой для u в точке x_2 , и, поскольку точка x_2 произвольна, выпуклость u тем самым доказана. ►

Как предел хорд, прямая L является правой касательной. В рассмотренном предельном процессе абсцисса x_2 точки P_2 стремится к x_1 , а сама точка P_2 — к некоторой точке на прямой L . Следовательно, $P_2 \rightarrow P_1$. Аналогичное рассуждение применимо при приближении к x_2 слева, и поэтому график u непрерывен и обладает правой и левой касательной в каждой точке. Далее эти касательные являются опорными прямыми, и их наклоны меняются монотонно. Поскольку монотонная функция имеет не более счетного множества точек разрыва, то нами доказана

Теорема 2. *Выпуклая функция имеет правую и левую производные в каждой точке. Эти производные являются монотонно не убывающими функциями, совпадающими друг с другом всюду, за исключением, быть может, счетного множества точек.*

Свойства выпуклых функций, указанные в теореме 2, являются также и достаточными для выпуклости. В частности, если u обладает второй производной, то она выпукла тогда и только тогда, когда $u'' \geq 0$.

Обычно в качестве определения выпуклости функции берется (8.7). При $t = 1/2$ из (8.7) получаем

$$u\left(\frac{x_1+x_2}{2}\right) \leq \frac{1}{2} u(x_1) + \frac{1}{2} u(x_2). \quad (8.8)$$

Геометрически (8.8) означает, что *середина* хорды лежит не ниже соответствующей точки графика u . Если функция u непрерывна, то отсюда вытекает, что график не может пересечь хорду и, следовательно, u выпукла. Можно показать, что вообще любая *бэровская функция*¹⁾, обладающая свойством (8.8), выпукла.

в) Неравенства для моментов

Покажем, что для любой случайной величины X функция

$$u(t) = \log E(|X|^t), \quad t \geq 0, \quad (8.9)$$

выпукла по t в каждом интервале, где она конечна. В самом деле, согласно неравенству Шварца (8.1),

$$E^2(|X|^t) \leq E(|X|^{t+h}) E(|X|^{t-h}), \quad 0 \leq h \leq t, \quad (8.10)$$

если только математические ожидания существуют. Положив в (8.10) $x_1 = t-h$, $x_2 = t+h$ и перейдя к логарифмам, видим, что функция u удовлетворяет условию (8.8) и поэтому выпукла.

Поскольку $u(0) \leq 0$, то тангенс угла наклона $t^{-1}u(t)$ прямой, соединяющей начало координат с точкой $(t, u(t))$, меняется монотонно и, следовательно, $(E(|X|^t))^{1/t}$ является неубывающей функцией от $t > 0$.

¹⁾ Каждая функция u , удовлетворяющая условию (8.8), либо выпукла, либо она колеблется на каждом интервале от $-\infty$ до ∞ . См.: Харди Г. Г., Литтлвуд Дж. Е. и Полиа Г., Неравенства. — ИЛ, М., 1948, особенно с. 114.

г) Неравенство Гёльдера

Пусть $p > 1$, $q > 1$ и $p^{-1} + q^{-1} = 1$. Тогда для неотрицательных функций φ и ψ имеет место неравенство

$$E(\varphi\psi) \leq (E(\varphi^p))^{1/p} (E(\psi^q))^{1/q}, \quad (8.11)$$

если только математические ожидания существуют.

(Неравенство (8.1) является частным случаем этого неравенства, получающимся из него при $p=q=2$. Для неравенств (8.2) и (8.3) имеются обобщения, аналогичные (8.11).)

Доказательство. При $x > 0$ функция $u = \log x$ вогнута, т. е. она удовлетворяет неравенству, обратному (8.7). Потенцируя это неравенство, приходим к соотношению

$$x_1^{1-t} x_2^t \leq (1-t)x_1 + tx_2, \quad (8.12)$$

справедливому при $x_1, x_2 > 0$. Как при доказательстве неравенства Шварца (второе доказательство), чтобы установить (8.11), достаточно ограничиться случаем, когда $E(\varphi^p) = E(\psi^q) = 1$. Положим $t = q^{-1}$ и $1-t = p^{-1}$. Подставив в (8.12) $x_1 = \varphi^p$, $x_2 = \psi^q$ и взяв математические ожидания, получаем $E(\varphi\psi) \leq 1$, что и требовалось доказать. ►

д) Неравенство Колмогорова

Пусть $X_1, \dots, X_n$ — независимые случайные величины с конечными дисперсиями и с $E(X_k) = 0$. Тогда для любого $x > 0$

$$P\{\max[|S_1|, \dots, |S_n|] > x\} \leq x^{-2} E(S_n^2). \quad (8.13)$$

Это важнейшее усиление неравенства Чебышева было выведено нами в 1; IX, 7 для дискретных случайных величин. Доказательство сохраняется без изменений, но мы перефразируем его так, чтобы сделалось очевидным, что неравенство Колмогорова применимо и в более общей обстановке к субмартингалам. Мы вернемся к этому вопросу в гл. VII, 9.

Доказательство. Положим $x^2 = t$. При фиксированном значении t и $j = 1, 2, \dots, n$ обозначим через A_j событие, заключающееся в том, что $S_j^2 > t$, но $S_v^2 \leq t$ для всех индексов $v < j$. Словами можно сказать, что событие A_j состоит в том, что j есть наименьший среди индексов k , для которых $S_k^2 > t$. Конечно, такой индекс j не обязательно существует. Объединение событий A_j — это в точности событие, заключенное в левой части неравенства Колмогорова. Так как события A_j взаимно исключают друг друга, то неравенство может быть переписано в форме

$$\sum_{j=1}^n P\{A_j\} \leq t^{-1} E(S_n^2). \quad (8.14)$$

Обозначим через 1_{A_j} индикатор события A_j , иначе, 1_{A_j} есть случайная величина, которая равна 1 на A_j и равна 0 на дополнении к A_j . Тогда $\sum 1_{A_j} \leq 1$, и поэтому

$$E(S_n^2) \geq \sum_{j=1}^n E(S_n^2 1_{A_j}). \quad (8.15)$$

Покажем, что

$$E(S_n^2 1_{A_j}) \geq E(S_j^2 1_{A_j}). \quad (8.16)$$

Поскольку $S_j^2 > t$ всякий раз, когда происходит A_j , то правая часть (8.16) $\geq tP(A_j)$. Таким образом (8.15) приводит к неравенству (8.14).

Для доказательства (8.16) заметим, что $S_n = S_j + (S_n - S_j)$, и, следовательно,

$$E(S_n^2 1_{A_j}) \geq E(S_j^2 1_{A_j}) + 2E((S_n - S_j) S_j 1_{A_j}). \quad (8.17)$$

Второе слагаемое в правой части (8.17) обращается в нуль, так как величины $S_n - S_j = X_{j+1} + \dots + X_n$ и $S_j 1_{A_j}$ независимы, и поэтому к их математическим ожиданиям применимо правило умножения. Таким образом, (8.17) доказывает утверждение (8.16). ►

§ 9. ПРОСТЫЕ УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ. СМЕСИ

В гл. III, 2, мы определили «условную плотность случайной величины Y при фиксированном значении другой случайной величины X » в том случае, когда совместное распределение X и Y имеет непрерывную плотность. Не претендуя на общность, мы введем теперь аналогичное понятие для более широкого класса распределений. (Систематически теория условных распределений и математических ожиданий развивается в § 10 и 11.)

Для произвольной пары интервалов A, B на прямой положим

$$Q(A, B) = P\{X \in A, Y \in B\}. \quad (9.1)$$

При этом обозначении для *частного распределения* X имеем

$$\mu\{A\} = Q(A, \mathcal{R}^1). \quad (9.2)$$

Если $\mu\{A\} > 0$, то условная вероятность события $\{Y \in B\}$ при условии $\{X \in A\}$ равна

$$P\{Y \in B | X \in A\} = \frac{Q(A, B)}{\mu\{A\}}. \quad (9.3)$$

(Если $\mu\{A\} = 0$, то эта условная вероятность не определена.) Воспользуемся формулой (9.3), когда A есть интервал $A_h =]x, x+h[$. Пусть $h \rightarrow 0+$. При выполнении соответствующих

условий регулярности предел

$$q(x, B) = \lim_{h \rightarrow 0} \frac{Q(A_h, B)}{\mu\{A_h\}} \quad (9.4)$$

существует для всех x и B . В этом случае по аналогии с гл. III, 2, мы обозначаем

$$q(x, B) = P\{Y \in B | X = x\} \quad (9.5)$$

и называем q «условной вероятностью события $\{Y \in B\}$ при условии $X = x$ ». Рассмотренная конструкция распространяет понятие условных вероятностей на случаи, когда «гипотезы» имеют нулевую вероятность. Никаких дальнейших трудностей не возникает, если функция q достаточно регулярна, однако мы не будем заниматься изучением соответствующих условий регулярности, поскольку в следующем параграфе рассматривается более общая схема. В частных случаях обычно оказывается достаточным простой наивный подход, и вид условного распределения нередко можно найти с помощью интуиции.

Примеры. а) Предположим, что пара случайных величин X, Y имеет совместную плотность $f(x, y)$. Для упрощения предположим, что f непрерывна и строго положительна. Тогда

$$q(x, B) = \frac{1}{f_1(x)} \int_B f(x, y) dy,$$

где $f_1(x) = q(x, -\infty, \infty)$ есть частная плотность X . Другими словами, при фиксированном значении x функция множества q имеет плотность, равную $f(x, y)/f_1(x)$.

б) Пусть X и Y — независимые случайные величины с распределениями F и G соответственно. Предположим для простоты, что $X > 0$ [т. е. $F(0) = 0$]. Рассмотрим произведение $Z = XY$. Имеем, очевидно,

$$P\{Z \leq t | X = x\} = G\left(\frac{t}{x}\right). \quad (9.6)$$

Интегрируя (9.6) по F , получим функцию распределения U величины Z [см. гл. II, (3.1)]. Это утверждение представляет собой частный случай формулы (9.8), см. ниже]. В частности, когда величина X распределена равномерно на интервале $\overline{0, 1}$, получаем

$$U(t) = \int_0^1 G\left(\frac{t}{x}\right) dx. \quad (9.7)$$

Формула (9.7) может быть использована как удобный исходный пункт в теории одновершинных распределений¹⁾. ►

¹⁾ Функция распределения U называется *одновершинной* (унимодальной) с модой в нуле, если на интервале $-\infty, 0$ она выпукла, а на интервале $\overline{0, \infty}$

Дальнейшие примеры содержатся в задачах 18, 19. ►

Следующая теорема, принадлежащая Л. Шеппу, является допускающим простую вероятностную интерпретацию. вариантом формального критерия, открытого А. Хинчиным.

Теорема. *Функция распределения U одновершинна тогда и только тогда, когда ее можно представить в виде (9.7). Иными словами, U одновершинна тогда и только тогда, когда она является функцией распределения произведения $Z = XY$ двух независимых случайных величин, из которых одна величина, скажем X , распределена равномерно на $[0, 1]$.*

Доказательство. Возьмем $h > 0$ и обозначим U_h функцию распределения, график которой является кусочно-линейной функцией, совпадающей с U в точках $0, \pm h, \dots$ [т. е. $U_h(nh) = U(nh)$ и функция U_h линейна в интервалах между nh и $(n+1)h$]. Из определения одновершинности непосредственно вытекает, что функция U одновершинна тогда и только тогда, когда одновершинны функции U_h . Далее, U_h имеет плотность u_h , являющуюся ступенчатой функцией, и каждую ступенчатую функцию с разрывами в точках nh можно записать в виде

$$\sum p_n \frac{1}{|n|h} f\left(\frac{x}{nh}\right), \quad (*)$$

где $f(x) = 1$ при $0 < x < 1$ и $f(x) = 0$ в остальных точках. Функция $(*)$ монотонна в интервалах $-\infty, 0$ и $0, \infty$ тогда и только тогда, когда $p_n \geq 0$ при всех n , и является плотностью, если $\sum p_n = 1$. Но когда $p_n \geq 0$ и $\sum p_n = 1$, $(*)$ представляет собой плотность распределения произведения $Z_h = XY_h$ — двух независимых случайных величин X и Y_h , из которых X распределена равномерно на $[0, 1]$, а Y_h имеет арифметическое распределение $P\{Y_h = nh\} = p_n$. Мы доказали, таким образом, что функция U_h одновершинна тогда и только тогда, когда ее можно представить в виде (9.7) с функцией G , замененной на функцию G_h , сосредоточенную в точках $0, \pm h, \dots$. Устремляя теперь h к нулю и используя теорему о монотонной сходимости, получаем нужный результат. (См. задачи 25, 26 и задачу 10 из гл. XV, 9.) ►

При выполнении соответствующих условий регулярности функция $q(x, B)$ при фиксированном x является вероятностным распределением по B и при фиксированном B представляет собой непрерывную функцию по x . При этом

$$Q(A, B) = \int_A q(x, B) \mu\{dx\}. \quad (9.8)$$

В самом деле, стоящий справа интеграл задает вероятностное распределение на плоскости, и, дифференцируя его по типу (9.4), получаем $q(x, B)$. Формула (9.8) показывает, как заданное распределение на плоскости можно представить в терминах условного и частного распределений. Пользуясь терминологией гл. II, 5, можно сказать, что формула (9.8) дает представление заданного распределения в виде смеси зависящего от параметра x семейства

могнута [см. 8, 6)]. Нуль может быть точкой разрыва U , но вне нуля одновершинность предполагает существование у U плотности u , которая монотонна и в $-\infty, 0$, и в $0, \infty$, (Интервалы постоянства не исключаются).

распределений $q(x, B)$ и распределения μ , играющего роль распределения рандомизированного параметра.

На практике описанный процесс нередко обращается. В качестве исходного берут «вероятностное (стохастическое) ядро» q , т. е. функцию $q(x, B)$ точки x и множества B_x такую, что при фиксированном x она является вероятностным распределением, а при фиксированном B — бэровской функцией. При любом заданном вероятностном распределении μ интеграл в (9.8) определяет вероятности подмножеств плоскости вида (A, B) и тем самым определяет некоторое вероятностное распределение на плоскости. Обычно формулу (9.8) выражают в терминах функций точек. Рассмотрим семейство функций распределения $G(\theta, y)$, зависящих от параметра θ , и некоторое вероятностное распределение μ . Определим новую функцию распределения с помощью следующей формулы:

$$U(y) = \int_{-\infty}^{\infty} G(x, y) \mu\{dx\}. \quad (9.9)$$

[Эта формула представляет собой частный случай формулы (9.8); она получается из нее при $A = -\infty, \infty$ и $q(x, -\infty, y) = G(x, y)$.] Смеси подобного рода встречались нам уже в I, гл. V, и обсуждались в гл. II, 5. В следующем параграфе будет показано, что q всегда можно интерпретировать как условное распределение вероятностей.

Примеры. в) Для всяких двух распределений F_1 и F_2 распределение $pF_1 + (1-p)F_2$ ($0 < p < 1$) является их смесью и представляется формулой (9.9) с распределением μ , сосредоточенным в двух точках.

г) *Случайные суммы.* Пусть $X_1, X_2, \dots$ — независимые случайные величины с одинаковым распределением F . Пусть далее N — независимая от X_j случайная величина, принимающая значения $0, 1, \dots$ с положительными вероятностями $p_0, p_1, \dots$. Рассмотрим случайную величину $S_N = X_1 + \dots + X_N$. Условное распределение S_N при условии $N=n$ есть F^{n*} , и поэтому распределение S_N дается формулой

$$U = \sum_{n=0}^{\infty} p_n F^{n*}, \quad (9.10)$$

представляющей собой частный случай (9.9). В рассматриваемом случае каждое условие $N=n$ имеет положительную вероятность p_n , и поэтому мы имеем здесь дело с условными вероятностными распределениями в строгом смысле. Другие примеры имеются в гл. II, 5—7. (См. задачи 21 и 24.)

§ 10 *). УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ

Вряд ли было разумным исследовать точные условия, при которых условные вероятности q могут быть получены с помощью процесса дифференцирования типа (9.4). Основные свойства условных вероятностей выражены в соотношении (9.8), представляющем вероятности множеств в терминах условных вероятностей, и проще всего использовать (9.8) в качестве *определения* условных вероятностей. Соотношение (9.8) не определяет q однозначно, так как, если для каждого множества B с точностью до множества μ -меры нуль $q(x, B) = \bar{q}(x, B)$, то (9.8) по-прежнему остается справедливым при замене q на $\bar{q}$. Однако указанная неопределенность неизбежна. Например, когда μ сосредоточено на интервале I , невозможно никакое естественное определение q для x , не лежащих в I . В силу самой природы вещей мы имеем дело со всем классом эквивалентных условных вероятностей и должны говорить скорее о *каком-то*, а не об определенном условном распределении вероятностей q . В конкретных же случаях требования регулярности диктуют обычно некоторый определенный выбор условного распределения вероятностей.

Для определенности мы будем рассматривать лишь события, задаваемые условиями типа $X \in A$ и $Y \in B$, где X и Y — фиксированные случайные величины и A, B — борелевские множества прямой. Посмотрим сначала, как можно понимать фразу «условная вероятность события $\{Y \in B\}$ относительно X ». Значение X может быть либо фиксированным числом, либо быть неопределенным. При второй из этих интерпретаций мы имеем функцию от X , т. е. случайную величину, которую будем обозначать $P\{B|X\}$ или $q(X, B)$ и т. д. Когда имеется в виду фиксированное значение X , скажем x , будет использоваться обозначение $P\{Y \in B|X=x\}$ или $q(x, B)$.

Определение 1. Пусть B — фиксированное множество. «Условной вероятностью $P\{Y \in B|X\}$ события $\{Y \in B\}$ относительно X » называется функция $q(X, B)$, такая, что для любого подмножества A прямой

$$P\{X \in A, Y \in B\} = \int_A q(x, B) \mu\{dx\}, \quad (10.1)$$

где μ — частное распределение X .

Когда x является атомом распределения μ , условие $X=x$ имеет положительную вероятность и условная вероятность $P\{Y \in B|X=x\}$ была в этом случае определена нами раньше формулой (9.3), в которой под A следует понимать множество, состоящее из единственной точки x . При этом формула (10.1) сводится к (9.3), так что наши прежние и настоящие определения и обозначения согласуются друг с другом.

*) Этот параграф можно опустить при первом чтении.

Мы видим, что условная вероятность $P\{Y \in B | X\}$ всегда существует. В самом деле, ясно, что

$$P\{X \in A, Y \in B\} \leq \mu(A). \quad (10.2)$$

Левая часть (10.2), рассматриваемая при фиксированном B как функция множества A , определяет конечную меру, и из (10.2) следует, что эта мера абсолютно непрерывна по отношению к μ (см. теорему Радона—Никодима в § 3). Это означает, что определенная нами мера обладает плотностью q и поэтому (10.1) справедливо.

До сих пор множество B было у нас фиксированным, но обозначение $q(x, B)$ было выбрано с тем расчетом, чтобы впоследствии сделать B переменным. Иными словами, мы хотим теперь рассматривать q как функцию от двух переменных—точки x и подмножества B прямой. При этом желательно, чтобы при фиксированном x функция q была вероятностной мерой. Это означает, что $q(x, \mathcal{R}) = 1$ и что для любой последовательности непересекающихся множеств $B_1, B_2, \dots$, в объединении дающих B , имеет место равенство

$$q(x, B) = \sum q(x, B_n). \quad (10.3)$$

Если стоящие справа члены представляют собой условные вероятности B_n , то их сумма является некоторой условной вероятностью множества B , однако у нас имеется дополнительное требование согласованности, состоящее в том, чтобы (10.3) выполнялось при нашем выборе q при всех x . [Отметим, что определение 1 не исключает абсурдных значений функции $q(x, B)$ в некоторых точках x типа $q(x, B) = 17$.] Как нетрудно видеть, функцию $q(x, B)$ действительно можно выбрать так, чтобы выполнялись все эти требования¹⁾. Это означает, что существует условное вероят-

¹⁾ Проще всего построить такую функцию q , определяя непосредственно значения $q(x, B)$ лишь для множеств B , являющихся интервалами в диadicеском подразбиении $\mathcal{R}^1$. Пусть, например, $B_1 = [0, \infty)$, $B_2 = (-\infty, 0]$. Возьмем в качестве $q(x, B_1)$ любую условную вероятность B_1 , подчиненную очевидному требованию $0 \leq q(x, B_1) \leq 1$. Теперь автоматически следует положить $q(x, B_2) = 1 - q(x, B_1)$. Разобьем далее B_1 на B_{11} и B_{12} и выберем $q(x, B_{11})$ так, чтобы выполнялись неравенства $0 \leq q(x, B_{11}) \leq q(x, B_1)$. Положим $q(x, B_{12}) = q(x, B_1) - q(x, B_{11})$. Процесс измельчения подразбиений продолжается далее неограниченно. Требование аддитивности (10.3) позволяет затем определить $q(x, B)$ для всех открытых множеств B и, следовательно, для всех борелевских множеств.

Описанная конструкция основывается лишь на существовании так называемой *сети*, т. е. разбиении пространства на конечное число непересекающихся множеств, каждое из которых в свою очередь разбивается на конечное число непересекающихся множеств, причем процесс продолжается таким образом, что каждая точка пространства является единственным пределом стягивающейся последовательности множеств, входящих в последовательные разбиения. Рассматриваемое утверждение имеет место, следовательно, и в $\mathcal{R}^n$ и во многих других пространствах.

ностное распределение Y относительно X в смысле следующего определения.

Определение 2. Условным вероятностным распределением Y относительно X называется функция q , зависящая от двух переменных—точки x и множества B , такая, что

(i) при фиксированном множестве B

$$q(X, B) = P\{Y \in B | X\} \quad (10.4)$$

является условной вероятностью события $\{Y \in B\}$ относительно X ,

(ii) при каждом фиксированном x q есть вероятностное распределение.

В действительности условное вероятностное распределение представляет собой семейство обычных вероятностных распределений, и поэтому вся теория переносится на них без изменений. Так, если q задано¹⁾, следующее определение вводит скорее новое обозначение, чем новое понятие.

Определение 3. Условное математическое ожидание $E(Y|X)$ есть функция от X , принимающая в точке x значение

$$E(Y|x) = \int_{-\infty}^{\infty} yq(x, dy). \quad (10.5)$$

При этом предполагается, что интеграл сходится (за исключением, возможно, множества точек x , имеющих вероятность нуль).

Условное математическое ожидание $E(Y|X)$ является функцией от X и, следовательно, представляет собой случайную величину. Для большей ясности значение его в фиксированной точке x иногда предпочитают обозначать $E(Y|X=x)$. Из определения $E(Y|X)$ непосредственно вытекает, что

$$E(Y) = \int_{-\infty}^{\infty} E(Y|x) \mu\{dx\} \text{ или } E(Y) = E(E(Y|X)). \quad (10.6)$$

§ 11 *). УСЛОВНЫЕ МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

Мы определили условные математические ожидания $E(Y|X)$ в терминах условных вероятностных распределений. Этот подход вполне удовлетворителен, когда имеют дело лишь с одной парой случайных величин X, Y . Однако когда приходится иметь дело с целыми семействами случайных величин, то неединственность условных распределений приводит к серьезным трудностям. Поэтому замечательно то, что на практике можно обойтись без этой громоздкой теории. Действительно, оказывается, что можно построить на удивление простую и гибкую теорию условных математических ожиданий без всякого обращения к условным распреде-

¹⁾ Более гибкое общее определение приводится в § 11.

*) Содержание этого параграфа будет использоваться лишь в связи с маргингалами в гл. VI, 12, и гл. VII, 9.

лениям. Для понимания этой теории лучше всего начать с более внимательного рассмотрения равенства (10.5).

Пусть A — борелевское множество на прямой и $1_A(X)$ — случайная величина, равная 1 при $X \in A$ и нулю в остальных случаях. Мы интегрируем обе части равенства (10.5) относительно частного распределения μ случайной величины X , взяв в качестве области интегрирования множество A . Результат можно записать в следующем виде:

$$E(Y 1_A(X)) = \int_A E(Y | x) \mu\{dx\} = \int_{-\infty}^{+\infty} 1_A(x) E(Y | x) \mu(dx). \quad (11.1)$$

Случайная величина X отображает выборочное пространство $\mathcal{G}$ на действительную прямую, и последний интеграл имеет отношение только к функциям и мерам на прямой. Случайная величина $Y 1_A(X)$, однако, определена в исходном выборочном пространстве, и поэтому необходима более подходящая система обозначений. Очевидно, что $1_A(X)$ есть индикатор множества B всех таких точек из $\mathcal{G}$, в которых X принимает значения из множества A . Как мы знаем из гл. IV, 3, множества B , которые таким образом соответствуют произвольному борелевскому множеству A на прямой, образуют σ -алгебру множеств в $\mathcal{G}$, которая называется *алгеброй, порожденной X* . Таким образом, (11.1) утверждает, что $U = E(Y | X)$ есть функция от X , которая удовлетворяет тождеству

$$E(Y 1_B) = E(U 1_B) \quad (11.2)$$

для любого множества B из σ -алгебры, порожденной X . Мы увидим, что это равенство может быть использовано как определение условных математических ожиданий и поэтому очень важно уметь истолковывать его надлежащим образом. Поясним природу этого равенства на простом примере.

Примеры. а) Рассмотрим плоскость с координатными величинами X и Y как выборочное пространство и предположим для простоты, что вероятности задаются посредством строго положительной непрерывной плотности $f(x, y)$. Случайная величина X принимает постоянное значение вдоль любой прямой, параллельной y -оси. Если A есть множество на x -оси, то соответствующее плоское множество B состоит из всех таких прямых, проходящих через точки множества A . Левая часть (11.2) представляет собой обычный интеграл от $y f(x, y)$ по этому множеству, и он может быть записан как повторный интеграл. Таким образом

$$E(Y 1_B) = \int_A dx \int_{-\infty}^{+\infty} y f(x, y) dy. \quad (11.3)$$

Правая часть (11.2) представляет собой обычный интеграл от функции $U(x) f_1(x)$, где f_1 — частная плотность X . В таком случае

(11.2) утверждает, что

$$U(x) = \frac{1}{f_1(x)} \int_{-\infty}^{+\infty} y f(x, y) dy \quad (11.4)$$

в соответствии с определением (10.5) условного математического ожидания и в соответствии с интуицией.

б) *Продолжение.* Покажем теперь, что равенство (11.2) определяет условное математическое ожидание U даже в том случае, когда плотности не существуют и вероятностное распределение на плоскости произвольно. При заданном борелевском множестве A на x -оси левая часть в (11.2) определяет число $\mu_1\{A\}$. Очевидно, что μ_1 есть мера на борелевских множествах x -оси. Другая мера задается посредством частного распределения μ случайной величины X , которое определено как $\mu\{A\} = E(1_A)$. Поэтому очевидно, что если $\mu\{A\} = 0$, то и $\mu_1\{A\} = 0$. Иными словами, μ_1 абсолютно непрерывна относительно μ , и по теореме Радона—Никодима из § 3 существует функция U , такая, что

$$\mu_1\{A\} = \int_A U(x) \mu(dx). \quad (11.5)$$

Это равенство отличается только обозначениями от (11.2). Конечно, (11.5) остается справедливым, если U изменить на множестве μ -меры 0, но эта неединственность присуща понятию условного математического ожидания. ►

Приведенный пример демонстрирует, что (11.2) может быть использовано для определения условного математического ожидания $U(X) = E(Y|X)$ для произвольной пары случайных величин X, Y в произвольном вероятностном пространстве [при условии, конечно, что $E(Y)$ существует]. Однако этот подход ведет много дальше. Например, для определения условного математического ожидания $E(Y|X_1, X_2)$ относительно случайных величин X_1, X_2 можно использовать (11.2) без изменений, за тем исключением, что B должно теперь быть произвольным множеством из σ -алгебры $\mathfrak{B}$, порожденной X_1 и X_2 (см. гл. IV, 3). Конечно, U должна быть функцией X_1, X_2 , но мы видели в гл. IV, 4, что класс бэровских функций от пары (X_1, X_2) совпадает с классом всех $\mathfrak{B}$ -измеримых функций. Таким образом, мы можем охватить все мыслимые случаи с помощью следующего определения, предложенного впервые Дубом.

Определение. Пусть $(\mathfrak{G}, \mathfrak{A}, P)$ — вероятностное пространство, и пусть $\mathfrak{B}$ — σ -алгебра множеств из $\mathfrak{A}$ (иначе $\mathfrak{B} \subset \mathfrak{A}$). Пусть Y — случайная величина с математическим ожиданием EY .

Случайная величина U называется *условным математическим ожиданием Y относительно $\mathfrak{B}$* , если U $\mathfrak{B}$ -измерима и (11.2) выполняется для всех множеств $B \in \mathfrak{B}$. В этом случае мы пишем $U = E(Y|\mathfrak{B})$.

В том случае, когда $\mathfrak{B}$ есть σ -алгебра, порожденная случайными величинами $X_1, \dots, X_r$, величина U сводится к бэровской функции от $X_1, \dots, X_r$ и будет обозначаться $E(Y|X_1, \dots, X_r)$.

Существование $E(Y|\mathfrak{B})$ устанавливается способом, указанным в примере б), который использует теорему Радона—Никодима.

Укажем основные свойства условного математического ожидания $U = E(Y|\mathfrak{B})$. Отметим, что соотношение (11.2) выполняется, если 1_{B_j} заменить линейной комбинацией индикаторов множеств B_j , принадлежащих $\mathfrak{B}$. Но в гл. IV, 3, мы видели, что любая $\mathfrak{B}$ -измеримая функция может быть равномерно приближена с помощью таких линейных комбинаций. Переходя к пределу, убеждаемся, что из (11.2) следует, что для любой $\mathfrak{B}$ -измеримой функции Z выполняется $E(YZ) = E(UZ)$. Заменяя Z на $Z1_B$ и сравнивая с определением (11.2), мы видим, что для любой $\mathfrak{B}$ -измеримой функции Z

$$E(YZ|\mathfrak{B}) = ZE(Y|\mathfrak{B}). \quad (11.6)$$

Это соотношение чрезвычайно важно.

Наконец, рассмотрим σ -алгебру $\mathfrak{B}_0 \subset \mathfrak{B}$ и пусть $U_0 = E(Y|\mathfrak{B}_0)$. При $B \in \mathfrak{B}_0$ мы можем толковать (11.2) относительно $\mathfrak{B}_0$ так же, как и относительно $\mathfrak{B}$, и поэтому мы находим, что

$$E(Y1_B) = E(U1_B) = E(U_01_B).$$

Таким образом, в силу определения $U_0 = E(U|\mathfrak{B}_0)$, и поэтому если $\mathfrak{B}_0 \subset \mathfrak{B}$, то

$$E(Y|\mathfrak{B}_0) = E(E(Y|\mathfrak{B})|\mathfrak{B}_0). \quad (11.7)$$

К примеру, $\mathfrak{B}$ может быть алгеброй, порожденной двумя случайными величинами X_1, X_2 , а $\mathfrak{B}_0$ — алгеброй, порожденной только величиной X_1 . Тогда (11.7) переписывается как $E(Y|X_1) = E(E(Y|X_1, X_2)|X_1)$.

Наконец, мы отметим, что из (11.2) следует, что для функции, тождественно равной 1, условное математическое ожидание равно 1 независимо от того, как выбрана алгебра $\mathfrak{B}$. Таким образом, из (11.6) вытекает, что $E(Z|\mathfrak{B}) = Z$ для всех $\mathfrak{B}$ -измеримых случайных величин Z .

Едва ли нужно говорить, что основные свойства математических ожиданий распространяются на условные математические ожидания.

§ 12. ЗАДАЧИ

1. Пусть X и Y — независимые случайные величины с функциями распределения F и G . Найдите функции распределения следующих величин ¹⁾: а) $X \cup Y$, б) $X \cap Y$, в) $2X \cup Y$, г) $X^2 \cup Y$.

¹⁾ Для двух чисел a и b $a \cup b = \max(a, b)$ обозначает наибольшее из них, а $a \cap b = \min(a, b)$ — наименьшее. Для функций f и g обозначает функцию, которая в точке x принимает значение $f(x) \cup g(x)$ (см. гл. IV, 1). Таким образом, $X \cup Y$ и $X \cap Y$ являются случайными величинами.

2. *Смеси.* Пусть X, Y, Z — независимые случайные величины, X и Y имеют распределения F и G , а $P\{Z=1\}=p, P\{Z=0\}=q (p+q=1)$. Найдите функции распределения следующих величин: а) $ZX + (1-Z)Y$, б) $ZX + (1-Z)(X \cup Y)$, в) $ZX + (1-Z)(X \cap Y)$.

3. Показать, что для непрерывной функции распределения F

$$\int_{-\infty}^{\infty} F(x) F(dx) = \int_0^1 y dy = \frac{1}{2}.$$

а) используя само определение первого интеграла (разбиение $-\infty, \infty$ на подынтервалы) и б) интерпретируя первый интеграл, как $E(F(X))$, и пользуясь тем, что $F(X)$ имеет равномерное распределение. Показать также, что имеет место общая формула

$$\int_{-\infty}^{\infty} F^k(x) G(dx) = \frac{n}{n+k}, \quad \text{где } G(x) = F^n(x).$$

4. Пусть $F(x, y)$ обозначает вероятностное распределение на плоскости. Положим $U(x, y) = 0$ при $x < 0$ и $y < 0$ и $U(x, y) = F(x, y)$ во всех других точках. Покажите, что U — монотонная функция каждой переменной, но U не является функцией распределения.

Указание: рассмотрите смешанные разности.

5. *Двумерное распределение с заданными частными распределениями*¹⁾. Пусть F и G — функции распределения в $\mathcal{R}^1$ и

$$U(x, y) = F(x) G(y) [1 + \alpha (1 - F(x)) (1 - G(y))],$$

где $|\alpha| \leq 1$. Доказать, что U есть функция распределения в $\mathcal{R}^2$ с частными распределениями F и G и что U имеет плотность тогда и только тогда, когда F и G имеют плотности.

Указание. Смешанные разности функции $w(x, y) = u(x)v(y)$ [определенные в (1.12)] имеют вид $\Delta u \Delta v$. Отметим также, что $\Delta(F^2) \leq 2\Delta F$.

6. Внутри единичного квадрата положим $U(x, y) = x$ при $x \leq y$ и $U(x, y) = y$ при $x \geq y$. Покажите, что U есть функция и что соответствующее ей распределение сосредоточено на биссектрисе (и, следовательно, сингулярно).

7. *Максимальное распределение Фреше с заданными частными распределениями.* Пусть F и G — функции распределения в $\mathcal{R}^1$ и $U(x, y) = F(x) \cap G(y)$. Доказать, что а) U есть функция распределения с частными распределениями F и G , б) если V — любая другая функция распределения с частными распределениями F и G , то $V \leq U$, в) распределение, отвечающее U , сосредоточено на множестве $\{(x, y): F(x-) \cup G(y-) \leq F(x) \cap G(y)\}$ и, следовательно, сингулярно. (Задача 6 является частным случаем этой задачи.)

8. Пусть U — равномерное распределение на $[-h, 0]$ и T — треугольное распределение на $[-h, h]$ [см. гл. II, 4 (б)]. Показать, что плотности распределений $F \star U$ и $F \star T$ равны соответственно

$$h^{-1} [F(x+h) - F(x)] \quad \text{и} \quad h^{-2} \int_0^h [F(x+y) - F(x-y)] dy.$$

9. Если независимые случайные величины X и Y имеют пуассоновские распределения с математическими ожиданиями pt и $qt, p+q=1$, то

$$P\{X - Y = k\} = e^{-t} \sqrt{\left(\frac{p}{q}\right)^k} I_{|k|}(2t \sqrt{pq}).$$

¹⁾ Эта задача содержит новый пример не нормального распределения с нормальными частными распределениями (см. задачи 2 и 3 в гл. III, 9), принадлежащий Е. Дж. Гумбелу.

где I_k — функция Бесселя, определенная в гл. II, (7.1).

10. Каждому распределению F , для которого интеграл

$$\varphi(a) = \int_{-\infty}^{\infty} e^{ax} F\{dx\}$$

существует при $-a < \alpha < a$, сопоставим новое распределение F^* посредством формулы $\varphi(\alpha) F^*\{dx\} = e^{\alpha x} F\{dx\}$. Пусть F_1 и F_2 — два распределения, обладающие указанным выше свойством, и $F = F_1 \star F_2$. Доказать, что (используя очевидные обозначения) $\varphi(\alpha) = \varphi_1(\alpha) \varphi_2(\alpha)$ и $F^* = F_1^* \star F_2^*$.

11. Пусть F — атомическое распределение с массами $p_1, p_2, \dots$ в точках $a_1, a_2, \dots$. Обозначим через p максимальное среди чисел $p_1, p_2, \dots$. Используя лемму 1 § 4а, показать, что

а) массы всех атомов распределения $F \star F$ строго меньше p , за исключением того случая, когда F сосредоточено на конечном числе атомов одинаковой массы;

б) для симметризованного распределения ${}^0F$ нуль является атомом с массой $p' = \sum p_i^2$; массы остальных атомов ${}^0F$ строго меньше p' .

12. *Случайные векторы в $\mathcal{R}^3$.* Пусть L — длина суммы двух независимых единичных векторов со случайными направлениями (т. е. их концевые точки распределены равномерно на единичной сфере). Покажите, что $P\{L \leq t\} = t^2/2$ при $0 < t < 2$ [См. пример 4, д).]

13. Пусть случайные величины X_k взаимно независимы и принимают значения 0 и 1 с вероятностью $1/2$ каждое. В примере 4, г) было показано, что случайная величина $X = \sum 2^{-k} X_k$ равномерно распределена на $[0, 1]$. Покажите, что $\sum 2^{-k} X_{2k}$ имеет сингулярное распределение.

14. а) Если F имеет плотность f и f^2 — интегрируемая функция, то плотность f_2 свертки $F \star F$ ограничена.

б) Используя теорему об аппроксимации в среднем из гл. IV, 2, покажите, что если f ограничена, то f_2 непрерывна.

[Если f неограничена в окрестности некоторой точки, то может оказаться, что f^{2n} неограничена при каждом n . См. пример XI, 3, а).]

15. Используя неравенство Шварца, покажите, что если X есть положительная случайная величина, то $E(X^{-p}) \geq (E(X^p))^{-1}$ при всех $p > 0$.

16. Пусть X и Y — случайные величины с плотностями f и g , такими, что $f(x) \geq g(x)$ при $x < a$ и $f(x) \leq g(x)$ при $x > a$. Доказать, что $E(X) \leq E(Y)$. Далее, если $f(x) = g(x) = 0$ при $x < 0$, то $E(X^k) \leq E(Y^k)$ для всех k .

17. Пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины с одинаковой функцией распределения F . Пусть N — положительная случайная величина, принимающая целые значения и имеющая производящую функцию $P(s)$. Если N не зависит от X_j , то случайная величина $\max\{X_1, \dots, X_N\}$ имеет распределение $P(F)$.

18. Пусть $X_1, \dots, X_n$ — взаимно независимые случайные величины с непрерывным распределением F . Положим $X = \max\{X_1, \dots, X_n\}$ и $Y = \min\{X_1, \dots, X_n\}$. Тогда

$$P\{X < x, Y > y\} = (F(x) - F(y))^n \text{ при } y < x$$

$$P\{Y > y | X = x\} = [1 - F(y)/F(x)]^{n-1}.$$

19. При каждом фиксированном $k \leq n$ (в обозначениях предыдущей задачи)

$$P\{X_k \leq x | X = t\} = \begin{cases} \frac{n-1}{n} \frac{f(x)}{f(t)} & \text{при } x < t, \\ 1 & \text{при } x \geq t. \end{cases}$$

Доказать это а) с помощью эвристического рассуждения, рассматривая событие $\{X_k = X\}$, и б) формально, используя (9.4).

20. Продолжение. Доказать, что

$$E(X_k | X = t) = \frac{n-1}{n} \frac{1}{F(t)} \int_{-\infty}^t y F\{dy\} + \frac{t}{n}.$$

21. *Случайные суммы.* В примере 9, в) случайную величину X_k положим равной 1 и -1 с вероятностями p и $q=1-p$. Если N имеет пуассоновское распределение с математическим ожиданием t , то распределение S_N совпадает с распределением, указанным в задаче 9.

22. *Смеси.* Пусть распределение G в (9.9) имеет математическое ожидание $m(x)$ и дисперсию $\sigma^2(x)$. Показать, что математическое ожидание и дисперсия смеси U равны

$$a = \int_{-\infty}^{\infty} m(x) \mu\{dx\}, \quad b = \int_{-\infty}^{\infty} \sigma^2(x) \mu\{dx\} + \int_{-\infty}^{\infty} (m^2(x) - a^2) \mu\{dx\}.$$

23. Применяя очевидные обозначения, имеем $E(E(Y|X)) = E(Y)$, но

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)).$$

Задача 22 является частным случаем этой задачи.

24. *Случайные суммы.* В обозначениях примера 9, в) $E(S_N) = E(N)E(X)$ и $\text{Var}(S_N) = E(N)\text{Var}(X) + (E(X))^2\text{Var}(N)$. Доказать эти соотношения непосредственно и показать, что они вытекают из результатов двух предыдущих задач.

Замечание. Следующие задачи относятся к сверткам *одновершинных распределений*, определенных в сноске на с. 186. Была выдвинута гипотеза, что свертка двух одновершинных распределений одновершинна. Противоречащий этой гипотезе пример был построен К. Л. Чжуном. Задача 25 содержит другой такой пример. Задача 26 показывает, что гипотеза верна для *симметричных*¹⁾ распределений (этот результат принадлежит А. Винтнеру).

25. Пусть $u(x) = 1$ при $0 < x < 1$ и $u(x) = 0$ в остальных точках. Положим

$$v(x) = \frac{e}{a} u\left(\frac{x}{a}\right) + \frac{1-e}{b} u\left(\frac{x}{b}\right),$$

где $0 < a < b$. Если e и a малы, а b велико, то плотность v одновершинна, но плотность $w = v * u$ не одновершинна.

Указание как избежать вычислений. В результате свертки двух равномерных плотностей получается треугольная плотность. Следовательно, $w(a) > e^2 a^{-1}$, $w(b) > e^2 b^{-1}$ и интеграл от w в пределах от b до $2b$ больше $(1-e)^2/2$. Отсюда вытекает, что плотность w должна иметь минимум между a и b .

26. Пусть F — равномерное, а G — одновершинное распределения. С помощью простого дифференцирования показать, что если F и G симметричны, то свертка $F * G$ одновершинна. Вывести отсюда (без дальнейших вычислений), что тот же результат получается, когда F есть произвольная смесь симметричных равномерных распределений и что, следовательно, *свертка симметричных одновершинных распределений одновершинна*.

¹⁾ С трудностями, возникающими в несимметричном случае, можно познакомиться в работе И. А. Ибрагимова, Теория вероятностей и ее применения, т. 1 (1956), с. 283—288.

Настоящая глава включена в книгу из-за прискорбной необходимости избегать повторений и взаимных ссылок в главах, предназначенных для независимого чтения. Например, теория устойчивых распределений будет изложена независимо на основе полугрупповых методов (гл. IX), анализа Фурье (гл. XVII) и, по крайней мере частично, на основе преобразований Лапласа (гл. XIII). Вынесение определений и примеров в отдельное место дает некоторую экономию и делает возможным тщательное исследование некоторых основных соотношений, не стесненное чистотой используемых методов.

Разнообразные темы, которым посвящена настоящая глава, не всегда логически связаны: теория очередей имеет мало общего с мартингалами или с устойчивыми распределениями. Материал главы не предназначен для последовательного чтения; лучше читать тот или иной параграф по мере появления необходимости. Параграфы 6—9 в известной степени связаны между собой, но не зависят от остальной части главы. В них излагается важный материал, не затрагиваемый в других местах книги.

§ 1. УСТОЙЧИВЫЕ РАСПРЕДЕЛЕНИЯ В $\mathcal{R}^1$

Устойчивые распределения играют все возрастающую роль в качестве естественного обобщения нормального распределения. Для их описания удобно ввести следующее сокращенное обозначение: если распределения двух случайных величин U и V совпадают, то мы будем писать

$$U \stackrel{d}{=} V. \quad (1.1)$$

В частности, соотношение $U \stackrel{d}{=} aV + b$ означает, что распределения величин U и V отличаются лишь параметрами расположения и масштабными параметрами [см. определение 1 в гл. V, 2]. Всюду в этом параграфе символы X , X_1 , X_2 , ... обозначают взаимно независимые случайные величины с одним и тем же распределением R , а $S_n = X_1 + \dots + X_n$.

Определение 1. Распределение R называется устойчивым

(в широком смысле), если оно не сосредоточено в нуле и для каждого n существуют постоянные $c_n > 0$ и γ_n , такие, что¹⁾

$$S_n \stackrel{d}{=} c_n X + \gamma_n. \quad (1.2)$$

R устойчиво в узком смысле (строго устойчиво), если оно удовлетворяет соотношению (1.2) с $\gamma_n = 0$.

Примеры будут приведены в § 2. Весьма поучителен элементарный вывод некоторых основных свойств устойчивых распределений, и мы дадим его здесь несмотря на то, что дальше будут даны другие доказательства этих свойств. Систематическое изложение теории устойчивых распределений в гл. IX и XVII не опирается на нижеследующий материал.

Теорема 1. Нормирующие постоянные имеют вид $c_n = n^{1/\alpha}$ с $0 < \alpha \leq 2$. Постоянная α будет называться характеристическим показателем распределения R .

Доказательство. Доказательство значительно упрощается при использовании симметризации. Если распределение R устойчиво, то распределение ${}^0R$ величины $X_1 - X_2$ тоже устойчиво, причем с теми же самыми нормирующими постоянными c_n . Достаточно, следовательно, ограничиться случаем симметричного устойчивого распределения R .

Заметим сначала, что S_{m+n} есть сумма независимых случайных величин S_m и $S_{m+n} - S_m$, распределения которых совпадают соответственно с распределениями величин $c_m X$ и $c_n X$. Таким образом, для симметричных устойчивых распределений имеем

$$c_{m+n} X \stackrel{d}{=} c_m X_1 + c_n X_2. \quad (1.3)$$

Сумму S_{rk} можно аналогичным образом разбить на r независимых частей по k членов в каждой, и поэтому $c_{rk} = c_r c_k$ при всех r и k . Отсюда по индукции заключаем, что

$$\text{если } n = r^v, \text{ то } c_n = c_r^v. \quad (1.4)$$

Далее положим $v = m + n$ и заметим, что в силу симметрии величин из (1.3) при $t > 0$

$$P\{X > t\} \geq \frac{1}{2} P\{X_2 > tc_v/c_n\}. \quad (1.5)$$

Отсюда следует, что при $v > n$ отношения c_n/c_v остаются ограниченными.

Для любого целого числа r существует единственное α , такое, что $c_r = r^{1/\alpha}$. Для того чтобы доказать равенство $c_n = n^{1/\alpha}$, достаточно показать, что из $c_\rho = \rho^{1/\beta}$ вытекает $\beta = \alpha$. В силу (1.4)

$$\text{если } n = r^j, \text{ то } c_n = n^{1/\alpha},$$

$$\text{если } v = \rho^k, \text{ то } c_v = v^{1/\beta}.$$

¹⁾ В другой форме это определение см. в задаче 1.

Но для любого $v = r^k$ существует $n = r^j$, такое, что $n < v \leq rn$. В таком случае

$$c_v = v^{1/\beta} \leq (rn)^{1/\beta} = r^{1/\beta} c_n^{\alpha/\beta}.$$

Так как отношения c_n/c_v ограничены, то отсюда следует, что $\beta \leq \alpha$. Меняя r и ρ ролями, мы аналогичным образом показываем, что $\beta \geq \alpha$ и, следовательно, $\beta = \alpha$.

Докажем, что $\alpha \leq 2$. Для этого вспомним, что нормальное распределение устойчиво с характеристическим показателем $\alpha = 2$. В этом случае (1.3) сводится к правилу сложения дисперсий, откуда следует, что любое устойчивое распределение с конечными дисперсиями необходимо соответствует значению $\alpha = 2$. Для завершения доказательства достаточно поэтому показать, что любое устойчивое распределение с $\alpha > 2$ должно иметь конечную дисперсию.

Для симметричных распределений соотношение (1.2) выполняется с $\gamma_n = 0$, и, следовательно, мы можем выбрать t так, что $P\{|S_n| > tc_n\} < 1/4$ для всех n . В силу симметрии отсюда следует, что $n[1 - R(tc_n)]$ остается ограниченным [см. V, (5.10)]. Тогда для всех $x > t$ и подходящей постоянной M выполняется $x^\alpha [1 - R(x)] < M$. Таким образом, вклад интервала $2^{k-1} < x \leq 2^k$ в интеграл, определяющий $E(X^2)$, ограничен величиной $M2^{(2-\alpha)k}$, а при $\alpha > 2$ это является общим членом сходящегося ряда. ►

Значительному упрощению теории устойчивых распределений способствует то приятное обстоятельство, что на практике центрирующими постоянными γ_n можно пренебречь. Это так, потому что мы вольны центрировать произвольным образом распределение R , т. е. мы можем заменить $R(x)$ на $R(x+b)$. Следующая теорема показывает, что, исключая случай $\alpha = 1$, мы можем употребить эту свободу на устранение γ_n из соотношения (1.2).

Теорема 2. Для всякого устойчивого распределения R с характеристическим показателем $\alpha \neq 1$ можно выбрать центрирующую постоянную b так, чтобы $R(x+b)$ было бы строго устойчиво.

Доказательство. $S_{m,n}$ есть сумма m независимых случайных величин, каждая из которых распределена одинаково с величиной $c_n X + \gamma_n$. Соответственно

$$S_{m,n} = \sum_{i=1}^d c_n S_m + m\gamma_n = c_n c_m X + c_n \gamma_m + m\gamma_n. \quad (1.6)$$

Поскольку m и n играют одну и ту же роль, это означает, что имеет место равенство

$$(c_n - n)\gamma_m = (c_m - m)\gamma_n. \quad (1.7)$$

При $\alpha = 1$ это утверждение не имеет смысла¹⁾, но при $\alpha \neq 1$ отсюда следует, что $\gamma_n = b(c_n - n)$ при всех n . Наконец, из (1.2) видно, что сумма S'_n n случайных величин, распределенных как $X' - b$, удовлетворяет условию $S'_n \stackrel{d}{=} c_n X'$.

Соотношение (1.3) было получено из (1.2) при единственном предположении, что $\gamma_n = 0$, и поэтому оно справедливо для всех строго устойчивых распределений. Это влечет, что

$$s^{1/\alpha} X_1 + t^{1/\alpha} X_2 \stackrel{d}{=} (s+t)^{1/\alpha} X \quad (1.8)$$

всякий раз, когда отношение s/t рационально. Простое использование свойства непрерывности²⁾ приводит к следующей теореме.

Теорема 3. Для строго устойчивых распределений R с характеристическим показателем α соотношение (1.8) выполняется при всех $s > 0$ и $t > 0$.

Для нормального распределения соотношение (1.8) просто устанавливает правило сложения для дисперсий. Вообще из (1.8) следует, что все линейные комбинации $a_1 X_1 + a_2 X_2$ принадлежат одному типу.

Та важная роль, которую играет нормальное распределение $\mathcal{N}$ в теории вероятностей, в значительной мере основана на центральной предельной теореме. Пусть $X_1, \dots, X_n$ — взаимно независимые величины с общим распределением F , нулевым математическим ожиданием и единичной дисперсией. Положим $S_n = X_1 + \dots + X_n$. Центральная предельная теорема³⁾ утверждает, что распределения случайных величин $S_n n^{-1/2}$ сходятся к $\mathcal{N}$. Если распределения не имеют дисперсий, то можно сформулировать аналогичные предельные теоремы, однако нормирующие постоянные должны быть выбраны другими. Весьма интересным является тот факт, что в качестве предельных распределений могут быть получены все устойчивые законы и только они. Введем следующую терминологию, которая облегчит дальнейшее обсуждение рассматриваемого вопроса.

Определение 2. Скажем, что распределение F независимых случайных величин X_k принадлежит области притяжения распределения R , если существуют нормирующие постоянные $a_n > 0$, b_n , такие, что распределение величины $a_n^{-1}(S_n - b_n)$ сходится к R .

Приведенное выше утверждение можно теперь переформулировать следующим образом: распределение R имеет область при-

¹⁾ Случай $\alpha = 1$ см. в задаче 4.

²⁾ По поводу непрерывности устойчивых распределений см. задачу 2.

³⁾ Центральная предельная теорема показывает, что нормальное распределение — это единственное устойчивое распределение, обладающее дисперсией.

тяжения тогда и только тогда, когда оно устойчиво. Действительно, из определения устойчивости непосредственно вытекает, что устойчивое распределение R принадлежит своей области притяжения. Тот факт, что никакое неустойчивое распределение не выступает в качестве предельного, можно обосновать с помощью рассуждения, использованного при доказательстве теоремы 1.

Полученные результаты имеют важные и неожиданные следствия. Рассмотрим, например, устойчивое распределение, удовлетворяющее соотношению (1.8) с $\alpha < 1$. Среднее $(X_1 + \dots + X_n)/n$ имеет одинаковое распределение с величиной $X_1 n^{-1+1/\alpha}$. Обратим внимание на то, что множитель $n^{-1+1/\alpha}$ стремится к бесконечности с ростом n . Выражаясь нестрого, можно сказать, что среднее n величин X_k оказывается значительно больше любого фиксированного слагаемого X_k . Это возможно лишь тогда, когда *максимальный член* $M_n = \max[X_1, \dots, X_n]$ растет исключительно быстро и играет основную роль в сумме S_n . Более длительный анализ подтверждает это заключение. В случае положительных случайных величин математическое ожидание отношения S_n/M_n стремится к $(1-\alpha)^{-1}$, и это верно также для любой последовательности $\{X_n\}$ с распределением, принадлежащим области притяжения рассматриваемого устойчивого закона (см. задачу 26 из гл. XIII, 11).

Исторические замечания. Основы общей теории устойчивых распределений были заложены П. Леви¹⁾ (1924), который нашел преобразования Фурье всех строго устойчивых распределений. (Устойчивые распределения в широком смысле первоначально назывались квазистойчивыми. Как мы видели, они существенно отличаются от строго устойчивых распределений лишь при $\alpha=1$. Случай $\alpha=1$ был исследован совместно Леви и Хинчиным.) Более простой новый подход к теории устойчивых распределений стал возможным после открытия безгранично делимых распределений. Этот новый подход (тоже основанный на анализе Фурье) также принадлежит П. Леви (1937). Интерес к теории был стимулирован замечательным исследованием А. Дёблина (1939), посвященным областям притяжения. В предложенных им критериях впервые появились правильно меняющиеся функции. Современная теория по-прежнему носит отпечаток этой основополагающей работы, несмотря на то что некоторыми авторами были внесены различные усовершенствования и получены новые результаты. Гл. XVII содержит изложение теории устойчивых распределений на основе методов Фурье, ставших теперь уже классическими. В гл. IX эта теория развивается с помощью прямого подхода, более близкого к современным методам теории марковских процессов. Значительные упрощения и объединение многих критериев оказались возможными благодаря систематическому использованию развитой Карамата теории правильно меняющихся функций. Усовершенствованный вариант этой теории излагается в гл. VIII, 8, 9.

¹⁾ Преобразования Фурье *симметричных* устойчивых распределений были указаны Коши, однако оставалось неясным, действительно ли они отвечают вероятностным распределениям. При $\alpha < 1$ этот вопрос был решен Пойа. Распределение Хольцмарка из примера 2, в) было известно астрономам, но не известно математикам.

§ 2. ПРИМЕРЫ

а) *Нормальное распределение* с нулевым математическим ожиданием строго устойчиво с $c_n = \sqrt{n}$.

б) *Распределение Коши* с произвольными параметрами расположения, имеющее плотность

$$\frac{1}{\pi} \frac{c}{c^2 + (x - \gamma)^2},$$

устойчиво с $\alpha = 1$. Это следует из свойства свертки для распределения Коши, см. гл. II, (4.6).

в) *Устойчивое распределение* с $\alpha = 1/2$. Распределение

$$F(x) = 2 \left[1 - \mathfrak{H} \left(\frac{1}{\sqrt{x}} \right) \right], \quad x > 0, \quad (2.1)$$

имеющее плотность

$$f(x) = \frac{1}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}x}, \quad x > 0 \quad (2.2)$$

$[f(x) = 0 \text{ при } x < 0]$, строго устойчиво с нормирующими постоянными $c_n = n^2$.

Это утверждение доказывается посредством элементарного интегрирования, однако предпочтительнее получить его как следствие того, что F имеет область притяжения. Действительно, пусть в симметричном случайном блуждании S_r обозначает момент r -го возвращения в начало координат. Очевидно, что S_r есть сумма r независимых одинаково распределенных случайных величин (времен ожидания между последовательными возвращениями). В конце I, III, (7.7) было показано, что

$$P\{S_r \leq r^2 t\} \rightarrow F(t), \quad r \rightarrow \infty. \quad (2.3)$$

Таким образом, F имеет область притяжения и потому устойчиво (продолжение в примере д)).

г) *Гравитационное поле звезд (распределение Хольцмарка)*. На языке астрономии задача формулируется так: вычислить x -компоненту гравитационной силы, создаваемой звездной системой в случайно выбранной точке O . В основе такой постановки задачи лежит предположение, что звездная система представляет собой «случайное скопление» точек со «случайно распределенными массами». Это предположение можно было бы выразить строго математически в терминах распределения Пуассона и т. п., однако, к счастью, для рассматриваемой задачи подобных тонкостей не требуется.

Будем рассматривать плотность звездной системы как свободный параметр; пусть X_λ обозначает x -компоненту гравитационной силы звездной системы с плотностью λ . Нас интересуют возможные типы распределений X_λ . Интуитивное понятие «случайное

скопление звезд» предполагает, что два независимых скопления с плотностями s и t могут быть объединены в одно скопление с плотностью $s+t$. С вероятностной точки зрения это приводит к требованию, чтобы сумма двух независимых случайных величин, распределенных как X_t и X_s , была распределена как X_{s+t} . Иначе это можно выразить равенством

$$X_s + X_t \stackrel{d}{=} X_{s+t}. \quad (2.4)$$

Считая, что замена единичной плотности плотностью λ приводит к изменению единичной длины на длину $1/\sqrt[3]{\lambda}$ и что гравитационная сила обратно пропорциональна квадрату расстояния, приходим к выводу, что величины X_t и $t^{2/3}X_1$ должны иметь одинаковое распределение. Это означает, что распределения величин X_t различаются лишь масштабным множителем и (2.4) сводится к (1.8) с $\alpha=3/2$. Иными словами, X_t имеет симметричное устойчивое распределение с характеристическим показателем $\alpha=3/2$. В дальнейшем мы увидим, что (с точностью до тривиального масштабного множителя) имеется в точности одно такое распределение. Таким образом, мы решили нашу задачу, не прибегая к более глубокой теории. Астроном Хольцмарк получил равносильный ответ, применяя другие методы (см. задачу 7), причем, что весьма замечательно, до появления работы П. Леви.

д) *Моменты первого прохождения в броуновском движении.* Введем сначала понятие одномерного диффузионного процесса. Диффузионный процесс—это такой процесс $X(t)$, приращения которого $X(s+t)-X(s)$ за непересекающиеся промежутки времени независимы и имеют симметричное нормальное распределение с дисперсией t . Будем предполагать известным, что траектории одномерного диффузионного процесса непрерывно зависят от времени. При $X(0)=0$ существует момент T_a , в который частица впервые достигает уровня $a>0$. Для нахождения функции распределения $F_a(t)=P(T_a \leq t)$ заметим, что понятие аддитивного процесса включает в себе полное отсутствие последействия (строго марковское свойство). Это означает, что приращение абсциссы $X(t+T_a)-a$ между моментами времени T_a и T_a+t не зависит от течения процесса до момента T_a . Далее, чтобы достигнуть уровня $a+b>a$, частица должна сначала достигнуть уровня a , и поэтому время ожидания $T_{a+b}-T_a$ до достижения уровня $a+b$ после достижения a не зависит от T_a и имеет распределение, одинаковое с распределением T_b . Иными словами, $F_a \star F_b = F_{a+b}$. Так как вероятности перехода зависят лишь от отношения x^2/t , то T_a имеет то же распределение, что и a^2T_1 . Это означает, что распределения F_a различаются лишь масштабным множителем и, следовательно, устойчивы с характеристическим показателем $\alpha=1/2$.

Проведенное рассуждение, основанное на соображениях раз-

мерности, доказывает устойчивость распределения момента первого прохождения уровня, но не дает его точной формы. Чтобы доказать *совпадение* F с *распределением* примера в), мы воспользуемся рассуждением, основанным на соображениях симметрии (так называемом принципе отражения). В силу непрерывности траекторий событие $\{X(t) > a\}$ может произойти лишь в том случае, когда уровень a был пересечен в некоторый момент $T_a < t$. Если $T_a = \tau < t$, то $X(\tau) = a$, и из симметрии ясно, что вероятность события $X(t) - X(\tau) > 0$ равна $1/2$. Следовательно,

$$P\{T_a < t\} = 2P\{X(t) > a\} = 2 \left[(1 - \mathfrak{N}) \left(\frac{a}{\sqrt{t}} \right) \right], \quad (2.5)$$

что и требовалось доказать (ср. (2.1)).

е) *Точки достижения в двумерном броуновском движении.* Двумерное броуновское движение образуется парой $(X(t), Y(t))$ независимых одномерных броуновских движений. Нас интересует точка (a, Z_a) , в которой траектория впервые достигает прямой $x = a > 0$. Отметим, что подобно тому, как это было в предыдущем примере, траектория может достичь прямой $x = a + b > a$ лишь после пересечения прямой $x = a$. Если взять в качестве нового начала координат точку (a, Z_a) , то станет ясно, что величина Z_{a+b} распределена так же, как сумма двух независимых случайных величин, распределенных одинаково с величинами Z_a и Z_b . Далее, очевидные соображения подобия приводят к заключению, что величина Z_a распределена так же, как величина aZ_1 , и поэтому Z_a имеет симметричное устойчивое распределение с характеристическим показателем $\alpha = 1$. Единственным таким распределением является распределение Коши, так что *точка достижения Z_a имеет распределение Коши.*

Приведенное поучительное рассуждение, основанное на соображениях размерности, не определяет масштабного множителя. Для его вычисления заметим, что $Z_a = Y(T_a)$, где T_a — момент первого достижения прямой $x = a$. Распределение величины T_a дается формулой (2.5), а $Y(t)$ имеет нормальное распределение с дисперсией t . Следовательно, величина Z_a имеет плотность, равную¹⁾

$$\int_0^{\infty} \frac{1}{t^{1/2} \sqrt{2\pi}} e^{-\frac{1}{2} x^2/t} \frac{a}{t^{3/2} \sqrt{2\pi}} e^{-\frac{1}{2} a^2/t} dt = \frac{a}{\pi (a^2 + x^2)}. \quad (2.6)$$

[Здесь мы имеем пример подчиненности процессов, к которому мы еще вернемся в гл. X, 7.]

ж) *Устойчивые распределения в экономике.* Примененные в двух последних примерах и основанные на соображениях размерности

¹⁾ Подстановка $y = \frac{1}{2} (x^2 + a^2)/t$ сводит подынтегральную функцию к функции e^{-y} .

рассуждения были использованы Б. Мандельбротом для доказательства того, что различные экономические процессы (в частности, распределения дохода) подчинены устойчивым (или «Леви — Парето») распределениям. До настоящего времени сила и ценность этой интересной теории, привлеченной внимание экономистов, проявились скорее в теоретической сфере, чем в прикладном аспекте. [По поводу кажущегося согласия хвостов распределения с многими эмпирическими явлениями, начиная от размера городов до частоты слов, см. гл. II, 4, з).]

з) *Произведения*. Существует много любопытных соотношений между устойчивыми распределениями с различными характеристическими показателями. Наиболее интересное из них можно сформулировать в виде следующего предложения. Пусть X и Y — независимые строго устойчивые случайные величины с характеристическими показателями α и β соответственно. Предположим, что величина Y положительна (при этом $\beta < 1$). Тогда *произведение $XY^{1/\alpha}$ имеет устойчивое распределение с характеристическим показателем $\alpha\beta$* . В частности, произведение нормальной случайной величины и квадратного корня из случайной величины с устойчивым распределением примера в) имеет распределение Коши.

Это утверждение вытекает как простое следствие из теоремы о подчиненных процессах¹⁾ [пример 7, в) гл. X]. Кроме того, его легко проверить с помощью анализа Фурье [задача 9 из гл. XVII, 12] и для положительных случайных величин с помощью преобразований Лапласа [гл. XIII, 7, д) и задача 10 из гл. XIII, 11].

§ 3. БЕЗГРАНИЧНО ДЕЛИМЫЕ РАСПРЕДЕЛЕНИЯ В $\mathcal{R}^1$

Определение 1. *Распределение F безгранично делимо, если для каждого n существует распределение F_n , такое, что $F = F_n^{*n}$.*

*Иными словами*²⁾, F безгранично делимо, если при каждом n

¹⁾ Для непосредственной проверки с минимальными вычислениями можно поступить следующим образом. Найдем распределение величины $Z = X_1 \sqrt[n]{Y_1} + X_2 \sqrt[n]{Y_2}$, вычислив сначала условное распределение Z при условии $Y_1 = y_1$, $Y_2 = y_2$. Это условное распределение есть функция суммы $y_1 + y_2$, и замена переменных $u = y_1 + y_2$, $v = y_1 - y_2$ показывает, что распределение Z отличается от распределения $X_1 \sqrt[n]{Y_1}$ лишь масштабным множителем. Аналогичные вычисления можно проделать в случае n таких слагаемых.

²⁾ Следует иметь в виду, что случайные величины $X_{k,n}$ вводятся только для простоты и наглядности. При фиксированном n случайные величины $X_{1,n}, \dots, X_{n,n}$ предполагаются взаимно независимыми, но при $m \neq n$ величины $X_{j,m}$ и $X_{k,n}$ не обязаны даже быть определенными на одном и том же вероятностном пространстве (иначе говоря, совместное распределение величин $X_{k,m}$ и $X_{k,n}$ может не существовать). Это замечание относится к схеме серий вообще.

его можно представить как распределение суммы $S_n = X_{1,n} + \dots + X_{n,n}$ n независимых случайных величин с одним и тем же распределением F_n .

Это определение относится к распределениям в любом конечномерном пространстве, однако здесь мы ограничимся рассмотрением лишь одномерных распределений. Следует отметить, что безграничная делимость является свойством типа, т. е. вместе с F все распределения, отличающиеся от F лишь параметрами расположения, безгранично делимы. Устойчивые распределения безгранично делимы и выделяются среди безгранично делимых тем, что для них F_n отличаются от F лишь параметрами расположения.

Примеры. а) Все гамма-распределения (включая показательные) в силу свойства свертки из (2.3) гл. II безгранично делимы. Безграничная делимость дискретного варианта гамма-распределений — «отрицательных биномиальных» распределений (включая геометрические распределения) была доказана в 1; гл. XII, 2, д).

б) Пуассоновское и обобщенное пуассоновское распределения безгранично делимы. Будет показано, что все безгранично делимые распределения являются пределами обобщенных пуассоновских распределений.

в) Связанное с бесселевыми функциями распределение (7.13) из гл. II безгранично делимо, однако это вовсе не очевидно; см. пример 7, г) гл. XIII.

г) Безгранично делимое распределение не может быть сосредоточено на конечном интервале, если оно не сосредоточено в одной точке. В самом деле, пусть F безгранично делимо. Если $|S_n| < a$ с вероятностью единица, то $|X_{k,n}| < an^{-1}$ и поэтому $\text{Var}(X_{k,n}) < a^2 n^{-2}$. Следовательно, дисперсия F , будучи меньше $a^2 n^{-1}$, должна быть равна нулю. ►

Возвращаясь к определению 1, посмотрим, что произойдет, если опустить требование одинаковой распределенности величин $X_{k,n}$ и потребовать лишь, чтобы при каждом n существовало n распределений $F_{1,n}, \dots, F_{n,n}$, для которых

$$F = F_{1,n} \star \dots \star F_{n,n}. \quad (3.1)$$

Такая общность приводит к новому явлению, которое лучше всего проиллюстрировать на примерах.

Примеры. д) Если F — безгранично делимое распределение, а U — произвольное распределение, то распределение $G = U \star F$ можно представить в виде (3.1) с $G_{1,n} = U$ и $G_{k,n} = F_{n-1}$ при $n > 1$. Роль первой компоненты здесь совершенно отлична от роли остальных компонент.

е) Рассмотрим сходящийся ряд $X = \sum X_k$ взаимно независимых случайных величин. Распределение F величины X , являясь сверт-

кой распределений величин $X_1, X_2, \dots, X_{n-1}$ и распределения остаточного члена $(X_n + X_{n+1} + \dots)$, представлено в виде (3.1). Распределения такого рода называются *бесконечными свертками* и будут изучаться в дальнейшем. Пример гл. I, 11, в) показывает, что равномерное распределение является бесконечной сверткой. ►

Отличительной чертой этих примеров является то обстоятельство, что вклад отдельной компоненты $X_{k,n}$ в сумму S_n существен, в то время как в случае одинаково распределенных компонент вклад каждой отдельной компоненты стремится к нулю. Мы хотим теперь связать безгранично делимые распределения с типичными предельными теоремами для сумм «большого числа малых компонент». Для этого необходимо дополнить нашу схему требованием, чтобы отдельные компоненты $X_{k,n}$ были асимптотически пренебрегаемыми в том смысле, что для любого $\varepsilon > 0$ при достаточно больших n выполнялись бы неравенства

$$P\{|X_{k,n}| > \varepsilon\} < \varepsilon \quad (k=1, \dots, n). \quad (3.2)$$

В терминологии гл. VIII, 2, это означает, что величины $X_{k,n}$ стремятся по вероятности к нулю равномерно по $k=1, \dots, n$. Системы случайных величин такого типа встречаются довольно часто, и поэтому удобно дать им специальное название.

Определение 2. Двойная последовательность случайных величин $X_{k,n}$ ($k=1, 2, \dots, n$; $n=1, 2, \dots$), в которой образующие n -ю строку (серию) случайные величины $X_{1,n}, \dots, X_{n,n}$ взаимно независимы, называется *схемой серий*¹⁾.

Если выполняется (3.2), то схема серий называется *нулевой* (или *схемой серий с асимптотически пренебрегаемыми компонентами*).

Более общим образом можно было бы рассматривать схемы серий с r_n величинами в n -й серии, где $r_n \rightarrow \infty$, однако в действительности общность при этом не увеличивается. (См. задачу 10.)

Пример. ж) Пусть $\{X_j\}$ — последовательность одинаково распределенных независимых случайных величин и $S_n = X_1 + \dots + X_n$. Нормированная сумма $S_n a_n^{-1}$ есть сумма величин n -й серии в схеме серий, в которой $X_{k,n} = X_k a_n^{-1}$. Эта последовательность серий является нулевой, если $a_n \rightarrow \infty$. Иного типа схема серий была рассмотрена при выводе пуассоновского распределения в гл. VI, 6. ►

В главах IX и XVII будет доказан следующий замечательный факт: *предельное распределение для сумм S_n величин, образующих n -ю серию в нулевой схеме серий, (если оно существует) безгранично делимо*. Если выполнено условие асимптотической пренебрегаемости (3.2), то неважно, одинаково или неодинаково распре-

¹⁾ В подлиннике: triangular array — треугольная схема. — Прим. перев.

делены компоненты $X_{k,n}$, и в (3.1) знак равенства можно заменить пределом: класс безгранично делимых распределений совпадает с классом предельных распределений для сумм величин, образующих n -ю серию в нулевых схемах серий.

Примеры прикладного характера. 3) *Дробовой эффект в вакуумных трубках.* Различные варианты и обобщения рассматриваемого ниже случайного процесса встречаются в физике и технике связи.

Рассмотрим флуктуации электрического тока, возникающие из-за случайных флуктуаций числа попадающих на анод электронов. Предполагается, что моменты попадания электронов на анод образуют пуассоновский процесс и что попавший электрон возбуждает ток, сила которого x единиц времени спустя равна $I(x)$. Сила тока в момент t изображается случайной величиной

$$X(t) = \sum_{k=1}^{\infty} I(t - T_k), \quad (3.3)$$

где T_k — предшествующие t моменты попадания электронов. (Иными словами, величины $t - T_1, T_2 - T_1, T_3 - T_2, \dots$ взаимно независимы и имеют одинаковое показательное распределение.)

Нетрудно проанализировать сумму (3.3) непосредственно методами теории случайных процессов, однако простой подход с помощью схемы серий может помочь лучшему пониманию сути вещей. Разобьем интервал $-\infty, t$ на маленькие подынтервалы с концами в точках $t_k = t - kh$ ($k = 0, 1, \dots$). В силу самого определения пуассоновского процесса вклад интервала t_k, t_{k-1} в сумму (3.3) можно приблизительно считать случайной величиной, принимающей значение 0 с вероятностью $1 - \alpha h$ и значение $I(t - t_k)$ с вероятностью αh . Математическое ожидание и дисперсия этой случайной величины равны соответственно $\alpha h I(kh)$ и $\alpha h (1 - \alpha h) I^2(kh)$. Положим $h = 1/\sqrt{n}$ и построим схему серий, в которой $X_{k,n}$ есть вклад k -го интервала. Сумма величин, образующих отдельную серию, имеет математическое ожидание $\alpha h \sum I(kh)$ и дисперсию $\alpha h (1 - \alpha h) \sum I^2(kh)$. Если ряду (3.3) можно придать какой-либо смысл, то распределения сумм по сериям должны будут сходиться к распределению $X(t)$ и мы приходим к следующим соотношениям

$$E(X(t)) = \alpha \int_0^{\infty} I(s) ds, \quad \text{Var}(X(t)) = \alpha \int_0^{\infty} I^2(s) ds. \quad (3.4)$$

Эти выводы легко подтвердить с помощью предельных теорем для схемы серий. Соотношения (3.4) хорошо известны как *теорема Кэмпбелла*. С современной точки зрения это неглубокий результат, однако он был получен в 1909 г. за десятилетия до того, как была развита систематическая теория. В свое время этот результат казался замечательным и его передоказывали многими разными способами.

и) *Загруженность линий связи.* Приведем еще один сходный с предыдущим пример, иллюстрирующий характер возможных обобщений. Рассмотрим телефонную станцию с бесконечным числом линий. Предположим, что приходящие вызовы образуют пуассоновский процесс и что каждый приходящий вызов направляется на свободную линию. Времена обслуживания имеют одинаковое распределение F и, как обычно, предполагаются независимыми друг от друга и от процесса поступления вызовов. Число занятых линий в момент t является случайной величиной $X(t)$, распределение которой можно найти, применяя метод, основанный на схеме серий. Как и в предыдущем примере, разобьем $\overline{0, t}$ на n интервалов длины $h = t/n$ и обозначим через $X_{k,n}$ число разговоров, которые начинаются между $n - kh$ и $n - (k-1)h$ и все еще продолжаются в момент t . Если n велико, то величина $X_{k,n}$ практически принимает только значения 0 и 1, причем последнее с вероятностью $\alpha h [1 - F(kh)]$. Поэтому математическое ожидание S_n представляет собой сумму этих вероятностей, и, переходя к пределу, мы выводим, что *число занятых линий имеет математическое ожидание*

$$E(X(t)) = \alpha \int_0^{\infty} [1 - F(s)] ds. \quad (3.5)$$

Отметим, что интеграл в (3.5) равен математическому ожиданию времени обслуживания. ►

Исторические замечания. Понятие безгранично делимости восходит к Б. де Финетти (1929). Преобразования Фурье безгранично делимых распределений с конечными дисперсиями были найдены А. Колмогоровым (1932). В 1934 г. П. Леви нашел преобразования Фурье произвольных безгранично делимых распределений. Им же безгранично делимые распределения изучались с точки зрения случайных процессов. Все последующие исследования происходили под сильным влиянием фундаментальной работы П. Леви. Первые чисто аналитические выводы общей формулы были даны в 1937 г. независимо Феллером и Хинчиным. Этими авторами было также доказано, что предельные распределения для нулевых схем серий безгранично делимы.

§ 4. ПРОЦЕССЫ С НЕЗАВИСИМЫМИ ПРИРАЩЕНИЯМИ

Безгранично делимые распределения тесно связаны со случайными процессами с независимыми приращениями. Под процессом с независимыми приращениями мы понимаем зависящее от непрерывного параметра t семейство случайных величин $X(t)$, такое, что для любого конечного множества значений параметра $t_1 < t_2 < \dots < t_n$ приращения $X(t_{k+1}) - X(t_k)$ взаимно независимы. На данном этапе нам не потребуются теория случайных процессов; мы просто допускаем, что если некоторое явление может быть описано вероятностным образом, то это описание приведет к безгранично делимым распределениям. Именно так мы уже рас-

считали некоторые процессы с независимыми приращениями в 1, гл. XVII, 1, и в примере гл. III, 8, а). Ограничимся для простоты рассмотрением числовых случайных величин $X(t)$, хотя теория переносится и на векторные величины.

Процесс имеет *стационарные приращения*, если распределение величины $X(s+t) - X(s)$ зависит лишь от длины t интервала и не зависит от s .

Разобьем интервал $s, s+t$ посредством $n+1$ равноотстоящих точек $s = t_0 < t_1 < \dots < t_n = s+t$ и положим $X_{k,n} = X(t_k) - X(t_{k-1})$. Приращение $X(s+t) - X(s)$ процесса со стационарными независимыми приращениями равно сумме n независимых одинаково распределенных случайных величин $X_{k,n}$ и, следовательно, имеет *безгранично делимое распределение*. Мы увидим, что верно и *обратное*. Именно, однопараметрическое семейство вероятностных распределений Q_t , определенное при $t > 0$, может служить распределением приращений $X(s+t) - X(s)$ процесса со стационарными независимыми приращениями тогда и только тогда, когда

$$Q_{s+t} = Q_s \star Q_t, \quad s, t > 0. \quad (4.1)$$

Семейство распределений, удовлетворяющее соотношению (4.1), называется *полугруппой* распределений (см. гл. IX, 2). Каждое безгранично делимое распределение может быть взято в качестве элемента Q_t ($t > 0$ произвольно) такой полугруппы.

Прежде чем переходить к нестационарному случаю, рассмотрим несколько типичных примеров.

Примеры. а) *Обобщенный пуассоновский процесс*. Пусть F — произвольное вероятностное распределение и $\alpha > 0$. Обобщенное пуассоновское распределение задается формулой

$$Q_t = e^{-\alpha t} \sum_{k=0}^{\infty} \frac{(\alpha t)^k}{k!} F^{k*}. \quad (4.2)$$

Нетрудно убедиться, что соотношение (4.1) здесь действительно выполняется. Пусть $X(t)$ — случайный процесс со стационарными независимыми приращениями и приращения $X(t) - X(0)$ имеют распределения Q_t . Когда распределение F сосредоточено в точке 1, этот процесс сводится к обычному пуассоновскому процессу и (4.2) можно записать в виде

$$P\{X(t) - X(0) = n\} = e^{-\alpha t} \frac{(\alpha t)^n}{n!}. \quad (4.3)$$

Общая модель (4.2) допускает следующую интерпретацию в терминах обычного пуассоновского процесса. Пусть $Y_1, Y_2, \dots$ — независимые случайные величины с одинаковым распределением F , и пусть $N(t)$ — обычный пуассоновский процесс: $P\{N(t) = n\} = e^{-\alpha t} (\alpha t)^n / n!$, не зависящий от Y_k . Распределение (4.2) совпадает

с распределением случайной суммы $Y_1 + \dots + Y_{N(t)}$. Иными словами, с n -скачком пуассоновского процесса связывается эффект Y_n , и приращение $X(t) - X(0)$ представляет собой сумму эффектов, соответствующих интервалу времени $0, t$. Рандомизированное случайное блуждание, изучавшееся в гл. II, 7, является обобщенным пуассоновским процессом, отвечающим тому случаю, когда величины Y_k принимают лишь значения ± 1 . Практические приложения будут проиллюстрированы в конце параграфа.

б) Броуновское движение или процесс Винера—Башелье. Так называется начинающийся в начале координат процесс (т. е. $X(0)=0$), приращения которого $X(s+t) - X(s)$ за непересекающиеся промежутки времени независимы и имеют нормальное распределение с нулевым математическим ожиданием и дисперсией t . Винером и Леви было показано, что траектории такого процесса с вероятностью 1 непрерывны¹⁾ и что это свойство выделяет нормальное распределение среди всех безгранично делимых распределений.

в) Устойчивые процессы. Соотношение (1.8) для строго устойчивых процессов равносильно соотношению (4.1) с $Q_t(x) = R(t^{-1/\alpha}x)$. Таким образом, распределение $R(t^{-1/\alpha}x)$ определяет переходные вероятности процесса со стационарными независимыми приращениями; при $\alpha=2$ получаем броуновское движение. ►

Основная теорема теории безгранично делимых распределений утверждает (см. гл. IX и XVII), что *самое общее решение уравнения (4.1) и, следовательно, самое общее безгранично делимое распределение можно представить как предел подходящей последовательности обобщенных пуассоновских распределений*. Результат этот довольно удивителен, если принять во внимание большую формальную разницу между примерами а) и б).

Распределение величины $X(t+s) - X(t)$ в случае процесса с нестационарными независимыми приращениями тоже представимо как распределение сумм по строкам в схеме серий $X_{k,n}$, однако теперь для выполнения (3.2) нужно потребовать выполнения некоторых условий непрерывности. Пример д) объясняет необходимость этого требования. Если наложить некоторые слабые ограничения, то *распределения величин $X(t+s) - X(t)$ будут необходимо безгранично делимы*.

Примеры. г) *Операционное время*. Простая замена времени нередко сводит общий процесс к более доступному для изучения процессу со стационарными приращениями. Если φ —произвольная возрастающая непрерывная функция, то замена $t \rightarrow \varphi(t)$ переводит процесс $X(t)$ в процесс $Y(t) = X(\varphi(t))$. Свойство независимости приращений, очевидно, сохраняется, и при подходящем выборе φ приращения нового процесса могут оказаться стационар-

¹⁾ Как всюду в этом томе, автор отвлекается от ограничений типа сепарабельности и т. п. — Прим. перев.

ными. На практике выбор функции f обычно подсказывается природой рассматриваемого процесса. Например, на *телефонной станции* вряд ли кто стал бы сравнивать ночной час с загруженным дневным часом — естественно измерять время в переменных единицах, таких, что ожидаемое число вызовов за единицу времени оставалось бы постоянным. Аналогично при растущем *страховом* бизнесе претензии будут поступать с возрастающей частотой, однако это отклонение от стационарности устраняется просто введением операционного времени для измерения частоты поступления претензий.

д) *Практические приложения*. Огромное количество практических задач можно свести к обобщенным пуассоновским процессам. Приведем несколько типичных примеров. (i) Накопленный *убыток* за счет автомобильных катастроф, пожаров, молний и т. п. По поводу применений к *теории страхования* см. пример 5, а). (ii) Общий улов рыболовецкого судна при поиске косяков рыбы (Дж. Нейман). (iii) Количество воды в *водохранилище* в зависимости от дождей и потребления воды. Аналогично рассматриваются другие задачи *хранения запасов*. (iv) *Камень, лежащий на дне реки*, находится без движения в течение столь долгих промежутков времени, что последовательные смещения допустимо считать практически мгновенными. Общее смещение камня за время $0, t$ можно рассматривать как обобщенный пуассоновский процесс. (Впервые этот вопрос изучался другими методами Альбертом Эйнштейном-младшим и Д. Пойа.) (v) *Телефонные вызовы*, или прибытия клиентов на пункт обслуживания. При соответствующих условиях общее время, необходимое для обслуживания требований, поступивших в интервале времени $0, t$, описывается обобщенным пуассоновским процессом. Замечательной особенностью этого процесса является тот факт, что значение $X(t)$ *не наблюдаемо в момент t* , поскольку оно определяется временами обслуживания, относящимися к будущему. (vi) Изменения энергии физических частиц при столкновениях рассмотрены в примере X, 1, б).

§ 5*). ОБОБЩЕННЫЕ ПУАССОНОВСКИЕ ПРОЦЕССЫ И ЗАДАЧИ О РАЗОРЕНИИ

Пусть $X(t)$ — обобщенный пуассоновский процесс. Это значит, что приращение $X(t+s) - X(s)$ за любой интервал времени длины t имеет распределение Q_t , задаваемое формулой (4.2). Пусть $c > 0$ и $z > 0$ фиксированы. *Разорением* мы называем событие

$$\{X(t) > z + ct\}. \quad (5.1)$$

*) Этот параграф посвящен специальной теме. Он имеет большой практический интерес, но материал его нигде не будет использоваться в дальнейшем, за исключением примеров, где эта тема изучается другими методами.

В дальнейшем s будет рассматриваться как постоянная, а $z > 0$ как свободный параметр. Обозначим через $R(z)$ вероятность того, что разорение никогда не произойдет. Предполагая, что вся задача имеет смысл, мы выведем формально, что вероятность $R(z)$ должна быть неубывающим решением функционального уравнения (5.2). Приведем сначала несколько примеров, показывающих степень разнообразия практических ситуаций, приводящих к рассматриваемой задаче.

Примеры. а) *Теория страхования*¹⁾. Значением $X(t)$ здесь является накопленное количество претензий, предъявленных страховой компании за интервал времени $[0, t]$. Предполагается, что поступление претензий описывается пуассоновским процессом и что отдельные претензии имеют распределение F . В принципе «претензии» могут быть как положительными, так и отрицательными (например, смерть клиента может освободить компанию от обязательства и увеличить ее резервы). На практике растущая развивающаяся нормально компания естественно измеряет время в операционных единицах, пропорционально полной сумме полученных страховых премий [см. пример 4, г)]. Поэтому можно предположить, что при отсутствии претензий резервы компании возрастают с постоянной скоростью c . Если z — начальный резерв компании в момент 0, то общий резерв компании в момент t представляется случайной величиной $z + ct - X(t)$, а «разорение» (крах компании) происходит, как только резерв становится отрицательным.

б) *Хранение запасов*. Идеальное водохранилище наполняется реками и дождями с постоянной скоростью c . В случайные моменты времени расходуется часть воды в количествах $X_1, X_2, \dots$. Рассматриваемая модель описывается обобщенным пуассоновским процессом, и если исходное количество воды в момент времени 0 равно z , то в момент t количество воды будет равно $z + ct - X(t)$ при условии, что водохранилище не осушится до этого времени. Задачам подобного рода посвящена огромная литература (см. монографии, указанные в конце книги).

в) *Обслуживание больных*²⁾. Будем рассматривать промежутки времени, которые врач тратит на обслуживание пациентов, как независимые случайные величины с показательным распределением

¹⁾ Этой теории (основанной Лундбергом) посвящена огромная литература. Обзор ее имеется в относительно недавней работе Г. Крамера (см.: Cramér H., On some questions connected with mathematical risk, Univ. Calif. Publications in Statistics, vol. 2, № 5 (1954), 99—125). Асимптотические оценки Крамера (полученные с помощью глубоких методов Винера — Хопфа) доказаны элементарными методами в примерах гл. XI, 7, а), и гл. XII, 5, г).

²⁾ Pyke R., The supremum and infimum of the Poisson process, *Ann. Math. Statist.*, 30 (1959), 568—576. Пайк рассматривает лишь чисто пуассоновский процесс, но зато получает более точные результаты (иными методами).

и средней длительностью α^{-1} . Когда обслуживание происходит без перерывов, уход обслуженных пациентов подчинен *обычному пуассоновскому процессу*. Пусть $X(t)$ — число уходов пациентов за время $0, t$. Предположим, что в момент 0 (начало работы) врача ожидают z пациентов и что последующие пациенты приходят в моменты времени $c^{-1}, 2c^{-1}, 3c^{-1}, \dots$. Врач не будет без дела, пока $X(t) \leq z + ct$. ►

Следующее формальное рассуждение приводит к уравнению, определяющему вероятность разорения $1 - R$. Пусть *первый* скачок траектории происходит в момент τ и имеет величину x . Чтобы разорение никогда не произошло, необходимо, чтобы $x \leq z + c\tau$ и чтобы при всех $t > \tau$ приращения $X(t) - x$ не превосходили $z - x + ct$. Поскольку приращения независимы, последнее из этих событий имеет вероятность, равную $R(z - x + c\tau)$. Интегрируя по всем возможным значениям τ и x , получаем

$$R(z) = \int_0^{\infty} \alpha e^{-\alpha\tau} d\tau \int_{-\infty}^{z+c\tau} R(z+c\tau-x) F\{dx\}. \quad (5.2)$$

Это и есть нужное нам уравнение, однако его можно упростить. Производя замену $s = z + c\tau$, имеем

$$R(z) = \frac{\alpha}{c} \int_z^{\infty} e^{-(\alpha/c)(s-z)} ds \int_{-\infty}^s R(s-x) F\{dx\}. \quad (5.3)$$

Из (5.3) следует, что функция R дифференцируема. Дифференцируя (5.3), получаем окончательное *интегро-дифференциальное уравнение*

$$R'(z) = \frac{\alpha}{c} R(z) - \frac{\alpha}{c} \int_{-\infty}^z R(z-x) F\{dx\}. \quad (5.4)$$

Отметим, что по определению $R(s) = 0$ при $s < 0$, и поэтому интеграл справа в (5.4) есть *свертка* $F \star R$. Мы вернемся к уравнению (5.4) в примерах 9, г); гл. XI, 7, а) и гл. XII, 5, г).

§ 6. ПРОЦЕССЫ ВОССТАНОВЛЕНИЯ

Основные понятия теории восстановления были введены в 1, гл. XIII в связи с рекуррентными событиями. Мы увидим, что введение непрерывного времени связано скорее с изменениями в обозначениях, чем с изменениями по существу. Важнейшее свойство рекуррентных событий состоит в том, что последовательные времена ожидания T_k являются взаимно независимыми случайными величинами с общим распределением F ; момент n -го осуществле-

ния события равен

$$S_n = T_1 + \dots + T_n. \quad (6.1)$$

Принимается также соглашение, что $S_n = 0$, и 0 считается моментом осуществления номер нуль.

Даже в вероятностных процессах, зависящих от непрерывного времени, нередко можно обнаружить одну или более последовательностей моментов времени, для которых имеет место (6.1). В таких случаях простыми методами удастся получить удивительно точные результаты. Аналитически мы имеем дело просто с суммами независимых положительных случайных величин, и единственным оправданием введения термина «процесс восстановления» является его частое использование в связи с другими процессами, а также то обстоятельство, что при применении этого термина молчаливо подразумевается использование мощного средства — уравнения восстановления¹⁾.

Определение 1. Последовательность случайных величин $\{S_n\}$ образует процесс восстановления, если величины S_n имеют вид (6.1), где T_k — взаимно независимые случайные величины с общим распределением F , удовлетворяющим условию²⁾ $F(0) = 0$.

Поскольку рассматриваемые случайные величины положительны, то математическое ожидание $\mu = E(T_k)$ имеет смысл, даже если соответствующий интеграл расходится (в этом случае $\mu = \infty$). Математическое ожидание μ называется *средним временем возвращения*. Как обычно в подобных обстоятельствах, для нашего исследования неважно, появляются ли величины T_k в некотором вероятностном процессе или последовательность $\{T_j\}$ сама определяет вероятностное пространство.

Во многих (но не во всех) приложениях величину T_j можно интерпретировать как «время ожидания», а суммы S_n являются тогда моментами *восстановления* (или *регенерации*).

Представляется интуитивно очевидным, что для любого конечного интервала $I = a, b$ число попадающих в I моментов восстановления S_n конечно с вероятностью единица и, следовательно, представляет собой вполне определенную величину N . Если событие $\{S_n \in I\}$ называть «успехом», то N есть общее число успехов в бесконечной последовательности испытаний с математическим ожиданием, равным

$$U\{I\} = \sum_{n=0}^{\infty} P\{S_n \in I\} = \sum_{n=0}^{\infty} F^n \star \{I\}. \quad (6.2)$$

¹⁾ Более сложное обобщение рекуррентных событий имеется в работе: Kingman J. F. C., The stochastic theory of regenerative events, *Zeit. Wahrscheinlichkeitstheorie*, 2, (1964), 180—224. Русский перевод см. сб. *Математика*, 11:2 (1967), стр. 106—152.

²⁾ Существование атома с массой $p < 1$ в нуле не должно иметь серьезных последствий,

Для изучения этой меры мы введем, как обычно, соответствующую ей функцию распределения:

$$U(x) = \sum_{n=0}^{\infty} F^{n*}(x). \quad (6.3)$$

Понятно, что при $x < 0$ $U(x) = 0$, но в нуле мера U имеет атом единичной массы.

В дискретном случае, который был рассмотрен в 1, XIII, мера U была сосредоточена на множестве целых чисел: u_k обозначало вероятность того, что одна из сумм S_n равна k . Так как это событие могло появиться только один раз, то u_k можно было также интерпретировать как ожидаемое число сумм S_n , равных k . В рассматриваемой ситуации $U\{I\}$ должно интерпретироваться как математическое ожидание, а не как вероятность, так как событие $\{S_n \in I\}$ может произойти при многих n .

Докажем, что $U(x) < \infty$. Из определения сверток распределений, сосредоточенных на $\overline{0, \infty}$, следует, что $F^{n*}(x) \leq F^n(x)$ и поэтому ряд в (6.3) сходится со скоростью геометрической прогрессии при всех x , при которых $F(x) < 1$. Остается случай распределений, сосредоточенных на конечном интервале, но в этом случае существует целое число r , такое, что $F^{r*}(x) < 1$. Члены ряда с номерами $n = r, 2r, 3r, \dots$ образуют сами ряд, который сходится, и так как члены ряда (6.3) монотонно зависят от n , то отсюда следует сходимость всего ряда (6.3).

Как и в дискретном случае, мера U тесно связана с *уравнением восстановления*

$$Z = z + F \star Z. \quad (6.4)$$

Это уравнение является сокращенной записью следующего уравнения:

$$Z(x) = z(x) + \int_0^x Z(x-y) F\{dy\}, \quad x > 0, \quad (6.5)$$

где интервал интегрирования замкнут. На самом деле пределы интегрирования можно заменить на $-\infty$ и ∞ при условии, конечно, что $z(x) = Z(x) = 0$ при $x < 0$. Мы будем придерживаться этого соглашения.

Следующая теорема содержит самый существенный факт, касающийся уравнения восстановления.

Теорема 1. Если функция z ограничена и обращается в нуль при $x < 0$, то свертка $Z = U \star z$, определяемая равенством

$$Z(x) = \int_0^x z(x-y) U\{dy\}, \quad (6.6)$$

является решением уравнения восстановления (6.5). Это решение

единственно в классе функций, обращающихся в нуль на $-\infty, 0$ и ограниченных на конечных интервалах.

Доказательство. Мы уже знаем, что определяющий U ряд (6.3) сходится при всех x . Вычисляя свертки обеих частей равенства (6.3) с z , убеждаемся в том, что Z ограничена на конечных интервалах и удовлетворяет уравнению восстановления. Разность V двух таких решений должна удовлетворять уравнению $V = F \star V$ и, следовательно, уравнениям $\dot{V} = F^{n*} \star V$ при всех n . Но $F^{n*}(x) \rightarrow 0$ при всех x , и если V ограничена, то отсюда вытекает, что $V(x) = 0$ при всех x . ►

Мы возвратимся к уравнению восстановления в гл. XI, где будут изучены асимптотические свойства U и Z . (Обобщенный вариант уравнения восстановления см. в § 10.)

Нужно отметить, что функция U сама удовлетворяет уравнению

$$U(x) = 1 + \int_0^x U(x-y) F\{dy\}, \quad x > 0, \quad (6.7)$$

которое представляет собой частный случай при $z=1$ уравнения восстановления (6.5). Это можно показать непосредственно с помощью вероятностного рассуждения, которое часто используется в теории восстановления¹⁾. Поскольку 0 считается моментом восстановления, ожидаемое число моментов восстановления в замкнутом интервале $\overline{0, x}$ равно единице плюс ожидаемое число таких моментов в полуоткрытом интервале $\overline{0, x}$. Интервал $\overline{0, x}$ содержит моменты восстановления, лишь когда $T_1 \leq x$; при условии, что $T_1 = y \leq x$, ожидаемое число моментов восстановления в интервале $\overline{0, x}$ равно $U(x-y)$. Интегрируя по y , получаем уравнение (6.7).

Полезны два простых обобщения процесса восстановления. Сначала по аналогии с недоверными рекуррентными событиями мы можем рассмотреть *несобственные распределения*. Дефект $q = 1 - F(\infty)$ интерпретируется при этом как вероятность обрыва. Выражаясь абстрактно, можно сказать, что к действительной прямой добавляется одна точка Ω , называемая «смертью», и T_k равно либо положительному числу, либо Ω . Для удобства ссылок введем теперь следующее неформальное

Определение 2²⁾. Процесс, отличающийся от обычного процесса восстановления лишь тем, что функция распределения F несоб-

¹⁾ Автор называет это рассуждение «renewal argument». — Прим. перев.

²⁾ В примере 7, е) имеется иллюстрация этого определения, а в задаче 4 — его обобщение.

ственная, называется обрывающимся или невозвратным процессом восстановления. Дефект $q = 1 - F(\infty)$ интерпретируется как вероятность обрыва.

Ради согласованности 0 считается моментом восстановления номер нуль. Вероятность того, что процесс переживает n -й момент восстановления, равна $(1 - q)^n$ и стремится к 0 при $n \rightarrow \infty$. Таким образом, с вероятностью единица обрывающийся процесс обрывается за конечное время. Полная масса распределения F^{n*} есть $(1 - q)^n$, и поэтому математическое ожидание числа моментов возвращения равно $U(\infty) = q^{-1} < \infty$. Число $U(\infty)$ можно назвать, таким образом, ожидаемым числом возобновлений процесса. Вероятность того, что $S_n \leq x$ и процесс обрывается на n -м моменте восстановления, равна $qF^{n*}(x)$. Нами доказана, таким образом,

Теорема 2. Для обрывающегося процесса восстановления функция qU является собственным вероятностным распределением длительности жизни процесса (возраста в момент смерти).

Второе обобщение соответствует рекуррентным событиям с запаздыванием и состоит в том, что исходному времени ожидания разрешается иметь другое распределение. В этом случае добавляется величина T_j с $j = 0$ и $S_0 = T_0 \neq 0$.

Определение 3. Процессом восстановления с запаздыванием называется последовательность $S_0, S_1, \dots$ вида (6.1), где T_k — взаимно независимые строго положительные (собственные или не-собственные) случайные величины, причем $T_1, T_2, \dots$ (но не T_0) имеют одинаковое распределение.

§ 7. ПРИМЕРЫ И ЗАДАЧИ

От дискретного случая легко перейти к изучению примеров процессов восстановления, связанных с самовосстанавливающимися агрегатами, счетчиками элементарных частиц, ростом популяций и т. д. Ниже рассматривается одна специальная задача, приводящая, однако, к интересным общим выводам.

Пример. а) Парадокс инспекции. При теоретическом рассмотрении самовосстанавливающихся агрегатов предполагают, что какая-либо часть оборудования, например электрическая батарея, работает после установки до момента выхода из строя. После этого она мгновенно заменяется другой подобной батареей и таким образом процесс продолжается без перерыва. Моменты замены образуют процесс восстановления, в котором T_k — продолжительность службы (lifetime) k -й батареи.

Допустим теперь, что с помощью выборочного обследования необходимо оценить истинное распределение времени службы батарей. Мы берем выборку из совокупности батарей, работающих в какой-либо данный момент времени $t > 0$ и наблю-

даем выбранные батареи до конца их службы. Пусть F — общее распределение вероятностей времени службы каждой из батарей. Можно было бы ожидать, что это же распределение имеет и (остающееся) время службы выбранных нами батарей. Но это не так. Если F — показательное распределение, то описываемая ситуация лишь словесно отличается от рассмотренной в гл. I, 4 (парадокс времени ожидания). Соответственно время службы обследуемых единиц имеет распределение, отличающееся от F . Тот факт, что исследуемая часть оборудования начинает наблюдаться с момента t , изменяет распределение срока ее службы по сравнению с «безусловным» (в частности, удваивает среднюю продолжительность службы). Мы увидим в гл. XI, (4.6), что это обстоятельство типично для всех процессов восстановления. Из сказанного вытекают серьезные для приложений следствия. Беспристрастный на первый взгляд план контроля может вести к ложным заключениям, поскольку то, что в действительности наблюдается, может не быть типичным для всей популяции. Будучи замеченным, этот феномен легко получил свое объяснение (см. гл. I, 4), но он тем не менее указывает на возможные опасности при выведении следствий из наблюдений и на необходимость взаимосвязи теории с практикой. Отметим, что никаких трудностей не возникнет, если наблюдения производить над прибором, заменяющим первый вышедший из строя после момента t . ►

Введем три случайные величины, важные для теории восстановления. В рассмотренном выше примере все эти три величины относятся к сроку службы участвующего в процессе прибора в момент $t > 0$ и могут быть описаны следующими, не нуждающимися в объяснении терминами: остающееся время службы, прошедшее время службы и полное время службы. Дадим теперь формальное определение.

Каждому $t > 0$ поставим в соответствие однозначно определяемый (случайный) индекс N_t , такой, что $S_{N_t} \leq t < S_{N_t+1}$. Тогда

а) *остающееся время ожидания* по определению равно $S_{N_t+1} - t$, т. е. равно времени, протекающему от t до следующего за t момента восстановления;

б) *прошедшее время ожидания* по определению равно $t - S_{N_t}$, т. е. равно времени, протекающему до t от первого момента восстановления, предшествующего t ;

в) сумма остающегося и прошедшего времен ожидания, равная $S_{N_t+1} - S_{N_t} = T_{N_t+1}$, есть длина интервала возвращения, накрывающего t .

Введенную терминологию нельзя считать установившейся — она меняется в зависимости от контекста. Например, в теории случайных блужданий остающееся время ожидания называют *точкой первого достижения* интервала t, ∞ . В рассмотренном выше при-

мере время ожидания называлось сроком службы. Все три введенные случайные величины будут изучаться в гл. XI, 4, и гл. XIV, 3.

Пуассоновский процесс был определен как процесс восстановления с показательным распределением промежутков времени ожидания T_j . Во многих задачах массового обслуживания естественно считать, что входящий поток образует пуассоновский процесс. В некоторых других процессах промежутки времени между поступлениями постоянны. Для объединения этих двух случаев в теории массового обслуживания стало модным рассматривать общий процесс восстановления с произвольными промежутками времени между поступлениями¹⁾.

Перейдем теперь к некоторым задачам довольно общего характера, связанным с процессами восстановления. Соответствующее процессу распределение вероятностей по-прежнему будем обозначать буквой F .

Начнем с того, что можно грубо описать как *«время ожидания W до большого пропуска»*. Процесс восстановления с временами ожидания T_j будем останавливать при первом появлении интервала времени длины ξ , в течение которого не происходит восстановлений. Выведем уравнение восстановления для функции распределения V времени ожидания W . Так как время ожидания W превышает ξ , то $V(t) = 0$ при $t < \xi$. При $t \geq \xi$ рассмотрим две взаимно исключающие возможности: либо $T_1 > \xi$, либо $T_1 = y \leq \xi$. В первом случае $W = \xi$. Во втором случае процесс начинается сначала, и условная вероятность того, что $\{W \leq t\}$ при условии $T_1 = y$, равна $V(t - y)$. Суммируя по всем имеющимся возможностям, получаем

$$V(t) = 1 - F(\xi) + \int_0^{\xi} V(t - y) F(dy), \quad t \geq \xi; \quad (7.1)$$

как было отмечено выше, $V(t) = 0$ при $t < \xi$. Это уравнение сводится к обычному уравнению восстановления

$$V = z + G \star V \quad (7.2)$$

с несобственной функцией распределения G и функцией z , определяемыми равенствами

$$G(x) = \begin{cases} F(x) & \text{при } x \leq \xi, \\ F(\xi) & \text{при } x \geq \xi; \end{cases} \quad (7.3)$$

$$z(x) = \begin{cases} 0 & \text{при } x < \xi, \\ 1 - F(\xi) & \text{при } x \geq \xi. \end{cases} \quad (7.4)$$

¹⁾ Общность эта в известной степени обманчива, поскольку трудно найти соответствующие практические примеры, за исключением примера автобуса, идущего без расписания по круговому маршруту. Иллюзия общности несколько умалется тем печальным фактом, что непуассоновский вход обычно не является также и марковским.

Наиболее важным частным случаем является случай *пропусков в пуассоновском процессе*, т. е. случай, когда $\bar{F}(t) = 1 - e^{-ct}$. Функция V здесь связана с теоремами о покрытии из гл. I, 9 [см. задачу 15 и пример гл. XIV, 2, а). О другом подходе см. в задаче 16.]

Примеры прикладного характера. б) *Пересечение потока автомобилей*¹⁾. Пусть автомобили движутся в один ряд с постоянной скоростью, причем последовательные пересечения автомобилями перехода (или переезда) образуют выборку из пуассоновского процесса (или из какого-либо другого процесса восстановления). Пешеход, появившийся на тротуаре (или автомобиль, подъехавший к перекрестку), начинает переход (переезд) после того, как убеждается, что в течение следующих ξ секунд времени, необходимых для перехода (переезда), не пройдет ни одного автомобиля. Обозначим через W время, необходимое для перехода (переезда), т. е. время ожидания на тротуаре (перекрестке) плюс ξ . Функция распределения V величины W удовлетворяет уравнению (7.1) с $F(t) = 1 - e^{-ct}$ [продолжение см. в примерах гл. XI, 7, б) и XIV, 2, а)].

в) *Счетчики Гейгера типа II*. Частицы попадают в счетчик в моменты, образующие пуассоновский процесс, и каждая из этих частиц (зарегистрированная или нет) делает счетчик нечувствительным на фиксированное время ξ . Если частица зарегистрирована, то он остается нечувствительным (блокированным) и его чувствительность восстанавливается только тогда, когда промежуток времени между какими-либо следующими друг за другом частицами впервые превзойдет ξ . Наша теория применима к распределению V длительности периода «нечувствительности» [см. I, XIII, 11, задача 14].

г) *Максимальный промежуток между возвращениями*. Рассмотрим процесс восстановления; пусть Z_t обозначает максимальную длину промежутков времени T_j , наблюдаемых²⁾ до момента t . Событие $\{Z_t \leq \xi\}$ происходит тогда и только тогда, когда до момента t каждый интервал времени длины ξ содержит момент восстановления, и поэтому в наших обозначениях $P\{Z_t > \xi\} = V(t)$. ►

Значительное число встречающихся в приложениях процессов восстановления могут быть охарактеризованы как *альтернирующие процессы* или *процессы с двумя фазами*. В зависимости от контекста эти две фазы могут быть названы фазами «активности» и

¹⁾ С относительно старой литературой и другими вариантами этой задачи (изучаемыми иными методами) можно познакомиться по работе Таппер J. C., *The delay to pedestrians crossing a road*, *Biometrika*, 38 (1951), 383—392.

²⁾ Более точно, пусть n — (случайный) индекс, такой, что $S_{n-1} \leq t < S_n$. Тогда $Z_t = \max [T_1, \dots, T_{n-1}, \xi]$. Случайные величины этого типа систематически изучались А. Ламперти,

«пассивности», «чувствительности» и «нечувствительности», «возбуждения» и «покоя». Активные и пассивные фазы процесса чередуются; их продолжительности предполагаются независимыми случайными величинами с общим распределением для каждой фазы.

Примеры. д) *Поломки с последующими простоями.* В качестве простейшего примера рассмотрим последовательные замены выходящей из строя детали в предположении, что каждая поломка сопровождается некоторым простоем (его можно интерпретировать как время, необходимое для обнаружения поломки, или же как время, нужное для ремонта). Последовательные периоды работы $T_1, T_2, \dots$ чередуются с последовательными «мертвыми» периодами $Y_1, Y_2, \dots$, и мы получаем собственный обычный процесс восстановления с временами возвращения $T_j + Y_j$. Этот процесс можно также рассматривать как процесс восстановления с запаздыванием с моментом первого восстановления, равным T_1 , и временами возвращения $Y_j + T_{j+1}$.

е) *Потерянные вызовы.* Рассмотрим одну телефонную линию; пусть поступающие на нее вызовы образуют пуассоновский процесс с распределением времен между вызовами, равным $G(t) = 1 - e^{-ct}$. Предположим, что длительности следующих за вызовами разговоров являются независимыми случайными величинами с общим распределением F . Линия может быть свободной или занятой, и вызовы, поступающие в течение периодов занятости теряются и не влияют на процесс. Мы имеем здесь процесс с двумя фазами с распределением времени возвращения, равным $F \star G$. (См. задачу 17, а также задачи 3, 4 из гл. XIV, 10.)

ж) *Прибывший последним обслуживается первым.* Иногда распределения различных времен ожидания априори неизвестны, но могут быть вычислены по другим данным. В качестве примера рассмотрим устройство для обработки информации, в которое новая информация поступает в соответствии с пуассоновским процессом, так что свободные периоды имеют показательное распределение. Время, которое требуется для обработки новой информации, поступающей в любой момент, имеет распределение G .

Периоды занятости и свободные периоды чередуются, но длительность периодов занятости зависит от того, как обрабатывается информация, поступающая в течение периода занятости. Бывают ситуации, когда лишь последняя информация представляет интерес. В таких случаях вновь поступившая информация сразу же начинает обрабатываться, а вся предыдущая информация отбрасывается. Распределение V продолжительности периодов занятости можно вычислить по уравнению восстановления (см. задачу 18).

и) *Счетчики Гейгера.* В счетчиках типа I за каждой регистрацией следует период нечувствительности фиксированной длины ξ . Попадания частиц в этот период никак не отмечаются. Процесс

здесь тот же, что и в примере д), причем величины T_j имеют показательное распределение, а величины Y_j равны ξ . В счетчиках типа II нерегистрируемые частицы тоже приводят к блокировке и ситуация аналогична предыдущей, с тем исключением, что распределения величин Y_j зависят от исходного процесса и должны вычисляться по уравнению восстановления (7.1) [пример в)]. ▶

§ 8. СЛУЧАЙНЫЕ БЛУЖДЕНИЯ

Пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины с одинаковым распределением F . Обозначим, как обычно,

$$S_0 = 0, \quad S_n = X_1 + \dots + X_n. \quad (8.1)$$

Будем говорить, что S_n есть положение в момент n частицы, совершающей случайное блуждание общего вида. Никаких новых теоретических идей в этом параграфе не появляется¹⁾ — просто будет введена терминология для краткого и наглядного описания процесса $\{S_n\}$. Например, для всякого интервала I (или какого-либо другого множества) событие $\{S_n \in I\}$ называется *попаданием* в I , и изучение последовательных попаданий в заданный интервал I выявляет важные характеристики флуктуаций последовательности $S_1, S_2, \dots$. Индекс n интерпретируется как временной параметр, и мы будем говорить о «моменте n ». В этом параграфе некоторые удивительные особенности случайных блужданий описываются в терминах последовательных рекордных значений. Полезность полученных результатов будет видна из их применений в § 9. Другой (независимый) подход будет рассмотрен в § 10.

«Вложенные» процессы восстановления

Скажем, что в момент $n > 0$ достигается рекордное значение, если

$$S_n > S_j, \quad j = 0, 1, \dots, n-1. \quad (8.2)$$

Такие индексы n могут существовать не для всех траекторий; если они существуют, то образуют конечную или бесконечную упорядоченную последовательность. Поэтому можно говорить о первом, втором и т. д. осуществлении (8.2). Моменты этих осуществлений являются (возможно, несобственными) случайными величинами. Теперь мы в состоянии ввести важные случайные величины, на использовании которых будет основываться значительная часть исследования случайных блужданий.

¹⁾ Выборочные пространства бесконечных случайных блужданий уже рассматривались в первом томе, однако там нам приходилось определять такие понятия, как «вероятность разорения», с помощью очевидных предельных переходов. Теперь эти очевидные переходы к пределу обоснованы теорией меры (см. гл. IV, 6).

Определение. Назовем k -м (верхним) лестничным индексом момент k -го осуществления неравенств (8.2). Значение S_n в k -й лестничный момент будем называть k -й лестничной высотой. (Обе введенные случайные величины, возможно, являются несобственными.)

Нижние лестничные случайные величины определяются аналогично с помощью неравенств, обратных неравенствам (8.2)¹⁾.

Термин «верхний» обычно будет излишним и будет использоваться лишь при желании особо это подчеркнуть или достигнуть большей ясности.

Лестничные точки на графике траектории $(S_0, S_1, \dots)$ — это те точки, в которых график доходит до недостигнувшейся ранее высоты (рекордное значение). На рис. 1 изображена траектория уходящего к минус бесконечности случайного блуждания $\{S_n\}$ с последним положительным положением при $n=31$. Здесь 5 верхних и 18 нижних лестничных точек этой траектории обозначены $\bullet$ и $\circ$ соответственно. Случайное блуждание с распределением Коши изображено на рис. 2 (см. стр. 227).

Пример. а) В «обычном» случайном блуждании F имеет атомы 1 и -1 с массами p и q . Верхние лестничные величины при $q > p$ будут несобственными, причем дефект равен $1 - p/q$ [см. 1; гл. XI, (3.9)]. Здесь k -я лестничная высота равна k и потому не упоминалась в первом томе; k -й лестничный индекс равен моменту первого достижения точки k . Распределение k -го лестничного индекса было найдено в 1; гл. XI, 4, е), а для частного случая $p=1/2$ — в теореме 2 в 1; гл. III, 4. ►

Первый лестничный индекс $\mathcal{T}_1$ равен моменту первого попадания в интервал $0, \infty$, а первая лестничная высота $\mathcal{H}_1$ равна $S_{\mathcal{T}_1}$. Продолжение случайного блуждания после момента $\mathcal{T}_1$ является вероятностной копией всего случайного блуждания. При условии, что $\mathcal{T}_1 = n$, появление второго лестничного индекса a в момент $k > n$ зависит только от $X_{n+1}, \dots, X_k$, и поэтому число испытаний между первым и вторым лестничными индексами представляет собой случайную величину $\mathcal{T}_2$, не зависящую от $\mathcal{T}_1$ и имеющую одинаковое с $\mathcal{T}_1$ распределение. Аналогичное рассуждение приводит к следующему общему результату: k -й лестничный индекс и k -ю лестничную высоту можно представить в виде

$$\mathcal{T}_1 + \dots + \mathcal{T}_k, \quad \mathcal{H}_1 + \dots + \mathcal{H}_k,$$

где $\mathcal{T}_j$ и $\mathcal{H}_j$ — взаимно независимые одинаково распределенные слу-

¹⁾ Заменяя в (8.2) строгие неравенства на неравенства $\geq$ и $\leq$, получаем слабые лестничные индексы. Это вызывающее некоторые осложнения различие несущественно, когда распределение F непрерывно. На рис. 1 слабые лестничные точки отмечены буквой w .

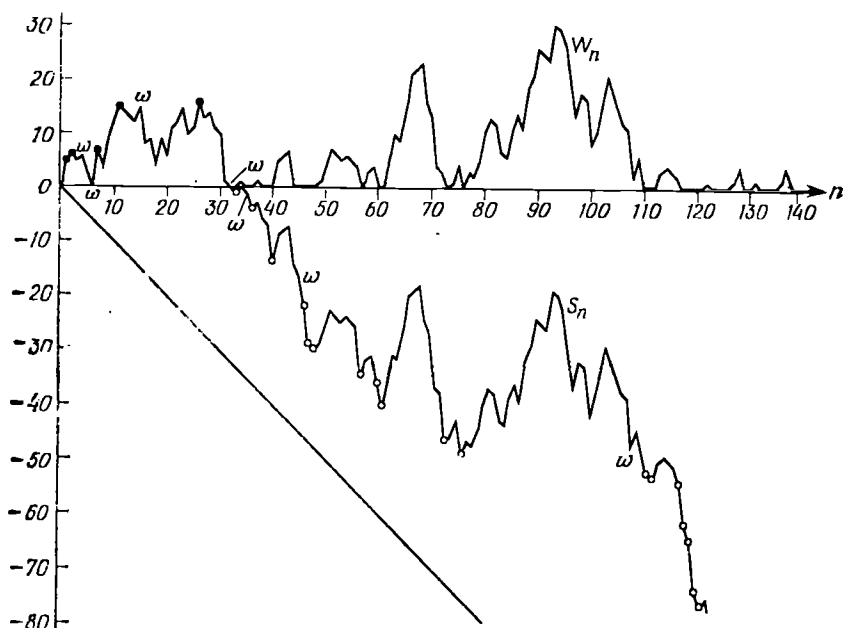


Рис. 1. Случайное блуждание и связанный с ним процесс ожидания. Случайные величины X_n в случайном блуждании $\{S_n\}$ имеют математическое ожидание —1 и дисперсию 16. Верхние и нижние лестничные точки обозначены ω и o соответственно. Пятая лестничная точка, имеющая координаты (26, 16), с большой вероятностью представляет максимум всего случайного блуждания.

[Буква ω указывает те точки, где рекордные значения достигаются второй или третий раз; эти точки являются *слабыми* лестничными точками, определяемыми соотношениями (8.2) с заменой строгих неравенств на нестрогие $\geq$.]

Всюду на графике S_n превосходит свое математическое ожидание, равное $-n$. В действительности *первым индексом, для которого $S_n \leq -n$, является $n = 135$* (при этом $S_{135} = -137$). Отмеченное обстоятельство согласуется с тем фактом, что математическое ожидание такого n равно бесконечности.

Случайные величины X_n имеют вид $X_n = B_n - A_n$, где B_n и A_n — взаимно независимые случайные величины, распределенные равномерно на множествах 1, 3, 5, 7, 9 и 2, 4, 6, 8, 10 соответственно. В примере 9,а) величины W_n представляют собой полное время ожидания n -го клиента, когда время между поступлениями требований принимает с одинаковой вероятностью значения 2, 4, 6, 8, 10, а время обслуживания равно каждому из чисел 1, 3, 5, 7 или 9 с вероятностью $1/5$. Распределение величины X_n имеет массы $(5-k)/25$ в точках $\pm 2k-1$, где $k=0, 1, 2, 3, 4$.

чайные величины. Иными словами, лестничные индексы и высоты образуют (возможно, обрывающийся) процесс восстановления.

Интуитивно представляется очевидным, что в случае обрывающихся процессов суммы S_n уходят к $-\infty$ и с вероятностью единица достигается максимум последовательности $\{S_n\}$. В следующем параграфе будет обнаружено, что лестничные случайные величины

являются мощным аппаратом при исследовании одного класса процессов, имеющего значительный практический интерес.

Пример. б) Точные выражения. Пусть F имеет плотность

$$f(x) = \begin{cases} \frac{ab}{a+b} e^{ax}, & x < 0, \\ \frac{ab}{a+b} e^{-bx}, & x > 0. \end{cases} \quad (8.3)$$

Такое случайное блуждание обладает тем редким свойством, что

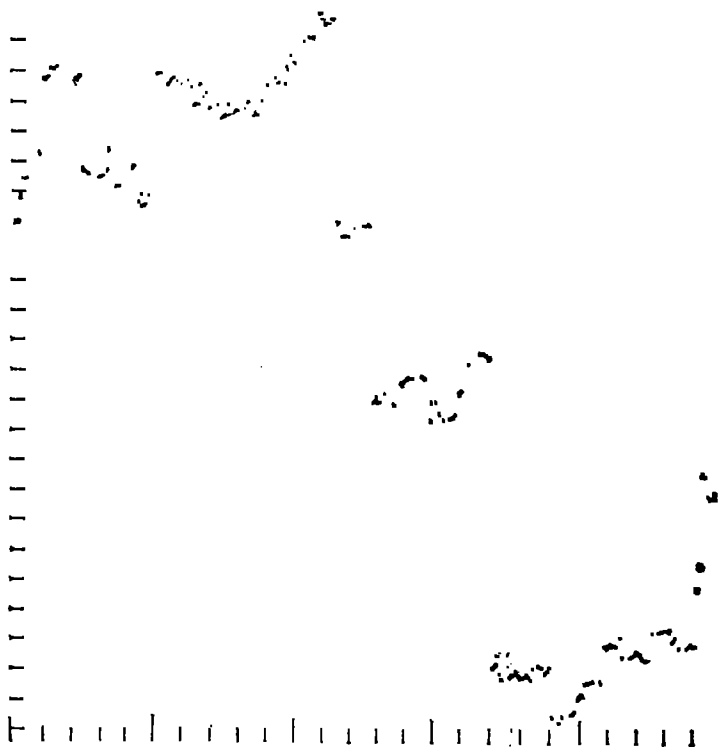


Рис. 2. Случайное блуждание, порожденное распределением Коши. (Распределение было урезано, чтобы исключить большие скачки.)

все связанные с ним распределения могут быть точно вычислены. Это блуждание играет большую роль в теории массового обслуживания, поскольку функция f есть свертка двух экспоненциальных плотностей, сосредоточенных на $0, \infty$ и $-\infty, 0$ соответственно. Иными словами, величины X_j могут быть представлены в виде разностей $X_j = \mathcal{B}_j - \mathcal{A}_j$ двух независимых положительных

показательно распределенных случайных величин. Не ограничивая общности, можем считать, что $a \leq b$.

Верхняя лестничная высота $\mathcal{H}_1$ имеет плотность ae^{-bx} ; величина $\mathcal{H}_1$ будет несобственной, и дефект ее равен $(b-a)/b$. Верхний лестничный момент $\mathcal{F}_1$ имеет производящую функцию $b^{-1}p(s)$, где

$$2p(s) = a + b - \sqrt{(a+b)^2 - 4abs}. \quad (8.4)$$

Дефект этой случайной величины тоже равен $(b-a)/b$.

Нижняя лестничная высота $\mathcal{H}_1^-$ при $x < 0$ имеет плотность ae^{ax} , и производящая функция нижнего лестничного момента $\mathcal{F}_1^-$ равна $a^{-1}p(s)$. Эта производящая функция в частном случае $a=b$ сводится к производящей функции $1 - \sqrt{1-s}$, которая известна из теории обычного случайного блуждания (или бросания монеты). [Доказательства и другие результаты см. в гл. XII, 4, 5, и XVIII, 3. См. также пример 4, д.)]

§9. ПРОЦЕССЫ МАССОВОГО ОБСЛУЖИВАНИЯ

Имеется чрезвычайно обширная литература¹⁾, посвященная различным вопросам, связанным с обслуживанием, управлением запасами, временем ожидания и т. п. Немало уже сделано для построения общей теории, однако наличие очень большого числа мало отличающихся друг от друга вариантов задач затемняет ситуацию, так что трудно разглядеть лес за деревьями. Мощь новых общих методов до сих пор недооценивается. Начнем с формального введения случайного процесса, определяемого с помощью

¹⁾ Литературные ссылки можно найти в соответствующих монографиях, указанных в библиографии. Было бы трудно дать короткий очерк развития теории массового обслуживания со справедливым указанием авторства. Наиболее выдающиеся работы, в которых были заложены новые методы, считаются теперь устаревшими в силу того прогресса, который они стимулировали. [Примером может служить интегральное уравнение Линдли в теории массового обслуживания (1952).] Другие работы заслуживают внимания благодаря исследованным в них (иногда очень сложным) частным задачам, однако нет возможности включить их в систематический обзор общей теории. В целом обширная литература в нескольких областях посвящена изучению различных вариантов и примеров, в то время как единой теории нет. Установление авторства затруднено также тем, что многие результаты были получены независимо разными авторами. Например, решение некоторого интегрального уравнения появилось в 1939 году в одной диссертации в Стокгольме со ссылкой на неопубликованные лекции, читанные Феллером в 1934 году. Это решение теперь известно в сочетании со многими именами. По истории развития теории см. обзорные статьи Д. Г. Кендалла, которые представляют и самостоятельный интерес: *Some problems in the theory of queues*; *Some problems in the theory of dams*, J. Roy. Statist. Soc., Series B, 13 (1951), 151—185, 19 (1957), 207—233; *Some recent work and further problems in the theory of queues*. Теория вероятностей и ее применения. IX, 1, 1964, 3—15. Внимание к общим методам отличает монографии А. А. Боровкова по теории массового обслуживания.

рекуррентной схемы, которая представляется на первый взгляд довольно искусственной. Дальнейшие примеры иллюстрируют широкую применимость этой схемы; впоследствии мы увидим, что весьма точные результаты можно получить, используя удивительно простые методы (см. гл. XII, 5).

Определение 1. Пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины с одинаковым (собственным) распределением F . Процессом ожидания, индуцированным последовательностью $\{X_k\}$, назовем последовательность случайных величин $W_0, W_1, \dots$, определяемую рекуррентно следующим образом: положим $W_0 = 0$ и

$$W_{n+1} = \begin{cases} W_n + X_{n+1} & \text{при } W_n + X_{n+1} \geq 0, \\ 0 & \text{при } W_n + X_{n+1} \leq 0. \end{cases} \quad (9.1)$$

В сокращенной записи (9.1) означает, что $W_{n+1} = (W_n + X_{n+1}) \cup 0$. Рис. 1 иллюстрирует это определение (См. стр. 226).

Примеры. а) Система с одним прибором. Пусть на «прибор» поступают «требования», причем поступления требований образуют собственный процесс восстановления с промежутками времени между поступлениями¹⁾, равными $A_1, A_2, \dots$ (требования нумеруются числами $0, 1, 2, \dots$ и поступают в моменты $0, A_1, A_1 + A_2, \dots$). Предположим, что на n -е требование тратится время обслуживания B_n . Будем считать, что случайные величины B_n не зависят от моментов поступлений требований и друг от друга и имеют одинаковое распределение. Прибор может быть либо «свободным», либо «занятым», причем в начальный момент 0 он свободен. Схема обслуживания требований такова. Если требование поступает в момент, когда прибор свободен, то оно сразу же начинает обслуживаться. В противном случае оно становится в очередь, и прибор непрерывно обслуживает требования в порядке их поступления²⁾ до тех пор, пока очередь не исчезнет и прибор не станет «свободным». Под длиной очереди мы понимаем число поступивших необслуженных требований (включая требование, которое обслуживается в данный момент). Временем обслуживания W_n n -го требования называется промежуток времени от его

¹⁾ Как правило, мы будем считать, что промежутки времени между поступлениями требований либо постоянны, либо имеют показательное распределение, однако, следуя современной моде, можно рассмотреть и произвольный процесс восстановления; см. сноску на стр. 221.

²⁾ Этот «порядок очереди» совершенно несуществен для длины очереди, длительности периодов занятости и аналогичных характеристик. Тот или иной порядок очереди имеет значение лишь с точки зрения обслуживания отдельных требований. Порядки очереди могут быть различными; «крайними случаями» являются следующие: требования обслуживаются в порядке их поступления, первым обслуживается последнее поступившее требование, обслуживается случайно выбранное из очереди требование. Вся картина меняется, если допустить возможность отказов.

поступления до начала его обслуживания. Общее время, затраченное на обслуживание n -го требования, равно $W_n + B_n$ (например, если первые времена обслуживания суть 4, 4, 1, 3, ..., а времена между поступлениями требований суть 2, 3, 2, 3, то требования с номерами 1, 2, ... становятся в очередь длины 1, 1, 2, 1 соответственно и отвечающие им номера ожидания равны 2, 3, 2, 2, ...).

Чтобы избежать тривиальных сложностей типа поступления одного требования в момент окончания обслуживания другого, мы предположим, что распределения A и B случайных величин A_n и B_n непрерывны. В этом случае длина очереди в любой момент вполне определена.

Построим схему последовательного вычисления времен ожидания W_n . По определению нулевое требование поступает в момент 0 на свободный прибор, и поэтому его время ожидания W_0 равно нулю. Предположим теперь, что n -е требование поступает в момент t и что время его ожидания W_n известно. Обслуживание n -го требования начинается в момент $t + W_n$ и завершается в момент $t + W_n + B_n$. Следующее $(n+1)$ -е требование поступает в момент $t + A_{n+1}$. В этот момент прибор свободен, если $W_n + B_n < A_{n+1}$, а в противном случае время ожидания $(n+1)$ -го требования равно $W_{n+1} = W_n + B_n - A_{n+1}$. Таким образом, *последовательность времен ожидания $\{W_n\}$ совпадает с процессом ожидания, индуцированным независимыми случайными величинами*

$$X_n = B_{n-1} - A_n, \quad n = 1, 2, \dots \quad (9.2)$$

б) *Управление запасами.* Для наглядности мы рассмотрим водохранилище, однако используемая модель применима и в других ситуациях, связанных с управлением запасами. Количество воды в водохранилище зависит от ее притока и от ее использования. Приток происходит за счет рек и дождей, а использование определяется поступлением требований (требования, конечно, удовлетворяются лишь в том случае, когда водохранилище не пусто).

Рассмотрим теперь количество воды¹⁾ $0, W_1, W_2, \dots$ в водохранилище в моменты $0, \tau_1, \tau_2, \dots$. Пусть X_n есть разность между величиной действительного притока и величиной теоретического (идеального) требования, поступившего в интервал времени τ_{n-1}, τ_n . Будем считать, что все изменения происходят мгновенно и возможны лишь в моменты времени $\tau_1, \tau_2, \dots$. В момент 0 имеем $W_0 = 0$. В общем случае изменение $W_{n+1} - W_n$ содержимого бассейна равно X_{n+1} , если только требование не превышает возможности. Следовательно, последовательность $\{W_n\}$ должна удовлетворять соотношениям (9.1), т. е. *последовательные*

¹⁾ Для простоты мы начинаем с пустого водохранилища. Переход к произвольным начальным условиям не вызывает затруднений [см. пример в)].

количества воды в бассейне образуют процесс ожидания, индуцированный последовательностью $\{X_n\}$, при условии, конечно, что случайные величины X_n независимы и одинаково распределены.

Задача (если не для прикладника, то во всяком случае для математика) состоит в том, чтобы найти условия, при выполнении которых величины X_k независимы и имеют одинаковое распределение F , а также найти возможные формы F . Обычно величины τ_k либо находятся на одинаковом расстоянии друг от друга, либо представляют собой моменты последовательных скачков пуассоновского процесса, но для наших целей достаточно предположить, что τ_k образуют процесс восстановления, индуцированный, скажем, последовательностью $A_1, A_2, \dots$. Наиболее часто используемые модели принадлежат одному из следующих двух классов:

(i) Приток поступает с постоянной скоростью c , требования $\mathcal{B}_n$ произвольны. При этом $X_n = cA_n - \mathcal{B}_n$, и мы должны предположить независимость X_n от «прошлого» $X_1, \dots, X_{n-1}$ (обычное требование независимости A_n и $\mathcal{B}_n$ здесь излишне: нет оснований предполагать некоррелированность требования $\mathcal{B}_n$ с длительностью A_n).

(ii) Использование запаса происходит с постоянной скоростью, приток произволен. Этот класс схож с предыдущим: отличие, по существу, лишь в том, что A_n и $\mathcal{B}_n$ поменялись ролями.

в) *Очереди на пригородные поезда*¹⁾. Пригородный поезд, имеющий r пассажирских мест, отправляется от станции каждый час. Будущие пассажиры появляются на станции и становятся в очередь. На каждый отходящий поезд садятся первые r человек, а остальные остаются в очереди. Предположим, что количество пассажиров, появляющихся на станции между последовательными отходами поездов, являются независимыми случайными величинами $A_1, A_2, \dots$ с одинаковым распределением. Пусть число пассажиров, стоящих в очереди сразу же после n -го отправления, равно W_n и допустим для простоты, что $W_0 = 0$. Тогда $W_{n+1} = W_n + A_{n+1} - r$, если $W_n + A_{n+1} - r > 0$, и $W_{n+1} = 0$ в противном случае. Таким образом, последовательность $\{W_n\}$ есть процесс ожидания, индуцированный величинами $X_n = A_n - r$. ►

Поставим теперь задачу описать процесс ожидания $\{W_n\}$ в терминах порожденного величинами X_k случайного блуждания. Как в § 8, положим $S_0 = 0$, $S_n = X_1 + \dots + X_n$ и будем придерживаться введенных там обозначений для лестничных величин. Для облегчения описания воспользуемся моделью примера а).

¹⁾ Boudreau P. E., Griffin J. S., Кас М., On elementary queuing problem, Amer. Math. Monthly, 69 (1962), 713—724. Эта статья имеет дидактический характер — она написана для неспециалистов, не знакомых с предметом. В ней используется другой подход, но приведенные там вычисления перекрываются вычислениями примера гл. XII, 4, в).

Пусть v — такой индекс, что $S_1 \geq 0, S_2 \geq 0, \dots, S_{v-1} \geq 0$, но $S_v < 0$. При этом требования с номерами $1, 2, \dots, v-1$ имеют положительные времена ожидания $W_1 = S_1, \dots, W_{v-1} = S_{v-1}$, а v -е требование является первым требованием (с положительным индексом), поступающим на свободный прибор (первое удачное требование). С момента прибытия v -го требования процесс идет как вероятностная копия всего процесса. С другой стороны, v есть просто индекс первой отрицательной суммы, т. е. v является первым нижним лестничным индексом. Обозначим его в соответствии с предыдущим через $\mathcal{F}_1^-$. Мы приходим, таким образом, к *первому выводу: нижние лестничные индексы соответствуют удачным требованиям, находящим прибор свободным*. Отсюда вытекает, что моменты поступлений удачных требований образуют процесс восстановления, у которого времена возвращений распределены одинаково с $\mathcal{F}_1^-$.

В практических случаях величина $\mathcal{F}_1^-$ должна быть собственной. В самом деле, в противном случае с вероятностью, равной дефекту $\mathcal{F}_1^-$, требование никогда не застанет прибор свободным (исключая нулевое требование), и с вероятностью единица будет последнее удачное требование, за которым будет следовать неоканчивающаяся очередь. Мы увидим, что $\mathcal{F}_1^-$ есть собственная случайная величина, если $E(\mathcal{B}_k) < E(\mathcal{A}_k)$.

Предположим теперь, что $(v-1)$ -е требование поступает в момент t . Ему приходится ждать время $W_{v-1} = S_{v-1}$, и обслуживание его завершается к моменту $t + W_{v-1} + \mathcal{B}_{v-1}$. Первое удачное требование (с номером v) поступает в момент $t + \mathcal{A}_v$, и до этого момента прибор оставался свободным в течение времени

$$\mathcal{A}_v - W_{v-1} - \mathcal{B}_{v-1} = -S_{v-1} - X_v = -S_v.$$

Согласно определению v , величина S_v есть первая нижняя лестничная высота $\mathcal{H}_1^-$. Поскольку с момента поступления v -го требования процесс начинается сначала, то мы приходим ко *второму выводу: длительности свободных периодов являются независимыми случайными величинами, распределенными так же, как $-\mathcal{H}_1^-$* (т. е. как время возвращения для нижних лестничных высот). Таким образом, требование с номером $\mathcal{F}_1^- + \dots + \mathcal{F}_r^-$ является r -м требованием, застающим прибор свободным. В момент его прибытия прибор уже свободен в течение времени $-\mathcal{H}_r^-$.

Теперь должно быть ясно, что между последовательными лестничными моментами *сегменты траектории процесса массового обслуживания $\{W_n\}$ конгруэнтны сегментам графика траектории случайного блуждания $\{S_n\}$* , причем они сдвинуты вертикально так, чтобы начинаться с точки на временной оси (рис. 1). Для аналитического описания этого факта обозначим временно через $[n]$ *последний* нижний лестничный индекс $\leq n$; иными словами, $[n]$ есть (случайный) индекс, такой, что $[n] \leq n$ и

$$S_{[n]} \leq S_j, \quad j = 0, 1, \dots, n. \quad (9.3)$$

Индекс $[n]$ определен однозначно с вероятностью 1 (напомним, что распределение величин X_j непрерывно). Имеем, очевидно,

$$W_n = S_n - S_{[n]}. \quad (9.4)$$

Соотношение (9.4) приводит к очень важному выводу, если величины $X_1, \dots, X_n$ рассмотреть в *обратном порядке*. Положим для краткости $X'_1 = X_n, \dots, X'_k = X_i$. Частные суммы последовательности $\{X'_k\}$ суть $S'_k = X'_1 + \dots + X'_k = S_n - S_{n-k}$. Из (9.4) вытекает, что максимальный член последовательности $0, S'_1, \dots, S'_n$ имеет индекс $n - [n]$ и равен W_n . Поскольку векторы $(X'_1, \dots, X'_n)$ и $(X_1, \dots, X_n)$ одинаково распределены, нами доказана важная

Теорема ¹⁾. *Случайные величины W_n и*

$$M_n = \max[0, S_1, \dots, S_n] \quad (9.5)$$

распределены одинаково.

Следствия из этой теоремы будут обсуждаться в гл. XII. Здесь мы покажем, что теорема позволяет сводить некоторые задачи о разорении к задачам теории массового обслуживания (несмотря на внешнее несходство этих задач).

Примеры. г) *Задачи о разорении.* В § 5 «разорение» было определено как событие, состоящее в том, что $X(t) > z + ct$ при некотором t , где $X(t)$ — обобщенный пуассоновский процесс с распределением (4.2). Обозначим $\tau_1, \tau_2, \dots$ моменты последовательных скачков процесса $X(t)$. Если разорение вообще происходит, то оно происходит также в некоторый момент τ_k , и поэтому достаточно рассматривать вероятность осуществления события $S_n = X(\tau_n) - c\tau_n > z$ при некотором n . Но по определению обобщенного пуассоновского процесса величина $X(\tau_n)$ есть сумма n независимых случайных величин Y_k с одним и тем же распределением F , в то время как τ_n есть сумма n независимых показательно распределенных величин A_k . Таким образом, по существу мы имеем дело со случайным блужданием, порожденным случайными величинами $X_k = Y_k - cA_k$, плотность распределения которых дается формулой

$$\frac{\alpha}{c} \int_c^{\infty} e^{(\alpha/c)(x-y)} F\{dy\}. \quad (9.6)$$

Разорение происходит тогда и только тогда, когда в этом случайном блуждании при некотором n происходит событие $\{S_n \geq z\}$. Нахождение вероятности разорения сводится, следовательно,

¹⁾ Впервые этот результат был обнаружен, по-видимому, Полачеком в 1952 г. и использован (в ином контексте) Спигером: The Wiener—Hopf equation whose kernel is a probability, Duke Math. J., 24 (1947), 327—344. По поводу доказательства Спигера см. задачу 21.

к отысканию распределений величин W_n , образующих индуцированный величинами X_k процесс ожидания.

д) Численная иллюстрация. Наиболее важным процессом массового обслуживания является процесс, в котором времена между поступлениями требований и времена обслуживания имеют показательные распределения с математическими ожиданиями $1/a$ и $1/b$ соответственно, причем $a < b$. Характеристики этого процесса описаны в примере 8, б) и можно сделать вывод, что время ожидания n -го требования имеет предельное распределение W с атомом массы $1 - a/b$ в начале координат и плотностью $\frac{b-a}{b} a e^{-(b-a)x}$ при $x > 0$. Математическое ожидание этого распределения равно $\frac{a}{b(b-a)}$. Плотность распределения свободных периодов прибора совпадает с плотностью первой нижней лестничной высоты и равна $a e^{-at}$. В рассматриваемом случае свободные периоды и времена между поступлениями требований распределены одинаково (это, однако, не является общим фактом для процессов массового обслуживания).

Номер N первого требования, застающего прибор свободным, имеет производящую функцию $p(s)/a$, где $p(s)$ — функция, задаваемая формулой (8.4). Рассмотрим теперь период занятости, начинающийся в момент 0, т. е. время до первого момента, когда прибор оказывается свободным. Поскольку этот период начинается с поступления требования номер 0, то случайная величина N равна числу требований, поступивших в течение первого периода занятости. Простой подсчет показывает, что случайная величина N имеет математическое ожидание $b/(b-a)$ и дисперсию $ab(a+b)/(b-a)^3$.

Пусть, наконец, T есть длительность периода занятости. Точное выражение плотности T дается формулой (6.16) гл. XIV, с $sr = a$ и $sq = b$. Это выражение содержит бесселеву функцию и не поддается простым вычислениям. Тем не менее моменты величины T могут быть найдены с помощью ее преобразования Лапласа, полученного иными методами в примерах гл. XIV, 4, а) и XIV, 6, б). Они равны

$$E(T) = \frac{1}{b-a} \text{ и } \text{Var}(T) = (a+b) \frac{1}{(b-a)^3}.$$

В рассматриваемом процессе ожидания периоды занятости чередуются со свободными периодами, и их математические ожидания равны соответственно $1/(b-a)$ и $1/a$. Следовательно, отношение $(b-a)/a$ есть мера доли времени, в течение которого прибор остается свободным. Более точно: если за время t прибор не был занят общее время $U(t)$, то $EU(t)/t \rightarrow (b-a)a$.

В таблице 1 математическое ожидание времени обслуживания принято за единицу, так что a представляет собой математическое ожидание числа требований, поступающих в течение вре-

Таблица 1

		$b=1$					
		$a=0,5$	$a=0,6$	$a=0,7$	$a=0,8$	$a=0,9$	$a=0,95$
Время ожидания (стационарный режим)	Математическое ожидание	1	1,5	2,3	4	9	19
	Дисперсия	3	5,3	10	24	99	399
Период занятости	Математическое ожидание	2	2,5	3,3	5	10	20
	Дисперсия	12	25	63	225	1900	15 600
Число требований в течение периода занятости	Математическое ожидание	2	2,5	3,3	5	10	20
	Дисперсия	6	15	44	180	1710	14 820

мени обслуживания одного требования. Как видно из таблицы, дисперсии периодов занятости очень велики. Следовательно, *нужно ожидать очень больших флуктуаций периодов занятости*. Таким образом, обычная доверчивость к математическим ожиданиям весьма опасна в практических приложениях. Для периода занятости с дисперсией 225 тот факт, что математическое ожидание его равно 5, мало о чем говорит.

Многомерный аналог рассмотренного нами процесса массового обслуживания гораздо более сложен. Основы его теории были заложены в фундаментальной работе Дж. Кифера и Дж. Вольфовица [On the theory of queues with many servers, *Trans. Amer. Math. Soc.*, 78 (1955), 1—18].

§ 10. ВОЗВРАТНЫЕ И НЕВОЗВРАТНЫЕ СЛУЧАЙНЫЕ БЛУЖДЕНИЯ

В настоящем параграфе дается классификация случайных блужданий. Эта классификация не зависит от материала § 8 и тесно связана с теорией восстановления, изложенной в § 6. Каждой функции распределения F на прямой поставим формально в соответствие функцию интервалов

$$U\{I\} = \sum_{k=0}^{\infty} F^{k*}\{I\}. \quad (10.1)$$

Ряд в (10.1) тот же, что и в (6.2), но если распределение F не сосредоточено на полупрямой, он может расходиться даже, когда I есть конечный интервал. Ниже будет показано, что сходимость или расходимость ряда (10.1) имеет глубокий смысл. Основные результаты настоящего параграфа просты, однако их формули-

ровки, к сожалению, усложняются необходимостью отдельного рассмотрения арифметических распределений¹⁾.

Обозначим для краткости I_h и $I_h + t$ интервалы $-h < x < h$ и $t-h < x < t+h$ соответственно.

Теорема 1. (i) Если распределение F неарифметическое, то либо $U\{I\} < \infty$ для всех конечных интервалов, либо $U\{I\} = \infty$ для всех интервалов.

(ii) Если F — арифметическое распределение с шагом λ , то либо $U\{I\} < \infty$ для всех конечных интервалов, либо $U\{I\} = \infty$ для всякого интервала, содержащего точку вида $n\lambda$.

(iii) Если $U\{I_h\} < \infty$, то

$$U\{I_h + t\} \leq U\{I_{2h}\} \quad (10.2)$$

при всех t и $h > 0$.

Для облегчения ссылок на два возможных случая введем следующее определение (в этом определении F наделяется прилагательными, которые в действительности относятся к соответствующему блужданию).

Определение. Распределение F называется невозвратным, если $U\{I\} < \infty$ для всех конечных интервалов, и возвратным в противном случае.

Имея чисто вероятностное значение, теорема 1 связана также с интегральным уравнением

$$Z = z + F \star Z, \quad (10.3)$$

аналогичным уравнению восстановления (6.4). Мы используем интегральное уравнение (10.3) как исходный пункт и докажем теорему 1 одновременно со следующим предложением.

Теорема 2. Пусть z — непрерывная функция, такая, что $0 \leq z(x) \leq \mu_0$ при $|x| < h$ и $z(x) = 0$ вне интервала I_h .

Если F — невозвратное распределение, то функция

$$Z(x) = \int_{-\infty}^{\infty} z(x-y) U\{dy\} \quad (10.4)$$

является равномерно непрерывным решением уравнения (10.3), удовлетворяющим условию

$$0 \leq Z(x) \leq \mu_0 U\{I_{2h}\}. \quad (10.5)$$

Максимум функции Z достигается в некоторой точке интервала I_h .

Доказательство теорем 1 и 2. (i) Предположим, что $U\{I_\alpha\} < \infty$

¹⁾ Распределение F арифметическое, если все его точки роста принадлежат множеству вида $0, \pm\lambda, \pm2\lambda, \dots$. Наибольшее число λ , обладающее этим свойством, называется шагом F (см. гл. V, 2).

при некотором $\alpha > 0$. Возьмем $h < \frac{1}{2}\alpha$, и пусть функция z равна нулю вне I_h , но не обращается в нуль тождественно. Будем решать уравнение (10.3) методом последовательных приближений, положив $Z_0 = z$ и

$$Z_n(x) = z(x) + \int_{-\infty}^{\infty} Z_{n-1}(x-y) F\{dy\}. \quad (10.6)$$

Обозначив

$$U_n\{I\} = F^0 \star \{I\} + \dots + F^n \star \{I\}, \quad (10.7)$$

имеем, очевидно,

$$Z_n(x) = \int_{-\infty}^{\infty} z(x-y) U_n\{dy\} \quad (10.8)$$

(интегрирование в действительности распространяется на интервал длины $\leq 2h$). Так определенная функция Z_n непрерывна. Покажем, пользуясь индукцией, что она достигает своего максимума μ_n в точке ξ_n , такой, что $z(\xi_n) > 0$. Для функции $Z_0 = z$ это верно тривиальным образом. Для перехода от $n-1$ к n достаточно заметить, что $z(x) = 0$ влечет за собой $Z_n(x) \leq \mu_{n-1}$, в то время как $\mu_n \geq Z_n(\xi_{n-1}) \geq Z_{n-1}(\xi_{n-1}) = \mu_{n-1}$.

Следовательно, интервал $I_h + \xi_n$ содержится в интервале I_{2h} , и поэтому в силу (10.8)

$$\mu_n \leq \mu_0 U\{I_{2h}\}. \quad (10.9)$$

Тем самым доказано, что функции Z_n равномерно ограничены. Поскольку $Z_0 \leq Z_1 \leq \dots$, то отсюда вытекает, что $Z_n \rightarrow Z$, причем функция Z удовлетворяет неравенствам (10.5).

В силу монотонной сходимости из (10.6) и (10.8) следует, что предел Z удовлетворяет интегральному уравнению (10.3) и имеет вид (10.4). Неравенство (10.5) выполняется вследствие (10.9). Верхняя граница зависит только от максимума μ_0 функции z , и мы можем свободно положить $z(x) = \mu_0$ для всех x внутри подходящего подынтервала I_η интервала I_h . В этом случае мы имеем из (10.8)

$$Z_n(x) \geq \mu_0 U_n\{I_\eta + x\}. \quad (10.10)$$

Неравенство (10.10) справедливо при всех $\eta < h$ и, следовательно, также при $\eta = h$ (так как интервал I_h замкнут). Два последних неравенства вместе доказывают, что (10.2) справедливо. Отсюда вытекает, что $U\{I\} < \infty$ для всех интервалов, длина которых не превышает h . Но каждый конечный интервал может быть разбит на конечное число интервалов, у которых длина $< h$, и поэтому $U\{I\} < \infty$ для всех конечных I . Наконец, беря разности значений в (10.4), получаем

$$|Z(x+\delta) - Z(x)| \leq U\{I_{2h}\} \cdot \sup |z(x+\delta) - z(x)|.$$

Следовательно, функция Z равномерно непрерывна, и, таким образом, все утверждения, касающиеся невозвратного распределения F , доказаны.

(ii) Предположим теперь, что $U\{I_\alpha\} = \infty$ при всех $\alpha > 0$. Из (10.10) следует, что $Z_n(x) \rightarrow \infty$ при всех x из некоторой окрестности нуля. Отсюда и из (10.6) вытекает, что если t — точка роста F , то $Z_n(x) \rightarrow \infty$ для всех x из некоторой окрестности t . Пользуясь индукцией, легко вывести, что то же самое верно для всех точек роста любой из функций распределения $F^{2*}, F^{3*}, \dots$. Предположим, что распределение F не арифметическое. Оно не может быть сосредоточенным на полупрямой, так как тогда мы имели бы $U\{I_\alpha\} < \infty$ (§ 6). В силу леммы 2 из гл. V, 4 а, точки роста функций $F^{2*}, F^{3*}, \dots$ образуют плотное множество на прямой и поэтому $Z_n(x) \rightarrow \infty$ при всех x . Отсюда вытекает, что $U_n\{I\} \rightarrow \infty$ для всех интервалов. С очевидными изменениями это же рассуждение применимо к арифметическим распределениям. Теоремы доказаны. ►

Мы еще вернемся к уравнению восстановления (10.3) в гл. XI, а сейчас выведем из теоремы 1 ряд следствий для случайных блужданий. Пусть $X_1, X_2, \dots$ — независимые случайные величины с общим распределением F . Положим $S_n = X_1 + \dots + X_n$. «Попаданием в I в момент n » мы называем событие $S_n \in I$.

Теорема 3 ¹⁾. Если распределение F невозвратно, то число попаданий в каждый конечный интервал конечно с вероятностью единица, а математическое ожидание числа таких попаданий равно $U\{I\}$.

Если F — возвратное неарифметическое распределение, то число попаданий в каждый конечный интервал бесконечно с вероятностью единица. Если F — возвратное арифметическое распределение с шагом λ , то число попаданий в каждую точку вида $n\lambda$ бесконечно с вероятностью единица.

Доказательство. Пусть распределение F невозвратно. Вероятность попадания в I после момента n не превосходит n -го остаточного члена ряда в (10.1), и поэтому для любого $\varepsilon > 0$ при достаточно больших n вероятность более чем n попаданий не превосходит ε . Тем самым первое утверждение теоремы доказано.

Предположим теперь, что F — возвратное неарифметическое распределение. Обозначим через $p_h(t)$ вероятность хотя бы одного попадания в $I_h + t$. Достаточно показать, что $p_h(t) = 1$ при всех $h > 0$ и t , поскольку отсюда с очевидностью следует, что любое

¹⁾ Эта теорема вытекает из второго закона нуля и единицы (см. гл. IV, 6). Если $\varphi(I+t)$ есть вероятность бесконечного числа попаданий в интервал $I+t$, то при фиксированном I функция φ может принимать лишь два значения: 0 и 1. С другой стороны, рассматривая первый шаг случайного блуждания, видим, что $\varphi = F \star \varphi$, откуда $\varphi = \text{const}$ (см. гл. XI, 9).

число попаданий в произвольный интервал имеет вероятность 1.

Прежде чем продолжить, отметим, что если $S_n = x$, то можно перенести начало координат в точку x с тем, чтобы сделать вывод, что вероятность последующего попадания в I_h равна $\rho_h(-x)$. В частности, если x принадлежит $I_h + t$, то вероятность последующего попадания в I_h не превосходит $\rho_{2h}(-t)$.

Докажем сначала, что $\rho_h(0) = 1$. Для произвольного, но фиксированного $h > 0$ обозначим через p_r вероятность по меньшей мере r попаданий в I_h . Тогда $p_1 + p_2 + \dots$ есть ожидаемое число попаданий в I_h , и оно бесконечно в силу определения возвратности. С другой стороны, из сделанного нами предварительного замечания следует, что $p_{r+1} \leq p_r \cdot \rho_{2h}(0)$. Вследствие расходимости ряда $\sum p_r$ имеем $\rho_{2h}(0) = 1$.

Переходя к произвольным интервалам $I_h + t$, предположим сначала, что F — неарифметическое распределение. По лемме 2 из гл. V, 4 а, каждый интервал содержит при некотором k точку роста распределения F^{k*} и поэтому вероятность $\rho_h(t)$ попадания в $I_h + t$ положительна при всех $h > 0$ и всех t . Мы уже видели, однако, что после попадания в $I_h + t$ возвращение в I_h достоверно, и в силу предыдущего замечания отсюда следует, что $\rho_{2h}(-t) = 1$. Так как h и t произвольны, то на этом доказательство для неарифметических распределений завершается. Для арифметических распределений применимы те же рассуждения. ►

При проверке сходимости или расходимости ряда (10.1), определяющего $U\{I\}$, обычно обращаются к предельным теоремам, которые дают информацию только для очень больших интервалов. В такой ситуации задачу облегчает следующий

Критерий. Если распределение F невозвратно, то величина $x^{-1}U\{I_x\}$ остается ограниченной при $x \rightarrow \infty$.

Утверждение является очевидным следствием (10.2), так как любой интервал I_{nh} можно разбить на n интервалов вида $I_h + t$. В качестве иллюстрации указанного метода будет доказана

Теорема 4. Распределение с математическим ожиданием μ является возвратным, если $\mu = 0$, и невозвратным, если $\mu \neq 0$.

Доказательство. Пусть $\mu = 0$. В силу закона больших чисел существует такое целое число n_ε , что $P\{|S_n| < \varepsilon n\} > 1/2$ для всех $n > n_\varepsilon$. В соответствии с этим $F^{n*}\{I_a\} > 1/2$ для $n_\varepsilon < n < a/\varepsilon$. Если $a > 2\varepsilon n_\varepsilon$, то существуют более чем $a/(2\varepsilon)$ целых чисел n , удовлетворяющих этому условию, и, следовательно, $U\{I_a\} > a/(4\varepsilon)$. Поскольку ε произвольно, то отсюда следует, что отношение $a^{-1}U\{I_a\}$ неограничено и, следовательно, F не может быть невозвратным.

Если $\mu > 0$, то из усиленного закона больших чисел вытекает,

что с вероятностью, сколь угодно близкой к единице, сумма S_n будет положительной для всех достаточно больших n . Поэтому вероятность бесконечного числа попаданий на отрицательную полупрямую $-\infty, 0$ равна 0, и, таким образом, распределение F невозвратно. ►

При возвратном случайном блуждании последовательность $\{S_n\}$ меняет знак бесконечно часто, и поэтому соответствующие такому блужданию верхние и нижние лестничные процессы возвратны. Можно считать удивительным, что обратное утверждение неверно. *Даже при невозвратном случайном блуждании последовательность $\{S_n\}$ может менять знак бесконечно часто* (с вероятностью единица). В действительности это имеет место в случае, когда F симметрично. Так как при невозвратном блуждании число попаданий в конечный интервал $-a, a$ конечно, то это означает (очень грубо говоря), что перемены знаков происходят из-за отдельных скачков весьма большой величины: величина $|S_n|$ перерастает все границы, но поразительное неравенство $X_{n+1} < -S_n - a$ осуществляется бесконечно часто, как бы велика ни была постоянная a .

Рис. 2 (стр. 227) показывает, как происходят большие скачки, однако он не дает полного представления о рассматриваемом явлении, поскольку распределение пришлось урезать, чтобы получить конечный график.

§ 11. ОБЩИЕ МАРКОВСКИЕ ЦЕПИ

Дискретные марковские цепи (I; гл. XV) довольно просто обобщаются на случай декартовых (и более общих) пространств состояний. В дискретном случае вероятности перехода задавались стохастической матрицей с элементами p_{ij} , строки которой образовывали вероятностные распределения. Теперь мы рассматриваем переходы из точки x в некоторый интервал или множество Γ в $\mathcal{R}^n$. Вероятность такого перехода мы обозначаем $K(x, \Gamma)$. Чтобы обосновать существование необходимых нам интегралов, мы должны наложить на K некоторые условия регулярности. Требования непрерывности хватило бы для большинства практических целей, однако, накладывая это требование, мы ничего не выигрываем по сравнению с более общим случаем.

Определение 1. *Стохастическим ядром K называется функция от двух переменных — точки и множества, такая, что (i) при любом фиксированном x $K(x, \Gamma)$ является вероятностным распределением по Γ , и (ii) для любого фиксированного интервала Γ $K(x, \Gamma)$ есть бэровская функция по x .*

Это определение относится к любому числу измерений (и даже к более общим пространствам).

Мы не требуем, чтобы ядро K было определено на всем пространстве. Если x и Γ являются подмножествами множества Ω , то мы говорим, что K *сосредоточено* на Ω . Иногда необходимо допустить *несобственные (дефектные)* распределения, и в этом случае ядро K называют *субстохастическим*. Нередко K можно представить в виде

$$K(x, \Gamma) = \int_{\Gamma} k(x, y) dy. \quad (11.1)$$

Функция k в (11.1) называется *плотностью стохастического ядра*. В соответствии с соглашением гл. V, (3.3), соотношение (11.1) кратко записывается в виде

$$K(x, dy) = k(x, y) dy.$$

[Строго говоря, k есть плотность по отношению к мере Лебега или длине; в случае произвольной меры m мы бы использовали запись $K(x, dy) = k(x, y) m\{dy\}$.]

Прежде чем давать формальное определение марковской цепи, подготовим соответствующий аналитический аппарат по аналогии со случаем дискретных цепей. Вероятность перехода из x в Γ за два шага положим по определению равной

$$K^{(2)}(x, \Gamma) = \int_{\Omega} K(x, dy) K(y, \Gamma). \quad (11.2)$$

Множество Ω в (11.2) либо есть все пространство, либо совпадает с тем множеством, на котором сосредоточено ядро K . Соотношение (11.2) утверждает, что первый шаг ведет из точки x в некоторую точку y , а второй шаг — из точки y в множество Γ . Основное допущение состоит в том, что при заданной промежуточной точке y прошлая история не влияет на дальнейшие переходы. Аналогичное можно сказать о $K^{(n)}$ — вероятности перехода за n шагов. Если положить $K^{(1)} = K$, то для произвольных целых положительных m и n должно быть

$$K^{(m+n)}(x, \Gamma) = \int_{\Omega} K^{(m)}(x, dy) K^{(n)}(y, \Gamma). \quad (11.3)$$

Когда $m=n=1$, эта формула сводится к (11.2). При $m=1$ и $n=1, 2, \dots$ формула (11.3) *последовательно определяет вероятности перехода* $K^{(n)}$. Обозначим для согласованности символом $K^{(0)}$ вероятностное распределение, сосредоточенное в точке x (кронековское ядро). Тогда формула (11.3) будет справедлива при всех $m \geq 0, n \geq 0$. Операция (11.3) над ядрами, нередко встречающаяся и вне теории вероятностей, называется *сверткой ядер*. Она во всех отношениях схожа с перемножением матриц.

Вряд ли стоит специально отмечать, что ядра $K^{(n)}$ являются стохастическими. Если K имеет плотность, то все ядра $K^{(n)}$ имеют

плотности, причем эти плотности связаны формулами свертки

$$k^{(m+n)}(x, z) = \int_{\Omega} k^{(m)}(x, y) k^{(n)}(y, z) dy. \quad (11.4)$$

Примеры. а) *Свертки.* Если $k(x, y) = f(x-y)$, где f — плотность вероятности, то свертки (11.4) сводятся к обычным сверткам. Аналогичное справедливо и в общем случае, если ядро K однородно в том смысле, что

$$K(x, \Gamma) = K(x+s, \Gamma+s),$$

где $\Gamma+s$ получено сдвигом множества Γ на s ; по поводу свертки на окружности см. теорему 3 в гл. VIII, 7.

б) *Потери энергии при столкновениях.* В физике последовательные столкновения частицы обычно рассматриваются как случайный процесс. Считают, что если энергия (масса) частицы до столкновения равна $x > 0$, то после столкновения она представляется случайной величиной Y , для которой $P\{Y \in \Gamma\} = K(x, \Gamma)$, где K — стохастическое ядро. Обычно предполагается, что возможны лишь потери и что отношение Y/x имеет функцию распределения G , не зависящую от x . При этом $P\{Y \leq y\} = G(y/x)$, и функция $G(y/x)$ определяет стохастическое ядро.

В аналогичной задаче о звездной радиации [пример гл. X, 2, в)] Амбарцумян рассматривал частный случай, когда $G(y) = y^\lambda$ при $0 \leq y \leq 1$, где λ — положительная постоянная. Плотность соответствующего ядра сосредоточена на множестве $0 < y < x$ и равна $\lambda y^{\lambda-1} x^{-\lambda}$. Нетрудно проверить, что плотность вероятности перехода за n шагов выражается в рассматриваемом случае формулой

$$k^{(n)}(x, y) = \frac{\lambda^n}{(n-1)!} \frac{y^{\lambda-1}}{x^\lambda} \left(\log \frac{x}{y} \right)^{n-1}, \quad 0 < y < x. \quad (11.5)$$

Значение $\lambda=1$ соответствует равномерному распределению (*углеродная доля* «распределена случайно»), и (11.5) сводится при этом к гл. I, (8.2) [продолжение в примере гл. X, 1, а)].

в) *Случайные цепи.* Рассмотрим цепь (или ломаную линию) в $\mathcal{R}^3$ со звеньями длины единица и со случайными углами между соседними звеньями. Многие варианты подобной схемы (часто довольно сложные) встречаются в химии полимеров. Мы ограничимся рассмотрением случая, когда последовательные углы являются независимыми случайными величинами.

Под *длиной* L цепи с концевыми точками A и B мы понимаем расстояние между A и B . В результате добавления одного звена к цепи длины x получается цепь длины $\sqrt{x^2 + 1 - 2x \cos \theta}$, где θ — случайный угол между новым звеном и прямой, проходящей через точки A и B . Рассмотрим два распределения θ , особенно интересных в химии.

(i) Пусть θ принимает значения 60° и 120° с вероятностью $1/2$

каждое. Тогда $\cos \theta = \pm 1/2$ и длина продолженной цепи определяется стохастическим ядром $K(x, \Gamma)$, наделяющим каждую из двух точек $\sqrt{x^2 \pm x + 1}$ вероятностью $1/2$. При фиксированном x распределение $K^{(n)}$ сосредоточено в 2^n точках.

(ii) Пусть направление нового звена выбирается «случайно», т. е. предположим, что величина $\cos \theta$ распределена равномерно на интервале $[-1, 1]$ [см. гл. I, 10]. Длина продолженной цепи L лежит между $|x-1|$ и $x+1$. Внутри этого промежутка по теореме косинусов

$$P(L < y) = P\{2x \cos \theta > x^2 + 1 + y^2\} = \frac{1}{2} - [x^2 + 1 + y^2]/4x.$$

Таким образом, длина продолженной цепи определяется стохастическим ядром с плотностью

$$k(x, y) = y/2x, \quad |x-1| < y < x+1.$$

Длина L_{n+1} цепи с $n+1$ звеньями имеет плотность $k^{(n)}(1, y)$ (см. задачу 23).

г) *Дискретные цепи Маркова.* Стохастическую матрицу (p_{ij}) можно рассматривать как плотность $k(i, j) = p_{ij}$ определенного на множестве Ω целых положительных чисел стохастического ядра по отношению к мере m , наделяющей каждую точку множества Ω единичной массой.

Абсолютные и стационарные вероятности

Последовательность случайных величин $X_0, X_1, \dots$ подчинена вероятностям перехода $K^{(n)}$, если $K^{(n)}(x, \Gamma)$ является условной вероятностью события $\{X_{m+n} \in \Gamma\}$ при условии $X_m = x$. *Вероятностное распределение величины X_n при заданном распределении γ_0 величины X_0 выражается формулой*

$$\gamma_n\{\Gamma\} = \int_{\Omega} \gamma_0\{dx\} K^{(n)}(x, \Gamma). \quad (11.6)$$

Определение 2. Распределение γ_0 называется *стационарным распределением ядра K* , если $\gamma_n = \gamma_0$ при всех n , т. е. если

$$\gamma_0\{\Gamma\} = \int_{\Omega} \gamma_0\{dx\} K(x, \Gamma). \quad (11.7)$$

Основные свойства стационарных распределений те же, что и в случае дискретных марковских цепей. При выполнении нежестких требований регулярности относительно K существует *единственное стационарное распределение*; оно является *асимптотическим распределением величин X_n при любом исходном распределении*. Иными словами, влияние начального состояния исчезает и система стремится к *устойчивому состоянию*, определяемому

стационарным решением. Этот результат является одной из форм эргодической теоремы (см. гл. VIII, 7).

Примеры. д) Процесс ожидания $\{W_n\}$, определенный в (9.1), является марковским процессом, сосредоточенным на замкнутой полупрямой $\overline{0, \infty}$. Вероятности перехода определены здесь лишь для $x, y \geq 0$, и при $\Gamma = 0$, y имеем $K(x, \Gamma) = F(y - x)$. Существование стационарной меры будет доказано в гл. VIII, 7.

е) Пусть $X_1, X_2, \dots$ — взаимно независимые положительные случайные величины с распределением F и с непрерывной плотностью f , сосредоточенной на $\overline{0, \infty}$. Определим случайные величины Y_k последовательно посредством формул

$$Y_1 = X_1, \quad Y_{n+1} = |Y_n - X_{n+1}|. \quad (11.8)$$

Последовательность $\{Y_n\}$ образует сосредоточенный на $\overline{0, \infty}$ марковский процесс с плотностями вероятностей перехода

$$k(x, y) = \begin{cases} f(x-y) + f(x+y) & \text{при } 0 < y < x, \\ f(x+y) & \text{при } y > x > 0. \end{cases} \quad (11.9)$$

Стационарная плотность g определяется уравнением

$$g(y) = \int_0^{\infty} g(x+y)f(x)dx + \int_0^y g(x)f(x+y)dx. \quad (11.10)$$

Если F имеет математическое ожидание μ , то функция

$$g(y) = \mu^{-1} [1 - F(y)] \quad (11.11)$$

является *стационарной плотностью вероятности*. В самом деле, с помощью простого интегрирования по частям можно убедиться, что g удовлетворяет уравнению (10.10)¹⁾. С другой стороны, из гл. V, (6.3), следует, что g есть плотность вероятности (см. задачу 22).

ж) *Одно техническое приложение*²⁾. Длинная линия передачи состоит из отдельных кусков кабеля, характеристики которых подвержены статистическим флуктуациям. Будем считать отклонения от идеального значения независимыми случайными величинами $Y_1, Y_2, \dots$ и предположим, что эффект их аддитивен. Перемена направления кабеля меняет знак его вноса в суммар-

¹⁾ Как обнаружить такого рода результат? Предположив, что функции f и g имеют производные, можно формально продифференцировать (10.10). Интегрируя далее по частям, приходим к уравнению $g'(y) = -g(0)f(y)$, из которого вытекает, что функция g должна иметь вид (11.11). Непосредственная проверка показывает теперь, что (11.11) является решением (11.10) без предположений о дифференцируемости f и g .

²⁾ Обобщение дискретной модели, использовавшейся фон Шеллингом, *Elektrische Nachr. — Technik*, 20 (1943), pp. 251 — 259.

ный эффект. Предположим, что отклонения Y_k симметричны, и обозначим $X_k = |Y_k|$. Эффективную конструкцию длинной линии передачи можно осуществить по следующему индуктивному правилу: $(n+1)$ -й кусок кабеля подсоединяется так, что соответствующая ему ошибка имеет знак, противоположный знаку накопленной ошибки предыдущих n кусков. Накопленные ошибки подчиняются тогда правилу (11.8), и (11.11) является не только стационарной плотностью, но и предельным распределением: *распределение ошибки линии, состоящей из n кусков кабеля, близко (при больших n) к распределению с плотностью (11.11)*. С другой стороны, если бы отдельные куски кабеля подсоединялись независимо, то была бы применима центральная предельная теорема и дисперсия ошибки росла бы линейно по n , т. е. линейно относительно длины кабеля. Таким образом, указанная простая процедура учета знака отдельных ошибок не дает разрастаться суммарной ошибке. ►

В приведенных примерах марковская последовательность $X_0, X_1, \dots$ определялась в терминах начального распределения γ_0 и вероятностей перехода K . Совместное распределение величин $(X_0, X_1, \dots, X_n)$ имеет вид

$$\gamma_0 \{dx_0\} K(x_0, dx_1) \dots K(x_{n-1}, dx_n).$$

Распределения такого вида уже обсуждались в гл. III, 8, и I, XV, 1. Мы имеем здесь типичный пример, показывающий преимущество определения абсолютных вероятностей в терминах условных вероятностей. Более систематичным был бы подход, использующий в качестве определения соотношение

$$P\{X_{n+1} \in \Gamma | X_0 = x_0, \dots, X_n = x_n\} = K(x_n, \Gamma). \quad (11.12)$$

Марковское свойство выражено здесь тем обстоятельством, что правая часть не зависит от $x_0, x_1, \dots, x_{n-1}$ и поэтому «предыстория» не влияет на будущее. Недостаток этого определения состоит в том, что оно вовлекло бы нас в рассмотрение вопросов существования условных вероятностей, их единственности и т. п.

(Марковские процессы с непрерывным параметром изучаются в гл. X.)

§ 12 *). МАРТИНГАЛЫ

Для первой ориентировки можно рассмотреть случайный процесс $\{X_n\}$, обладающий тем свойством, что совместное распределение величин $(X_1, \dots, X_n)$ имеет строго положительную непрерывную плотность p_n . При этом элементарный метод гл. III, 2, позволяет всюду определить условные плотности и условные ма-

*) Мартингалы очень важны, и поэтому им посвящен отдельный параграф. Однако нигде в этой книге мартингалы не используются в качестве аппарата.

тематические ожидания. Предположим также, что рассматриваемые величины X_n и Y_n имеют математические ожидания.

Последовательность $\{X_n\}$ назовем *абсолютно беспристрастной* (безобидной), если при $n=1, 2, \dots$

$$E(X_1)=0, \quad E(X_{n+1}|X_1, \dots, X_n)=0. \quad (12.1)$$

Последовательность $\{Y_n\}$ является *мартингалом*, если

$$E(Y_{n+1}|Y_1, \dots, Y_n)=Y_n, \quad n=1, 2, \dots \quad (12.2)$$

(Ниже будет дано более гибкое определение.)

Между этими двумя типами последовательностей имеется простая связь. Пусть $\{X_n\}$ — абсолютно беспристрастная последовательность. Положим

$$Y_n = X_1 + \dots + X_n + c, \quad (12.3)$$

где c — постоянная. Тогда

$$E(Y_{n+1}|X_1, \dots, X_n) = Y_n. \quad (12.4)$$

Задающие условие величины X_i в (12.4) могут быть заменены на Y_k , и поэтому (12.4) эквивалентно (12.2). С другой стороны, если последовательность $\{Y_n\}$ — мартингал, то положим $X_1 = Y_1 - E(Y_1)$ и $X_{n+1} = Y_{n+1} - Y_n$. Так, построенная последовательность $\{X_n\}$ абсолютно беспристрастна и, кроме того, для рассматриваемых величин выполняется (12.3) с $c = E(Y_1)$. Таким образом, *последовательность $\{Y_n\}$ образует мартингал тогда и только тогда, когда она имеет вид (12.3), где $\{X_n\}$ — абсолютно беспристрастная последовательность.*

Понятие мартингала было введено П. Леви. Затем Дж. Л. Дуб обнаружил неожиданные возможности аппарата мартингалов и развил их теорию. В гл. VII, 9, будет показано, что при выполнении нежестких условий ограниченности образующие мартингал величины Y_k сходятся к пределу. Этот факт имеет важное значение в современной теории вероятностных процессов.

Примеры. а) В классических играх случайные величины X_k независимы и имеют нулевое математическое ожидание: $E(X_k)=0$. Такая игра абсолютно безобидна ¹⁾, и частные суммы $S_n = X_1 + \dots + X_n$ образуют мартингал. Рассмотрим теперь обычную игру в орлянку, в которой игрок выбирает ставку по некоторому правилу, учитывающему его успехи в предыдущих турах. Последовательные выигрыши не являются уже независимыми, но игра по-прежнему абсолютно безобидна. Смысл безобидной игры состоит в том, что знание прошлой истории игры не должно давать игроку возможности увеличить свой капитал. Наглядно это

¹⁾ О практической ограниченности этого понятия говорилось в I, X, 3. Напомним, что существуют «безобидные» игры, в которых с вероятностью, большей 1 — ε , выигрыш игрока в n -й игре превосходит, скажем, $n/\log n$.

означает, что абсолютно безобидная игра должна оставаться таковой при любой системе игры, т. е. при любых правилах пропуска отдельных туров. Мы увидим, что это именно так и есть.

б) *Урновая схема Пойа* [1, V, 2, в)]. Из урны, содержащей b черных и r красных шаров, выбирается наудачу один шар. Затем этот шар возвращается и, кроме того, в урну добавляется s шаров того же цвета, что и выбранный шар. Пусть $Y_0 = b/(b+r)$, и пусть Y_n есть доля черных шаров, оказавшихся в урне в результате n -го извлечения. Последовательность $\{Y_n\}$ является мартингалом. Теорема о сходимости обеспечивает здесь существование предельного распределения [см. примеры гл. VII, 4, а), и гл. VII, 9, а)].

в) *Конкордантные¹⁾ функции*. Пусть $\{X_n\}$ — марковский процесс с вероятностями перехода, задаваемыми стохастическим ядром K . Существование математических ожиданий величин X_n не предполагается. Функция u называется *конкордантной по отношению к K* , если

$$u(x) = \int K(x, dy) u(y). \quad (12.5)$$

Рассмотрим случайные величины $Y_k = u(X_k)$ и предположим, что все моменты Y_k существуют (так будет, например, если функция u ограничена). Соотношение (12.5) означает, что $E(Y_{k+1} | X_k = x) = u(x)$, откуда $E(Y_{k+1} | X_k) = Y_k$. Следовательно, поскольку последовательность $\{X_k\}$ — марковская, имеет место (12.4), а так как величины Y_k являются функциями от X_k , то (12.4) влечет за собой (12.2) [см. гл. V, 10, а)]. Таким образом, последовательность $\{Y_n\}$ является *мартингалом*. Этот результат имеет большое значение в граничной теории марковских процессов, поскольку существование предела последовательности $\{Y_n\}$ обычно влечет существование предела исходной последовательности $\{X_n\}$ [см. примеры е) и гл. VII, 9, в)].

г) *Отношения правдоподобия*. Пусть $X_1, X_2, \dots$ — последовательность случайных величин, и пусть известно, что совместное распределение величин $(X_1, \dots, X_n)$ имеет либо плотность p_n , либо плотность q_n , однако неизвестно, какую именно. Чтобы прийти к какому-то решению, в статистике вводятся новые случайные величины

$$Y_n = \frac{q_n(X_1, \dots, X_n)}{p_n(X_1, \dots, X_n)}. \quad (12.6)$$

Ясно, что если выполнены достаточно сильные условия регулярности и если истинными плотностями являются p_n , то наблюдаемые значения величин $X_1, \dots, X_n$ в среднем будут группироваться вокруг тех точек, где p_n относительно велико. Когда дело обстоит именно так, величина Y_n , вероятно, будет малой или боль-

¹⁾ Этот термин был введен Г. Хантом.

шой, смотря по тому, является ли истинной плотностью p_n или q_n . Таким образом, асимптотическое поведение величин $\{Y_n\}$ представляет интерес в теории статистических решений.

Предположим для простоты, что плотности p_n строго положительны и непрерывны. Если истинными плотностями являются p_n , то условная плотность распределения величины X_{n+1} при заданных $X_1, \dots, X_n$ равна отношению p_{n+1}/p_n и, следовательно,

$$E(Y_{n+1} | X_1 = x_1, \dots, X_n = x_n) = \int_{-\infty}^{\infty} \frac{q_{n+1}(x_1, \dots, x_n, y)}{p_{n+1}(x_1, \dots, x_n, y)} \cdot \frac{p_{n+1}(x_1, \dots, x_n, y)}{p_n(x_1, \dots, x_n)} dy. \quad (12.7)$$

В подынтегральном выражении функции p_{n+1} сокращаются, а знаменатель второй дроби не зависит от y . Интегрирование q_{n+1} дает в результате частную плотность q_n . Таким образом, интеграл в (12.7) равен q_n/p_n , и поэтому выполняется (12.4). Следовательно, при сделанных предположениях отношения правдоподобия Y_n образуют мартингал. ►

Условия, относительно которых берутся условные математические ожидания в (12.2), не являются вполне удачными, поскольку нередко приходится заменять задающие условия величины $Y_1, \dots, Y_n$ некоторыми функциями от них [примером этого является (12.4)]. Еще больший дефект такого задания условий обнаруживается в примере а). Рассматриваемый процесс (скажем, игра в орлянку или рулетку) описывается последовательностью случайных величин Z_n , и выигрыш игрока в $(n+1)$ -м туре есть функция от $Z_1, \dots, Z_{n+1}$, а также, возможно, от некоторых других величин. Наблюдаемое прошлое представляется величинами $(Z_1, \dots, Z_n)$, которые могут содержать в себе больше информации, чем предыдущие выигрыши. Например, если игрок не участвует в турах с номерами 1, 3, 5, ..., то знание его выигрышей до момента $2n$ в лучшем случае эквивалентно знанию величин $Z_2, Z_4, \dots, Z_{2n}$. При этом дополнительное знание величин $Z_1, Z_3, \dots$ в принципе может давать некоторое преимущество, и абсолютная безобидность в таком случае должна иметь в качестве условий величины $Z_1, \dots, Z_n$. Таким образом, может оказаться необходимым рассматривать условные ожидания по отношению к нескольким множествам случайных величин, и, чтобы охватить все возможные ситуации, лучше всего использовать условные ожидания по отношению к произвольным σ -алгебрам событий.

Рассмотрим последовательность $\{Y_n\}$ случайных величин на произвольном вероятностном пространстве и обозначим $\mathfrak{A}_n$ σ -алгебру событий, порожденную величинами $(Y_1, \dots, Y_n)$ [см. гл. V, 10, а)]. Определяющее мартингалы соотношение (12.2) можно записать в виде $E(Y_{n+1} | \mathfrak{A}_n) = Y_n$. Мы хотим дать более общее определение мартингалов в форме последнего соотношения с той разницей,

что σ -алгебра $\mathfrak{A}_n$ заменяется более богатыми событиями σ -алгеброй $\mathfrak{B}_n$. В большинстве случаев σ -алгебра $\mathfrak{B}_n$ будет порождаться величинами $Y_1, \dots, Y_n$ и некоторыми другими величинами, зависящими от прошлого. Идея состоит в том, чтобы любая зависящая от прошлого случайная величина была измерима относительно $\mathfrak{B}_n$, и в этом смысле σ -алгебра $\mathfrak{B}_n$ представляет ту информацию, которая содержится в прошлой истории процесса. Поскольку информация о прошлом растет со временем, то мы предположим, что σ -алгебры $\mathfrak{B}_n$ возрастают:

$$\mathfrak{B}_1 \subset \mathfrak{B}_2 \subset \dots \quad (12.8)$$

Определение 1. Пусть $Y_1, Y_2, \dots$ — случайные величины, математические ожидания которых существуют, и пусть $\mathfrak{B}_1, \mathfrak{B}_2, \dots$ — σ -алгебры событий, удовлетворяющие условию (12.8).

Последовательность $\{Y_n\}$ образует мартингал по отношению к $\{\mathfrak{B}_n\}$ тогда и только тогда, когда

$$E(Y_{n+1} | \mathfrak{B}_n) = Y_n. \quad (12.9)$$

[В силу неединственности условных математических ожиданий равенство (12.9) следует понимать так: «Существует вариант условного математического ожидания, для которого выполняется (12.9)». Это замечание относится и к дальнейшему.]

Отметим, что из (12.9) вытекает $\mathfrak{B}_n$ -измеримость величины Y_n , и этот факт имеет два важных следствия. Поскольку $\mathfrak{B}_n \supset \mathfrak{B}_{n-1}$, то из основного тождества (11.7) гл. V для повторных математических ожиданий следует, что

$$E(Y_{n+1} | \mathfrak{B}_{n-1}) = Y_{n-1}.$$

Индукция показывает, что определение (12.9) влечет за собой более сильные соотношения

$$E(Y_{n+1} | \mathfrak{B}_k) = Y_k, \quad k = 1, 2, \dots, n. \quad (12.10)$$

Отсюда, в частности, следует, что каждая подпоследовательность $Y_{v_1}, Y_{v_2}, \dots$ мартингала $\{Y_n\}$ снова является мартингалом.

Далее отметим, что $\mathfrak{B}_n$ содержит σ -алгебру, порожденную величинами $Y_1, \dots, Y_n$, и аналогичное рассуждение показывает, что последовательность $\{Y_n\}$ также является мартингалом по отношению к последовательности σ -алгебр $\{\mathfrak{A}_n\}$. Таким образом, (12.9) влечет (12.2).

Пример. д) Пусть σ -алгебры $\mathfrak{B}_n$ удовлетворяют условию (12.8), и пусть Y — случайная величина, математическое ожидание которой существует. Положим $Y_n = E(Y | \mathfrak{B}_n)$. Величины Y_n $\mathfrak{B}_n$ -измеримы и поэтому удовлетворяют соотношению (12.9). Следовательно, последовательность $\{Y_n\}$ есть мартингал относительно $\{\mathfrak{B}_n\}$. ►

Возвратимся теперь к примеру а) и покажем невозможность систем игры довольно общего типа. Пусть $\{Y_n\}$ — мартингал по

отношению к $\{\mathfrak{B}_n\}$. Для описания возможности пропускать n -й тур введем *решающую функцию* e_n ; e_n есть $\mathfrak{B}_{n-1}$ -измеримая¹⁾ случайная величина, принимающая лишь значения 0 и 1. Если $e_n = 0$, то игрок пропускает n -й тур; при $e_n = 1$ он делает ставку, и в этом случае его выигрыш в n -м туре равен $Y_n - Y_{n-1}$. Обозначив накопленный выигрыш вплоть до n -го тура включительно через Z_n , имеем

$$Z_n = Z_{n-1} + e_n [Y_n - Y_{n-1}]. \quad (12.11)$$

Из соображений индукции ясно, что величины Z_n имеют математические ожидания. Кроме того, величины Z_{n-1} , e_n и Y_{n-1} $\mathfrak{B}_{n-1}$ -измеримы, и поэтому [см. гл. V, (11.6)]

$$E(Z_n | \mathfrak{B}_{n-1}) = Z_{n-1} + e_n [E(Y_n | \mathfrak{B}_{n-1}) - Y_{n-1}]. \quad (12.12)$$

Так как последовательность $\{Y_n\}$ — мартингал, то выражение в квадратных скобках обращается в нуль, и, следовательно, последовательность $\{Z_n\}$ тоже является мартингалом. Тем самым мы доказали теорему Халмоша, утверждающую

Невозможность систем игры. Мартингал $\{Y_n\}$ переводится всякой последовательностью решающих функций $e_1, e_2, \dots$ снова в мартингал $\{Z_n\}$.

Наиболее важным частным случаем этого результата является случай так называемой *произвольной остановки*. Система игры с произвольной остановкой — это такая система, когда в первых N турах игрок принимает участие, а от всех последующих уклоняется; N -й тур является последним. Случайная величина N (*момент остановки*) удовлетворяет условию $\{N > k\} \in \mathfrak{B}_k$ (в обозначениях теоремы $e_k = 1$ при $N > k - 1$ и $e_k = 0$ при $N \leq k - 1$). Мы имеем, стало быть, такое

Следствие. Произвольная остановка переводит мартингал в мартингал.

Примеры. е) Простое случайное блуждание на прямой начинается в начале координат. Частица переходит вправо на один шаг с вероятностью p и влево на один шаг с вероятностью $q = 1 - p$. Если S_n — положение частицы в момент n , то, как легко видеть, величины $Y_n = (q/p)^{S_n}$ образуют мартингал, причем $E(Y_n) = 1$ и $E(Y_0) = 1$ [частный случай примера в)].

В задаче о разорении случайное блуждание останавливается в тот момент, когда оно впервые достигает одну из точек — a или b (a, b — целые положительные числа). В таком модифицированном процессе $-a \leq S_n \leq b$, и с вероятностью единица S_n

¹⁾ Это условие обеспечивает принятие решения на основе предшествующих наблюдений. Никакая математическая теория не может опровергнуть возможности предвидения будущего, однако, из наших моделей мы обязаны эту возможность исключить.

становится равной и продолжает оставаться равной b , либо $-a$. Обозначим соответствующие вероятности x и $1-x$. Поскольку величины S_n ограничены,

$$E(Y_n) = x \left(\frac{q}{p}\right)^b + (1-x) \left(\frac{q}{p}\right)^{-a}.$$

Но $E(Y_n) = 1$, так как математические ожидания образующих мартингал величин совпадают. Следовательно, $x(q/p)^b + (1-x) \times (q/p)^{-a} = 1$, и тем самым мы получили линейное уравнение для определения вероятности x остановки процесса в точке b , которая иными методами была вычислена в 1, XIV, 2. Исползованный метод не работает при $p = q = 1/2$, однако в этом случае последовательность S_n образует мартингал, и аналогичное рассуждение показывает, что $x = a/(a+b)$. Хотя результат, к которому мы пришли, известен и весьма элементарен, исползованный метод иллюстрирует возможные применения теории мартингалов.

ж) *О системах игры.* Пусть $\{X_n\}$ — последовательность независимых случайных величин, такая, что X_n принимает значения $\pm 2^n$ с вероятностью $\frac{1}{2}$ каждое. Чтобы решить, участвовать ли ему в n -м туре, игрок бросает монету. Вероятность того, что впервые он начнет участвовать в игре в n -м туре, равна 2^{-n} , и в этом случае его выигрыш равен $\pm 2^n$. При этом *выигрыш игрока при его первом вступлении в игру является случайной величиной без математического ожидания.* Теорема о системах игры связана, следовательно, с тем обстоятельством, что мы не меняем временного параметра. ►

Нередко приходится иметь дело с абсолютными величинами и неравенствами, и в этих случаях удобно иметь название для процесса, удовлетворяющего условию (12.9) с заменой равенства неравенством.

Определение 2. Последовательность $\{Y_n\}$ называется *субмартингалом*¹⁾, если она удовлетворяет определению (12.9) с заменой знака равенства на знак $\geq$.

Отсюда незамедлительно вытекает, что каждый субмартингал удовлетворяет более сильным условиям:

$$E(Y_{n+1} | \mathfrak{B}_k) \geq Y_k, \quad k = 1, \dots, n. \quad (12.13)$$

Лемма. Пусть u — выпуклая функция и последовательность $\{Y_n\}$ является мартингалом. Тогда если математические ожидания величин $u(Y_n)$ существуют, то последовательность $\{u(Y_n)\}$ является субмартингалом. В частности, последовательность $\{|Y_n|\}$ есть субмартингал.

Утверждение леммы непосредственно вытекает из неравенства Иенсена [гл. V, (8.6)], которое применимо к условным математи-

¹⁾ Более старый термин «нижний полумартингал» вышел из употребления.

ческим ожиданиям так же, как и к обычным. Согласно этому неравенству, имеем

$$E(u(Y_{n+1}) | \mathfrak{B}_n) \geq u(E(Y_{n+1} | \mathfrak{B}_n)). \quad (12.14)$$

Остается заметить, что правая часть в (12.14) равна $u(Y_n)$.

Аналогично устанавливается, что если $\{Y_n\}$ — субмартингал и u — выпуклая *неубывающая* функция, то последовательность $\{u(Y_n)\}$ тоже является субмартингалом, если только математические ожидания величин $u(Y_n)$ существуют.

§ 13. ЗАДАЧИ

1. Определение (1.2) устойчивого распределения эквивалентно следующему: распределение R устойчиво тогда и только тогда, когда для произвольных постоянных c_1, c_2 найдутся такие постоянные c и γ , что

$$c_1 X_1 + c_2 X_2 \stackrel{d}{=} cX + \gamma.$$

2. *Всякое устойчивое распределение непрерывно.* Достаточно доказать это утверждение для симметричного распределения R . Из (1.3) выведите: если R имеет атом массы p в точке $s > 0$, то R должно иметь атом с массой $\geq p^2$ в каждой точке вида $s(c_m + c_n)/c_{m+n}$ (см. гл. V, 4а). Кроме того, единственный атом массы $p < 1$ в начале координат индуцирует атом массы p^2 в начале координат для свертки $R \star R$, тогда как устойчивость требует атом массы p .

3. Для устойчивости распределения F достаточно выполнения (1.2) при $n=2$ и 3 (П. Леви).

Указание. Произведения вида s/c_2^k , где $j, k=0, \pm 1, \pm 2, \dots$, либо плотны в интервале $0, \infty$, либо представляют собой степени некоторого фиксированного числа c . Нужно показать, что вторая возможность в рассматриваемом случае исключена.

Замечание. Весьма любопытно, что выполнения (1.2) только при $n=2$ еще *недостаточно* для устойчивости F ; см. пример гл. XVII, 3, е) и задачу 10 из гл. IX, 10.

4. Для устойчивого распределения с характеристическим показателем $\alpha=1$ центрирующие постоянные в определении (1.2) удовлетворяют соотношению $\gamma_{mn} = m\gamma_n + n\gamma_m$ [см. (1.6)]. Аналогом (1.8) служит

$$s(X_1 + \gamma \log s) + t(X_2 + \gamma \log t) \stackrel{d}{=} (s+t)(X + \gamma \log(s+t)).$$

5. Если F и G — устойчивые распределения с одним и тем же характеристическим показателем α , то $F \star G$ — тоже устойчивое распределение с характеристическим показателем α . Найти центрирующие постоянные γ_n для $F \star G$, выразив их через соответствующие постоянные для F и G .

6. Из неравенства симметризации гл. V, (5.11), вытекает, что выражение $n[1 - R(c_n x)]$, где R — симметричное устойчивое распределение, ограничено. Вывести отсюда, что R имеет все абсолютные моменты порядка меньше α [использовать гл. V, (6.3)]. С помощью симметризации последнее утверждение переносится на несимметричное R .

7. *Иной вывод распределения Хольцмарка.* Рассмотрим шар радиуса r с центром в начале координат и n звезд (точек), расположенных в нем наудачу и независимо друг от друга. Пусть каждая звезда имеет единичную массу. Обозначим через $X_1, \dots, X_n$ x -компоненты гравитационных сил, соответствующие отдельным звездам, и положим $S_n = X_1 + \dots + X_n$. Устремим r и n к бес-

консчности так, чтобы $\frac{4}{3} p^2 n^{-1} \rightarrow \lambda$. Показать, что распределение величины S_n стремится к симметричному устойчивому распределению с характеристическим показателем $3/2$.

8. Показать, что предыдущая задача, по существу, не изменится, если массу каждой звезды считать случайной величиной с единичным математическим ожиданием и массы различных звезд предполагать взаимно независимыми и не зависящими также от их расположения.

9. *Распределение Хольцмарка в случае четырех измерений.* Четырехмерным аналогом распределения Хольцмарка является симметричное устойчивое распределение с характеристическим показателем $4/3$ (в четырехмерном пространстве гравитационная сила меняется обратно пропорционально кубу расстояния).

10. Схема серий $\{X_{k,n}\}$ с r_n членами в n -й серии может быть преобразована в эквивалентную ей схему с n членами в n -й серии путем добавления фиктивных величин, принимающих лишь значение нуль, и повторения некоторых серий определенное число раз.

11. Пусть $\{X_{k,n}\}$ — нулевая схема серий, причем величины $X_{1,n}, \dots, X_{n,n}$ имеют одинаковое распределение F_n . Стремится ли к нулю вероятность $P\{\max(|X_{1,n}|, \dots, |X_{n,n}|) > \varepsilon\}$?

12. Найти плотность определяемой формулой (6.3) функции восстановления U , когда F имеет плотность а) $f(x) = e^{-x}$ и б) $f(x) = xe^{-x}$.

13. Рассматривается обрывающийся процесс восстановления с функцией распределения F , имеющей плотность $p = ce^{-ct}$. Найти распределения длительности жизни процесса и числа моментов восстановления.

14. *Обобщенный обрывающийся процесс восстановления.* Вместо того чтобы обрывать процесс мгновенно с вероятностью q , позволим ему (тоже с вероятностью q) продолжаться в течение случайного времени с собственным распределением F_0 и затем остановим. Иными словами, моменты восстановления имеют вид $T_1 + \dots + T_n + Y$, где величина Y имеет отличное от величин T распределение. Показать, что распределение V длительности процесса удовлетворяет уравнению восстановления

$$V = qF_0 + F \star V \quad (F(\infty) = 1 - q). \quad (*)$$

15. Показать, что процесс в задаче о «времени ожидания до больших пропусков» сводится к частному случаю процесса, рассмотренного в предыдущей задаче. Привести (7.1) к виду (*).

16. *Процесс Пуассона и теоремы о покрытии.* Напомним, что, согласно примеру гл. III, 3, г), в случае пуассоновского процесса условное распределение моментов восстановления на интервале $0, t$ при условии, что их ровно n , равномерно. Отсюда в силу теоремы о покрытиях (теорема 3 из гл. I, 9) следует, что вероятность $1 - V(t)$ того, что все пропуски по длине меньше ξ , представима в виде

$$1 - V(t) = e^{-ct} \sum_{n=1}^{\infty} \frac{(ct)^{n-1}}{(n-1)!} \varphi_n(t). \quad (**)$$

а) Проверить, что формула (**) действительно определяет решение уравнения (7.1) при $F(x) = 1 - e^{-cx}$.

б) Вывести теорему о покрытиях, считая известным, что определяемое формулой (**) решение уравнения (9.1) единственно. [Такое доказательство является примером доказательства с помощью рандомизации, см. задачу 5 в I, XII, 6.]

17. *Время ожидания первого потерянного вызова* в примере 7, е) можно интерпретировать как полное время жизни обрывающегося процесса, который получается при остановке исходного процесса в момент первого вызова, посту-

пильшего в течение периода занятости¹⁾. Покажите, что распределение H продолжительности периода занятости обрывающегося процесса определяется равенством $H\{dt\} = e^{-at} F\{dt\}$, а времена возвращения имеют распределения $G \star H$. (См. также задачи 3, 4 из гл. XIV, 10.)

18. Пусть V — распределение периода занятости из примера 7, ж) («прибывший последний обслуживается первым»). Показать, что V удовлетворяет уравнению восстановления $V(t) = A(t) + B \star V(t)$, где A и B — несобственные распределения, определяемые формулами $A\{dx\} = e^{-cx} G\{dx\}$ и $B\{dx\} = [1 - G(x)] ce^{-cx} dx$ (показать, что здесь мы имеем дело с обобщенным обрывающимся процессом в смысле задачи 14: процесс восстановления, порожденный распределением B , сопровождается ненарушаемым мертвым периодом, распределение которого пропорционально распределению A).

19. *Малые пропуски в пуассоновском процессе*²⁾. Скажем, что в момент S_n происходит «совпадение», если два момента восстановления S_{n-1} и S_n оказываются на расстоянии $\leq \xi$. Выразить распределение времени ожидания до первого совпадения в терминах распределения V из предыдущего примера.

20. *Обобщение предыдущей задачи о малых пропусках*³⁾. Рассмотрим стандартный процесс восстановления $S_1 = T_1 + \dots + T_n$ с временем ожидания T_j . Написать уравнение восстановления для распределения времени ожидания до первого осуществления события $\{T_n \leq Y_n\}$, где величины Y_n не зависят от процесса и друг от друга и имеют общее распределение G .

21. Пусть $a_1, \dots, a_n$ — конечная числовая последовательность и $S_k = a_1 + \dots + a_k$ — ее частные суммы. Определим последовательно $v_1 = a_n \cup 0$, $v_2 = (v_1 + a_{n-1}) \cup 0$, ..., $v_n = (v_{n-1} + a_1) \cup 0$. Пользуясь индукцией, доказать, что $v_n = \max[0, s_1, \dots, s_n]$. Вывести отсюда теорему § 9.

22. Положим в примере 11, е) $f(x) = 1$ при $0 < x < 1$. Доказать, что функция $g(y) = 2(1-y)$ является стационарной плотностью и что $k^{(n)}(x, y) = g(y)$ при $n \geq 2$. Если $f(x) = ae^{-ax}$, то $g(x) = f(x)$.

23. Определим плотность стохастического ядра на $\overline{0, 1}$, положив

$$k(x, y) = \frac{1}{2} (1-x)^{-1} \quad \text{при } 0 < x < y < 1,$$

$$k(x, y) = \frac{1}{2} x^{-1} \quad \text{при } 0 < y < x < 1.$$

Найти стационарную плотность (она удовлетворяет простому дифференциальному уравнению). Дать вероятностную интерпретацию.

24. Марковский процесс сосредоточен на $\overline{0, 1}$, и при $X_n = x$ величина X_{n+1} равномерно распределена на $\overline{1-x, 1}$. Показать, что функция $2x$ является стационарной плотностью (Т. Угахери).

25. Марковский процесс, сосредоточенный на $\overline{0, \infty}$, определяется следующим образом. Если $X_n = x$, то X_{n+1} равномерно распределена на $\overline{0, 2x}$, $n = 0, 1, \dots$. Показать, используя индукцию, что плотность стохастического ядра, соответствующего переходам за n шагов, равна

$$k^{(n)}(x, y) = \frac{1}{2^n (n-1)!} \left(\lg \frac{2^n x}{y} \right)^{n-1}, \quad \text{если } 0 < y < 2^n x,$$

и $k^{(n)}(x, y) = 0$ в ином случае.

¹⁾ Этот элементарный подход заменяет сложные процедуры, которые описаны в литературе, и приводит к простым точным результатам. Решения и оценки в явном виде см. гл. XI, 6.

²⁾ По поводу других вариантов этой задачи (исследуемых иным методом) см: Gilbert E. N., Pollak H. O., Coincidences in Poisson patterns, *Bell System Tech. J.*, 36 (1957), 1005–1033.

³⁾ Подобно этому можно обобщить задачу о «больших пропусках». Ответ получается аналогичным.

В первой части этой главы показано, что некоторые известные глубокие теоремы анализа можно удивительно легко получить путем вероятностных рассуждений. В § 7 и 8 обсуждаются варианты закона больших чисел, а § 9 содержит частный случай теоремы о сходимости мартингалов и стоит в стороне от основного содержания.

§ 1. ОСНОВНАЯ ЛЕММА. ОБОЗНАЧЕНИЯ

Рассмотрим сначала одномерное распределение G с математическим ожиданием θ и дисперсией σ^2 . Если $X_1, \dots, X_n$ — независимые случайные величины с распределением G , то их арифметическое среднее $M_n = (X_1 + \dots + X_n) n^{-1}$ имеет математическое ожидание θ и дисперсию $\sigma^2 n^{-1}$. При больших n эта дисперсия мала и M_n , как правило, близко к θ . Отсюда следует, что для любой непрерывной функции u значение $u(M_n)$ будет как правило, близко к $u(\theta)$. Это замечание объясняет суть (слабого) закона больших чисел. Оно несколько обобщается в приводимой ниже лемме, которая, несмотря на свою простоту, станет источником ценной информации.

Рассмотрим при $n = 1, 2, \dots$ семейство распределений $F_{n, \theta}$ с математическим ожиданием, равным θ , и дисперсией $\sigma_n^2(\theta)$. Здесь θ — параметр, изменяющийся в конечном или бесконечном интервале. Для математических ожиданий введем обозначение

$$E_{n, \theta}(u) = \int_{-\infty}^{+\infty} u(x) F_{n, \theta}\{dx\}. \quad (1.1)$$

Лемма 1. *Допустим, что функция u ограничена и непрерывна и что $\sigma_n^2(\theta) \rightarrow 0$ при любом θ . Тогда*

$$E_{n, \theta}(u) \rightarrow u(\theta). \quad (1.2)$$

Эта сходимость равномерна в любом замкнутом интервале, где $\sigma_n^2(\theta) \rightarrow 0$ равномерно и функция u равномерно непрерывна.

Доказательство. Очевидно, что

$$|E_{n, \theta}(u) - u(\theta)| \leq \int_{-\infty}^{+\infty} |u(x) - u(\theta)| F_{n, \theta}\{dx\}. \quad (3)$$

Существуют δ , зависящее от θ , и ε , такие, что при $|x - \theta| < \delta$ подинтегральное выражение меньше ε . Вне этой окрестности подинтегральное выражение меньше некоторой постоянной M , и по неравенству Чебышева [гл. V, неравенство (7.2)] вероятность попадания в область $|x - \theta| > \delta$ меньше чем $\sigma_n^2(\theta) \delta^{-2}$. Таким образом, правая часть не превосходит 2ε , если n столь велико, что $\sigma_n^2(\theta) < \varepsilon \delta^2 / M$. Эта граница не зависит от θ , если $\sigma_n^2(\theta) \rightarrow 0$ равномерно и если u равномерно непрерывна. ►

Примеры. а) Если $F_{n, \theta}$ — биномиальное распределение, сосредоточенное в точках k/n ($k = 0, \dots, n$), то $\sigma_n^2(\theta) = \theta(1 - \theta)n^{-1} \rightarrow 0$ и

$$\sum_{k=0}^n u\left(\frac{k}{n}\right) \binom{n}{k} \theta^k (1 - \theta)^{n-k} \rightarrow u(\theta) \quad (1.4)$$

равномерно при $0 \leq \theta \leq 1$. Следствия этого соотношения обсуждаются в § 2.

б) Если $F_{n, \theta}$ — распределение Пуассона, приписывающее вероятность $e^{-n\theta} (n\theta)^k / k!$ точке k/n , то $\sigma_n^2(\theta) = \theta/n$ и

$$e^{-n\theta} \sum_{k=0}^{\infty} u\left(\frac{k}{n}\right) \frac{(n\theta)^k}{k!} \rightarrow u(\theta) \quad (1.5)$$

равномерно в каждом конечном интервале значений θ . Эта формула верна и для n , отличных от целых чисел (продолжение см. в § 5 и 6).

в) Взяв в качестве $F_{n, \theta}$ гамма-распределение с математическим ожиданием θ и дисперсией θ/n , получаем, что

$$\frac{1}{(n-1)!} \int_0^{\infty} u(x) \left(\frac{nx}{\theta}\right)^{n-1} e^{-nx/\theta} \frac{n dx}{\theta} \rightarrow u(\theta) \quad (1.6)$$

равномерно в каждом конечном интервале. Эта формула [с заменой $(n-1)!$ на $\Gamma(n)$] также верна для нецелых значений n . В § 6 будет показано, что формула (1.6) представляет собой формулу обращения для преобразований Лапласа.

г) Статистики часто сталкиваются с положением, описанным в начале этого параграфа, но рассматривают математическое ожидание θ как неизвестный параметр, который нужно оценить по результатам наблюдений. На языке статистики соотношение (1.2) означает, что $u(M_n)$ является асимптотически несмещенной оценкой для неизвестного параметра $u(\theta)$ [оценка была бы несмещенной, если бы правая и левая части в (1.2) совпадали]. ►

Мы увидим, что каждый из этих примеров приводит к важным результатам, имеющим самостоятельный интерес, но для дальнейшего развития теории необходимы некоторые приготовления.

Обозначения для разностей

В нескольких следующих параграфах мы будем применять удобные обозначения, принятые в исчислении конечных разностей. Пусть дана конечная или бесконечная числовая последовательность $a_0, a_1, \dots$. Разностный оператор Δ определяется равенством $\Delta a_i = a_{i+1} - a_i$. Это равенство порождает новую последовательность $\{\Delta a_i\}$. Применяя оператор Δ еще раз, мы получим последовательность с элементами

$$\Delta^2 a_i = \Delta a_{i+1} - \Delta a_i = a_{i+2} - 2a_{i+1} + a_i.$$

Продолжая подобным образом, мы можем определить r -ю степень Δ^r рекуррентной формулой $\Delta^r = \Delta \Delta^{r-1}$. Легко проверить, что

$$\Delta^r a_i = \sum_{j=0}^r \binom{r}{j} (-1)^{r-j} a_{i+j}. \quad (1.7)$$

Для завершенности мы определим Δ^0 как тождественный оператор: $\Delta^0 a_i = a_i$. Тогда равенство (1.7) справедливо при всех $r \geq 0$ (разумеется, если $a_0, \dots, a_n$ — конечная последовательность, то изменение r ограничено).

Можно избежать многих утомительных вычислений, если один раз вывести одно любопытное соотношение, справедливое для двух произвольных последовательностей $\{a_i\}$ и $\{c_i\}$. Это соотношение позволяет выразить разности $\Delta^r a_i$ через разности $\Delta^r c_i$ и наоборот.

Для его вывода умножим (1.7) на $\binom{v}{r} c_r$ и просуммируем по $r = 0, \dots, v$. Коэффициент при a_{i+j} равен

$$\sum_{r=j}^v \binom{v}{r} \binom{r}{j} (-1)^{r-j} c_r = (-1)^{v-j} \binom{v}{j} \sum_{k=0}^{v-j} \binom{v-j}{k} (-1)^{v-j-k} c_{j+k}$$

(в правой части введен новый индекс суммирования $k = r - j$). Последняя сумма равна $\Delta^{v-j} c_j$, и таким образом нами получена формула взаимности

$$\sum_{i=0}^v c_i \binom{v}{i} \Delta^i a_i = \sum_{j=0}^v a_{i+j} \binom{v}{j} (-1)^{v-j} \Delta^{v-j} c_j. \quad (1.8)$$

Примеры. а) (Формула обращения.) Рассмотрим последовательность $\{a_i\}$ с $a_i = 1$ при всех i . Тогда $\Delta^0 a_i = 1$, а все другие разности равны нулю, так что (1.8) сводится к равенству

$$c_0 = \sum_{j=0}^v \binom{v}{j} (-1)^{v-j} \Delta^{v-j} c_j. \quad (1.9)$$

Если вместо $\{c_j\}$ рассмотреть последовательность $\{c_{k+j}\}$, то видно, что (1.9) остается справедливым, если c_0 и c_j заменить соответ-

венно на c_k и c_{k+j} . Таким образом, (1.9) представляет собой формулу обращения, выражающую члены последовательности через их разности. Величина v выбирается здесь произвольным образом.

б) Фиксируем $0 < \theta < 1$ и положим $c_r = \theta^r$. Тогда

$$\Delta^k c_r = \theta^r (1 - \theta)^k (-1)^k$$

и формула (1.8) принимает вид

$$\sum_{r=0}^v \theta^r \binom{v}{r} \Delta^r a_i = \sum_{j=0}^v a_{i+j} \binom{v}{j} \theta^j (1 - \theta)^{v-j}. \quad (1.10) \blacktriangleright$$

Мы будем часто иметь дело с последовательностями вида $a_k = u(x + kh)$, которые получаются с помощью какой-либо функции u при фиксированных x и шаге $h > 0$. В этих случаях удобно заменить разностный оператор Δ на *разностное отношение* $\Delta_h = h^{-1} \Delta$.

Тогда

$$\Delta_h u(x) = [u(x+h) - u(x)]/h \quad (1.11)$$

и вообще

$$\Delta_h^r u(x) = h^{-r} \sum_{j=0}^r \binom{r}{j} (-1)^{r-j} u(x+jh). \quad (1.12)$$

В частности, $\Delta_h^0 u(x) = u(x)$.

§ 2. ПОЛИНОМЫ БЕРНШТЕЙНА. АБСОЛЮТНО МОНОТОННЫЕ ФУНКЦИИ

Мы возвращаемся к предельному соотношению (1.4). В левой части стоит полином, который называется *полиномом Бернштейна степени n* , соответствующим данной функции u . Чтобы подчеркнуть эту связь с функцией u , будем обозначать полином $B_{n,u}$. Таким образом,

$$B_{n,u}(\theta) = \sum_{j=0}^n u(jh) \binom{n}{j} \theta^j (1 - \theta)^{n-j}, \quad (2.1)$$

где для удобства мы ввели $h = n^{-1}$. Сравнивая (2.1) с (1.10), видим, что $B_{n,u}$ можно записать в другой форме:

$$B_{n,u}(\theta) = \sum_{r=0}^n \binom{n}{r} (h\theta)^r \Delta_h^r u(0), \quad h = \frac{1}{n}. \quad (2.2)$$

Поразительно много глубоких выводов может быть получено из наблюдения, что представления (2.1) и (2.2) для полиномов Берн-

штейна эквивалентны. Однако, прежде чем двигаться в этом направлении, мы переформулируем для порядка результат, полученный в примере 1, а).

Теорема 1. Если функция u непрерывна в замкнутом интервале $\overline{0, 1}$, то полиномы Бернштейна $B_{n, n}(\theta)$ равномерно сходятся к $u(\theta)$.

Иными словами, для любого заданного $\varepsilon > 0$ и всех достаточно больших n

$$|B_{n, n}(\theta) - u(\theta)| < \varepsilon, \quad 0 \leq \theta \leq 1. \quad (2.3)$$

Знаменитая теорема Вейерштрасса об аппроксимации утверждает возможность равномерной аппроксимации некоторыми полиномами. Приведенная выше теорема сильнее, так как в ней явно указаны аппроксимирующие полиномы. Доказательство принадлежит С. Н. Бернштейну.

В качестве первого применения двойственного представления полиномов Бернштейна мы получим характеристизацию функций, которые могут быть представлены степенными рядами с положительными коэффициентами:

$$u(x) = p_0 + p_1 x + p_2 x^2 + \dots, \quad p_j \geq 0, \quad 0 \leq x < 1. \quad (2.4)$$

Очевидно, что такая функция обладает производными любого порядка и

$$u^{(n)}(x) \geq 0, \quad 0 < x < 1. \quad (2.5)$$

Для многих проблем в анализе важен тот факт, что и обратное утверждение справедливо, т. е. любая функция, обладающая свойством (2.5), допускает представление (2.4) в виде степенного ряда. Это было впервые обнаружено С. Н. Бернштейном, но обычные доказательства не отличаются простотой и не основаны на интуиции. Мы покажем, что представление (2.2) для полиномов Бернштейна приводит к простому доказательству и полезному результату, состоящему в том, что свойства (2.4) и (2.5) эквивалентны третьему свойству, а именно

$$\Delta_h^k u(0) \geq 0, \quad k = 0, \dots, n-1, \quad h = \frac{1}{n}. \quad (2.6)$$

Все это представляет интерес для теории вероятностей, поскольку если $\{p_r\}$ — распределение вероятностей на множестве целых чисел, то формула (2.4) определяет соответствующую производящую функцию (см. 1; гл. XI). Именно поэтому мы рассматриваем задачу *характеризации производящих функций*. Среди всех наших функций u они выделяются очевидным условием нормировки $u(1) = 1$. Однако пример $u(x) = (1-x)^{-1}$ показывает, что функции, представимые рядом (2.4), не обязаны быть ограниченными.

Теорема 2. Для непрерывной функции u , определенной для $0 \leq x < 1$, свойства (2.4), (2.5) и (2.6) равносильны.

Функции с этим свойством называются абсолютно монотонными функциями на $\overline{0, 1}$.

Доказательство. Мы проведем доказательство в два этапа и рассмотрим сначала только вероятностные производящие функции. Иными словами, мы предположим сейчас, что u непрерывна в замкнутом интервале $\overline{0, 1}$ и что $u(1) = 1$.

Очевидно, что (2.4) влечет (2.5). Если (2.5) справедливо, то u и все ее производные являются монотонными функциями. Из монотонности u следует, что $\Delta u(x) \geq 0$, а из монотонности u' следует, что $\Delta u(x)$ монотонно зависит от x и поэтому $\Delta^2 u(x) \geq 0$. Используя индукцию, заключаем, что (2.5) влечет (2.6), а также $\Delta^n u(0) \geq 0$.

Предположим, что (2.6) выполнено при $k=0, \dots, n$. Ввиду представления (2.2) полином $B_{n,u}$ имеет неотрицательные коэффициенты, а из (2.1) следует, что $B_{n,u}(1) = 1$. Таким образом, $B_{n,u}$ — вероятностная производящая функция, а по теореме непрерывности из 1; гл. XI, 6 (или из гл. VIII, 6 ниже) предел $u = \lim B_{n,u}$ сам по себе является вероятностной производящей функцией. (Предположение $u(1) = 1$ гарантирует, что сумма коэффициентов равна единице.) На этом завершается доказательство для ограниченных функций.

Если функция u не ограничена вблизи 1, то положим

$$v(x) = u\left(\frac{m-1}{m}x\right) / u\left(\frac{m-1}{m}\right), \quad 0 \leq x \leq 1, \quad (2.7)$$

где m — произвольное целое число. К функции v применимо предыдущее доказательство. Поэтому каждое из свойств (2.5) и (2.6) влечет за собой справедливость разложения (2.4) в степенной ряд по крайней мере при $0 < x \leq (m-1)/m$. Вследствие единственности представления в виде степенного ряда и произвольности m отсюда вытекает, что (2.4) выполняется при $0 \leq x < 1$. ►

§ 3. ПРОБЛЕМА МОМЕНТОВ

В последней теореме мы столкнулись с последовательностями $\{a_k\}$, у которых все разности $\Delta^r a_k$ были положительными. В данном параграфе мы будем иметь дело с классом последовательностей, в некотором смысле родственных предыдущим, — разности членов этих последовательностей чередуются по знаку, т. е. речь идет о таких последовательностях $\{c_n\}$, что

$$(-1)^r \Delta^r c_k \geq 0, \quad r = 0, 1, \dots \quad (3.1)$$

Такие последовательности называются *вполне монотонными*.

Пусть F — вероятностное распределение на $0, 1$. Обозначим через $E(u)$ интеграл от u по отношению к F . Момент порядка k распределения F определяется равенством

$$c_k = E(X^k) = \int_0^1 x^k F \{dx\} \quad (3.2)$$

(здесь подразумевается, что интервал интегрирования замкнут).

Вычисляя последовательные разности, находим, что

$$(-1)^r \Delta^r c_k = E(X^k (1-X)^{r-k}), \quad (3.3)$$

и поэтому последовательность моментов $\{c_k\}$ является вполне монотонной. Пусть теперь u — произвольная непрерывная функция на $0, 1$. Вычислим интеграл от выражения (2.1) для полинома Бернштейна $B_{n,n}$ по отношению к F . В силу (3.3) получим

$$E(B_{n,u}) = \sum_{j=0}^n u(jh) \binom{n}{j} (-1)^{n-j} \Delta^{n-j} c_j = \sum_{j=0}^n u(jh) p_j^{(n)}, \quad (3.4)$$

где $h=1/n$ и где для сокращения мы обозначили

$$p_j^{(n)} = \binom{n}{j} (-1)^{n-j} \Delta^{n-j} c_j. \quad (3.5)$$

Если выбрать $u(x)=1$, то получим $B_{n,u}(x)=1$ при всех x и, следовательно, сумма $p_j^{(n)}$ равна единице. Это означает, что при каждом значении n числа $p_j^{(n)}$ задают распределение вероятностей, приписывающее массу $p_j^{(n)}$ точке $jh=j/n$ (здесь всюду $j=0, \dots, n$). Мы обозначим это вероятностное распределение через F_n , а вычисленное относительно F_n математическое ожидание — через E_n . Тогда (3.4) сводится к равенству $E_n(u) = E(B_{n,u})$. В силу равномерной сходимости $B_{n,u} \rightarrow u$ отсюда следует, что

$$E_n(u) \rightarrow E(u). \quad (3.6)$$

До сих пор мы понимали под $\{c_k\}$ последовательность моментов, соответствующих данному распределению F . Можно, однако, считать $\{c_k\}$ произвольной вполне монотонной последовательностью и снова определить $p_j^{(n)}$ по формуле (3.5). Величины $p_j^{(n)}$ по определению неотрицательны, и мы покажем, что их сумма равна c_0 . Действительно, в силу основной формулы (1.8)

$$\sum_{j=0}^n u(jh) p_j^{(n)} = \sum_{r=0}^n c_r \binom{n}{r} h^r \Delta^r u(0). \quad (3.7)$$

Для функции u , тождественно равной единице, правая часть (3.7) сводится к c_0 , что доказывает наше утверждение.

Таким образом, мы видим, что любая вполне монотонная последовательность $\{c_k\}$, удовлетворяющая простейшему условию нормировки $c_0 = 1$, определяет распределение вероятностей $\{p_i^{(n)}\}$, причем математическое ожидание $E_n(u)$ относительно этого распределения задается формулой (3.7). Интересно посмотреть, что произойдет при $n \rightarrow \infty$. Пусть для простоты u будет полиномом степени N . Так как $h = 1/n$, то несложно обнаружить, что $\Delta^r u(0) \rightarrow u^{(r)}(0)$.

Кроме того, $n(n-1)\dots(n-r+1)h^r \rightarrow 1$. Ряд в правой части формулы (3.7) содержит по меньшей мере $N+1$ слагаемых, и теперь можно сделать вывод, что для каждого полинома степени N при $n \rightarrow \infty$

$$E_n(u) \rightarrow \sum_{r=0}^N \frac{c_r}{r!} u^{(r)}(0). \quad (3.8)$$

В частности, если $u(x) = x^r$, то

$$E_n(X^r) \rightarrow c_r. \quad (3.9)$$

Другими словами, момент порядка r распределения вероятностей F_n стремится к c_r . Поэтому весьма правдоподобно, что должно существовать вероятностное распределение F , r -й момент которого совпадает с c_r . Сформулируем это в виде теоремы.

Теорема 1. *Моменты c_r распределения вероятностей F образуют вполне монотонную последовательность с $c_0 = 1$. Обратно, произвольная вполне монотонная последовательность $\{c_r\}$, подвергнутая нормировке $c_0 = 1$, совпадает с последовательностью моментов единственного распределения вероятностей.*

Этот результат, принадлежащий Хаусдорфу, по праву считается глубоким и мощным. Систематическое использование функционального анализа постепенно приводило к упрощению доказательств, однако даже самые лучшие чисто аналитические доказательства остаются относительно сложными. Настоящий подход является новым и служит хорошей иллюстрацией того, как вероятностные рассуждения могут упрощать и обнажать интуитивный смысл сложных аналитических построений. Мы не только докажем сформулированную теорему, но и дадим явную формулу для F .

Соотношение (3.8) указывает нам, что для любого полинома u математические ожидания $E_n(u)$ сходятся к конечному пределу. Как следует из теоремы (2.3) о равномерной аппроксимации, то же самое справедливо для любой непрерывной на $0, 1$ функции u . Обозначим через $E(u)$ предел $E_n(u)$. Таким образом, соотношение (3.6) справедливо при любых условиях, однако если в предположения теоремы входит произвольная вполне монотонная последо-

вательность $\{c_k\}$, то мы должны доказать¹⁾, что существует распределение вероятностей F , такое, что предел $E(u)$ совпадает с математическим ожиданием u , вычисленным относительно распределения F .

Для данных $0 \leq t \leq 1$ и $\varepsilon > 0$ обозначим через $u_{t, \varepsilon}$ непрерывную функцию на $[0, 1]$, которая обращается в нуль при $x \geq t + \varepsilon$, равна 1 при $x \leq t$ и линейна при x между t и $t + \varepsilon$. Пусть $U_\varepsilon(t) = E_\varepsilon(u_{t, \varepsilon})$. Если $t < \tau$, то, очевидно, $u_{t, \varepsilon} \leq u_{\tau, \varepsilon}$ и максимум разности $u_{\tau, \varepsilon} - u_{t, \varepsilon}$ не превосходит $(\tau - t)/\varepsilon$. Отсюда при фиксированном $\varepsilon > 0$ следует, что $U_\varepsilon(t)$ есть непрерывная неубывающая функция t . С другой стороны, при фиксированном t величина $U_\varepsilon(t)$ при $\varepsilon \rightarrow 0$ может только убывать. Следовательно, $U_\varepsilon(t)$ стремится к пределу, который мы обозначим через $F(t)$. Функция $F(t)$ не убывает при изменении от 0 до 1. Она автоматически непрерывна справа²⁾, хотя это не так важно, поскольку этого можно достичь всегда, изменяя значение функции F в точках ее скачков.

Для функции распределения F распределения (3.5) мы получаем тривиальным образом из определения $u_{t, \varepsilon}$ при $\delta > \varepsilon$:

$$E_n(u_{t-\delta, \varepsilon}) \leq F_n(t) \leq E_n(u_{t, \varepsilon}). \quad (3.10)$$

При $n \rightarrow \infty$ два крайних члена неравенства (3.10) стремятся к $U_\varepsilon(t - \delta)$ и $U_\varepsilon(t)$ соответственно. Если t и $t - \delta$ суть точки непрерывности F , мы, полагая $\varepsilon \rightarrow 0$, делаем вывод о том, что все предельные точки последовательности $\{F_n(t)\}$ лежат между $F(t - \delta)$ и $F(t)$. Наконец, устремляя $\delta \rightarrow 0$, заключаем, что $F_n(t) \rightarrow F(t)$ в каждой точке t непрерывности F . Соотношение $F_n(t) \rightarrow F(t)$ означает, что

$$\sum_{j \leq nt} \binom{n}{j} (-1)^{n-j} \Delta^{n-j} c_j \rightarrow F(t). \quad (3.11)$$

Интегрируя по частям выражение для момента порядка r распределения F_n при $r \geq 1$, получаем

$$E_n(X^r) = 1 - r \int_0^1 x^{r-1} F_n(x) dx. \quad (3.12)$$

Левая часть здесь стремится, как показано в (3.9), к c_r , а так как $F_n \rightarrow F$, то отсюда следует, что r -й момент F совпадает с c_r . Это завершает доказательство теоремы 1. ►

¹⁾ Здесь мы могли бы закончить доказательство, так как утверждение содержится в любом из двух результатов, доказанных в других местах этой книги: а) в теореме Рисса о представлении из гл. V, 1, поскольку $E(u)$ очевидным образом представляет собой положительный линейный функционал с нормой 1; б) в основной теореме о сходимости из гл. VIII, 1. Доказательство здесь частично совпадающее с доказательством в гл. V, 1) дано для того, чтобы сделать эту главу независимой от других и чтобы прийти к формуле обращения.

²⁾ Для проверки см. доказательство теоремы о представлении Рисса в гл. V, 1.

Отметим, что для произвольной функции распределения, соответствующей распределению вероятностей на $\overline{0, 1}$, соотношение (3.11) представляет собой формулу обращения, выражающую F в терминах своих моментов. Сформулируем соответствующее утверждение.

Теорема 2. Для вероятностного распределения F из теоремы 1 предельная формула (3.11) выполняется в каждой точке непрерывности.

Чтобы избежать неправильных представлений, следует подчеркнуть, что положение существенно меняется при переходе к распределениям, не сосредоточенным ни на каком конечном интервале. В общем случае *распределение не определяется однозначно своими моментами*.

Пример. Логарифмически нормальное распределение не определяется своими моментами. Про положительную случайную величину X говорят, что она имеет логарифмически нормальное распределение, если величина $\log X$ нормально распределена. При стандартном нормальном распределении плотность случайной величины X равна

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-1} e^{-\frac{1}{2}(\log x)^2}, \quad x > 0,$$

и $f(x) = 0$, $x \leq 0$. Введем при $-1 \leq a \leq 1$

$$f_a(x) = f(x) [1 + a \sin(2\pi \log x)]. \quad (3.13)$$

Мы утверждаем, что плотность вероятности f_a имеет в точности те же самые моменты, что и f . Поскольку $f_a \geq 0$, достаточно показать, что

$$\int_0^{\infty} x^k f(x) \sin(2\pi \log x) dx = 0, \quad k = 0, 1, \dots$$

Подстановки $\log x = t = y + k$ сводят этот интеграл к интегралам

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2 + kt} \sin(2\pi t) dt = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}k^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} \sin(2\pi y) dy,$$

а последний интеграл здесь равен нулю, так как подынтегральное выражение представляет собой четную функцию. (Этот интересный пример предложен К. Хейди.)

Полученный отрицательный результат не должен порождать необоснованного пессимизма, так как источник беспокойства легко устраняется введением подходящих условий регулярности. Наилучший результат содержится в теореме Карлемана, утверждающей, что распределение F на $-\infty, \infty$ *однозначно определяется своими моментами*, если

$$\sum \mu_{2n}^{-1/(2n)} = \infty, \quad (3.14)$$

т. е. если ряд расходится. В этой книге мы докажем более слабое утверждение, а именно, что F однозначно определяется своими моментами, если степенной ряд $\sum \mu_{2n} t^{2n}/(2n)!$ сходится в некотором интервале (см. § 6 и гл. XV, 4). Оба критерия ограничивают скорость роста μ_{2n} . Знание конечного числа моментов $\mu_0, \mu_1, \dots, \mu_n$ даже в самой общей ситуации приводит

к полезным неравенствам для F , аналогичным выведенным в гл. V, 7, для $n=1^1$).

§ 4*). ПРИМЕНЕНИЕ К СИММЕТРИЧНО ЗАВИСИМЫМ СЛУЧАЙНЫМ ВЕЛИЧИНАМ

Мы выведем здесь один красивый результат, принадлежащий Б. де Финетти. Он может служить типичным примером того, сколь легко теорема 1 из § 3 приводит к удивительным результатам.

Определение. Случайные величины $X_1, \dots, X_n$ называются симметрично зависимыми, если все $n!$ перестановок $(X_{k_1}, \dots, X_{k_n})$ имеют одно и то же n -мерное распределение. Случайные величины, образующие бесконечную последовательность $\{X_n\}$, называются симметрично зависимыми, если $X_1, \dots, X_n$ симметрично зависимы при любом n .

Как мы увидим из примеров, конечные и бесконечные последовательности существенно отличаются друг от друга. Мы рассмотрим здесь частный случай бесконечной последовательности $\{X_n\}$ симметрично зависимых величин, принимающих только значения 0 и 1. В следующей ниже теореме утверждается, что распределение такого процесса $\{X_n\}$ получается рандомизацией биномиального распределения. Как обычно, мы положим $S_n = X_1 + \dots + X_n$ и будем называть событие $\{X_k = 1\}$ «успехом».

Теорема. Каждой бесконечной последовательности симметрично зависимых случайных величин, принимающих только значения 0 и 1, соответствует такое сосредоточенное на $\overline{0, 1}$ распределение F , что

$$\begin{aligned} P\{X_1=1, \dots, X_k=1, X_{k+1}=0, \dots, X_n=0\} = \\ = \int_0^1 \theta^k (1-\theta)^{n-k} F\{d\theta\} \end{aligned} \quad (4.1)$$

и

$$P\{S_n = k\} = \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} F\{d\theta\}. \quad (4.2)$$

Доказательство. Для краткости обозначим левую часть (4.1)

¹) Первые сильные результаты были получены Марковым и Стилтесом около 1884 г. Современная литература на эту тему неисчерпаема. См., например, Wald A., Trans. Amer. Math. Soc., 46 (1939), 280—306; Royden H. L., Appl. Math. Statist., 24 (1953), 361—376 [приведены границы для $F(x) - F(-x)$]. Общий обзор см. в монографии: Shohat J. A., Tamarkin J. D., The problem of moments, New York, 1943 (Math. Surveys № 1). См. также Karlin S., Studden W. (1966).

*) Материал не используется в дальнейшем.

через $p_{k,n}$, $0 \leq k \leq n$. Положим $c_0 = 1$, и при $n = 1, 2, \dots$

$$c_n = p_{n,n} = P\{X_1 = 1, \dots, X_n = 1\}. \quad (4.3)$$

Тогда из вероятностных соображений

$$p_{n-1,n} = p_{n-1,n-1} - p_{n,n} = -\Delta c_{n-1}, \quad (4.4)$$

и поэтому

$$p_{n-2,n} = p_{n-2,n-1} - p_{n-1,n} = \Delta^2 c_{n-2}. \quad (4.5)$$

Продолжая этот процесс, получаем при $k \leq n$

$$p_{k,n} = p_{k,n-1} - p_{k+1,n} = (-1)^{n-k} \Delta^{n-k} c_k. \quad (4.6)$$

Все эти величины неотрицательны, и поэтому последовательность $\{c_n\}$ вполне монотонна. Отсюда следует, что c_r есть момент порядка r некоторого вероятностного распределения F и поэтому (4.1) представляет собой просто расшифровку соотношения (4.6). Утверждение (4.2) вытекает отсюда, так как имеется ровно $\binom{n}{k}$ способов появления k успехов в n испытаниях.

Обобщения. Нетрудно применить то же самое рассуждение к случайным величинам, допускающим три значения, однако в этом случае мы будем иметь два свободных параметра и вместо (4.2) получим смесь триномиальных распределений, а F будет двумерным распределением. Вообще теорема и ее доказательство легко могут быть приспособлены к случайным величинам, принимающим только конечное число значений. Этот факт, естественно, позволяет высказать предположение, что в наиболее общем случае последовательность симметрично зависимых случайных величин получается из последовательности независимых случайных величин рандомизацией по некоторому параметру. В специальных случаях трудностей не возникает, однако общая задача трудна тем, что «параметры» не определены отчетливо и могут выбираться непостижимым образом. Несмотря на это, были доказаны весьма общие варианты этой теоремы¹⁾.

Эта теорема позволяет получить условие применимости закона больших чисел и центральной предельной теоремы к симметрично зависимым случайным величинам (см. задачу 21 в гл. VIII, 10).

Первый из приводимых ниже примеров показывает, что в отдельных случаях теорема может приводить к удивительным результатам. Примеры б) и в) показывают, что теорема неверна для конечных последовательностей.

¹⁾ Hewitt E., Savage L. J., Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.*, 80 (1956), 470—501. Изучение с применением мартигалов см. Loève (1963). См. также: Bühlmann H., Austauschbare stochastische Variablen und ihre Grenzwertsätze, Univ. of California Publications in Statistics, 3, № 1 (1960), 1—36.

Примеры. а) В схеме Пойа (см. 1 гл. V, 2) урна содержит первоначально b черных и r красных шаров. После каждого извлечения выбранный шар возвращается и в урну добавляется c шаров того же цвета. Таким образом, вероятность получить черный шар в каждом из n первых испытаний равна

$$c_n = \frac{b(b+c) \dots (b+(n-1)c)}{(b+r) \dots (b+r+(n-1)c)} = \frac{\Gamma\left(\frac{b}{c}+n\right) \Gamma\left(\frac{b+r}{c}\right)}{\Gamma\left(\frac{b+r}{c}+n\right) \Gamma\left(\frac{b}{c}\right)}. \quad (4.7)$$

Положим $X_n = 1$ или 0 в зависимости от того, появился при n -м испытании черный или красный шар. Легко вычислить (см. 1, гл. V, 2), что эти случайные величины симметрично зависимы и потому значение c_n равно n -му моменту некоторого распределения F . По виду формула (4.7) напоминает бета-интеграл (2.5) из гл. II. Анализ показывает, что F есть бета-распределение (4.2) из гл. II с параметрами $\mu = b/c$ и $\nu = r/c$. Можно установить, используя снова бета-интеграл, что (4.1) согласуется с (2.3) из 1, гл. V, а (4.2) с (2.4) из той же главы первого тома.

б) Рассмотрим 6 размещений двух различных шаров в трех клетках. Припишем каждому вероятности $1/6$. Пусть X_i равно 1 или 0 в зависимости от того, занята i -я клетка или пуста. Величины X_i симметрично зависимы, но теорема к ним неприменима. В самом деле, из (4.3) получаем $c_0 = 1$, $c_1 = 1/2$, $c_2 = 1/6$, $c_3 = 0$, и на этом последовательность обрывается. Если бы она была начальным отрезком вполне монотонной последовательности $\{c_n\}$, мы должны были бы иметь $c_4 = c_5 = \dots = 0$. Но тогда было бы $\Delta^4 c_1 = -1/6 < 0$, что невозможно.

в) Пусть случайные величины $X_1, \dots, X_n$ независимы, имеют одно и то же распределение вероятностей и $S_n = X_1 + \dots + X_n$. Положим $Y_k = X_k - n^{-1}S_n$ для $k = 1, \dots, n-1$. Случайные величины $(Y_1, \dots, Y_{n-1})$ симметрично зависимы, но их совместное распределение нельзя представить в форме, соответствующей теореме де Финетти. ►

§ 5*) ОБОБЩЕННАЯ ФОРМУЛА ТЕЙЛОРА И ПОЛУГРУППЫ

Предыдущие три параграфа посвящены следствиям из предельного соотношения в примере 1, а), содержащего биномиальное распределение. Мы перейдем теперь к примеру 1, б), включающему распределение Пуассона. Так как последнее представляет собой предельную форму биномиального распределения, то можно надеяться, что наш подход допускает распространение на настоящую ситуацию. Исходным пунктом этого подхода было тождество (1.10), в котором биномиальное распределение появляется справа. Если положить $\theta = x/v$ и устремить v к бесконечности, то биномиаль-

*) Этот параграф при первом чтении можно пропустить.

ное распределение будет стремиться к распределению Пуассона с математическим ожиданием x и (1.10) перейдет в¹⁾

$$\sum_{r=0}^{\infty} \frac{x^r}{r!} \Delta^r a_i = e^{-x} \sum_{j=0}^{\infty} \frac{x^j}{j!} a_{i+j}. \quad (5.1)$$

Мы используем это тождество при $i=0$ и $a_j = u(jh)$, где u есть произвольная ограниченная непрерывная функция, определенная на $\overline{0, \infty}$, а h есть положительная постоянная. При $x = \theta h$ соотношение (5.1) превращается в

$$\sum_{r=0}^{\infty} \frac{\theta^r}{r!} \Delta_h^r u(0) = e^{-\theta/h} \sum_{j=0}^{\infty} u(jh) \frac{(\theta/h)^j}{j!}. \quad (5.2)$$

В правой части мы узнаем математическое ожидание u , вычисленное относительно распределения Пуассона, а, как мы знаем из примера 1, б), это математическое ожидание стремится к $u(\theta)$. Чтобы записать этот результат в более естественной форме, мы заменим $u(\theta)$ на $u(\theta+t)$, где $t > 0$ произвольно. Таким образом, нами доказана

Теорема. Для любой ограниченной непрерывной функции u на $\overline{0, \infty}$

$$\sum_{r=0}^{\infty} \frac{\theta^r}{r!} \Delta_h^r u(t) \rightarrow u(t+\theta), \quad (5.3)$$

где²⁾ $\theta > 0$ и $h \rightarrow 0+$.

Эта привлекательная теорема была впервые доказана Э. Хилле с помощью значительно более глубоких методов. Левая часть (5.3) представляет собой разложение Тейлора для u с той лишь разницей, что производные заменяются разностными отношениями. Для аналитических функций левая часть приближается к ряду Тейлора, но теорема применима к недифференцируемым функциям. В этом смысле (5.3) представляет собой обобщение разложения в ряд Тейлора и открывает нам его новые черты.

Существует другое истолкование формулы (5.3), приводящее к так называемой экспоненциальной формуле теории полугрупп (см. теорему 2 в гл. X, 9). Левая часть в (5.3) содержит формальный экспоненциальный ряд, который естественно использовать для того, чтобы определить оператор $\exp \theta \Delta_h$. Соотношение

¹⁾ Для непосредственного доказательства тождества (5.1) достаточно заменить $\Delta^r a_i$ определяющим его выражением (1.7). Левая часть тогда становится двойной суммой, а правая часть получается с помощью очевидной перестановки членов ряда.

²⁾ Нужно отметить, что рассуждение остается справедливым, если и θ , и R отрицательны при условии, что u определена на всей прямой,

(5.3) принимает при этом сокращенную форму

$$\exp \theta \Delta_h u(t) \rightarrow u(t + \theta). \quad (5.4)$$

Чтобы записать эту формулу в более удобном виде, введем оператор сдвига¹⁾ $T(\theta)$, переводящий функцию u в функцию u_θ , определяемую равенством $u_\theta(t) = u(t + \theta)$. Тогда $T(0) = 1$, т. е. это единичный оператор и

$$\Delta_h = h^{-1} [T(h) - 1]. \quad (5.5)$$

На операторном языке формула (5.3) принимает вид

$$e^{\theta h^{-1} [T(h) - 1]} \rightarrow T(\theta). \quad (5.6)$$

Основная информация, доставляемая этой формулой, состоит в том, что все семейство операторов $T(\theta)$ определяется поведением $T(h)$ при малых h .

Оглядываясь назад, мы видим, что наш вывод формулы (5.6) применим к значительно более широкому классу операторов. Правая часть в (5.2) представляет собой просто линейную комбинацию значений u и может быть интерпретирована как интерполяционная формула для u . Аналогичное интерполяционное выражение имеет смысл для *любого* семейства операторов $\{T(\theta)\}$, определенных при $\theta > 0$. В самом деле, при фиксированных θ и $h > 0$ оператор

$$A_h(\theta) = e^{-\theta h^{-1}} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\theta}{h}\right)^k T(kh) \quad (5.7)$$

является линейной комбинацией операторов $T(kh)$. Коэффициенты (веса) определяются в соответствии с распределением Пуассона и обладают тем свойством, что при $h \rightarrow 0$ преобладающую роль играют окрестности точки θ , а дополнительные множества имеют веса, стремящиеся к нулю. Это делает правдоподобным предположение, что при разумных определениях сходимости и непрерывности мы будем иметь $A_h(\theta) \rightarrow T(\theta)$ для *любого непрерывного семейства операторов* $T(\theta)$. В частности, если операторы $T(\theta)$ образуют *полугруппу*, то $T(kh) = (T(h))^k$ и интерполяционный оператор $A_h(\theta)$ имеет вид левой части в (5.6). Неудиви-

¹⁾ Это доказательство, принадлежащее М. Риссу, имеется (при несколько более общих условиях) в книге: Хилле Э., Филлипс Р., Функциональный анализ и полугруппы, ИЛ, М., 1962. Как можно понять, авторы не учли, что ссылка на линейную интерполяционную формулу (5.7) как на пуассоновскую *рандомизацию параметра полугруппы* и ссылка на неравенство Чебышева упрощают изложение. Вероятностные соображения отмечались Д. Кендаллом и были в полной мере использованы в статье: Chung K. L., On the exponential formulas of semi group theory, *Math. Scandinavica*, 10 (1962), 153—162.

тельно поэтому, что «экспоненциальная формула» (5.6) *верна вообще для непрерывных полугрупп ограниченных операторов*. Мы вернемся к доказательству этого в гл. X, 9.

§ 6. ФОРМУЛЫ ОБРАЩЕНИЯ ДЛЯ ПРЕОБРАЗОВАНИЯ ЛАПЛАСА

Предыдущий параграф и пример 1, б) основаны на следующей частной форме закона больших чисел. Пусть X — случайная величина, распределенная по закону Пуассона с параметром $\lambda\theta$. Тогда при больших λ вероятность события $|X - \lambda\theta| > \lambda\epsilon$ мала. Поэтому для $P\{X \leq \lambda x\}$ при $\lambda \rightarrow \infty$

$$e^{-\lambda\theta} \sum_{k \leq \lambda x} \frac{(\lambda\theta)^k}{k!} \rightarrow \begin{cases} 0, & \text{если } \theta > x, \\ 1, & \text{если } \theta < x. \end{cases} \quad (6.1)$$

Выражение в левой части получается из (5.2) при условии, что u принимает только значения 0 и 1. Поэтому утверждение (6.1) содержится в теореме § 5. Использование этой формулы в анализе мы проиллюстрируем сейчас на примере преобразований Лапласа (систематически они изучаются в гл. XIII).

Пусть F — распределение вероятностей, сосредоточенное на $0, \infty$. Преобразованием Лапласа функции F называется функция φ , определенная для $\lambda > 0$ формулой

$$\varphi(\lambda) = \int_0^{\infty} e^{-\lambda\theta} F\{d\theta\}. \quad (6.2)$$

Производные $\varphi^{(k)}(\lambda)$ существуют и их можно вычислить формальным дифференцированием:

$$(-1)^k \varphi^{(k)}(\lambda) = \int_0^{\infty} e^{-\lambda\theta} \theta^k F\{d\theta\}. \quad (6.3)$$

Из этих равенств и из формулы (6.1) видно, что в каждой точке непрерывности распределения F

$$\sum_{k \leq \lambda x} \frac{(-1)^k}{k!} \lambda^k \varphi^{(k)}(\lambda) \rightarrow F(x). \quad (6.4)$$

Это очень полезная *формула обращения*. Она показывает, в частности, что *распределение F однозначно определяется своим преобразованием Лапласа*.

Таковыми же рассуждениями получается множество близких формул обращения, применимых при различных обстоятельствах. Например, (1.6) представляет собой формулу обращения для интегралов Лапласа типа

$$\omega(\lambda) = \int_0^{\infty} e^{-\lambda x} u(x) dx. \quad (6.5)$$

Проведем формальное дифференцирование, как в (6.3). Тогда (1.6) означает, что для ограниченной и непрерывной функции u

$$\frac{(-1)^{n-1}}{(n-1)!} \left(\frac{n}{\theta} \right)^n \omega^{(n-1)} \left(\frac{n}{\theta} \right) \rightarrow u(\theta) \quad (6.6)$$

равномерно в каждом конечном интервале.

[Эти формулы обращения справедливы при значительно более общих условиях. Однако сейчас представляется нежелательным затемнять простоту рассуждений балластом новой терминологии. Абстрактный аналог соотношения (6.6) появится в гл. XIII, 9.]

Если распределение F обладает моментами $\mu_1, \dots, \mu_{2n}$, то преобразование Лапласа удовлетворяет неравенствам

$$\sum_{k=0}^{2n-1} \frac{(-1)^k \mu_k \lambda^k}{k!} < \varphi(\lambda) < \sum_{k=0}^{2n} \frac{(-1)^k \mu_k \lambda^k}{k!}. \quad (6.7)$$

Эти неравенства часто используются. Чтобы доказать их, мы будем исходить из хорошо известных неравенств¹⁾

$$\sum_{k=0}^{2n-1} \frac{(-1)^k t^k}{k!} < e^{-t} < \sum_{k=0}^{2n} \frac{(-1)^k t^k}{k!}, \quad t > 0, \quad (6.8)$$

Заменяя t на λt и интегрируя относительно F , получим (6.7). Отсюда следует, в частности, что

$$\varphi(\lambda) = \sum_{k=0}^{\infty} \frac{(-1)^k \mu_k \lambda^k}{k!} \quad (6.9)$$

в любом интервале $0 < \lambda < \lambda_0$, в котором сходится ряд, стоящий справа. Если такой интервал существует, то, согласно теории аналитических функций, ряд в (6.9) однозначно определяет $\varphi(\lambda)$ при всех $\lambda > 0$. Следовательно, последовательность моментов $\mu_0, \mu_1, \dots$ однозначно определяет распределение F , если ряд в (6.9) сходится в некотором интервале $|\lambda| < \lambda_0$. Этот полезный критерий справедлив и для распределений, не сосредоточенных на $0, \infty$. Но в общем случае доказательство основано на применении характеристических функций (см. гл. XV, 4).

§ 7*. ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ ДЛЯ ОДИНАКОВО РАСПРЕДЕЛЕННЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Всюду в этом параграфе мы используем обозначение $S_n = X_1 + \dots + X_n$. Самый старый вариант закона больших чисел состоит в утверждении, что для независимых случайных величин X_k с одним и тем же распределением, математическим ожида-

¹⁾ Простое дифференцирование и индукция показывают, что разность между любыми двумя членами в (6.8) является монотонной функцией от t .

²⁾ Тема этого параграфа имеет отношение к первому периоду развития теории вероятностей, но никакой особой важности для остального содержания книги она не представляет. Мы останавливаемся на ней по причине ее исторического и методологического интереса, а также потому, что имеется много работ, посвященных обращению закона больших чисел.

нием μ и конечной дисперсией ¹⁾

$$P \left\{ \left| \frac{S_n}{n} - \mu \right| > \varepsilon \right\} \rightarrow 0 \quad (7.1)$$

при $n \rightarrow \infty$ и любом фиксированном $\varepsilon > 0$. В начале этой главы отмечалось, что (7.1) вытекает из неравенства Чебышева. Чтобы получить более сильный результат, мы выведем неравенство (разновидность неравенства Чебышева), применимое в случае, когда не существует математических ожиданий. Определим новые случайные величины X'_k , получаемые «урезанием» X_k на произвольном, но фиксированном уровне $\pm s_n$, равенствами

$$X'_k = \begin{cases} X_k, & \text{когда } |X_k| \leq s_n, \\ 0, & \text{когда } |X_k| > s_n. \end{cases} \quad (7.2)$$

Положим

$$S'_n = X'_1 + \dots + X'_n, \quad m'_n = E(S'_n) = nE(X'_1). \quad (7.3)$$

Тогда очевидно, что

$$P \{ |S_n - m'_n| > t \} \leq P \{ |S'_n - m'_n| > t \} + P \{ S_n \neq S'_n \}, \quad (7.4)$$

так как из наступления события, написанного слева, вытекает наступление хотя бы одного из событий, стоящих справа.

Это неравенство справедливо также для зависимых случайных величин с различными распределениями, но здесь мы рассматриваем только одинаково распределенные и независимые случайные величины. Полагая $t = nx$ и применяя неравенство Чебышева к оценке первого члена в правой части (7.4), мы получаем следующий результат.

Лемма. Пусть X_k независимы и имеют одинаковое распределение F . Тогда при $x > 0$

$$P \left\{ \left| \frac{1}{n} S_n - E(X'_1) \right| > x \right\} \leq \frac{1}{n^2 x^2} E(X'^2_1) + nP \{ |X_1| > s_n \}. \quad (7.5)$$

Используя это неравенство, можно доказать закон больших чисел в форме Хинчина, который состоит в утверждении, что (7.1) выполняется при всех $\varepsilon > 0$, если X_k имеют конечное математическое ожидание μ . Доказательство этого факта было бы, по существу, повторением доказательства, приведенного в 1, гл. X, 2, для дискретного случая. Поэтому мы перейдем сразу к более сильному утверждению, указывающему необходимые и достаточные условия. Для формулировки теоремы введем при $t > 0$

$$\tau(t) = [1 - F(t) + F(-t)]t \quad (7.6)$$

¹⁾ Соотношение (7.1) эквивалентно (см. гл. VIII, 2) $n^{-1}S_n - \mu \xrightarrow{P} 0$, где знак $\xrightarrow{P}$ означает сходимость по вероятности,

и

$$\sigma(t) = \frac{1}{t} \int_{-t}^t x^2 F\{dx\} = -\tau(t) + \frac{2}{t} \int_0^t x\tau(x) dx \quad (7.7)$$

(равенство двух выражений в (7.7) проверяется путем интегрирования по частям).

Теорема 1. (Обобщенный слабый закон больших чисел.) Пусть случайные величины X_k независимы и имеют одно и то же распределение F . Для того чтобы существовали такие постоянные μ_n , что при любом $\varepsilon > 0$

$$P\left\{\left|\frac{S_n}{n} - \mu_n\right| > \varepsilon\right\} \rightarrow 0, \quad (7.8)$$

необходимо и достаточно, чтобы ¹⁾ $\tau(t) \rightarrow 0$ при $t \rightarrow \infty$. При выполнении (7.8) можно взять

$$\mu_n = \int_{-n}^n x F\{dx\}. \quad (7.9)$$

Доказательство. а) *Достаточность.* Определим μ_n равенством (7.9). Мы используем урешение (7.2), положив в нем $s_n = n$. Тогда $\mu_n = E(X_1)$ и по предыдущей лемме левая часть соотношения (7.8) будет меньше чем $\varepsilon^{-2}\sigma(n) + \tau(n)$, а это последнее выражение стремится к 0 при $\tau(t) \rightarrow 0$. Таким образом, достаточность доказана.

б) *Необходимость.* Предположим, что (7.8) выполняется. Рассмотрим, так же как в гл. V, 5, величины ${}^0X_k$, полученные непосредственно симметризацией X_k . Сумма этих величин ${}^0S_n$ может быть получена путем симметризации $S_n - n\mu$. Пусть a — медиана случайных величин X_k . Используя неравенства (5.6), (5.10) и (5.7) из гл. V в указанном порядке, мы получим

$$\begin{aligned} 2P\{|S_n - n\mu| > n\varepsilon\} &\geq P\{|{}^0S_n| > 2n\varepsilon\} \geq \\ &\geq \frac{1}{2} [1 - \exp(-nP\{|X_1| > 2n\varepsilon\})] \geq \\ &\geq \frac{1}{2} \left[1 - \exp\left(-\frac{1}{2}nP\{|X_1| > 2n\varepsilon + |a|\}\right)\right]. \end{aligned}$$

В силу (7.8) самая левая часть неравенства стремится к 0. Отсюда следует, что выражение в показателе экспоненциальной функции в самой правой части стремится к 0, а это явно невозможно без того, чтобы $\tau(t) \rightarrow 0$. ▶

Условие $\tau(t) \rightarrow 0$ будет выполнено, когда F имеет математическое ожидание μ . «Урезанные» моменты μ_n будут в этом случае

¹⁾ Из (7.7) следует, что $\tau(t) \rightarrow 0$ влечет $\sigma(t) \rightarrow 0$. Обратное также справедливо, см. задачу 11. Другое доказательство теоремы 1 см. в гл. XVII, 2, а),

стремиться к μ , и поэтому (7.8) будет равносильно классическому закону больших чисел (7.1). Однако классический закон больших чисел в форме (7.1) выполняется и для некоторых случайных величин, не имеющих математического ожидания. Например, если F такое симметричное распределение, что $t[1-F(t)] \rightarrow 0$, то $P\{|n^{-1}S_n| > \varepsilon\} \rightarrow 0$. А математическое ожидание существует только в том случае, когда функция $1-F(t)$ интегрируема для значений t между 0 и ∞ , что представляет собой более сильное условие.

(Интересно отметить, что усиленный закон больших чисел выполняется только для величин, имеющих математические ожидания. См. теорему 4 из § 8.)

Практический смысл закона больших чисел обсуждался в 1; X. Особое внимание было уделено при этом классической теории «безобидных игр». Мы могли видеть, в частности, что даже в том случае, когда математические ожидания существуют, участник «безобидной игры» может прочно занимать проигрышное положение. С другой стороны, анализ петербургской игры показал, что классическая теория применима также к некоторым играм с бесконечным математическим ожиданием, за тем исключением, когда «безобидный вступительный взнос» зависит от номера испытания. Следующая теорема выражает этот факт более точно.

Рассмотрим независимые положительные случайные величины X_k с одинаковым распределением F [таким, что $F(0)=0$]. Величины X_k можно интерпретировать как возможные выигрыши, а a_n — как полный вступительный взнос за n испытаний. Положим

$$\mu(s) = \int_0^s x F\{dx\}, \quad \rho(s) = \frac{\mu(s)}{s[1-F(s)]}, \quad (7.10)$$

Теорема 2. Для того чтобы существовали такие постоянные a_n , что

$$P\{|a_n^{-1}S_n - 1| > \varepsilon\} \rightarrow 0, \quad (7.11)$$

необходимо и достаточно, чтобы $\rho(s) \rightarrow \infty$ при $s \rightarrow \infty$. В этом случае существуют числа s_n , такие, что

$$n\mu(s_n) = s_n \quad (7.12)$$

и 7.11) выполняется при $a_n = n\mu(s_n)$.

Доказательство. а) *Достаточность.* Предположим, что $\rho(s) \rightarrow \infty$. Если n велико, то функция $n\mu(s)/s$ принимает значения > 1 , но она стремится к 0 при $s \rightarrow \infty$. Эта функция непрерывна справа, и предел слева не превосходит предела справа. Если s_n есть нижняя граница всех таких s , что $n\mu(s)s^{-1} \leq 1$, то отсюда следует справедливость (7.12).

Положим $\mu_n = \mu(s_n) = E(X_1')$. Применяя неравенство (7.5) леммы при $x = \varepsilon\mu_n$, получаем

$$P\left\{\left|\frac{S_n}{n\mu_n} - 1\right| > \varepsilon\right\} \leq \frac{1}{\varepsilon^2 n \mu_n^2} E(X_1'^2) + n[1-F(s_n)]. \quad (7.13)$$

¹ Мы увидим в гл. VIII, 9 (теорема 2), что $\rho(s) \rightarrow \infty$ тогда и только тогда, когда $\mu(s)$ медленно меняется на бесконечности. Соотношение (7.11) означает, что $a_n^{-1}S_n \xrightarrow{P} 1$ (См. VIII, 2).

Интегрируя по частям, сведем $E(X_1^2)$ к интегралу от функции $x(1-F(x))$. По предположению эта функция есть $o(\mu(x))$, и поэтому $E(X_1^2) = o(s_n \mu_n)$. Это означает ввиду (7.12), что первый член в правой части неравенства (7.13) стремится к 0. Аналогично из (7.12) и определения $\rho(s)$ в (7.10) следует, что $n[1-F(s_n)] \rightarrow 0$. Таким образом, из неравенства (7.13) получаем соотношение (7.11), где $a_n = n\mu_n$.

б) *Необходимость*. Предположим теперь, что (7.11) выполняется, и используем «урезание» (7.2) с $s_n = 2a_n$. Поскольку $E(X_1^2) \leq s_n \mu_n$, то из основного неравенства (7.5) при $x = \varepsilon a_n / n$ мы получаем

$$P\{S_n > n\mu_1 + \varepsilon a_n\} \leq \frac{2}{\varepsilon^2} \cdot \frac{n\mu_n}{a_n} + n[1-F(2a_n)]. \quad (7.14)$$

Поскольку мы имеем дело с положительными величинами, то

$$P\{S_n < 2a_n\} \leq P\left\{\max_{k \leq n} X_k \leq 2a_n\right\} = F^n(2a_n). \quad (7.15)$$

По предположению левая часть (7.15) стремится к 1 и отсюда следует, что $n[1-F(2a_n)] \rightarrow 0$ (так как $x \leq e^{-(1-x)}$ при $x \leq 1$). Если бы $n\mu_n/a_n$ стремилось к нулю, то же самое должно было быть верным и для правой части неравенства (7.14), но тогда это явно противоречило бы предположению (7.11). Такое рассуждение годится также и для подпоследовательностей и показывает, что $n\mu_n/a_n$ не стремится к нулю, а это в свою очередь означает, что $\rho(2a_n) \rightarrow \infty$.

Выбирая a_n так, чтобы $2a_n < x \leq 2a_{n+1}$, получаем, что $\rho(x) \rightarrow \infty$ при $x \rightarrow \infty$. Тогда $\rho(x) \geq (2a_n) a_n / a_{n+1}$, и очевидно, что из (7.11) с необходимостью вытекает ограниченность отношений a_{n+1}/a_n . ►

§ 8*). УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

Пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины с одинаковым распределением F , и пусть $E(X_k) = 0$. Как обычно, обозначим $S_n = X_1 + \dots + X_n$. Слабый закон больших чисел утверждает, что для любого $\varepsilon > 0$

$$P\{n^{-1}|S_n| > \varepsilon\} \rightarrow 0. \quad (8.1)$$

Этот факт не исключает возможности, что значения $n^{-1}S_n$ могут оказаться произвольно большими для бесконечно многих n . Например, в симметричном случайном блуждании вероятность того, что частица возвратится в начало координат на n -м шаге, стремится к нулю, а в то же время несомненно, что осуществится бесконечное число таких возвращений. На практике мало интересуются вероятностью в (8.1) для отдельных больших значений n . Гораздо более интересен вопрос о том, сделается ли величина $n^{-1}|S_n|$ в конце концов малой и будет оставаться такой в дальнейшем, иначе говоря, будет ли $n^{-1}|S_n| < \varepsilon$ одновременно для всех $n \geq N$. В соответствии с этим мы найдем вероятность события, которое заключается в том, что $n^{-1}S_n \rightarrow 0^1$.

* Этот параграф может быть опущен при первом чтении.

¹ Из закона нуля-единицы (см. гл. IV, б) следует, что эта вероятность равна 0 или 1, но мы не будем этим пользоваться,

Если это событие имеет вероятность единица, то мы говорим, что *последовательность* $\{X_k\}$ *подчиняется усиленному закону больших чисел*.

Следующая теорема показывает, что этот закон имеет место, когда $E(X_1) = 0$. [То, что утверждение гораздо сильнее слабого закона больших чисел, следует из того факта, что соотношение (8.1) выполняется, как мы видели, для некоторых последовательностей $\{X_k\}$, не имеющих математического ожидания. В противоположность этому существование математического ожидания является необходимым условием для усиленного закона больших чисел. Действительно, обсуждаемая в конце этого параграфа теорема, обратная усиленному закону, показывает, что при отсутствии математического ожидания средние $n^{-1}|S_n|$ обречены бесконечно часто превышать любую заданную границу a .]

Теорема 1. (Усиленный закон больших чисел.) Пусть $X_1, X_2, \dots$ — независимые, одинаково распределенные случайные величины с $E(X) = 0$. Тогда с вероятностью единица $n^{-1}S_n \rightarrow 0$.

Доказательство этой теоремы основывается на методе урезания и в действительности проходит для последовательностей случайных величин с различными распределениями. Чтобы избежать повторений, мы отложим на время доказательство теоремы 1 и подготовимся к нему с помощью другой теоремы, которая имеет широкий круг применений.

Теорема 2. Пусть $X_1, X_2, \dots$ — независимые случайные величины с произвольными распределениями. Предположим, что $E(X_k) = 0$ при всех k и что

$$\sum_{k=1}^{\infty} E(S_k^2) < \infty \quad (8.2)$$

Тогда последовательность $\{S_n\}$ сходится с вероятностью единица к конечному пределу S .

Доказательство. Обратимся к бесконечномерному выборочному пространству, определенному величинами X_k . Пусть $A(\epsilon)$ обозначает событие, которое заключается в том, что неравенство $|S_n - S_m| > \epsilon$ выполняется для некоторых произвольно больших значений n, m . Событие, состоящее в том, что $\{S_n\}$ не сходится, является монотонным пределом событий $A(\epsilon)$ при $\epsilon \rightarrow 0$, и поэтому достаточно доказать, что $P\{A(\epsilon)\} = 0$. Пусть событие $A_m(\epsilon)$ означает, что $|S_n - S_m| > \epsilon$ при некотором $n > m$. Тогда $A(\epsilon)$ является пределом при $m \rightarrow \infty$ невозрастающей последовательности событий $A_m(\epsilon)$, и поэтому достаточно доказать, что $P\{A_m(\epsilon)\} \rightarrow 0$. Наконец, пусть $A_{m,n}(\epsilon)$ при $n > m$ будет событием, заключающимся в том, что $|S_k - S_m| > \epsilon$ для некоторого $m < k < n$.

В силу неравенства Колмогорова

$$P\{A_{m,n}(\varepsilon)\} \leq \varepsilon^{-2} \text{Var}(S_n - S_m) = \varepsilon^{-2} \sum_{k=m+1}^n E(X_k^2). \quad (8.3)$$

При $n \rightarrow \infty$ мы приходим к заключению, что

$$P\{A_m(\varepsilon)\} \leq \varepsilon^{-2} \sum_{k=m+1}^{\infty} E(X_k^2), \quad (8.4)$$

где правая часть стремится к нулю при $m \rightarrow \infty$. ►

Доказанная теорема имеет множество приложений. Следующий ее вариант будет использован для доказательства усиленного закона больших чисел.

Теорема 3. Пусть $X_1, X_2, \dots$ — независимые случайные величины с произвольными распределениями. Предположим, что $E(X_k) = 0$ для всех k . Если $b_1 < b_2 < \dots \rightarrow \infty$ и если

$$\sum b_k^{-2} E(X_k^2) < \infty, \quad (8.5)$$

то ряд $\sum b_k^{-1} X_k$ сходится с вероятностью единица и

$$b_n^{-1} S_n \rightarrow 0. \quad (8.6)$$

Доказательство. Первое утверждение теоремы есть прямое следствие теоремы 2, отнесенной к случайным величинам $b_k^{-1} X_k$. Следующая лемма, содержащая широко известный результат, показывает, что соотношение (8.6) имеет место в каждой точке, в которой сходится ряд $\sum b_k^{-1} X_k$. На этом доказательство заканчивается. ►

Лемма 1. («Лемма Кронекера».) Пусть $\{x_k\}$ — произвольная числовая последовательность и $0 < b_1 < b_2 < \dots \rightarrow \infty$. Если ряд

$\sum_{k=1}^{\infty} b_k^{-1} x_k$ сходится, то

$$\frac{x_1 + \dots + x_n}{b_n} \rightarrow 0. \quad (8.7)$$

Доказательство. Обозначим остаточный член нашего сходящегося ряда через ρ_n . Тогда при $n = 1, 2, \dots$ $x_n = b_n(\rho_{n-1} - \rho_n)$ и, следовательно,

$$\frac{x_1 + \dots + x_n}{b_n} = -\rho_n + \frac{1}{b_n} \sum_{k=1}^{n-1} \rho_k (b_{k+1} - b_k) + \frac{\rho_0}{b_n}. \quad (8.8)$$

Предположим, что $|\rho_k| < \varepsilon$ при $k \geq r$. Так как $b_n \rightarrow \infty$, то вклад первых r слагаемых в сумму стремится к нулю, тогда как оставшиеся слагаемые в сумме не превышают $\varepsilon(b_n - b_r)/b_n < \varepsilon$. Таким образом, (8.7) справедливо. ►

Прежде чем вернуться к усиленному закону больших чисел, мы докажем еще одну лемму, которая имеет чисто аналитический характер.

Лемма 2. Пусть величины X_k имеют одинаковое распределение F . Тогда при любом $a > 0$

$$\sum P\{|X_k| > ak\} < \infty \quad (8.9)$$

в том и только том случае, когда существует $E(X_k)$.

Доказательство. В соответствии с леммой 2 из гл. V, 6, математическое ожидание $E(X_1)$ существует тогда и только тогда, когда

$$\int_0^{\infty} [1 - F(x) + F(-x)] dx < \infty. \quad (8.10)$$

Ряд в (8.9) можно рассматривать как риманову интегральную сумму, а так как подынтегральная функция монотонна, то каждое из соотношений (8.9) и (8.10) следует одно из другого. ►

Наконец, мы в состоянии доказать усиленный закон больших чисел¹⁾.

Доказательство теоремы 1. Используем метод урезаний и определим новые случайные величины следующим образом:

$$\begin{aligned} X'_k &= X_k, & X''_k &= 0, & \text{если } |X_k| \leq k, \\ X'_k &= 0, & X''_k &= X_k, & \text{если } |X_k| > k. \end{aligned} \quad (8.11)$$

Так как математические ожидания конечны, то из леммы 2 можно вывести при $a=1$, что

$$\sum P\{X''_k \neq 0\} < \infty. \quad (8.12)$$

Отсюда следует, что с вероятностью единица только конечное число величин X''_k будет отлично от 0. Таким образом, с вероятностью единица $S''_n/n \rightarrow 0$, где $S''_n = \sum_k X''_k$.

Далее мы покажем, что

$$\sum k^{-2} E(X'_k) < \infty. \quad (8.13)$$

Тогда в силу теоремы 3 с вероятностью единица

$$n^{-1} [S'_n - E(S'_n)] \rightarrow 0. \quad (8.14)$$

Но $E(X'_k) \rightarrow 0$, и поэтому очевидно, что

$$n^{-1} E(S'_n) = n^{-1} \sum E(X'_k) \rightarrow 0. \quad (8.15)$$

Для завершения доказательства нам остается только проверка

¹⁾ О доказательстве усиленного закона прямыми методами см. в задаче 12.

утверждения (8.13). Имеем

$$E(X_k^2) = \sum_{j=1}^k \int_{j-1 < |x| < j} x^2 F\{dx\}. \quad (8.16)$$

Отсюда следует, что

$$\sum_{k=1}^{\infty} k^{-2} E(X_k^2) = \sum_{j=1}^{\infty} \int_{j-1 < |x| < j} x^2 F\{dx\} \sum_{k=j}^{\infty} k^{-2}. \quad (8.17)$$

Внутренняя сумма меньше чем $2/j$, и поэтому правая часть меньше чем

$$\sum_{j=1}^{\infty} \int_{j-1 < |x| < j} |x| F\{dx\} = \int_{-\infty}^{+\infty} |x| F\{dx\}. \quad (8.18)$$

На этом доказательство закончено. ▶

Мы видели, что слабый закон больших чисел в форме (7.8) применим к некоторым последовательностям случайных величин, не имеющих математического ожидания. В противоположность этому для усиленного закона существование математического ожидания $E(X_1)$ необходимо. В подтверждение этого следующая теорема показывает, что при отсутствии конечного математического ожидания последовательность арифметических средних S_n/n не ограничена с вероятностью единица.

Теорема 4. (Обращение усиленного закона больших чисел.) Пусть случайные величины $X_1, X_2, \dots$ независимы и одинаково распределены. Если $E(|X_1|) = \infty$, то для любой числовой последовательности $\{c_n\}$ с вероятностью единица

$$\limsup |n^{-1}S_n - c_n| = \infty. \quad (8.19)$$

Доказательство. Пусть событие A_k состоит в том, что $|X_k| > ak$. События A_k взаимно независимы, и по лемме 2 отсутствие математических ожиданий имеет следствием расходимость ряда $\sum P\{A_n\}$. В силу второй леммы Бореля—Кантелли [см. I, VIII, 3] это означает, что с вероятностью единица осуществится бесконечно много событий A_k и поэтому последовательность $|X_k|/k$ не ограничена с вероятностью единица. Но так как $X_k = S_k - S_{k-1}$, то ограниченность $|S_n|/n$ должна была бы влечь ограниченность $|X_k|/k$, и поэтому мы заключаем, что последовательность арифметических средних S_n/n не ограничена с вероятностью единица.

Все это доказывает утверждение (8.19) в частном случае $c_k = 0$, а общий случай может быть сведен к разобранному с помощью симметризации. Так же как в гл. V, § 5, обозначим через ${}^0X_k$ симметризованные величины X_k . Из неравенства симметризации (5.1) гл. V следует, что $E(|{}^0X_k|) = \infty$, и поэтому последовательность арифметических средних ${}^0S_n/n$ не ограничена с вероятностью единица. Однако величины ${}^0S_n$ можно получить путем симметризации $S_n - c_n$, и, следовательно, вероятность того, что $(S_n - c_n)/n$ остается ограниченной, равна нулю. ▶

§ 9*) ОБОБЩЕНИЕ ДЛЯ МАРТИНГАЛОВ

Неравенство Колмогорова из гл. V, 8, является основным инструментом в доказательствах § 8. Детальный анализ этих доказательств показывает, что предположение о независимости случайных величин было использовано только при выводе некоторых неравенств для математических ожиданий, и поэтому основные результаты могут быть перенесены на мартингалы и субмартингалы. Такое обобщение имеет важное значение для многих приложений, оно проливает свет на природу наших теорем.

Напомним (см. гл. VI, 12), что конечная или бесконечная последовательность случайных величин U_r составляет *субмартингал*, если для всех r

$$E(U_r | \mathcal{B}_k) \geq U_k \quad \text{при } k=1, 2, \dots, r-1, \quad (9.1)$$

где $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \dots$ есть возрастающая последовательность σ -алгебр событий. Если все неравенства в (9.1) заменить на равенства, то последовательность $\{U_r\}$ называется *мартингалом*. [При любом значении r автоматически выполняются все $r-1$ условий (9.1), если эти условия справедливы для значения $k=r-1$.] Напомним также, что если $\{X_k\}$ есть последовательность независимых случайных величин с $E(X_k)=0$, то частные суммы S_n образуют мартингал; более того, если существуют дисперсии, то последовательность $\{S_n^2\}$ образует субмартингал.

Теорема 1. (Неравенство Колмогорова для положительных субмартингалов.) Пусть $U_1, \dots, U_n$ — положительные случайные величины. Предположим, что соотношение (9.1) выполнено при $r \leq n$. Тогда при $t > 0$

$$P\{\max_{k \leq n} U_k > t\} \leq t^{-1} E(U_n). \quad (9.2)$$

Если $\{U_k\}$ есть произвольный мартингал, то величины $|U_k|$ образуют субмартингал (см. лемму в гл. VI, 12). Отсюда вытекает, что теорема 1 имеет важное следствие.

Следствие. (Неравенство Колмогорова для мартингалов.) Если случайные величины $U_1, \dots, U_n$ составляют мартингал, то при $t > 0$

$$P\{\max_{k \leq n} |U_k| > t\} \leq t^{-1} E(|U_n|). \quad (9.3)$$

Доказательство теоремы 1. Мы повторим дословно доказательство неравенства Колмогорова из гл. V, 8, д), полагая $S_k^2 = U_k$. Предположение о том, что S_k представляет собой суммы независимых случайных величин, было использовано только при установлении неравенства V, (8.16), которое теперь может быть записано как

$$E(U_n 1_{A_j}) \geq E(U_j 1_{A_j}). \quad (9.4)$$

*) Содержание этого параграфа нигде в дальнейшем не используется.

Далее, случайная величина 1_{A_j} $\mathcal{B}_j$ -измерима и поэтому

$$E(U_n 1_{A_j} | \mathcal{B}_j) = 1_{A_j} E(U_n | \mathcal{B}_j) \geq U_j 1_{A_j} \quad (9.5)$$

[см. V, (10,9)]. Вычисляя в (9.5) математические ожидания левой и правой частей, получаем (9.4). ►

Мы обратимся к обобщению теоремы 2 о бесконечных свертках из § 8, хотя на этом пути мы получим только частный случай общей теоремы о сходимости мартингалов. Действительно, Дуб показал, что приводимая ниже теорема остается справедливой, если условие $E(S_n^2) < \infty$ заменить более слабым требованием, чтобы $E(|S_n|)$ оставалось ограниченным. Доказательство этой более общей теоремы, однако, сложно, и мы приводим здесь наш вариант ввиду большой важности теоремы и простоты доказательства. [Обобщение см. в задаче 13.]

Теорема 2. (Теорема о сходимости мартингалов.) Пусть $\{S_n\}$ — бесконечная последовательность, образующая мартингал с $E(S_n^2) < C < \infty$ при всех n . Тогда существует такая случайная величина S , что с вероятностью единица $S_n \rightarrow S$. Кроме того, $E(S_n) = E(S)$ при всех n .

Доказательство. Мы повторим доказательство теоремы 2 из § 8. Предположение о том, что S_k являются суммами независимых случайных величин, использовалось только в (8.3) для доказательства

$$E((S_n - S_m)^2) \rightarrow 0, \quad n, m \rightarrow \infty. \quad (9.6)$$

Далее, из гл. VI, 12, мы знаем, что по свойству мартингалов $E(S_n | \mathcal{S}_m) = S_m$ при $n > m$. В силу основного свойства V, (10.9) условных математических ожиданий получаем

$$E(S_n S_m | \mathcal{S}_m) = S_m^2, \quad n > m. \quad (9.7)$$

Вычисляя математические ожидания и воскрешая в памяти формулу V, (10.10) для повторных математических ожиданий, мы заключаем, что $E(S_n S_m) = E(S_m^2)$, и, следовательно,

$$E(S_n^2 - S_m^2) = E(S_n^2) - E(S_m^2), \quad n > m.$$

Но по лемме из гл. VI, 12, величины S_n^2 образуют субмартингал, и поэтому последовательность $\{E(S_n^2)\}$ является монотонно возрастающей. По предположению теоремы она ограничена и, следовательно, имеет конечный предел. Это влечет справедливость (9.6). Из свойства мартингала вытекает, что $E(S_n)$ не зависит от n , а из ограниченности последовательности $\{E(S_n^2)\}$ следует равенство $E(S) = E(S_n)$ [пример VIII, 1, д)].

Как прямое следствие мы получаем аналог теоремы 3 из § 8.

Теорема 3. Пусть $\{X_n\}$ — такая последовательность случай-

ных величин, что при всех n

$$E(X_n | \mathcal{B}_{n-1}) = 0. \quad (9.8)$$

Если $b_1 < b_2 < \dots \rightarrow \infty$ и

$$\sum b_k^{-2} E(X_k^2) < \infty, \quad (9.9)$$

то с вероятностью единица

$$\frac{X_1 + \dots + X_n}{b_n} \rightarrow 0. \quad (9.10)$$

Доказательство. Легко видеть, что случайные величины

$$U_n = \sum_{k=1}^n b_k^{-1} X_k \quad (9.11)$$

образуют мартингал и что математическое ожидание $E(U_n^2)$ ограничено сверху суммой ряда (9.9). Поэтому в силу предыдущей теоремы обеспечена сходимость $\{U_n\}$ с вероятностью единица, и по лемме Кронекера отсюда вытекает утверждение (9.10). ►

Примеры. а) *Урновая схема Пуа* уже рассматривалась в примере 12, б) из гл. VI и в примере 4, а) этой главы. Обозначим через Y_n долю черных шаров при n -м испытании. Уже было показано, что $\{Y_n\}$ — мартингал. Теперь мы видим, что с вероятностью единица существует предел $Y = \lim Y_n$. С другой стороны, вероятность извлечь при n -м испытании черный шар получается рандомизацией биномиального распределения. Таким образом, если S_n — общее число черных шаров, извлеченных при первых n испытаниях, то распределение величины $n^{-1}S_n$ сходится к бета-распределению F , найденному в примере 4, а). Поэтому предельная случайная величина Y имеет бета-распределение F .

б) *Ветвящиеся процессы.* В ветвящемся процессе, описанном в 1; гл. XII, 5, число X_n индивидуумов n -го поколения имеет математическое ожидание $E(X_n) = \mu^n$ [см. 1; гл. XII, формула (4.9)]. При условии, что $(n-1)$ -е поколение содержит v индивидуумов, (условное) математическое ожидание величины X_n равно μ^v и не зависит от размера предшествующих поколений. Таким образом, последовательность $\{S_n\}$, где $S_n = X_n / \mu^n$, образует мартингал. Нетрудно установить, что если $E(X_1^2) < \infty$, то $E(S_n^2)$ остается ограниченным в предположении $\mu > 1$ (см. задачу 7 из 1; гл. XII, 6.) Мы получаем удивительный результат, что S_n с вероятностью единица сходится к пределу S_∞ . Отсюда следует, в частности, что распределение величины S_n сходится к распределению величины S_∞ . Эти результаты принадлежат Т. Харрису.

в) *Гармонические функции.* Для ясности мы рассмотрим специальный пример, хотя последующие рассуждения применимы к более общим марковским цепям и координатным функциям (пример 12, в) из гл. VI).

Пусть D обозначает единичный круг, т. е. совокупность таких точек $x = (x_1, x_2)$, что $x_1^2 + x_2^2 \leq 1$. Для любой точки $x \in D$ обозначим через C_x наибольшую окружность с центром в x , содержащуюся в D . Рассмотрим марковский процесс $\{Y_n\}$ в D , определенный следующим образом. При данном $Y_n = x$ случайная величина Y_{n+1} распределена равномерно на C_x ; начальное положение $Y_0 = y$ считается известным. Переходные вероятности определяются стохастическим ядром K , которое при фиксированном x сосредоточено на C_x и сводится там к равномерному распределению. Функция u , заданная на D , будет конкордантной, если значение $u(x)$ равно среднему значению u на C_x . Рассмотрим теперь гармоническую функцию u , непрерывную в *замкнутом* круге D . Тогда $\{u(Y_n)\}$ — ограниченный мартингал и потому $Z = \lim u(Y_n)$ существует с вероятностью единица. Так как координатные переменные x_j являются гармоническими функциями, то из сказанного следует, что с вероятностью единица Y_n сходится к пределу $Y \in D$. Легко видеть, что процесс не может сходиться к внутренней точке круга D . Поэтому с вероятностью единица Y_n *сходится к точке Y на границе круга D* .

В обобщенном виде рассуждения подобного типа используются при изучении асимптотических свойств марковских процессов и при доказательстве общих теорем о гармонических функциях, таких, как теорема Фату (теорема о существовании радикальных граничных значений почти всюду)¹. ►

§ 10. ЗАДАЧИ

1. Если функция u ограничена и непрерывна на $\overline{0, \infty}$, то при $n \rightarrow \infty$

$$\sum_{k=0}^{\infty} \binom{n+k}{k} \frac{t^k}{(1+t)^{n+k+1}} u\left(\frac{k}{n+1}\right) \rightarrow u(t)$$

равномерно в каждом конечном интервале.

Указание. Вспомните отрицательное биномиальное распределение (1, гл. VI, 8). Не нужно никаких вычислений.

2. Если u имеет непрерывную производную u' , то производные $B'_{n,u}$ полиномов $B_{n,u}$ равномерно сходятся к u' .

3. *Полиномы Бернштейна в $\mathcal{R}^2$* . Если $u(x, y)$ непрерывна в треугольнике $x \geq 0, y \geq 0, x+y \leq 1$, то имеет место равномерная сходимость

$$\sum u\left(\frac{j}{n}, \frac{k}{n}\right) \frac{n!}{j!k!(n-j-k)!} x^j y^k (1-x-y)^{n-j-k} \rightarrow u(x, y).$$

4. Функцию u , непрерывную на $\overline{0, 1}$, можно равномерно аппроксимиро-

¹) Brelot M., Doob J. L., Ann. Inst. Fourier, 13 (1963), 395—415. [Русский перевод см. сб. Математика, 11:3 (1967), 101—116.]

вать полиномами четной степени. Если $u(0) = 0$, то же самое верно для полиномов нечетной степени¹⁾.

5. Если u непрерывна в интервале $0, \infty$ и $u(\infty)$ существует, то u может быть равномерно аппроксимирована линейными комбинациями функций e^{-px} .

6. Ниже указаны три последовательности моментов. Найдите вероятности $p_k^{(n)}$, входящие в (3.5). Используя предельное соотношение (3.11), найдите соответствующее распределение вероятностей

$$(a) \mu_n = p^n, \quad 0 < p < 1,$$

$$(b) \mu_n = \frac{1}{n+1},$$

$$(c) \mu_n = \frac{2}{n+2}.$$

7²⁾. Пусть p — полином степени v . Покажите, что при $n > v$ разность $\Delta^n p$ тождественно равна нулю. Покажите, что B_n, p — полином степени не более v (несмотря на то что в записи он выглядит как полином степени $n > v$).

8. Покажите, что если F имеет плотность, то соотношение (6.4) можно получить интегрированием соотношения (6.6)

9. Закон больших чисел для стационарных последовательностей. Пусть $\{X_k\}$ ($k=0, \pm 1, \pm 2, \dots$) — стационарная последовательность. Определим X'_k урезанием, таким же, как в (7.2). Если $E(X_k) = 0$ и $E(X'_0 X'_n) \rightarrow 0$ при $n \rightarrow \infty$, то

$$P\{n^{-1} | X_1 + \dots + X_n | > \varepsilon\} \rightarrow 0.$$

10. Пусть случайные величины X_k независимы, а случайные величины X'_k заданы формулой (7.2). Предположим, что $a_n \rightarrow 0$ и

$$\sum_{k=1}^n P\{|X_k| > s_n\} \rightarrow 0, \quad a_n^{-2} \sum_{k=1}^n E(X_k'^2) \rightarrow 0.$$

Докажите, что $P\left\{\left|S_n - \sum_{k=1}^n E(X'_k)\right| > \varepsilon a_n\right\} \rightarrow 0$.

11. (К теореме 1 § 7.) Покажите, что из $\sigma(t) \rightarrow 0$ следует $\tau(t) \rightarrow 0$. Указание. Докажите, что $\tau(x) - \tau(2x)/2 < \varepsilon$ для достаточно больших x . Применяя это неравенство последовательно для значений $x = t, 2t, 4t, \dots$, получите $\tau(t) < 2\varepsilon$.

12. (Прямое доказательство усиленного закона больших чисел.) Используя обозначения, принятые в доказательстве теоремы 1 в § 8, положим $Z_r = \max |S'_k|$ при $2^r < k \leq 2^{r+1}$. С помощью неравенства Колмогорова покажите, что

$$\sum P(Z_r > \varepsilon 2^r) < \infty,$$

откуда будет следовать усиленный закон больших чисел. (Это доказательство обходится без обращения к теоремам 2 и 3 § 8.)

13. (Теорема сходимости для субмартингалов.) Докажите, что теорема 2 § 8 применима к субмартингалам $\{U_k\}$, если $U_k > 0$ для всех k .

14. Распространите вариант неравенства Чебышева из примера 7, а) гл. V на мартингалы³⁾.

¹⁾ По известной теореме Мюнца равномерная аппроксимация линейными комбинациями степеней $1, x^{n_1}, x^{n_2}, \dots$ возможна в том и только том случае, когда ряд $\sum p_k^{-1}$ расходится.

²⁾ Использование этого результата значительно упрощает классическое решение проблемы моментов (см. книгу Шохата и Тамаркина).

³⁾ Marshall A. W., A one-sided analog of Kolmogorov's inequality, Ann. Math. Statist., 31 (1960), 483—487.

Основные результаты этой главы содержатся в § 1, 3 и 6. Параграфы 4, 5 и 7 можно рассматривать как источники интересных примеров. Выбор примеров объясняется их важностью для тех или иных целей.

Два последних параграфа посвящены правильно меняющимся функциям (в смысле Карамата). Эта интересная тема приобретает все большее значение. Однако она не освещена в учебниках и ее изложение не было приспособлено к случаю функций распределения. Огромное количество разрозненных вычислений в теории вероятностей можно упразднить при использовании асимптотических соотношений § 9. Последние по сравнению с простым § 8 имеют более «технический» характер.

§ 1. СХОДИМОСТЬ МЕР

Излагаемая ниже теория не зависит от размерности пространства. Формулировки в тексте для удобства даны для одномерного случая. Однако с учетом соглашений гл. III, 5, они без изменений переносятся на многомерный случай.

Два из приводимых ниже примеров типичны для явлений, с которыми нам придется иметь дело.

Примеры. а) Возьмем произвольное распределение вероятностей F и положим $F_n(x) = F(x - n^{-1})$. В точках x , в которых F непрерывна, имеем $F_n(x) \rightarrow F(x)$. В то же время в точках разрыва $F_n(x) \rightarrow F(x-)$. Тем не менее мы будем говорить, что $\{F_n\}$ сходится к F .

б) Положим на этот раз $F_n(x) = F(x + n)$, где F — непрерывная функция распределения. Теперь $F_n(x) \rightarrow 1$ при всех x ; предел существует, но не является вероятностной функцией распределения. Здесь $F_n\{I\} \rightarrow 0$ для любого ограниченного интервала I , но только не тогда, когда I совпадает со всей прямой.

в) Пусть $F_n(x) = F(x + (-1)^n n)$. Тогда $F_{2n}(x) \rightarrow 1$, но $F_{2n+1}(x) \rightarrow 0$. Таким образом, функции распределения $F_n(x)$ не сходятся ни к какой функции, и тем не менее $F_n\{I\} \rightarrow 0$ для каждого ограниченного интервала.

Основные понятия и обозначения

Нам будет нужно выделить три класса непрерывных функций. В одномерном случае ¹⁾ $C(-\infty, \infty)$ обозначает класс все ограниченных непрерывных функций, $C[-\infty, \infty]$ обозначает подкласс функций, имеющих конечные пределы $u(-\infty)$ и $u(\infty)$, наконец, $C_0(-\infty, \infty)$ обозначает подкласс функций, «исчезающих на бесконечности», т. е. функций с $u(\pm\infty) = 0$.

Мы скажем, что I есть *интервал непрерывности* для распределения вероятностей F , если I — открытый интервал, концы которого не являются атомами F ²⁾. Вся прямая рассматривается как интервал непрерывности. В этом параграфе мы используем сокращенные обозначения

$$E_n(u) = \int_{-\infty}^{+\infty} u(x) F_n\{dx\}, \quad E(u) = \int_{-\infty}^{+\infty} u(x) F\{dx\}. \quad (1.1)$$

Всюду в этом параграфе F_n обозначает собственное распределение вероятностей, но распределению F будет позволено быть и несобственным (т. е. его полная масса может быть меньше 1; см. определение в гл. V, 1). ►

Определение ³⁾. Последовательность $\{F_n\}$ сходится к (возможно, несобственному) распределению F , если

$$F_n\{I\} \rightarrow F\{I\} \quad (1.2)$$

для каждого ограниченного интервала непрерывности F . Это обозначается $F_n \rightarrow F$ или $F = \lim F_n$.

Сходимость называется собственной, если F — собственное распределение.

Для ясности мы будем иногда говорить о несобственной схо-

¹⁾ По поводу перехода к многомерному случаю заметим, что в одномерном случае $C[-\infty, \infty]$ — это класс непрерывных функций на прямой, «компактифицированной» добавлением точек $\pm\infty$. Для определения $C[-\infty, \infty]$ в $\mathcal{R}^2$ «расширяются» обе координатные оси, так что требуется существование при каждом x пределов $u(x, \pm\infty)$ и $u(\pm\infty, x)$. Сам по себе этот класс не очень интересен, но в него входят функции распределения. В определении $C_0(-\infty, \infty)$ требуется, чтобы $u(x, \pm\infty) = u(\pm\infty, x) = 0$.

²⁾ В многомерном случае требуется, чтобы граница I имела вероятность нуль.

³⁾ Для читателей, интересующихся общей теорией меры, мы сделаем следующее замечание. Определения и теоремы этого параграфа переносятся без изменений на ограниченные меры в произвольных локально компактных пространствах, если заменить «интервалы непрерывности» на «открытые множества с границей меры нуль». Ограниченным интервалам соответствуют подмножества компактных множеств. Наконец, C_0 — класс всех непрерывных функций, исчезающих на бесконечности, т. е. $u \in C_0$ тогда и только тогда, когда u непрерывна и $|u| < \varepsilon$ вне некоторого компактного множества. Другие классы не играют особой роли в этом параграфе.

димости, желая подчеркнуть, что предельное распределение F несобственное.

Для облегчения ссылок сформулируем два простых критерия собственной сходимости.

Критерий 1. Сходимость $F_n \rightarrow F$ является собственной тогда и только тогда, когда каждому $\varepsilon > 0$ соответствуют такие числа a и N , что $F_n \{\overline{-a, a}\} > 1 - \varepsilon$ при $n > N$.

Доказательство. Без ограничения общности можно предположить, что интервал $\overline{-a, a}$ является интервалом непрерывности для предельного распределения F . Из условия теоремы вытекает, что $F \{\overline{-a, a}\} > 1 - \varepsilon$ и, следовательно, F не может быть несобственным. Поэтому условие достаточно. Обратно, если F есть распределение вероятностей, то можно выбрать a столь большим, что $F \{\overline{-a, a}\} > 1 - \varepsilon/2$. Поэтому для достаточно больших n $F_n \{\overline{-a, a}\} > 1 - \varepsilon$, и, следовательно, условие необходимо. ►

Критерий 2. Последовательность $\{F_n\}$ распределений сходится к собственному вероятностному распределению F тогда и только тогда, когда (1.2) выполняется для каждого ограниченного или неограниченного интервала I , который является интервалом непрерывности распределения F .

(Отсюда вытекает, что в случае собственной сходимости $F_n \rightarrow F$ имеет место сходимость $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности функции F .)

Доказательство. Можно предположить, что F_n сходится (к, возможно, несобственному) распределению F . Очевидно, что F является собственным тогда и только тогда, когда (1.2) выполняется для $I = \overline{-\infty, \infty}$. Следовательно, условие критерия достаточно. Предположим теперь, что F — собственное распределение вероятностей. При $x > -a$ интервал $\overline{-\infty, x}$ есть объединение полуинтервалов $\overline{-\infty, a}$ и $\overline{a, x}$. Используя предыдущий критерий, можно поэтому заключить, что для достаточно больших a и n $F_n \{\overline{-\infty, x}\}$ отличается от $F \{\overline{-\infty, x}\}$ меньше чем на 3ε . Аналогичное рассуждение применимо и к интервалу $\overline{x, \infty}$, и можно сделать вывод, что (1.2) выполняется для всех полупрямых. ►

Мы определяли сходимость посредством соотношения (1.2), но следующая теорема показывает, что мы могли бы использовать в качестве определяющего соотношение (1.3).

Теорема 1. (i) Для сходимости $F_n \rightarrow F$ необходимо и достаточно, чтобы¹⁾

$$E_n(u) \rightarrow E(u) \text{ для всех } u \in C_0(-\infty, \infty). \quad (1.3)$$

(ii) Если сходимость собственная, то $E_n(u) \rightarrow E(u)$ для всех ограниченных непрерывных функций.

Доказательство. а) Нам удобно начать с утверждения относительно собственной сходимости. Итак, предположим, что F есть распределение вероятностей и $F_n \rightarrow F$. Пусть u такая непрерывная функция, что $|u(x)| < M$ для всех x . Пусть интервал непрерывности A распределения F столь велик, что $F(A) > 1 - \epsilon$. Для дополнения A' мы получаем в таком случае, что $F_n\{A'\} < 2\epsilon$ для всех достаточно больших n .

Так как u равномерно непрерывна на конечных интервалах, то можно разбить A на интервалы $I_1, \dots, I_n$, предполагая их столь малыми, что колебание u на каждом из них меньше ϵ . Эти I_k можно выбрать так, что они будут интервалами непрерывности для F . Внутри A мы можем аппроксимировать функцию u ступенчатой функцией σ , которая принимает постоянное значение на каждом из интервалов I_k , и при этом $|u(x) - \sigma(x)| < \epsilon$ для всех $x \in A$. На дополнении A' полагаем $\sigma(x) = 0$. Тогда $|u(x) - \sigma(x)| < M$ при $x \in A'$ и

$$|E(u) - E(\sigma)| \leq \epsilon F\{A\} + MF\{A'\} \leq \epsilon + M\epsilon. \quad (1.4)$$

Аналогично при достаточно большом n

$$|E_n(u) - E_n(\sigma)| \leq \epsilon F_n\{A\} + MF_n\{A'\} \leq \epsilon + 2M\epsilon. \quad (1.5)$$

Теперь $E_n(\sigma)$ — это конечная линейная комбинация значений $F_n\{I_k\}$, стремящихся к $F\{I_k\}$. Следовательно, $E_n(\sigma) \rightarrow E(\sigma)$ и поэтому при достаточно больших n

$$|E(\sigma) - E_n(\sigma)| < \epsilon. \quad (1.6)$$

Комбинируя три последних неравенства, получаем

$$|E(u) - E_n(u)| \leq |E(u) - E(\sigma)| + |E(\sigma) - E_n(\sigma)| + |E_n(\sigma) - E_n(u)| < 3(M+1)\epsilon, \quad (1.7)$$

а поскольку ϵ произвольно, то отсюда вытекает, что $E_n(u) \rightarrow E(u)$.

Это рассуждение не проходит в случае несобственной сходимости, потому что тогда значение $F_n\{A'\}$ не обязано быть малым. Однако в этом случае мы рассмотрим только функции $u \in C_0(-\infty, \infty)$, а интервал A может быть выбран столь большим, что $|u(x)| < \epsilon$

¹⁾ Если $E_n(u) \rightarrow E(u)$ для всех u из некоторого класса функций, то говорят, что F_n сходится к F «слабо по отношению к этому классу». Таким образом, сходимость в смысле определения 1 эквивалентна слабой сходимости по отношению к $C_0(-\infty, \infty)$.

при $x \in A'$. Тогда $|u(x) - \sigma(x)| < \varepsilon$ при всех x , а в формулах (1.4), (1.5) выполняются строгие неравенства с ε в их правой части. Таким образом, (1.2) влечет (1.3).

б) Докажем¹⁾, что из сходимости $E_n(u) \rightarrow E(u)$ вытекает $F_n \rightarrow F$. Пусть I — интервал непрерывности F длины L . Обозначим через I_δ «концентрически расположенный относительно I » интервал длины $L + \delta$, где δ выбрано столь малым, что $F\{I_\delta\} < F\{I\} + \varepsilon$. Пусть u — непрерывная функция, которая внутри I принимает постоянное значение 1, которая равна 0 вне I_δ и такая, что всюду $0 \leq u(x) \leq 1$. Тогда $E_n(u) \geq F_n\{I\}$ и $E(u) \leq F\{I_\delta\} \leq F\{I\} + \varepsilon$. Но при достаточно большом n имеет место неравенство $E_n(u) < E(u) + \varepsilon$ и поэтому

$$F_n\{I\} \leq E_n(u) < E(u) + \varepsilon \leq F\{I_\delta\} + \varepsilon < F\{I\} + 2\varepsilon.$$

Используя те же рассуждения для интервала I_δ длины $L - \delta$, мы получим противоположное неравенство $F_n\{I\} > F\{I\} - 2\varepsilon$. Отсюда $F_n \rightarrow F$, что и утверждалось. ►

Желательно иметь критерий для сходимости, который не предполагал бы знания предельного распределения. Этот пробел устраняется следующей теоремой.

Теорема 2. Для того чтобы последовательность $\{F_n\}$ вероятностных распределений сходилась к (возможно, несобственному) предельному распределению, необходимо и достаточно, чтобы для любой функции $u \in C_0(-\infty, \infty)$ последовательность математических ожиданий $E_n(u)$ сходилась к конечному пределу.

Доказательство ²⁾. Необходимость покрывается теоремой 1. Для доказательства достаточности мы должны, забегая вперед, сослаться на теорему о выборе — теорему 1 из § 6. (Она элементарна, однако лучше рассматривать ее вместе со всем кругом связанных с ней вопросов.)

В соответствии с этой теоремой всегда можно найти подпоследовательность $\{F_{n_k}\}$, которая сходится к (возможно, несобственному) предельному распределению Φ . Обозначим через $E^*(u)$ математическое ожидание u относительно Φ . Пусть $u \in C_0(-\infty, \infty)$. Тогда по теореме 1 $E_{n_k}(u) \rightarrow E^*(u)$. Но с таким же успехом $E_{n_k}(u) \rightarrow \lim E_n(u)$. Следовательно, $E^*(u) = \lim E_n(u)$, где u — произвольная функция из $C_0(-\infty, \infty)$. Еще одно применение теоремы 1 приводит к результату $F_n \rightarrow \Phi$, который требовалось доказать. ►

¹⁾ Было бы проще использовать доказательство теоремы 2, но приводимое здесь доказательство больше соответствует интуиции.

²⁾ Теорема тривиальна, если принять во внимание теорему Рисса о представлении (см. замечание 1 в гл. V, 1). В самом деле, $\lim E_n(u)$ определяет линейный функционал, а согласно этой теореме, предел есть математическое ожидание u относительно некоторого распределения F .

Примеры. г) *Сходимость моментов.* Если распределения F_n сосредоточены на $\overline{0, 1}$, то способ определения μ_k вне этого интервала несуществен. Поэтому в формулировке теоремы достаточно предполагать функцию μ_k непрерывной на $\overline{0, 1}$. Каждая такая функция может быть равномерно аппроксимирована полиномом (см. гл. VII, 2). Поэтому теорема может быть сформулирована по-новому следующим образом.

Последовательность распределений F_n , сосредоточенных на $\overline{0, 1}$, сходится к пределу F тогда и только тогда, когда при каждом k последовательность моментов $E_n(X^k)$ сходится к некоторому числу μ_k . В этом случае $\mu_k = E(X^k)$, т. е. μ_k будет k -м моментом распределения F , и сходимость будет собственной, так как $\mu_0 = 1$. (См. VII, 3.)

д) *Сходимость моментов (продолжение).* В общем случае математические ожидания F_n не обязаны сходиться, даже если $F_n \rightarrow F$ в собственном смысле. Например, если F_n приписывает массу n^{-1} точке n^2 и массу $1 - n^{-1}$ точке 0, то F_n сходится к распределению, сосредоточенному в нуле, но $E_n(X) \rightarrow \infty$. Однако имеется следующий полезный критерий. Если $F_n \rightarrow F$ и при некотором $p > 0$ математические ожидания $E_n(|X|^p)$ ограничены, то F является собственным распределением вероятностей. В самом деле, часть $E_n(|X|^p)$, соответствующая области $|x| \geq a$, больше $a^p(1 - F_n(\overline{-a, a}))$. Последняя величина будет меньше M , если $1 - F(\overline{-a, a}) \leq a^{-p}M$. Так как a можно взять произвольно большим, то F будет собственным распределением. Слегка усиливая проведенные рассуждения, можно показать, что абсолютные моменты $E_n(|X|^p)$ порядка $\alpha < p$ сходятся к $E(|X|^p)$.

е) *Сходимость плотностей.* Если распределения вероятностей F_n имеют плотности f_n , то последние не обязаны сходиться, даже если $F_n \rightarrow F$ и F имеет непрерывную плотность. Пусть, например, $f_n(x) = 1 - \cos 2\pi nx$ при $0 < x < 1$ и $f_n(x) = 0$ в других случаях. Тогда распределения F_n сходятся к равномерному распределению с плотностью $f(x) = 1$ для $0 < x < 1$, но f_n не сходятся к f . С другой стороны, если $f_n \rightarrow f$ и f — плотность вероятности, то $F_n \rightarrow F$, где F — собственное распределение с плотностью f . В самом деле, лемма Фату [(2.9) гл. IV] влечет за собой неравенство $\liminf F_n\{I\} \geq F\{I\}$ для каждого интервала непрерывности I . Если для некоторого I выполняется строгое неравенство, то оно выполняется для каждого более широкого интервала непрерывности, в частности для $-\infty, \infty$. Так как это невозможно, то (1,2) выполняется. ►

Пусть мы имеем дело с функциями u_t , зависящими от некоторого параметра t [такими, как $\sin tx$ или $v(t+x)$]. Тогда часто бывает полезно знать, что при достаточно больших n соотноше-

ние $|E_n(u_t) - E(u_t)| < \varepsilon$ выполняется одновременно для всех t . Мы докажем, что это свойство выполняется, если семейство функций u_t *равностепенно непрерывно*, т. е. если каждому $\varepsilon > 0$ соответствует δ , не зависящее от t и такое, что $|u_t(x_2) - u_t(x_1)| < \varepsilon$ при $|x_2 - x_1| < \delta$.

Следствие. Предположим, что $F_n \rightarrow F$ в собственном смысле. Пусть $\{u_t\}$ — семейство *равностепенно непрерывных функций*, зависящих от параметра t . Пусть при всех t $|u_t| < M < \infty$, где M — некоторая постоянная. Тогда $E_n(u_t) \rightarrow E(u_t)$ *равномерно по t* .

Доказательство. В теореме 2 при доказательстве того, что $E_n(u) \rightarrow E(u)$, было использовано разбиение интервала A на интервалы, на каждом из которых колебание u меньше ε . В настоящих условиях это разбиение может быть выбрано одним и тем же для всех t , после чего утверждение становится очевидным. ►

Пример. ж) Пусть $u_t(x) = u(tx)$, где u дифференцируема и $|u'(x)| \leq 1$. По теореме о среднем

$$|u_t(x_2) - u_t(x_1)| \leq |t| \cdot |x_2 - x_1|,$$

так что при t , лежащем в конечном интервале $-\overline{a}, \overline{a}$, рассматриваемое семейство *равностепенно непрерывно*. Следовательно, $E_n(u_t) \rightarrow E(u_t)$ *равномерно в каждом конечном интервале изменения t* . ►

§ 2. СПЕЦИАЛЬНЫЕ СВОЙСТВА

В соответствии с определением 1 из гл. V, 2 — два распределения U и V принадлежат одному и тому же *типу*, если одно от другого отличается только параметрами расположения и масштаба, т. е. если

$$V(x) = U(Ax + B), \quad A > 0. \quad (2.1)$$

Мы покажем теперь, что сходимость распределений сохраняет типы в том смысле, что изменения параметров расположения допредельных распределений не меняют типа предельного распределения. Именно этот факт дает возможность говорить об «асимптотических нормальных последовательностях» без точного указания соответствующих параметров. Более точно, имеет место

Лемма 1. Пусть U и V — два распределения, ни одно из которых не сосредоточено в единственной точке. Пусть последовательность $\{F_n\}$ вероятностных распределений такова, что при некоторых постоянных $a_n > 0$ и $\alpha_n > 0$

$$F_n(a_n x + b_n) \rightarrow U(x), \quad F_n(\alpha_n x + \beta_n) \rightarrow V(x) \quad (2.2)$$

во всех точках непрерывности. Тогда

$$\frac{\alpha_n}{a_n} \rightarrow A > 0, \quad \frac{\beta_n - b_n}{a_n} \rightarrow B \quad (2.3)$$

и выполняется (2.1). Обратно, если верно (2.3), то каждое из соотношений (2.2) влечет за собой другое и соотношение (2.1).

Доказательство. Основываясь на соображениях симметрии, предположим, что выполняется, например, первое из соотношений (2.2). Положим для упрощения обозначений $G_n(x) = F_n(a_n x + b_n)$, $\rho_n = \alpha_n/a_n$ и $\sigma_n = (\beta_n - b_n)/a_n$. Итак, пусть $G_n \rightarrow U$. Если

$$\rho_n \rightarrow A, \quad \sigma_n \rightarrow B, \quad (2.4)$$

то очевидно, что

$$G_n(\rho_n x + \sigma_n) \rightarrow V(x), \quad (2.5)$$

где $V(x) = U(Ax + B)$. Мы должны доказать, что без предположения (2.4) соотношение (2.5) может быть неверным.

Поскольку V не сосредоточено в единственной точке, существуют по меньшей мере два значения x' и x'' , такие, что последовательности $\{\rho_n x' + \sigma_n\}$ и $\{\rho_n x'' + \sigma_n\}$ остаются ограниченными. Отсюда вытекает ограниченность последовательностей $\{\rho_n\}$ и $\{\sigma_n\}$, и, следовательно, можно найти последовательность целых чисел n_k , такую, что $\rho_{n_k} \rightarrow A$ и $\sigma_{n_k} \rightarrow B$. Но тогда $V(x) = U(Ax + B)$ и, следовательно, $A > 0$, иначе V не будет вероятностным распределением. Отсюда вытекает, что пределы A и B одни и те же для всех подпоследовательностей и поэтому (2.4) справедливо. ►

Пример. Лемма не применима, если V сосредоточено в одной точке. Так, например, если $\rho_n \rightarrow \infty$ и $\sigma_n = (-1)^n$, то условие (2.4) не выполняется, однако соотношение (2.5) справедливо для распределения V , сосредоточенного в нуле. ►

Два типа последовательностей $\{F_n\}$ распределений вероятностей встречаются так часто, что заслуживают специального названия. Для большей ясности обозначений мы дадим формальные определения в терминах случайных величин X_n . Однако, по существу, эти понятия связаны только с соответствующими распределениями F_n . Стало быть, определения имеют смысл без всякого упоминания о каком бы то ни было вероятностном пространстве.

Определение 1. X_n сходится по вероятности к нулю, если

$$P\{|X_n| > \varepsilon\} \rightarrow 0 \quad (2.6)$$

при любом $\varepsilon > 0$. Будем обозначать это символом $X_n \xrightarrow{P} 0$.

По определению $X_n \xrightarrow{P} X$ означает то же самое, что $X_n - X \xrightarrow{P} 0$.

Заметим, что (2.6) выполняется тогда и только тогда, когда распределения F_n сходятся к распределению, сосредоточенному в нуле. В общем случае, однако, сходимость $F_n \rightarrow F$ не позволяет сделать никаких заключений относительно сходимости последовательности $X_1, X_2, \dots$. Например, если величины X_j взаимно независимы и имеют одно и то же распределение F , то $F_n \rightarrow F$, но последовательность $\{X_n\}$ не сходится по вероятности.

Приводимая ниже простая лемма часто используется, хотя это и не всегда отмечается явно. (Так, метод «урезания» из 1; гл. X, неявно опирается на эту лемму.)

Лемма 2. Обозначим распределения X_n и Y_n через F_n и G_n соответственно. Допустим, что $X_n - Y_n \xrightarrow{p} 0$ и что $G_n \rightarrow G$. Тогда и $F_n \rightarrow G$.

В частности, если $X_n \xrightarrow{p} X$, то $F_n \rightarrow F$, где F — распределение X .

Доказательство. Если $X_n \leq x$, то или $Y_n \leq x + \varepsilon$, или $X_n - Y_n \leq -\varepsilon$. Вероятность последнего события стремится к нулю. Следовательно, при всех достаточно больших n $F_n(x) \leq G_n(x + \varepsilon) + \varepsilon$. Точно так же устанавливается аналогичная оценка для F_n снизу. ►

Определение 2. Последовательность $\{X_n\}$ называется стохастически ограниченной, если для каждого $\varepsilon > 0$ найдется такое a , что

$$P\{|X_n| > a\} < \varepsilon \quad (2.7)$$

для всех достаточно больших n .

Это определение применимо и к случайным векторам X_n , и к соответствующим им многомерным распределениям.

Последовательность, сходящаяся в собственном смысле, очевидно, будет стохастически ограниченной. В то же время несобственная сходимость исключает стохастическую ограниченность. Мы получаем, следовательно, простой, но полезный критерий: если распределения F_n сходятся, то предел F будет собственным распределением тогда и только тогда, когда последовательность $\{F_n\}$ стохастически ограничена.

Вместе с $\{X_n\}$ и $\{Y_n\}$ стохастически ограничена и последовательность $\{X_n + Y_n\}$. В самом деле, событие $|X_n + Y_n| > 2a$ влечет событие: или $|X_n| > a$, или $|Y_n| > a$. Поэтому

$$P\{|X_n + Y_n| > 2a\} \leq P\{|X_n| > a\} + P\{|Y_n| > a\}. \quad (2.8)$$

§ 3. РАСПРЕДЕЛЕНИЯ КАК ОПЕРАТОРЫ

Свертка $U = F \star u$ функции точки u и распределения F была определена в гл. V, 4. Вводя семейство функций u_t формулой $u_t(x) = u(t - x)$, мы можем выразить значение $U(t)$ как математическое ожидание:

$$U(t) = \int_{-\infty}^{+\infty} u(t - y) F\{dy\} = E(u_t). \quad (3.1)$$

Мы используем эту запись для того, чтобы вывести критерий собственной сходимости. В этом критерии рассматривается класс $C[-\infty, \infty]$ непрерывных функций, имеющих пределы $u(\pm\infty)$. Эти функции равномерно непрерывны.

Теорема 1. *Последовательность распределений вероятностей F_n сходится к распределению F в собственном смысле тогда и только тогда, когда для каждой функции $u \in C[-\infty, \infty]$ свертки $U_n = F_n \star u$ равномерно сходятся к некоторому пределу U . В случае сходимости $U = F \star u$.*

Доказательство. Условие необходимо, так как равномерная непрерывность u влечет равностепенную непрерывность семейства $\{u_t\}$ и по следствию из § 1 U_n равномерно сходится к $F \star u$. Обратно, условие теоремы влечет за собой сходимость математических ожиданий $E_n(u)$. Мы видели в § 1, что отсюда вытекает сходимость $F_n \rightarrow F$, нужно только доказать, что распределение F собственное. Используем для этой цели критерий 1 из § 1.

Если u монотонно возрастает от 0 до 1, то же самое можно сказать и о каждой из U_n . В силу равномерной сходимости существует такое N , что $|U_n(x) - U_N(x)| < \varepsilon$ при $n > N$ и всех x . Выберем a столь большим, что $U_N(-a) < \varepsilon$. Функция U_N определяется формулой свертки (3.1); заменяя область интегрирования в (3.1) на $-\infty < y \leq -2a$, мы видим, что при $n > N$

$$2\varepsilon > U_n(-a) \geq u(a) F_n(-2a).$$

Так как $u(x)$ с ростом x стремится к единице, то при достаточно больших n и a $F_n(-a)$ может стать сколь угодно малым. По соображениям симметрии то же самое рассуждение применимо к $1 - F(a)$. Следовательно, распределение F является собственным в силу критерия 1.

Для демонстрации силы нашего последнего результата мы выведем одну важную теорему анализа, доказательство которой становится особенно простым при данном вероятностном подходе. (Типичное применение см. в задаче 10.)

Пример. а) *Общие теоремы об аппроксимации.* Свяжем с произвольным распределением вероятностей G семейство распределений G_h , отличающихся от G только масштабным параметром: $G_h(x) = G(h^{-1}x)$. При $h \rightarrow 0$ распределение G_h стремится к распределению, сосредоточенному в нуле, и потому по предыдущей теореме для каждой функции $u \in C[-\infty, \infty]$ имеет место равномерная сходимость $G_h \star u \rightarrow u$ (1).

¹⁾ Для непосредственной проверки заметим, что

$$G_h \star u(t) - u(t) = \int_{-\infty}^{+\infty} [u(t-y) - u(t)] G\{dy/h\}.$$

По данному $\varepsilon > 0$ выберем δ так, чтобы колебания u на каждом интервале длины 2δ были меньше ε . Часть интеграла, взятая по множеству $|y| \leq \delta$, меньше ε , а часть интеграла по множеству $|y| \geq \delta$ стремится к 0, так как распределение G_h приписывает множеству $|y| \geq \delta$ массу, которая стремится к нулю при $h \rightarrow 0$.

Если G имеет плотность g , то свертка $G_h \star u$ задается формулой

$$G_h \star u(t) = \int_{-\infty}^{+\infty} u(y) g\left(\frac{t-y}{h}\right) \frac{1}{h} dy. \quad (3.2)$$

Если g имеет ограниченную производную, то этим же свойством обладает и G_h и можно дифференцировать (3.2) под знаком интеграла. Беря в качестве g нормальную плотность, мы получим следующую лемму об аппроксимации. ►

Лемма об аппроксимации. Для каждой функции $u \in C[-\infty, \infty]$ найдется бесконечно дифференцируемая функция $v \in C[-\infty, \infty]$, такая, что $|u(x) - v(x)| < \varepsilon$ для всех x . ►

В настоящем контексте желательно заменить символ свертки $\star$ более простым обозначением, которое подчеркивало бы, что в (3.1) распределение F играет роль оператора, переводящего u в U . Этот оператор мы будем обозначать готической буквой $\mathfrak{F}$. Мы условимся, что равенство $U = \mathfrak{F}u$ имеет тот же смысл, что и $U = F \star u$. Истинное преимущество этого кажущегося педантичным подхода станет видимым только тогда, когда в том же контексте будут рассматриваться другие типы операторов. Тогда будет удобно видеть сразу, играет ли то или иное распределение свою первоначальную вероятностную роль или же оно служит только аналитическим оператором. После этого объяснения мы вводим

Соглашение об обозначениях. С каждым распределением F мы связываем оператор $\mathfrak{F}$, отображающий $C[-\infty, \infty]$ на себя. Этот оператор ставит в соответствие функции u ее преобразование $\mathfrak{F}u = F \star u$. Всюду, где это возможно, распределения и связанные с ними операторы будут обозначаться соответствующими латинскими и готическими буквами.

Как обычно, $\mathfrak{F}\mathfrak{U}u$ обозначает результат применения $\mathfrak{F}$ к $\mathfrak{U}u$, так что $\mathfrak{F}\mathfrak{U}$ обозначает оператор, соответствующий свертке $F \star G$ двух распределений вероятностей. В частности, $\mathfrak{F}^n$ обозначает оператор, связанный с $F^n \star$, т. е. с n -кратной сверткой F с собой.

Пример. б) Пусть H_a — атомическое распределение, сосредоточенное в точке a . Тогда $\mathfrak{H}_a$ обозначает оператор сдвига $\mathfrak{H}_a u(x) = u(x-a)$. В частности, $\mathfrak{H}_0$ — тождественный оператор: $\mathfrak{H}_0 u = u$.

Определим далее норму $\|u\|$ ограниченной функции u равенством

$$\|u\| = \sup |u(x)|. \quad (3.3)$$

При этом обозначении утверждение « u_n равномерно сходится к u » превращается в простое соотношение $\|u_n - u\| \rightarrow 0$. Заметим, что норма удовлетворяет легко проверяемому неравенству треугольника: $\|u+v\| \leq \|u\| + \|v\|$.

Оператор T называют *ограниченным*, если существует такая

постоянная a , что $\|Tu\| \leq a\|u\|$. Наименьшее число с этим свойством называют *нормой* T и обозначают $\|T\|$. В этих обозначениях основные свойства линейных операторов, связанных с функциями распределения, формулируются следующим образом.

Они положительны, т. е. из $u \geq 0$ вытекает $\mathfrak{F}u \geq 0$. Они имеют норму 1, откуда следует

$$\|\mathfrak{F}u\| \leq \|u\|. \quad (3.4)$$

Наконец, они перестановочны, т. е. $\mathfrak{F}\mathfrak{G} = \mathfrak{G}\mathfrak{F}$.

Определение¹⁾. Пусть $\mathfrak{F}_n$ и $\mathfrak{F}$ — операторы, связанные с распределениями вероятностей F_n и F . Мы пишем $\mathfrak{F}_n \rightarrow \mathfrak{F}$ в том и только том случае, когда

$$\|\mathfrak{F}_n u - \mathfrak{F}u\| \rightarrow 0 \quad (3.5)$$

для каждого $u \in C[-\infty, \infty]$.

Иными словами, $\mathfrak{F}_n \rightarrow \mathfrak{F}$, если $F_n \star u \rightarrow F \star u$ равномерно. Теорему 1 можно теперь сформулировать иначе.

Теорема 1а. $F_n \rightarrow F$ в собственном смысле тогда и только тогда, когда $\mathfrak{F}_n \rightarrow \mathfrak{F}$.

Нижеследующая лемма является основной. Она содержит некоторое неравенство и показывает целесообразность новой терминологии.

Лемма 1. Для операторов, связанных с распределениями вероятностей, имеем

$$\|\mathfrak{F}_1 \mathfrak{F}_2 u - \mathfrak{G}_1 \mathfrak{G}_2 u\| \leq \|\mathfrak{F}_1 u - \mathfrak{G}_1 u\| + \|\mathfrak{F}_2 u - \mathfrak{G}_2 u\|. \quad (3.6)$$

Доказательство. Оператор в левой части равен $(\mathfrak{F}_1 - \mathfrak{G}_1) \mathfrak{F}_2 + (\mathfrak{F}_2 - \mathfrak{G}_2) \mathfrak{G}_1$. Поэтому (3.6) вытекает из неравенства треугольника и того факта, что $\mathfrak{F}_2$ и $\mathfrak{G}_1$ имеют нормы не больше 1. Отметим, что доказательство применимо и к несобственным распределениям вероятностей. ►

Из (3.6) непосредственно следует

Теорема 2. Пусть последовательности $\{F_n\}$ и $\{G_n\}$ распределений вероятностей сходятся в собственном смысле к F и G соответственно. Тогда

$$F_n \star G_n \rightarrow F \star G. \quad (3.7)$$

(Сходимость является собственной в силу определения свертки $F \star G$. Теорема неверна в случае несобственной сходимости. См. задачу 9.)

¹⁾ В терминологии банаховых пространств (3.5) называется *сильной сходимостью*. Заметим, что отсюда не вытекает сходимости в смысле $\|\mathfrak{F}_n - \mathfrak{F}\| \rightarrow 0$. Например, если F_n сосредоточено в точке $1/n$ и $\mathfrak{F}$ — тождественный оператор, то $\mathfrak{F}_n u(x) - \mathfrak{F}u(x) = u(x - n^{-1}) - u(x)$ и (3.5) верно. Однако $\|\mathfrak{F}_n - \mathfrak{F}\| = 2$, так как существует функция v с $|v| \leq 1$, такая, что $v(0) = 1$ и $v(-n^{-1}) = -1$.

В качестве второго применения докажем, что теорема 1 остается справедливой, если сузить рассматриваемый класс функций u до класса особенно «хороших» функций, имеющих производные всех порядков. На этом пути мы получим более глубокий

Критерий 1. Пусть F_n — распределение вероятностей. Если для каждой бесконечно дифференцируемой¹⁾ функции $v \in C[-\infty, \infty]$ последовательность $\{\mathfrak{F}_n v\}$ равномерно сходится, то существует такое собственное распределение вероятностей F , что $F_n \rightarrow F$.

Доказательство. Как было показано в примере а), для любой $u \in C[-\infty, \infty]$ и любого $\varepsilon > 0$ найдется бесконечно дифференцируемая функция v , такая, что $\|u - v\| < \varepsilon$. По неравенству треугольника

$$\|\mathfrak{F}_n u - \mathfrak{F}_m u\| \leq \|\mathfrak{F}_n u - \mathfrak{F}_n v\| + \|\mathfrak{F}_n v - \mathfrak{F}_m v\| + \|\mathfrak{F}_m v - \mathfrak{F}_m u\|. \quad (3.8)$$

Первый и последний члены в правой части меньше ε , а средний член меньше ε при всех достаточно больших n и m . Таким образом, последовательность $\{\mathfrak{F}_n u\}$ сходится равномерно и по теореме 1 $F_n \rightarrow F$. ►

Точно такие же рассуждения с участием E_n и E , определенных в (1.1), дает

Критерий 2. Пусть F_n и F — собственные распределения вероятностей. Если для всех бесконечно дифференцируемых и обращающихся в нуль на бесконечности функций v $E_n(v) \rightarrow E(v)$, то $F_n \rightarrow F$.

Основное неравенство (3.6) по индукции распространяется на свертки более чем двух распределений. Для облегчения ссылок мы выделим этот очевидный результат в виде леммы.

Лемма 2. Пусть $\mathfrak{A} = \mathfrak{F}_1 \dots \mathfrak{F}_n$ и $\mathfrak{B} = \mathfrak{G}_1 \dots \mathfrak{G}_n$, где $\mathfrak{F}_j$ и $\mathfrak{G}_j$ — операторы, соответствующие распределениям вероятностей. Тогда

$$\|\mathfrak{A}u - \mathfrak{B}u\| \leq \sum_{j=1}^n \|\mathfrak{F}_j u - \mathfrak{G}_j u\|. \quad (3.9)$$

В частности,

$$\|\mathfrak{F}^n u - \mathfrak{G}^n u\| \leq n \cdot \|\mathfrak{F}u - \mathfrak{G}u\|. \quad (3.10)$$

(Применения см. в задачах 14 и 15.)

§ 4. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Центральная предельная теорема устанавливает условия, при которых суммы независимых случайных величин распределены асимптотически нормально. Ее роль и значение частично объяснялись в I; гл. X, 1. В нескольких случаях мы применяли ее [последний раз в примере 11, ж) гл. VI]. Она занимает почетное

¹⁾ Под этим подразумевается, что все производные v существуют и принадлежат $C[-\infty, \infty]$.

место в теории вероятностей благодаря своему «возрасту» и той роли, которую она играла в развитии теории вероятностей и все еще играет в приложениях. Естественно поэтому использовать центральную предельную теорему для опробования различных методов, имеющих в нашем распоряжении. По этой причине мы дадим несколько доказательств. Более систематическое изучение (включая необходимые и достаточные условия) будет проведено в гл. IX, XV и XVI. Теперешнее обсуждение отвлекает нас от развития основной темы. Его цель — показать преимущества операторной терминологии на ярком и значительном примере. Кроме того, многим читателям понравится легкий путь к доказательству центральной предельной теоремы в простейшей ситуации. Мы начнем с частного случая, хотя это и приведет к некоторым повторениям.

Теорема 1. (Одинаковые распределения в $\mathcal{R}^1$.) Пусть $X_1, X_2, \dots$ — взаимно независимые случайные величины с одним и тем же распределением F . Допустим, что

$$E(X_k) = 0, \quad \text{Var}(X_k) = 1. \quad (4.1)$$

При $n \rightarrow \infty$ распределение нормированных сумм

$$S_n^* = \frac{X_1 + \dots + X_n}{\sqrt{n}} \quad (4.2)$$

стремится к нормальному распределению $\mathfrak{N}$ с плотностью $p(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$.

В чисто аналитических терминах это означает, что для распределения F с нулевым математическим ожиданием и единичной дисперсией

$$F^n * (x/\sqrt{n}) \rightarrow \mathfrak{N}(x). \quad (4.3)$$

Для доказательства теоремы нам потребуется следующая

Лемма. Если $\mathfrak{F}_n$ — оператор, соответствующий $F_n(x) = F(x/\sqrt{n})$, то для каждой функции $u \in C[-\infty, \infty]$, имеющей три ограниченные производные,

$$n[\mathfrak{F}_n u - u] \rightarrow \frac{1}{2} u'' \quad (4.4)$$

равномерно на всей прямой.

Доказательство. Поскольку $E(X^2) = 1$, можно определить собственное распределение F_n^* равенством

$$F_n^* \{dy\} = n y^2 F_n \{dy\} = n y^2 F \{\sqrt{n} dy\}. \quad (4.5)$$

Замена переменных $\sqrt{n}y = s$ показывает, что F_n^* сходится к распределению, сосредоточенному в нуле. Ввиду (4.1) мы можем

представить разность левой и правой частей (4.4) в виде

$$n [\mathfrak{F}_n u(x) - u(x)] - \frac{1}{2} u''(x) = \\ = \int_{-\infty}^{+\infty} \left[\frac{u(x-y) - u(x) + y u'(x)}{y^2} - \frac{1}{2} u''(x) \right] F_n^\# \{dy\}. \quad (4.6)$$

Из формулы Тейлора, примененной к числителю дроби под интегралом, видно, что при $|y| < \varepsilon$ подынтегральное выражение не превосходит $\frac{1}{6} |y| \cdot \|u'''\| < \varepsilon \|u'''\|$, а при всех y не превосходит $\|u''\|$. Так как $F_n^\#$ сходится к распределению, сосредоточенному в нуле, то при всех достаточно больших n абсолютная величина интеграла не превосходит $\varepsilon (\|u''\| + \|u'''\|)$. Таким образом, левая часть (4.7) равномерно стремится к нулю. ►

Доказательство теоремы 1. Пусть $\mathfrak{G}$ и $\mathfrak{G}_n$ — операторы, связанные соответственно с нормальными распределениями $\mathfrak{N}$ и $\mathfrak{N}(x/\sqrt{n})$. Тогда по основному неравенству (3.10)

$$\|\mathfrak{F}_n u - \mathfrak{G} u\| = \|\mathfrak{F}_n u - \mathfrak{G}_n u\| \leq n \|\mathfrak{F}_n u - \mathfrak{G}_n u\| \leq \\ \leq n \|\mathfrak{F}_n u - u\| + n \|\mathfrak{G}_n u - u\|. \quad (4.7)$$

По предыдущей лемме правая часть стремится к нулю, и, следовательно, по критерию I § 3 $\mathfrak{F}_n \rightarrow \mathfrak{G}$.

Пример. а) *Центральная предельная теорема для случая бесконечной дисперсии.* С методологической точки зрения интересно отметить, что доказательство теоремы 1 применимо без изменений к некоторым распределениям с бесконечной дисперсией (в предположении, что нормирующие константы выбраны надлежащим образом). Например, если случайные величины X_k имеют плотность $f(x)$, такую, что $f(x) = 2|x|^{-3} \log|x|$ при $|x| \geq 1$ и $f(x) = 0$ при $|x| \leq 1$, то суммы $(X_1 + \dots + X_n)/(\sqrt{n} \log n)$ имеют в пределе нормальное распределение (доказательство требует лишь очевидных видоизменений). Необходимые и достаточные условия сходимости к нормальному распределению указаны в гл. IX, 7, и XVII, 5. ►

Приведенный выше метод доказательства получил широкое распространение. Хорошее упражнение предлагается в задаче 16. Здесь мы используем этот метод доказательства центральной предельной теоремы в более общем случае. Сформулированная ниже теорема касается двумерного случая, но ее можно распространить и на $\mathcal{R}'$.

Теорема 2. (Многомерный случай.) Пусть $\{X_n\}$ — последовательность взаимно независимых двумерных случайных величин с одним и тем же распределением F . Допустим, что математи-

ческие ожидания равны нулю и что матрица ковариаций равна

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \quad (4.8)$$

При $n \rightarrow \infty$ распределение $(X_1 + \dots + X_n)/\sqrt{n}$ сходится к двумерному нормальному распределению с нулевым математическим ожиданием и матрицей ковариаций C .

Доказательство. Если использовать матричные обозначения гл. III, 5, то почти без изменений применимо предыдущее доказательство. Так как нижние индексы уже заняты, то мы будем обозначать точки плоскости с помощью векторов-строк $x = (x^{(1)}, x^{(2)})$. В этих обозначениях $u(x)$ — функция двух переменных. Ее частные производные будут помечаться нижними индексами. Так, $u' = (u_i, u_2)$ — вектор-строка и $u'' = (u_{jk})$ — симметричная (2×2) -матрица. Формула Тейлора принимает вид

$$u(x-y) = u(x) - y u'(x) + \frac{1}{2} y u''(x) y^T + \dots, \quad (4.9)$$

где y^T получается транспонированием y , т. е. y^T — вектор-столбец с компонентами $y^{(1)}$ и $y^{(2)}$. Определим по аналогии с (4.5) собственное распределение вероятностей формулой

$$F_n^\# \{dy\} = nq(y) F \{\sqrt{n} dy\}, \quad \text{где } 2q(y) = \frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2}.$$

Так же как и в предыдущем доказательстве, $F_n^\#$ сходится к распределению, сосредоточенному в начале координат. Соотношение (4.6) заменяется на

$$\begin{aligned} n [\mathfrak{F}_n u(x) - u(x)] - \frac{1}{2} m(x) = \\ = \int_{\mathcal{A}^2} \frac{u(x-y) - u(x) + y u'(x) - y u''(x) y^T}{q(y)} F_n^\# \{dy\}, \end{aligned} \quad (4.10)$$

где ¹⁾

$$m(x) = E(y u''(x) y^T) = u_{11}(x) \sigma_1^2 + 2u_{12}(x) \rho\sigma_1\sigma_2 + u_{22}(x) \sigma_2^2. \quad (4.11)$$

(Здесь E обозначает математическое ожидание по отношению к F). В силу (4.9) подынтегральное выражение стремится к нулю, и, так же как и в предыдущей лемме, мы заключаем, что $n[\mathfrak{F}_n u - u] \rightarrow m/2$ равномерно. Доказательство теоремы не требует изменений. ►

Примеры. б) *Случайное блуждание в d -мерном пространстве.* Пусть $X_1, X_2, \dots$ — независимые случайные векторы с одним и тем же распределением, которое может быть описано следующим

¹⁾ Очевидно, что $m(x)$ — след матрицы $C u''$ т. е. сумма ее диагональных элементов). В одномерном случае $m(x) = \sigma^2 u''(x)$.

образом. Вектор $\mathbf{X}_k$ имеет случайное направление (см. I, 10), и длина L — случайная величина с $E(L^2) = 1$. По причинам симметрии матрица ковариаций C диагональна, с диагональными элементами $\sigma_j^2 = 1/d$. Распределение нормированной суммы $S_n/\sqrt{n}$ сходится к нормальному распределению с матрицей ковариаций C . Распределение длины вектора $S_n/\sqrt{n}$ сходится поэтому к распределению квадратного корня из суммы квадратов независимых нормальных величин. Как было сказано в гл. II, 3, последнее распределение имеет плотность

$$w_d(r) = \frac{d^{\frac{1}{2}}}{\frac{1}{2} d^{-1} \Gamma\left(\frac{1}{2} d\right)} e^{-\frac{1}{2} dr^2} r^{d-1}. \quad (4.12)$$

Этот результат показывает влияние числа измерений; он применим, в частности, к примеру 10, д) гл. I.

в) *Случайное рассеивание популяций.* В качестве практического применения предыдущего примера рассмотрим распространение популяции дубов в доисторическое время. Если бы новые растения возникали только из семян, упавших с материнского дерева, то сеянцы были бы расположены около него. Тогда расстояние дерева n -го поколения от исходного было бы распределено приблизительно нормально. При этих условиях площадь, покрытая потомками некоторого дерева, была бы, грубо говоря, пропорциональна возрасту дерева. Наблюдения показывают, что действительное положение вещей несовместимо с этой гипотезой. Биологи приходят к выводу, что наблюдаемое рассеивание в большой степени определяется птицами, переносящими семена на большие расстояния¹⁾. ►

Обобщим теперь теорему 1 на случай по-разному распределенных величин. Может возникнуть впечатление, что условия введены искусственно с единственной целью сохранить путь доказательства. В действительности оказывается, что эти условия необходимы и для центральной предельной теоремы с классической нормировкой, использованной в (4.17) (см. гл. XV, 6).

Теорема 3. (Линдберг.)²⁾ Пусть $\mathbf{X}_1, \mathbf{X}_2, \dots$ — взаимно неза-

¹⁾ Skellam J. G., *Biometrika*, 38 (1951), 196—218.

²⁾ Lindeberg J. W., *Math. Zeit.*, 15 (1922), 211—235. Частные случаи были известны ранее, но Линдберг дал первую общую формулировку, содержащую теорему 1. Необходимость условия Линдберга при классической нормировке была доказана Феллером (там же, 40 (1935), см. гл. XV, 6).

Метод Линдберга казался сложным и был практически заменен методом характеристических функций, развитым П. Леви. Троттер (*Archiv der Math.*, 9 (1959), 226—234) показал, что современные методы позволяют изложить доказательство Линдберга простым и интуитивно понятным способом. В доказательствах этого параграфа использованы идеи Троттера.

висимые одномерные случайные величины с распределениями $F_1, F_2, \dots$. Пусть

$$E(X_k) = 0, \quad \text{Var}(X_k) = \sigma_k^2 \quad (4.13)$$

и

$$s_n^2 = \sigma_1^2 + \dots + \sigma_n^2. \quad (4.14)$$

Предположим, что при каждом $t > 0$

$$\frac{1}{s_n^2} \sum_{k=1}^n \int_{|y| > ts_n} y^2 F_k \{dy\} \rightarrow 0, \quad (4.15)$$

или, что то же самое,

$$\frac{1}{s_n^2} \sum_{k=1}^n \int_{|y| < ts_n} y^2 F_k \{dy\} \rightarrow 1. \quad (4.16)$$

Тогда распределение нормированной суммы

$$S_n^* = \frac{X_1 + \dots + X_n}{s_n} \quad (4.17)$$

сходится к нормальному распределению $\mathfrak{N}$ с нулевым математическим ожиданием и единичной дисперсией.

Условие Линдберга (4.15) гарантирует, что дисперсии σ_k^2 отдельных слагаемых малы по сравнению с суммой s_n^2 в том смысле, что при любом $\varepsilon > 0$ и всех достаточно больших n

$$\frac{\sigma_k}{s_n} < \varepsilon, \quad k = 1, \dots, n. \quad (4.18)$$

В самом деле, очевидно, что σ_k^2/s_n^2 меньше, чем t^2 плюс левая часть (4.15). Положив $t = \varepsilon/2$, видим, что (4.15) влечет за собой (4.18).

Теорема 3 обобщается на многомерный случай способом, указанным в теореме 2. См. также задачи 17—20.

Доказательство. Поставим в соответствие каждому распределению F_k нормальное распределение G_k с нулевым математическим ожиданием и той же самой дисперсией σ_k^2 . Распределение $F_k(xs_n)$ случайной величины X_k/s_n зависит и от k , и от n . Соответствующий оператор обозначим через $\mathfrak{F}_{k,n}$. Аналогично оператор $\mathfrak{G}_{k,n}$ соответствует нормальному распределению $G_k(xs_n)$. В силу (3.9) нам достаточно доказать, что

$$\sum_{k=1}^n \|\mathfrak{F}_{k,n} u - \mathfrak{G}_{k,n} u\| \rightarrow 0 \quad (4.19)$$

для каждой функции $u \in C[-\infty, \infty]$ с тремя ограниченными производными. Мы поступим так же, как при доказательстве теоремы 1.

Однако теперь (4.6) заменится n соотношениями

$$\mathfrak{F}_{k,n} u(x) - u(x) - \frac{\sigma_k^2}{2s_n^2} u''(x) = \int_{-\infty}^{+\infty} \left[\frac{u(x-y) - u(x) + y u'(x) - \frac{1}{2} u''(x)}{y^2} \right] \cdot y^2 F_k \{s_n dy\}. \quad (4.20)$$

Разбивая область интегрирования на части $|y| \leq \varepsilon$ и $|y| > \varepsilon$ и используя те же самые оценки, что и в (4.6), получаем

$$\left\| \mathfrak{F}_{k,n} u - u - \frac{\sigma_k^2}{2s_n^2} u'' \right\| \leq \varepsilon \|u'''\| \frac{\sigma_k^2}{s_n^2} + \|u''\| \cdot \int_{|y| > \varepsilon} y^2 F_k \{s_n dy\}. \quad (4.21)$$

Условие Линдеберга (4.15) с $t = \varepsilon$ гарантирует теперь, что при всех достаточно больших n

$$\sum_{k=1}^n \left\| \mathfrak{F}_{k,n} u - u - \frac{\sigma_k^2}{2s_n^2} u'' \right\| \leq \varepsilon (\|u'''\| + \|u''\|). \quad (4.22)$$

Для нормальных распределений G_k условие Линдеберга (4.15) является простым следствием (4.18). Поэтому неравенство (4.22) остается верным, если в нем $\mathfrak{F}_{k,n}$ заменить на $\mathfrak{G}_{k,n}$. Соединяя оба этих неравенства, получаем (4.19). Доказательство закончено. ►

Примеры. г) *Равномерные распределения.* Пусть X_k распределено равномерно, с плотностью $1/(2a_k)$, в промежутке между $-a_k$ и a_k . Тогда $\sigma_k^2 = a_k^2/3$. Легко видеть, что условия теоремы выполнены, если числа a_k ограничены, а $a_1^2 + \dots + a_n^2 \rightarrow \infty$. В самом деле, в этом случае сумма (4.15) равна нулю при всех достаточно больших n . С другой стороны, если $\sum a_k^2 < \infty$, то s_n остаются ограниченными и (4.15) не может выполняться. В этом случае центральная предельная теорема не применима (вместо этого мы приходим к бесконечным сверткам, которые будут изучены в § 5).

Менее очевидный случай, где центральная предельная теорема не применима, — это случай $a_k^2 = 2^k$. Тогда $3s_n^2 = 2^{n+1} - 2 < 2a_n^2$ и, очевидно, левая часть (4.15) больше $1/2$ для $t < 1/100$. Эти примеры показывают, что назначение условия (4.15) — обеспечить асимптотическую пренебрегаемость отдельных X_k : вероятность того, что какое-нибудь из слагаемых X_k будет того же порядка, что и сумма S_n , должна стремиться к нулю.

д) *Ограниченные величины.* Допустим, что X_k равномерно ограничены, т. е. все распределения F_k сосредоточены на некотором конечном интервале $[-a, a]$. Условие Линдеберга (4.15) выполняется тогда и только тогда, когда $s_n \rightarrow \infty$. ►

е) Пусть F — распределение вероятностей с нулевым математическим ожиданием и единичной дисперсией. Выберем последо-

вательность положительных чисел σ_k и положим $F_n(x) = F(x/\sigma_n)$ (так что дисперсия F_k равна σ_k^2). Условие Линдберга удовлетворяется тогда и только тогда, когда $s_n \rightarrow \infty$ и $\sigma_n/s_n \rightarrow 0$. В самом деле, мы знаем, что эти условия являются необходимыми. С другой стороны, левая часть (4.15) сводится к выражению

$$s_n^{-2} \sum_{k=1}^n \sigma_k^2 \int_{|x| < ts_n/\sigma_k} x^2 F \{dx\}.$$

При сформулированных условиях s_n/σ_k стремится к бесконечности равномерно по $k=1, \dots, n$. Поэтому при достаточно большом n все интегралы, входящие в сумму, будут меньше ε . Таким образом, сумма будет меньше εs_n^2 , и поэтому соотношение (4.15) будет справедливо.

Методологически интересно заметить, что изложенный метод доказательства применим и к некоторым последовательностям случайных величин, *не имеющих математических ожиданий*. Конечно, нормирующие константы должны быть другими. Мы вернемся к этим вопросам в гл. XV, 6, где мы также подвергнем дальнейшему изучению природу условия Линдберга. (См. задачи 19, 20.)

Мы закончим настоящий экскурс вариантом центральной предельной теоремы для *случайных сумм*. Идея состоит в следующем. Если в теореме 1 заменить фиксированное число n слагаемых пуассоновской случайной величиной N с математическим ожиданием n , то кажется правдоподобным, что распределение S_N будет асимптотически нормально. Сходные ситуации возникают в статистике и физике, когда число наблюдений не фиксируется заранее.

Мы рассмотрим только суммы вида $S_N = X_1 + \dots + X_N$, где X_j и N — взаимно независимые случайные величины. Мы предположим, что X_j имеют одно и то же распределение F с нулевым математическим ожиданием и дисперсией 1. Используя обозначение § 2, мы покажем, что верна

Теорема 4¹. (Случайные суммы.) Пусть $N_1, N_2, \dots$ — положительные целочисленные случайные величины, такие, что

$$\frac{N_n}{n} \xrightarrow{P} 1. \quad (4.23)$$

Тогда распределение $S_{N_n}/\sqrt{n}$ сходится к $\mathfrak{N}$.

¹) Обобщение на зависимые X_j см. у Биллингсли: Billingsley P., Limit theorems for randomly selected partial sums, Ann. Math. Statist., 33 (1963), 85—92. Если отбросить (4.23), то получаются предельные теоремы новой формы, см.: Robbins H. E., The asymptotic distribution of the sum of a random number of random variables, Bull. Amer. Math. Soc., 54 (1948), 1151—1161.

Обобщения центральной предельной теоремы на другие типы зависимых величин читатель может найти в книге М. Лозва (о симметрично зависимых величинах см. задачу 21).

Интересно отметить, что величина $\frac{1}{\sqrt{n}} S_{N_n}$ не обязана иметь единичную дисперсию. Действительно, теорема применима к случаям, где $E(N_n) = \infty$, и, даже если математическое ожидание существует, из (4.23) не следует сходимость $\frac{1}{n} E(N_n) \rightarrow 1$. Нормировка, приводящая к единичной дисперсии, может быть невозможна, а если она и возможна, то усложняет доказательство.

Доказательство. Чтобы избежать двойных индексов, положим $P\{N_n = k\} = a_k$, подразумевая, что a_k зависит от n . Оператор, связанный с S_{N_n} , задается формальным степенным рядом $\sum a_k \mathfrak{F}^k$. Пусть $\mathfrak{F}_n$, как и в доказательстве теоремы 1, — оператор, соответствующий $F(x/\sqrt{n})$. Так как $F_n^{\star} \rightarrow \mathfrak{N}$, то достаточно доказать, что

$$\sum_{k=1}^{\infty} a_k \mathfrak{F}_n^k u - \mathfrak{F}_n^n u \rightarrow 0 \quad (4.24)$$

равномерно для каждой функции $u \in C[-\infty, \infty]$ с тремя ограниченными производными.

Используя очевидное разложение на множители и основное неравенство (3.9), видим, что

$$\|\mathfrak{F}_n^k u - \mathfrak{F}_n^n u\| \leq \|\mathfrak{F}_n^{k-n} u - u\| \leq |k-n| \cdot \|\mathfrak{F}_n u - u\|. \quad (4.25)$$

В силу (4.23) сумма коэффициентов a_k с $|k-n| > \varepsilon n$ меньше ε при всех достаточно больших n . Для этих n норма левой части (4.24) не превосходит

$$\sum_{k=1}^{\infty} a_k \|\mathfrak{F}_n^k u - \mathfrak{F}_n^n u\| \leq 2\varepsilon \cdot \|u\| + 2\varepsilon \cdot n \|\mathfrak{F}_n u - u\|. \quad (4.26)$$

Из доказательства леммы видно, что правая часть не превосходит $2\varepsilon \|u\| + 3\varepsilon \|u''\|$ при всех достаточно больших n , так что (4.24) выполняется в смысле равномерной сходимости. ►

§ 5*) БЕСКОНЕЧНЫЕ СВЕРТКИ

Нижеследующие теоремы даны как по причине их внутреннего интереса, так и потому, что они хорошо иллюстрируют применение наших критериев. Более сильные варианты можно найти в гл. VII, 8, XI, 9, и XVII, 10.

Мы обозначаем через $X_1, X_2, \dots$ взаимно независимые случайные величины с распределениями $F_1, F_2, \dots$. Предполагается, что $E(X_j) = 0$ и что $\sigma_k^2 = E(X_k^2)$ существует.

Теорема. Если $\sigma^2 = \sum \sigma_k^2 < \infty$, то распределения¹⁾ G_n частных сумм $X_1 + \dots + X_n$ сходятся к распределению вероятностей G с нулевым математическим ожиданием и дисперсией σ^2 .

^{*}) Этот параграф не используется в дальнейшем.

¹⁾ В гл. VII, 8, было показано, что и сами случайные величины S_n сходятся к пределу.

Доказательство. Чтобы установить существование собственного предела G , достаточно (по теореме 1 § 3) показать, что для любой бесконечно дифференцируемой функции $u \in C[-\infty, \infty]$ последовательность функций $\mathfrak{F}_1 \mathfrak{F}_2 \dots \mathfrak{F}_n u$ сходится равномерно при $n \rightarrow \infty$. Теперь при $n > m$ очевидное разложение на множители дает

$$\|\mathfrak{F}_1 \dots \mathfrak{F}_n u - \mathfrak{F}_1 \dots \mathfrak{F}_m u\| \leq \|\mathfrak{F}_{m+1} \dots \mathfrak{F}_n u - u\|. \quad (5.1)$$

Так как $E(X_k) = 0$, то верно тождество

$$\mathfrak{F}_k u(x) - u(x) = \int_{-\infty}^{+\infty} [u(x-y) - u(x) + y u'(x)] F_k(dy). \quad (5.2)$$

Из формулы Тейлора видим, что подынтегральное выражение не превосходит $\|u''\| y^2$. Поэтому $\|\mathfrak{F}_k u - u\| \leq \sigma_k^2 \|u''\|$. По основному неравенству (3.9) величина (5.1) не превосходит $(\sigma_{m+1}^2 + \dots + \sigma_n^2) \times \|u''\|$. Следовательно, существует собственное распределение G , такое, что $G_n \rightarrow G$. Так как G_n имеет дисперсию $\sigma_1^2 + \dots + \sigma_n^2$, то второй момент G существует и не больше σ^2 . По критерию примера 1, д) отсюда вытекает, что распределение G имеет математическое ожидание, равное нулю. Наконец, G есть свертка G_n и предельного распределения для суммы $X_{n+1} + \dots + X_{n+k}$. Следовательно, дисперсия G не может быть меньше дисперсии G_n . Этим заканчивается доказательство. ►

Примеры. а) В примере I, 11, в) случайный выбор точки между 0 и 1 осуществляется посредством последовательных бросаний монеты. В терминологии настоящего параграфа это означает представление равномерного распределения в виде бесконечной свертки. Пример I, 11, г) показывает, что бесконечная свертка соответствующих распределений с четными номерами есть сингулярное распределение. (См. XVII, 10.)

б) Пусть Y_k — независимые случайные величины с $E(Y_k) = 0$ и $E(Y_k^2) = 1$. Тогда распределения частных сумм ряда $\sum b_k Y_k$ сходятся, если $\sum b_k^2 < \infty$. Этот факт был использован при обсуждении свойств нормальных случайных процессов в гл. III, 7.

в) *Применение к процессам размножения.* Пусть X_n — положительная случайная величина с плотностью $\lambda_n e^{-\lambda_n t}$. Тогда $E(X_n) = 1/\lambda_n$, и в случае, когда $m = \sum \lambda_n^{-1} < \infty$, наша теорема применима к центрированным случайным величинам $X_n - \lambda_n^{-1}$. Это замечание приводит к вероятностной интерпретации расходящегося процесса чистого размножения (описанного в I, гл. XVII, 3, 4). «Частица» движется последовательными скачками. Времена пребывания в отдельных состояниях $X_1, X_2, \dots$ — независимые случайные величины, распределенные по показательному закону. Сумма $S_n = X_1 + \dots + X_n$ представляет собой момент n -го скачка. Если $\lim E(S_n) = m < \infty$, то распределение S_n сходится к собственному пределу G . Тогда величина $G(t)$ равна

вероятности того, что до момента t произойдет бесконечно много скачков.

г) Применение к флуктуационным шумам, телефонным вызовам и тому подобному см. в задаче 22. ►

§ 6. ТЕОРЕМЫ О ВЫБОРЕ

Один из стандартных методов доказательства сходимости числовой последовательности состоит в том, что сначала устанавливается существование хотя бы одной предельной точки, а затем—ее единственность. Аналогичный способ применим и к распределениям. Аналог утверждения о существовании предельной точки дается следующей важной теоремой, приписываемой обычно Хелли. Как и все теоремы этого параграфа, она верна в пространстве любого числа измерений (специальный случай был использован в I; гл. XI,6).

Теорема 1. (i) Каждая последовательность $\{F_k\}$ распределений вероятностей в $\mathcal{R}^r$ содержит подпоследовательность $F_{n_1}, F_{n_2}, \dots$, сходящуюся (в собственном или несобственном смысле) к некоторому пределу F .

(ii) Для того чтобы все такие пределы были собственными, необходимо и достаточно, чтобы последовательность $\{F_n\}$ была стохастически ограниченной (см. определение 2 в § 2).

(iii) Для того чтобы $F_n \rightarrow F$, необходимо и достаточно, чтобы предел любой сходящейся подпоследовательности был равен F .

Доказательство основано на следующей лемме.

Лемма. Пусть $a_1, a_2, \dots$ —произвольная последовательность точек. Каждая последовательность $\{u_n\}$ числовых функций содержит подпоследовательность $u_{n_1}, u_{n_2}, \dots$, сходящуюся во всех точках a_j (может быть, к $\pm\infty$).

Доказательство. Мы используем «диагональный метод» Кантора. Найдется последовательность номеров $v_1, v_2, \dots$, такая, что последовательность значений $u_{v_k}(a_1)$ сходится. Чтобы избежать сложных индексов, положим $u_k^{(1)} = u_{v_k}$, так что $\{u_k^{(1)}\}$ —подпоследовательность $\{u_n\}$, сходящаяся в точке a_1 . Из этой подпоследовательности выделим другую $u_1^{(2)}, u_2^{(2)}, \dots$, сходящуюся в точке a_2 . Продолжая по индукции, мы построим при каждом n последовательность $u_1^{(n)}, u_2^{(n)}, \dots$, сходящуюся в точке a_n и содержащуюся в предыдущей последовательности. Рассмотрим теперь «диагональную» последовательность $u_1^{(1)}, u_2^{(2)}, u_3^{(3)}, \dots$. Эта последовательность, за исключением ее $n-1$ первых членов, содержится в n -й последовательности $u_1^{(n)}, u_2^{(n)}, \dots$ и потому сходится в a_n . Так как это верно при каждом n , то диагональная последовательность $\{u_n^{(n)}\}$ сходится во всех точках $a_1, a_2, \dots$. Лемма доказана. ►

Доказательство теоремы 1. (i) Выберем в качестве $\{a_j\}$ какую-либо всюду плотную последовательность и выберем подпоследовательность $\{F_{n_k}\}$, которая сходится в каждой точке a_j . Обозначим пределы через $G(a_j)$. Для любой точки x , не принадлежащей множеству $\{a_j\}$, определим $G(x)$ как точную нижнюю грань всех значений $G(a_j)$ для $a_j > x$. Таким образом определенная функция G возрастает от 0 до 1, но не обязана быть непрерывной справа: можно лишь утверждать, что значение $G(x)$ лежит между пределами $G(x+)$ и $G(x-)$. Однако можно перопределить G в точках разрыва, так чтобы получить непрерывную справа функцию F , которая совпадает с G во всех точках непрерывности. Пусть x — такая точка. Существуют такие две точки $a_i < x < a_j$, что

$$G(a_j) - G(a_i) < \varepsilon, \quad G(a_i) \leq F(x) \leq G(a_j). \quad (6.1)$$

В силу монотонности F_n имеем $F_{n_k}(a_i) \leq F_{n_k}(x) \leq F_{n_k}(a_j)$. Устремляя $k \rightarrow \infty$, мы выводим из (6.1), что предельное значение последовательности $\{F_{n_k}(x)\}$ не может отличаться от $F(x)$ более чем на ε , и поэтому $F_{n_k}(x) \rightarrow F(x)$ во всех точках непрерывности.

(ii) Напомним теперь, что сходящаяся последовательность распределений сходится собственно в том и только том случае, когда она стохастически ограничена. Коль скоро (i) установлено, другие утверждения почти очевидны. ►

Теорема о выборе является крайне важной. Приводимая ниже известная теорема теории чисел может дать представление о ее удивительной силе. Она также служит напоминанием о том, что вероятностная терминология не должна затемнять более широкий характер развиваемой теории.

Примеры. а) *Теоретико-числовая теорема о равномерном распределении дробных долей*¹⁾. Пусть α — иррациональное число и α_n — дробная часть $n\alpha$. Обозначим через $N_n(x)$ число членов среди $\alpha_1, \alpha_2, \dots, \alpha_n$, не превосходящих x . Тогда $n^{-1}N_n(x) \rightarrow x$ при всех $0 < x < 1$.

Доказательство. Мы будем рассматривать распределения и функции, заданные на окружности единичной длины. Иными словами, координаты складываются по mod 1. (Эта идея уже была объяснена в гл. II, 8. Заметим, что такое удобное понятие, как функция распределения, теряет на окружности смысл, однако понятие распределения как меры сохраняется). Пусть F_n — атомическое распределение, сосредоточенное в n точках $\alpha, 2\alpha, \dots, n\alpha$

¹⁾ Эту теорему обычно приписывают Г. Вейлю, хотя она открыта независимо друг от друга Бодем (Bohl) и Серпинским (Sierpiński). Из книги Hardy G. H., Wright E. M., *Theory of numbers*, Oxford, 1945, 378—381, можно понять трудности доказательства при вероятностном подходе к теореме.

и приписывающее каждой из них вероятность $1/n$. По теореме о выборе существует последовательность номеров $n_1, n_2, \dots$, такая, что $F_{n_k} \rightarrow F$, где F — *собственное* распределение вероятностей (так как окружность — ограниченное множество). Беря свертку F_{n_k} с произвольной непрерывной функцией u , получим

$$\frac{1}{n_k} [u(x - \alpha) + u(x - 2\alpha) + \dots + u(x - n_k \alpha)] \rightarrow v(x). \quad (6.2)$$

Ясно, что замена x на $x - \alpha$ не меняет асимптотического поведения левой части. Следовательно, $v(x) = v(x - \alpha)$ для всех x . Отсюда вытекает в свою очередь, что $v(x) = v(x - k\alpha)$ для $k = 1, 2, \dots$. По следствию леммы 2 гл. V, 4, точки $\alpha, 2\alpha, \dots$ образуют всюду плотное множество. Поэтому $v = \text{const}$. Мы показали, таким образом, что для любой непрерывной функции u свертка $F \star u$ является константой. Отсюда следует, что распределение F должно приписывать одну и ту же вероятность интервалам равной длины, т. е. $F(I)$ должно быть равно длине интервала I . Так как никакие другие предельные распределения существовать не могут, то вся последовательность $\{F_n\}$ сходится к этому распределению, что доказывает теорему. Мы назовем распределение F *равномерным распределением* на окружности.

б) *Сходимость моментов*. Пусть F_n и F — распределения, имеющие конечные моменты всех порядков, которые мы обозначим $\mu_k^{(n)}$ и μ_k соответственно. Как мы знаем из гл. VII, 3, различные функции распределения могут иметь одну и ту же последовательность моментов. Поэтому мы не всегда можем сделать по поведению $\mu_k^{(n)}$ вывод, что $F_n \rightarrow F$. Однако если F — *единственное распределение, имеющее моменты* $\mu_1, \mu_2, \dots$, и если $\mu_k^{(n)} \rightarrow \mu_k$ при $k = 1, 2, \dots$, то $F_n \rightarrow F$. В самом деле, результат примера 1, д) показывает, что всякая сходящаяся подпоследовательность $\{F_n\}$ имеет предел F .

в) *Сепарабельность*. Назовем для краткости распределение рациональным, если оно сосредоточено в конечном числе рациональных точек и приписывает каждой из них рациональный вес. Любое распределение F можно представить в виде предела последовательности $\{F_n\}$ рациональных распределений. Мы можем считать, что математическое ожидание F_n равно нулю, так как этого всегда можно добиться, добавляя новый атом и сколь угодно мало изменяя остальные веса. Но множество всех рациональных распределений счетно и его можно записать в форме последовательности $G_1, G_2, \dots$. Таким образом, *существует последовательность $\{G_n\}$ распределений с нулевыми математическими ожиданиями и конечными дисперсиями, такая, что каждое распределение F является пределом некоторой подпоследовательности $\{G_{n_k}\}$* . ►

Теореме 1 была придана форма, наиболее полезная для теории вероятностей. Но эта форма излишне ограничительна. Доказа-

тельство опиралось на тот факт, что последовательность $\{F_n\}$ монотонных функций с $F_n(-\infty)=0$, $F_n(\infty)=1$ содержит сходящуюся подпоследовательность. Но это утверждение остается верным, если условие $F_n(\infty)=1$ заменить менее жестким условием, а именно, что при любом x числовая последовательность $\{F_n(x)\}$ ограничена. Предел F будет при этом конечным, но он не обязан быть ограниченным. Индуцированная функцией F мера будет принимать конечные значения на интервалах $-\infty, x$, но может быть бесконечна на $-\infty, \infty$. Аналогично можно изменить условия на $-\infty$. В результате мы приходим к следующему обобщению теоремы 1, в котором символ $\mu_n \rightarrow \mu$ употребляется в том смысле, что это соотношение верно для всех конечных интервалов.

Теорема 2. Пусть $\{\mu_n\}$ — такая последовательность мер, что при каждом x числовая последовательность $\mu_n \{-x, x\}$ ограничена. Тогда существует мера μ и последовательность $n_1, n_2, \dots$, такие, что $\mu_{n_k} \rightarrow \mu$.

Аналоги теоремы о выборе имеют место для многих классов функций. Особенно полезна следующая теорема, называемая обычно или теоремой Асколи, или теоремой Арцела.

Теорема 3. Пусть $\{u_n\}$ — последовательность равномерно непрерывных функций¹⁾ и $|u_n| \leq 1$. Тогда найдется подпоследовательность $\{u_{n_k}\}$, сходящаяся к непрерывному пределу u . Эта сходимость равномерна в каждом конечном интервале.

Доказательство. Выберем снова всюду плотную последовательность точек a_j и подпоследовательность $\{u_{n_k}\}$, сходящуюся в точках $a_1, a_2, \dots$. Обозначим ее предел в точке a_j через $u(a_j)$. Тогда

$$|u_{n_r}(x) - u_{n_s}(x)| \leq |u_{n_r}(x) - u_{n_r}(a_j)| +$$

$$+ |u_{n_s}(x) - u_{n_s}(a_j)| + |u_{n_r}(a_j) - u_{n_s}(a_j)|. \quad (6.3)$$

По предположению последний член справа стремится к 0. Вследствие равномерной непрерывности для любой точки x найдется такая точка a_j , что

$$|u_n(x) - u_n(a_j)| < \varepsilon \quad (6.4)$$

для всех n , и для любого конечного интервала I можно найти надлежащее конечное множество таких точек a_j . Отсюда вытекает, что правая часть в (6.3) будет меньше 3ε для всех достаточно больших r и s равномерно относительно $x \in I$. Таким образом, существует предел $u(x) = \lim u_{n_r}(x)$, и вследствие (6.4) имеет место неравенство $|u(x) - u(a_j)| \leq \varepsilon$, откуда вытекает непрерывность u . ▶

¹⁾ Для каждого $\varepsilon > 0$ найдется такое $\delta > 0$, что из $|x' - x''| < \delta$ вытекает $|u_n(x') - u_n(x'')| < \varepsilon$ для всех n .

§ 7*). ЭРГОДИЧЕСКИЕ ТЕОРЕМЫ ДЛЯ ЦЕПЕЙ МАРКОВА

Пусть K — стохастическое ядро, сосредоточенное на конечном или бесконечном интервале Ω (в соответствии с определением 1 гл. VI, 11, это означает, что K есть функция двух переменных, точки x и множества Γ ; при фиксированном Γ K — бэровская функция x , а при фиксированном $x \in \Omega$ — распределение вероятностей на Ω). В многомерных пространствах интервал Ω может быть заменен областью более общего вида; при этом теория не требует никаких изменений.

В гл. VI, 11, было показано, что существует марковская цепь $(X_0, X_1, \dots)$ с переходными вероятностями, определяемыми K . Распределение γ_0 начальной величины X_0 может быть выбрано произвольно. После этого распределения $X_1, X_2, \dots$ определяются последовательно по формуле

$$\gamma_n(\Gamma) = \int_{\Omega} \gamma_{n-1}(dx) K(x, \Gamma). \quad (7.1)$$

В частности, если γ_0 сконцентрировано в точке x_0 , то $\gamma_n(\Gamma) = K^{(n)}(x_0, \Gamma)$ совпадает с вероятностью перехода из x_0 в Γ .

Определение 1. Назовем меру α строго положительной на Ω , если $\alpha(I) > 0$ для любого открытого интервала $I \subset \Omega$. Назовем ядро K строго положительным, если $K(x, I) > 0$ для каждого x и каждого открытого интервала, входящего в Ω .

Определение 2. Назовем ядро эргодическим, если существует строго положительное распределение вероятностей α , такое, что $\gamma_n \rightarrow \alpha$ при любом начальном распределении γ_0 .

Это равносильно требованию сходимости

$$K^{(n)}(x, I) \rightarrow \alpha(I) > 0 \quad (7.2)$$

для каждого интервала непрерывности распределения α . Приведенное определение такое же, как и для дискретного случая (1; гл. XV). Его смысл был объяснен в примерах гл. VI, 11.

Так как с наиболее общими стохастическими ядрами могут быть связаны различные патологии, то мы ограничимся рассмотрением ядер, непрерывно зависящих от x . Простейший способ выразить это свойство связан с преобразованиями множества непрерывных функций, порождаемыми K . Пусть u — функция, ограниченная и непрерывная на рассматриваемом интервале Ω . Положим $u_0 = u$ и определим по индукции

$$u_n(x) = \int_{\Omega} K(x, dy) u_{n-1}(y). \quad (7.3)$$

* Этот материал рассматривается как по причине его важности, так и потому, что он дает яркую иллюстрацию применения теоремы о выборе. Явным образом он в дальнейшем не используется.

Это преобразование функций двойственно с преобразованием мер (7.1). Отметим, что в обоих случаях индексы (всюду в этом параграфе) указывают на преобразование, индуцированное K .

Свойство регулярности, которое мы собираемся потребовать от K , состоит, грубо говоря, в том, что функция u_1 должна быть не хуже, чем u_0 . Нижеследующее определение дает точное выражение нашему требованию, но выглядит довольно формальным. Примеры покажут, однако, что это определение применимо в типичных ситуациях.

Определение 3. Назовем ядро K регулярным, если семейство преобразований u_k равностепенно непрерывно¹⁾ при равномерно непрерывной на Ω функции u_0 .

Примеры. а) *Свертки.* Если F есть распределение вероятностей, то свертка

$$u_n(x) = \int_{-\infty}^{+\infty} u_{n-1}(x-y) F\{dy\}$$

представляет собой частный случай преобразования (7.3). (Само собой понятно, что здесь $K(x, I) = F\{I - x\}$.) Это преобразование регулярно, поскольку в силу равномерной непрерывности u_0 существует такое δ , что из $|x' - x''| < \delta$ следует $|u_0(x') - u_0(x'')| < \varepsilon$ и по индукции получаем $|u_n(x') - u_n(x'')| < \varepsilon$ для всех n .

б) Пусть Ω — единственный интервал, и пусть K имеет плотность k , непрерывную в замкнутом единичном квадрате. Тогда K регулярно. Действительно,

$$|u_n(x') - u_n(x'')| \leq \int_0^1 |k(x', y) - k(x'', y)| \cdot |u_{n-1}(y)| dy. \quad (7.4)$$

По индукции заключаем, что если $|u_0| < M$, то и $|u_n| < M$ при всех n . В силу равномерной непрерывности k найдется такое δ , что

$$|k(x', y) - k(x'', y)| < \varepsilon/M \text{ при } |x' - x''| < \delta.$$

Поэтому при всех n $|u_n(x') - u_n(x'')| < \varepsilon$. ►

Условие строгой положительности в приводимой ниже теореме излишне ограничительно. Его основное назначение состоит в том, чтобы устранить осложнения, связанные с разложимостью и периодичностью цепей (с ними мы сталкивались в I; гл. XV).

Теорема 1. Каждое строго положительное регулярное ядро K на ограниченном замкнутом интервале Ω является эргодическим.

Теорема неверна для неограниченных Ω , так как в этом слу-

¹⁾ См. сноску на стр. 310. Наша «регулярность» аналогична «полной непрерывности» в теории операторов в гильбертовом пространстве.

чае предел в (7.2) может быть тождественно равен нулю. Общий результат может быть сформулирован в терминах стационарных мер. Напомним, что мера α называется *стационарной* относительно K , если $\alpha_1 = \alpha_2 = \dots = \alpha$, т. е. если все ее преобразования (7.1) одинаковы.

Теорема 2. *Строго положительное регулярное ядро K будет эргодическим тогда и только тогда, когда оно обладает строго положительным стационарным распределением вероятностей α .*

Доказательство теоремы 1. Пусть v_0 — непрерывная функция и v_1 — ее преобразование (7.3). Доказательство использует тот очевидный факт, что для строго положительного ядра K максимум v_1 строго меньше максимума v_0 (исключая только случай постоянной v_0).

Рассмотрим теперь последовательность преобразований u_n какой-либо непрерывной функции u_0 . Так как Ω замкнуто, то функция u_0 равномерно непрерывна на Ω , и, следовательно, последовательность $\{u_n\}$ равностепенно непрерывна. Таким образом, по теореме 3 § 6 существует подпоследовательность $\{u_{n_k}\}$, равномерно сходящаяся к непрерывной функции v_0 . Тогда u_{n_k+1} сходится к преобразованию v_1 функции v_0 . Числовая последовательность максимумов m_n функций u_n монотонна, и потому $m_n \rightarrow m$. Так как сходимость к v_0 и v_1 равномерна, то и v_0 , и v_1 имеют один и тот же максимум m . Следовательно, $v_0(x) = m$ при всех x . Так как этот предел не зависит от выбора подпоследовательности $\{u_{n_k}\}$, мы приходим к заключению, что $u_n \rightarrow m$ равномерно.

Пусть теперь γ_0 — произвольное распределение вероятностей на Ω . Обозначим через E_n математическое ожидание, взятое по отношению к γ_n [преобразование γ_n определяется по (7.1)]. Сравнение (7.1) и (7.3) показывает, что

$$E_n(u_0) = E_0(u_n) \rightarrow E_0(m) = m.$$

Из сходимости $E_n(u_0)$ для любой непрерывной функции u_0 вытекает существование вероятностной меры α , такой, что $\gamma_n \rightarrow \alpha$ (см. теорему 2 § 1; сходимость будет собственной, так как все распределения γ_n сосредоточены на конечном интервале). Из (7.1) следует, что α — стационарное распределение относительно K . Строгая положительность α непосредственно вытекает из строгой положительности K . ►

Доказательство теоремы 2. Обозначим через E математическое ожидание, взятое по отношению к данному стационарному распределению вероятностей α . Пусть u_0 — произвольная функция из $C[-\infty, \infty]$ и u_k — ее преобразование. В силу стационарности имеем $E(u_0) = E(u_1) = \dots$

Далее, $E(|u_k|)$ убывает с ростом k , так что $\lim E(|u_k|) = m$ существует.

Так же как в предыдущем доказательстве, выберем такую подпоследовательность, что $u_{n_k} \rightarrow v_0$. Тогда $u_{n_k+1} \rightarrow v_1$, где v_1 — преобразование v_0 . По теореме об ограниченной сходимости отсюда следует, что $E(u_{n_k}) \rightarrow E(v_0)$ и $E(|u_{n_k}|) \rightarrow E(|v_0|)$. Таким образом,

$$E(v_1) = E(v_0) = E(u_0) \quad \text{и} \quad E(|v_1|) = E(|v_0|) = m.$$

Из свойства строгой положительности K и последнего неравенства вытекает, что непрерывная функция v_0 не меняет знака. Когда $E(u_0) = 0$, то $v_0(x) = 0$ тождественно. Из сказанного следует, что при любой начальной функции u_0 мы имеем $v_0(x) = E(u_0)$ при всех x . Этим доказано, что $u_n(x) \rightarrow E(u_0)$, а последнее соотношение равносильно сходимости $K^{(n)}(x, \Gamma) \rightarrow \alpha(\Gamma)$ на всех интервалах непрерывности. ►

Применим теперь изложенную теорию к *сверткам на окружности* длины 1, т. е. к преобразованиям вида

$$u_{n+1}(x) = \int_0^1 u_n(x-y) F\{dy\}, \quad (7.5)$$

где F — распределение вероятностей на окружности и сложение производится по mod 1 [см. гл. II, 8, и пример (6.а)]. Это преобразование можно записать в форме (7.3) с $\Omega = \overline{0, 1}$ и $K^{(n)}(x, \Gamma) = F^{n*}\{x - \Gamma\}$. Если F строго положительно, то применима теорема 1. Однако мы докажем следующий более общий аналог центральной предельной теоремы.

Теорема 3¹⁾. Пусть F — распределение вероятностей на окружности. Допустим, что оно не сосредоточено в вершинах правильного многоугольника. Тогда F^{n*} сходится к распределению с постоянной плотностью.

Доказательство. Достаточно показать, что для произвольной непрерывной функции u_0 преобразования u_n сходятся к постоянной m (зависящей от u_0). В самом деле, как показывает вторая часть доказательства теоремы 1, из этого вытекает, что F^{n*} сходятся к некоторому распределению вероятностей α на окружности. Так как $\alpha \star u_0$ равно постоянной для любой непрерывной функции u_0 , то отсюда следует, что α совпадает с равномерным распределением.

Чтобы показать, что $u_n \rightarrow m$, мы используем первую часть доказательства теоремы 1. Мы только заново должны доказать,

¹⁾ Аналог этой теоремы для числовой прямой указан в задачах 23 и 24. Обобщение на случай различных распределений см. Lévy P., Bull. Soc. Math. France, 67 (1939), 1—41; Dvoretzky A. and Wolfowitz J., Duke Math. J., 18 (1951), 501—507.

что максимум преобразования v_1 любой непрерывной функции v_0 строго меньше, чем максимум v_0 (за исключением случая, когда v_0 равно константе). Поэтому для доказательства теоремы достаточно установить справедливость следующего утверждения. *Если v_0 — непрерывная функция, такая, что $v_0 \leq m$ и $v_0(x) < m$ для всех x в интервале I длины $\lambda > 0$, то существует такое r , что $v_r(x) < m$ при всех x .*

Так как вращения не изменяют величины максимума, то мы можем без ограничения общности предположить, что 0 есть точка роста F . Если b — другая точка роста, то $0, b, 2b, \dots, rb$ суть точки роста F^{r*} . Поэтому можно выбрать b и r таким образом, что каждый интервал длины λ содержит по крайней мере одну из этих точек (см. лемму 1 и следствие в гл. V, 4а). По определению

$$v_r(x) = \int_0^1 v_0(x-y) F^{r*} \{dy\}. \quad (7.6)$$

Каждой точке x возможно поставить в соответствие такую точку роста функции F^{r*} (скажем, y), что $x-y$ входит в I . Тогда $v_0(x-y) < m$ и потому $v_r(x) < m$. Так как x произвольно, то требуемое утверждение доказано. ►

Замечание. Доказательство легко видоизменить так, чтобы установить следующее предложение: *если F сконцентрировано в вершинах правильного многоугольника, одна из вершин которого расположена в точке 0 , то F^{n*} сходится к атомическому распределению, атомам которого приписан один и тот же вес. Сходимость не обязана иметь места, если 0 не находится в числе атомов.*

Пример. в) Пусть F сосредоточено в двух иррациональных точках a и $a + \frac{1}{2}$. Тогда F^{n*} сосредоточено в точках na и $na + \frac{1}{2}$ и сходимость невозможна. ►

§ 8. ПРАВИЛЬНО МЕНЯЮЩИЕСЯ ФУНКЦИИ

Понятие правильного изменения функций (введенное Карамата в 1930 г.) оказалось во многих отношениях плодотворным и находит все возрастающее число применений в теории вероятностей. Причины этого отчасти объяснены в приводимой ниже лемме, которая, несмотря на свою простоту, играет здесь основную роль. Примеры этого параграфа демонстрируют интересные вероятностные результаты, а задачи 29, 30 содержат известный результат, касающийся устойчивых распределений. Он элементарным путем выводится из леммы. (См. также задачу 31).

Мы часто сталкиваемся с монотонными функциями U , получающимися из некоторого распределения вероятностей F интегрированием функции $y^x F\{dy\}$ на $0, x$ или x, ∞ [см., например, (4.5), (4.15), (4.16)]. Обычной заменой параметров можно перейти

от функции U к семейству функций вида $a_t U(tx)$. Мы должны изучить их поведение при $t \rightarrow \infty$. Если предел $\psi(x)$ существует, то достаточно брать нормирующие множители a_t вида $a_t = \psi(1)/U(t)$ (при условии $\psi(1) > 0$). Следующая лемма является поэтому более общей, чем это может показаться с первого взгляда. Она показывает, что класс возможных пределов удивительно мал.

Лемма 1. Пусть U — положительная монотонная функция на $0, \infty$, такая, что при $t \rightarrow \infty$

$$\frac{U(tx)}{U(t)} \rightarrow \psi(x) \leq \infty \quad (8.1)$$

на всюду плотном множестве A точек x . Тогда

$$\psi(x) = x^p, \quad (8.2)$$

где $-\infty \leq p \leq \infty$.

«Бессмысленный» символ x^∞ вводится только для того, чтобы избежать исключений. Его следует интерпретировать, конечно, как ∞ при $x > 1$ и как 0 при $x < 1$. Аналогично $x^{-\infty}$ равно ∞ или 0 в зависимости от того, будет $x < 1$ или $x > 1$ (см. задачу 25).

Доказательство. Тождество

$$\frac{U(tx_1 x_2)}{U(t)} = \frac{U(tx_1 x_2)}{U(tx_2)} \cdot \frac{U(tx_2)}{U(t)} \quad (8.3)$$

показывает, что если конечный положительный предел в (8.1) существует при $x = x_1$ и при $x = x_2$, то он существует и при $x = x_1 x_2$ и

$$\psi(x_1 x_2) = \psi(x_1) \psi(x_2). \quad (8.4)$$

Допустим сначала, что для некоторой точки x_1 $\psi(x_1) = \infty$. Тогда по индукции заключаем, что $\psi(x_1^n) = \infty$ и $\psi(x_1^{-n}) = 0$ при всех n . Так как ψ монотонна, то отсюда следует, что или $\psi(x) = x^\infty$, или $\psi(x) = x^{-\infty}$. Остается доказать лемму для случая конечных ψ (см. задачу 25). В силу монотонности ψ мы можем доопределить ее всюду по непрерывности справа. Тогда (8.4) выполняется при всех x_1 и x_2 . Но это уравнение лишь обозначениями отличается от многократно использовавшегося нами характеристического уравнения для показательной функции. В самом деле, полагая $x = e^\xi$ и $\psi(e^\xi) = u(\xi)$, мы преобразуем (8.4) к виду $u(\xi_1 + \xi_2) = u(\xi_1) u(\xi_2)$. Из 1; гл. XVII, 6, мы знаем, что все решения этого уравнения, ограниченные на каждом конечном интервале, имеют вид $u(\xi) = e^{p\xi}$. Это, однако, то же самое, что $\psi(x) = x^p$. ►

Функция U , удовлетворяющая условиям леммы 1 с конечными p , будет называться правильно меняющейся на бесконечности. Это определение может быть распространено на немонотонные функции. Положим

$$U(x) = x^p L(x). \quad (8.5)$$

Отношение $U(tx)/U(t)$ будет стремиться к x^p тогда и только тогда, когда

$$\frac{L(tx)}{L(t)} \rightarrow 1, \quad t \rightarrow \infty, \quad (8.6)$$

для любого $x > 0$. Функции с этим свойством называются *медленно меняющимися*. Таким образом, преобразование (8.5) сводит понятие правильного изменения функции к понятию медленного изменения. Удобно использовать это обстоятельство для формального определения правильно меняющейся функции.

Определение. Положительная (но не обязательно монотонная) функция L , определенная на $\overline{0, \infty}$, называется *медленно меняющейся на бесконечности* в том и только том случае, когда она удовлетворяет условию (8.6).

Функция U называется *правильно меняющейся с показателем* p ($-\infty < p < \infty$) тогда и только тогда, когда она имеет вид (8.5) при том, что L является медленно меняющейся функцией.

Это определение используется для введения понятия *правильного изменения в нуле*: U называется *правильно меняющейся в нуле* в том и только том случае, когда $U(x^{-1})$ правильно меняется на бесконечности. Таким образом, новой теории в этом случае не возникает.

Свойство функции быть правильно меняющейся зависит только от поведения на бесконечности, и поэтому не обязательно, чтобы функция $L(x)$ была положительной (или даже вообще определенной) для всех $x > 0$.

Пример. а) Все степени $|\log x|$ медленно меняются и в нуле, и на бесконечности. Аналогично функция, которая стремится к *положительному* пределу, является медленно меняющейся.

Функция $(1+x^2)^p$ правильно меняется на бесконечности с показателем $2p$. Функция e^x не является правильно меняющейся на бесконечности, но она удовлетворяет условиям леммы 1 с показателем $p = \infty$. Наконец, функция $2 + \sin x$ не удовлетворяет (8.1). ►

Для облегчения ссылок мы перефразируем лемму 1, придав ей форму теоремы.

Теорема. Монотонная функция U правильно меняется на бесконечности тогда и только тогда, когда (8.1) выполняется на всюду плотном множестве точек и предел ψ конечен и положителен в некотором интервале¹⁾.

¹⁾ Понятие правильно меняющихся функций можно обобщить следующим образом. Предложение о существовании предела в (8.1) мы заменяем требованием, что каждая последовательность $\{t_n\}$, стремящаяся к бесконечности, содержит подпоследовательность, такую, что $U(t_{n_k}x)/U(t_{n_k})$ стремится к конечному положительному пределу. Мы говорим при этом, что функция U *изменяется «доминируемо»*. См. конец § 10.

Эта теорема переносится на немонотонные функции, но лишь при том предположении, что сходимость имеет место во *всех* точках.

Следующая лемма служит для развития представлений о правильном изменении функций. Эта лемма оказывается прямым следствием общего вида (9.9) медленно меняющихся функций.

Лемма 2. Если L медленно меняется на бесконечности, то

$$x^{-\varepsilon} < L(x) < x^{\varepsilon} \quad (8.7)$$

при любом фиксированном $\varepsilon > 0$ и всех достаточно больших x .

Предельный переход в (8.6) равномерен на конечных интервалах $0 < a < x < b$.

Мы заключаем это обозрение следующим часто используемым критерием.

Лемма 3. Пусть

$$\frac{\lambda_{n+1}}{\lambda_n} \rightarrow 1, \quad a_n \rightarrow \infty.$$

Если U — монотонная функция, такая, что предел

$$\lim \lambda_n U(a_n x) = \chi(x) \leq \infty \quad (8.8)$$

существует на всюду плотном множестве и функция χ конечна и положительна в некотором интервале, то U правильно меняется и $\chi(x) = cx^{\rho}$, где $-\infty < \rho < \infty$.

Доказательство. Мы можем предположить, что $\chi(1) = 1$ и что (8.8) выполняется при $x = 1$ (этого можно добиться тривиальным изменением масштаба). Для каждого данного t определим n как наименьшее целое число, для которого $a_{n+1} > t$. Тогда $a_n \leq t < a_{n+1}$ и для неубывающей функции U

$$\frac{U(a_n x)}{U(a_{n+1})} \leq \frac{U(tx)}{U(t)} \leq \frac{U(a_{n+1}x)}{U(a_n)}. \quad (8.9)$$

Для невозрастающей функции U верны обратные неравенства. Так как $\lambda_n U(a_n) \rightarrow 1$, то крайние члены стремятся к $\chi(x)$ в каждой точке, где выполняется (8.8). Поэтому утверждение леммы вытекает из последней теоремы. ►

В качестве типичных примеров применений мы дадим известную теорему Р. А. Фишера—Б. В. Гнеденко и один новый результат.

Примеры. б) *Распределение максимума.* Пусть случайные величины X_k взаимно независимы и имеют одно и то же распределение F . Положим

$$X_n^* = \max[X_1, \dots, X_n].$$

Выясним, существуют ли нормирующие множители a_n , такие,

что случайные величины X_n^*/a_n имеют предельное распределение G . Мы исключим при этом два тривиальных случая. Если F имеет наибольшую точку роста ξ , то распределение X_n^* тривиальным образом сходится к распределению, сосредоточенному в ξ . С другой стороны, всегда можно выбрать нормирующие множители a_n столь быстро возрастающими, что величины X_n^*/a_n сходятся по вероятности к нулю. Остальные случаи охватываются следующим утверждением. ►

Пусть $F(x) < 1$ при всех x . Для того, чтобы при некотором выборе нормирующих постоянных a_n распределения G_n случайных величин X_n^*/a_n сходились к распределению G , не сосредоточенному в 0, необходимо и достаточно, чтобы функция $1-F$ была правильно меняющейся с показателем $\rho < 0$. В этом случае

$$G(x) = e^{-cx^\rho} \quad (8.10)$$

при $x > 0$ и $G(x) = 0$ при $x < 0$ (очевидно, $c > 0$).

Доказательство. Если предельное распределение G существует, то

$$F^n(a_n x) \rightarrow G(x) \quad (8.11)$$

во всех точках непрерывности. Переходя к логарифмам и вспоминая, что $\log(1-z) \sim -z$ при $z \rightarrow 0$, получаем

$$n[1 - F(a_n x)] \rightarrow -\log G(x). \quad (8.12)$$

Так как в некотором интервале $0 < G(x) < 1$, то последняя лемма гарантирует правильное изменение $1-F$. Обратно, если $1-F$ меняется правильно, то возможно определить постоянные a_n так, что $n[1 - F(a_n)] \rightarrow 1$. В этом случае левая часть (8.12) стремится к x^ρ (см. задачу 26). ►

Пример. в) *Свертки.* Из определения (8.6) очевидным образом следует, что сумма двух медленно меняющихся функций есть снова функция, медленно меняющаяся. Мы докажем следующее

Утверждение. Пусть F_1 и F_2 — две функции распределения, такие, что

$$1 - F_i(x) = \frac{1}{x^\rho} L_i(x), \quad (8.13)$$

где L_i — медленно меняющиеся функции. Тогда свертка $G = F_1 \star F_2$ имеет правильно меняющийся «хвост», так что

$$1 - G(x) \sim \frac{1}{x^\rho} (L_1(x) + L_2(x)). \quad (8.14)$$

Доказательство. Пусть X_1 и X_2 — независимые случайные величины с распределениями F_1 и F_2 . Пусть $t' = (1 + \delta)t > t$. Событие $X_1 + X_2 > t$ осуществляется всякий раз, когда одна из ве-

личин X_i больше t' , а другая больше $-\delta t$. При $t \rightarrow \infty$ вероятность последнего события стремится к 1 и, следовательно, при любом $\varepsilon > 0$ и достаточно большом t

$$1 - G(t) \geq [(1 - F_1(t')) + (1 - F_2(t'))](1 - \varepsilon). \quad (8.15)$$

С другой стороны, если положить $t'' = (1 - \delta)t$ при $0 < \delta < \frac{1}{2}$, то из события $X_1 + X_2 > t$ вытекает, что или хотя бы одна из величин X_i превышает t'' , или обе больше δt . Ввиду соотношения (8.13) становится ясно, что вероятность последнего события (асимптотически) пренебрежимо мала по сравнению с вероятностью того, что $X_i > t''$. Отсюда следует, что для достаточно больших t

$$1 - G(t) \leq [(1 - F_1(t'')) + (1 - F_2(t''))](1 + \varepsilon). \quad (8.16)$$

Поскольку δ и ε могут быть выбраны произвольно малыми, то неравенства (8.15) и (8.16) вместе дают утверждение (8.14). ►

Индукцией по r получаем интересное

Следствие. Если $1 - F(x) \sim x^{-\rho} L(x)$, то $1 - F^{r*}(x) \sim r x^{-\rho} L(x)$.

Это утверждение¹⁾ в тех случаях, где оно применимо, дополняет центральную предельную теорему, давая информацию о «хвостах» распределений. (О применениях к устойчивым распределениям см. задачи 29 и 30. Аналогичную теорему для обобщенного распределения Пуассона см. в задаче 31.)

§ 9*). АСИМПТОТИЧЕСКИЕ СВОЙСТВА ПРАВИЛЬНО МЕНЯЮЩИХСЯ ФУНКЦИЙ

Цель этого параграфа — изучить связь между «хвостами» и урезанными моментами распределений с правильно меняющимися «хвостами». Основной результат состоит в том, что если $1 - F(x)$ и $F(-x)$ меняются правильно, то тем же свойством обладают урезанные моменты. Это утверждение содержится в теореме 2, которая дает большие сведения, чем то, что нам нужно для теории устойчивых распределений. Эту теорему можно доказать непосредственно, а можно рассматривать как следствие теоремы 1, которая включает в себя замечательную характеризацию правильно меняющихся функций, данную Карамата²⁾.

¹⁾ Частные случаи были отмечены С. Портом.

²⁾ Материал этого параграфа используется только в теории устойчивых распределений. Однако применение теоремы 2 устранило бы многочисленные длинные вычисления, встречающиеся в литературе.

³⁾ Karamata J., Sur un mode de croissance régulière, *Mathematica (Cluj)*, 4 (1930), 38—53. Несмотря на частые ссылки на эту работу, более нового изложения ее результатов, по-видимому, не существует. Последние обобщения и применения к тауберовым теоремам см. в статье W. Feller, One-sided analogues of Karamata's regular variation, в мемориальном томе *L'Enseignement Mathématique*. (1968), посвященном Карамата.

Лучше всего, как нам кажется, дать полное изложение теории, в частности, потому, что тогда многие рассуждения могут быть значительно упрощены¹⁾.

Введем следующие сокращения:

$$Z_p(x) = \int_0^x y^p Z(y) dy, \quad Z_p^*(x) = \int_x^\infty y^p Z(y) dy. \quad (9.1)$$

Мы покажем теперь, что в случае правильно меняющейся функции Z эти функции асимптотически ведут себя по отношению к Z так же, как в простом случае $Z(x) = x^a$.

Асимптотическое поведение Z_p на бесконечности не зависит от поведения Z вблизи нуля. Поэтому мы можем, не ограничивая общности, допустить, что Z обращается в нуль всюду в некоторой окрестности нуля, так что интеграл, определяющий Z_p , имеет смысл при всех p .

Лемма. Пусть $Z > 0$ — медленно меняющаяся функция. Тогда интегралы в (9.1) сходятся на бесконечности при $p < -1$ и расходятся при $p > -1$.

Если $p \geq -1$, то Z_p правильно меняется с показателем $p+1$. Если $p < -1$, то Z_p^* правильно меняется с показателем $p+1$. Последнее верно и для $p+1=0$, если только Z_{-1}^* существует.

Доказательство. При заданных положительных x и ε выберем η таким, что при $y \geq \eta$

$$(1-\varepsilon)Z(y) \leq Z(xy) \leq (1+\varepsilon)Z(y). \quad (9.2)$$

Предположим, что интегралы в (9.1) сходятся. Из

$$Z_p^*(tx) = x^{p+1} \int_t^\infty y^p Z(xy) dy \quad (9.3)$$

следует при $t > \eta$, что

$$(1-\varepsilon)x^{p+1}Z_p^*(t) \leq Z_p^*(tx) \leq (1+\varepsilon)x^{p+1}Z_p^*(t).$$

Поскольку ε произвольно, то мы делаем вывод, что при $t \rightarrow \infty$

$$\frac{Z_p^*(tx)}{Z_p^*(t)} \rightarrow x^{p+1}. \quad (9.4)$$

Это доказывает правильное изменение функции Z_p^* . Более того, так как Z_p^* — убывающая функция, то $p+1 \leq 0$. Таким образом, интегралы в (9.1) расходятся при невыполнении условия $p \leq -1$.

Предположим теперь, что указанные интегралы расходятся.

¹⁾ Наше доказательство теоремы 1, будучи новым, использует тем не менее идеи Караматы.

Тогда при $t > \eta$

$$Z_p(tx) = Z_p(\eta x) + x^{p+1} \int_{\eta}^t y^p Z(x, y) dy$$

и, следовательно,

$$(1 - \varepsilon) x^{p+1} Z_p(t) \leq Z_p(tx) - Z_p(\eta x) \leq (1 + \varepsilon) x^{p+1} Z_p(t).$$

Разделив на $Z_p(t)$ и устремив $t \rightarrow \infty$, мы получаем, как и выше, что $Z_p(tx)/Z_p(t)$ стремится к x^{p+1} . Поэтому в случае расходимости интегралов Z_p правильно изменяется, а расходимость возможна лишь тогда, когда $p \geq -1$. ►

Следующая теорема показывает, что из факта правильного изменения Z вытекает правильное изменение Z_p и Z_p^* ; обратное также верно, исключая тот случай, когда Z_p или Z_p^* — медленно меняющиеся функции. Кроме того, мы получаем полезный критерий правильного изменения этих функций. Части а) и б) теоремы посвящены функциям Z_p^* и Z_p соответственно. Они существуют во всех отношениях в параллель друг другу, но лишь часть а) широко используется в теории вероятностей.

Теорема 1. а) Если Z правильно меняется с показателем γ и Z_p^* существует, то

$$\frac{t^{p+1} Z(t)}{Z_p^*(t)} \rightarrow \lambda, \quad (9.5)$$

где $\lambda = -(p + \gamma + 1) \geq 0$.

Обратно, если (9.5) выполняется с $\lambda > 0$, то Z и Z_p^* изменяются правильно с показателями $\gamma = -\lambda - p - 1$ и $-\lambda$ соответственно. Если (9.5) выполняется с $\lambda = 0$, то функция Z_p^* меняется медленно (а о Z ничего нельзя сказать).

б) Если Z меняется правильно с показателем γ и если $p \geq -\gamma - 1$, то

$$\frac{t^{p+1} Z(t)}{Z_p(t)} \rightarrow \lambda \quad (9.6)$$

с $\lambda = p + \gamma + 1$.

Обратно, если (9.6) выполняется с $\lambda > 0$, то Z и Z_p меняются правильно с показателями $\lambda - p - 1$ и λ соответственно. Если (9.6) выполняется с $\lambda = 0$, то Z_p меняется медленно.

Доказательство. Доказательство для двух частей одно и то же, и мы проведем его для части а). Положим

$$\frac{y^p Z(y)}{Z_p^*(y)} = \frac{\eta(y)}{y}. \quad (9.7)$$

Числитель в левой части равен производной знаменателя, взятой

со знаком минус. Поэтому при $x > 1$ мы получаем

$$\log \frac{Z_p^*(t)}{Z_p^*(tx)} = \int_t^{tx} \eta(y) \frac{dy}{y} = \eta(t) \int_1^x \frac{\eta(ts)}{\eta(t)} \frac{ds}{s}. \quad (9.8)$$

Предположим теперь, что Z правильно меняется с показателем γ . По предыдущей лемме Z_p^* меняется правильно с показателем $\lambda = \gamma + p + 1$, и поэтому обе части (9.7) правильно меняются с показателем -1 . Таким образом, η есть медленно меняющаяся функция. Поэтому при $t \rightarrow \infty$ подынтегральное выражение последнего интеграла в (9.8) стремится к $1/s$. К сожалению, нам неизвестно, ограничена ли функция η , и поэтому мы можем лишь утверждать, что нижний предел интеграла $\geq \log x$ в силу леммы Фату [см. (2.9) в гл. IV]. Вследствие правильного изменения Z_p^* левая часть (9.8) стремится к $\lambda \log x$ и поэтому

$$\limsup \eta(t) \leq \lambda.$$

Но отсюда вытекает ограниченность η , и, следовательно, мы можем выбрать последовательность $t_n \rightarrow \infty$ так, что $\eta(t_n) \rightarrow c < \infty$. Отсюда вследствие того, что η медленно меняется, вытекает $\eta(t_n s) \rightarrow c$ для всех s , и эта сходимость является ограниченной. Таким образом, правая часть в (9.8) приближается к $c \log x$ и, следовательно, $c = \lambda$. Отсюда следует, что предел c не зависит от последовательности $\{t_n\}$ и поэтому $\eta(t) \rightarrow \lambda$. Это доказывает, что (9.5) справедливо.

Обратное доказывается легче. Предположим, что $\eta(t) \rightarrow \lambda \geq 0$. Тогда обе части в (9.8) стремятся к $\lambda \log x$ и, следовательно, отношение $Z_p^*(t)/Z_p^*(tx)$ сходится к x^λ , как и утверждалось. Отсюда и из (9.5) при $\lambda > 0$ следует, что Z правильно меняется с показателем $-\lambda - p - 1$. ▶

Отметим (хотя мы и не будем его использовать) интересное

Следствие. *Функция Z является медленно меняющейся тогда и только тогда, когда она может быть представлена в виде*

$$Z(x) = a(x) \exp \left(\int_1^x \frac{\varepsilon(y)}{y} dy \right), \quad (9.9)$$

где $\varepsilon(x) \rightarrow 0$ и $a(x) \rightarrow c < \infty$ при $x \rightarrow \infty$.

Доказательство. Легко проверить, что правая часть (9.9) представляет собой медленно меняющуюся функцию. Обратно, предположим, что Z меняется медленно. Используя (9.6) при $p = \gamma = 0$, получаем

$$\frac{Z(t)}{Z_0(t)} = \frac{1 + \varepsilon(t)}{t},$$

где $\varepsilon(t) \rightarrow 0$. Числитель в левой части является производной

знаменателя, и, производя интегрирование, мы получаем

$$Z_0(x) = Z_0(1) \cdot x \cdot \exp\left(\int_1^x \frac{e(t)}{t} dt\right),$$

что равносильно (9.9), поскольку в силу (9.6) $Z(x) \sim Z_0(x) x^{-1}$. ►

Займемся применением теоремы 1 к урезанным моментам вероятностного распределения F . Мы можем рассмотреть каждый хвост в отдельности или же объединить их, перейдя от F к $F(x) - F(-x)$. Поэтому достаточно изучить распределения F , сосредоточенные на $\overline{0, \infty}$. Для таких распределений F определим урезанные моменты U_ξ и V_η равенствами

$$U_\xi(x) = \int_0^x y^\xi F\{dy\}, \quad V_\eta(x) = \int_x^\infty y^\eta F\{dy\}. \quad (9.10)$$

Должно быть понятно, что когда второй интеграл сходится, то первый интеграл стремится к ∞ при $x \rightarrow \infty$. Потребуем, чтобы $\xi > 0$ и $-\infty < \eta < \xi$. В частности, $V_0 = 1 - F$ есть хвост распределения F .

Мы докажем обобщение части (а) теоремы 1; часть (б) обобщается подобным же образом.

Теорема 2¹). *Предположим, что $U_\xi(\infty) = \infty$.*

(i) Пусть или U_ξ , или V_η правильно меняются. Тогда существует предел

$$\lim_{t \rightarrow \infty} \frac{t^{\xi-\eta} V_\eta(t)}{U_\xi(t)} = c, \quad 0 \leq c \leq \infty. \quad (9.11)$$

Этот предел представляется единственным образом в форме

$$c = \frac{\xi - \alpha}{\alpha - \eta}, \quad \eta \leq \alpha \leq \xi, \quad (9.12)$$

где $\alpha = \eta$, если $c = \infty$.

(ii) Обратно, если (9.11) выполняется с $0 < c < \infty$, то автоматически $\alpha \geq 0$ и существует медленно меняющаяся функция L , такая, что

$$U_\xi(x) \sim (\alpha - \eta) x^{\xi-\alpha} L(x), \quad V_\eta(x) \sim (\xi - \alpha) x^{\eta-\alpha} L(x), \quad (9.13)$$

где символ $\sim$ указывает на то, что отношение двух частей стремится к 1.

(iii) Утверждение остается справедливым, когда $c = 0$ или $c = \infty$, при условии, что символ $\sim$ интерпретируется очевидным образом.

¹) Обобщение см. в задачах 34 и 35.

Например, если (9.11) выполняется с $c=0$, то $\alpha=\zeta$ и функция U_ζ медленно меняется, но относительно V_η мы знаем только то, что $V_\eta(x)=o(x^{\eta-\zeta}L(x))$. В этом случае V_η не обязана быть правильно меняющейся функцией (см. задачу 31). Однако медленное изменение—это как раз тот единственный случай, при котором правильное изменение одной из функций U_ζ или V_η не влечет за собой правильного изменения другой.

Доказательство. (i) Запишем V_η в виде

$$V_\eta(x) = \int_x^\infty y^{\eta-\zeta} U_\zeta(dy). \quad (9.14)$$

Интегрируя по частям в пределах от x до $t > x$, получаем

$$V_\eta(x) - V_\eta(t) = -x^{\eta-\zeta} U_\zeta(x) + t^{\eta-\zeta} U_\zeta(t) + (\zeta - \eta) \int_x^t y^{\eta-\zeta-1} U_\zeta(y) dy.$$

Два последних члена в правой части равенства положительны, и поэтому интеграл должен сходиться при $t \rightarrow \infty$. Отсюда вследствие монотонности U_ζ вытекает, что $y^{\eta-\zeta} U_\zeta(t) \rightarrow 0$ и, следовательно,

$$V_\eta(x) = -x^{\eta-\zeta} U_\zeta(x) + (\zeta - \eta) \int_x^\infty y^{\eta-\zeta-1} U_\zeta(y) dy \quad (9.15)$$

или

$$\frac{x^{\zeta-\eta} V_\eta(x)}{U_\zeta(x)} = -1 + \frac{\zeta - \eta}{x^{\eta-\zeta} U_\zeta(x)} \int_x^\infty y^{\eta-\zeta-1} U_\zeta(y) dy. \quad (9.16)$$

Предположим теперь, что U_ζ правильно меняется. Так как $U_\zeta(\infty) = \infty$, то ее показатель необходимо $\leq \zeta$, и мы обозначим его $\zeta - \alpha$. (Поскольку интеграл в (9.16) сходится, то необходимо $\alpha \geq \eta$.) Из соотношения (9.5) с $Z = U_\zeta$ и $\rho = \eta - \zeta - 1$ следует, что правая часть в (9.16) стремится к

$$-1 + (\zeta - \eta)/(\alpha - \eta) = (\zeta - \alpha)/(\alpha - \eta),$$

если $\lambda \neq 0$, и стремится к ∞ , если $\lambda = 0$. Таким образом, мы показали, что если U_ζ правильно меняется с показателем $\zeta - \alpha$, то (9.11) выполняется с $c \geq 0$, которое задается равенством (9.12).

Предположим теперь, что V_η правильно меняется. Ее показатель не превосходит η , и мы обозначим его $\eta - \alpha$. Используем то же самое рассуждение, заменив только (9.15) аналогичным соотношением

$$U_\zeta(x) = -x^{\zeta-\eta} V_\eta(x) + (\zeta - \eta) \int_0^x y^{\zeta-\eta-1} V_\eta(y) dy. \quad (9.17)$$

Используя (9.6) с $Z=V_\eta$ и $p=\xi-\eta-1$, получаем, что (9.11) выполняется с c , которое задано равенством (9.12), где $\alpha \geq 0$.

(ii) Для доказательства обратного утверждения предположим справедливость (9.11) и представим c в виде (9.12). Предположим сначала, что $0 < c < \infty$. Тогда из (9.16) мы видим, что

$$\frac{x^{\eta-\xi} U_\xi(x)}{\int_x^\infty y^{\eta-\xi-1} U_\xi(y) dy} \rightarrow \frac{\xi-\eta}{c+1} = \alpha - \eta. \quad (9.18)$$

Из теоремы 1, а) прямо вытекает, что U_ξ правильно меняется с показателем $\xi - \alpha > 0$, и тогда из (9.11) следует, что V_η правильно меняется с показателем $\eta - \alpha$. Следовательно, U_ξ и V_η могут быть представлены в форме (9.13), где $\alpha \geq 0$.

Если $c=0$, то те же доводы показывают, что U_ξ медленно меняется, но соотношение (9.11) не дает оснований для вывода о том, что V_η меняется правильно.

Наконец, если (9.11) выполняется с $c=\infty$, то мы выводим из (9.18), что

$$\frac{x^{\xi-\eta} V_\eta(x)}{\int_0^x y^{\xi-\eta-1} V_\eta(y) dy} \rightarrow 0, \quad (9.19)$$

и по теореме 1, б) отсюда следует, что V_η медленно меняется. ►

§ 10. ЗАДАЧИ

1. *Альтернативное определение сходимости.* Пусть F_n и F — некоторые распределения вероятностей. Покажите, что $F_n \rightarrow F$ (в собственном смысле) тогда и только тогда, когда для любых данных $\varepsilon > 0$, $h > 0$ и t существует $N(\varepsilon, h, t)$, такое, что при $n > N(\varepsilon, h, t)$

$$F(t-h) - \varepsilon < F_n(t) < F(t+h) + \varepsilon. \quad (10.1)$$

2. *Несобственная сходимость.* Если F — несобственное распределение, то из (10.1) вытекает, что $F_n \rightarrow F$ в несобственном смысле. Обратное *неверно*. Покажите, что *собственная* сходимость может быть *определена* требованием, что (10.1) выполняется для $n \geq N(\varepsilon, h)$ независимо от t .

3. Пусть $\{F_n\}$ сходится в собственном смысле к распределению, не сосредоточенному в одной точке. Тогда последовательность $\{F_n(a_n x + b_n)\}$ сходится к распределению, сосредоточенному в нуле в том и только том случае, когда $a_n \rightarrow \infty$ и $b_n = o(a_n)$.

4. Пусть $X_1, X_2, \dots$ — независимые случайные величины с одинаковым распределением F , и пусть $S_n = X_1 + \dots + X_n$. Предположим, что предельное распределение величин $a_n^{-1} S_n - b_n$ является собственным и не сосредоточено в одной точке. Если $a_n > 0$, то

$$a_n \rightarrow \infty, \quad a_n/a_{n-1} \rightarrow 1.$$

Указание. Используя теорему 2 из § 3, покажите, что a_{2n}/a_n стремится к конечному пределу. Достаточно рассмотреть симметричные распределения. (Предельное распределение устойчиво, см. гл. VI, 1.)

5. Пусть $\{u_n\}$ — последовательность ограниченных монотонных функций, сходящаяся поточечно к ограниченной и непрерывной предельной функции (которая автоматически монотонна). Докажите, что сходимость равномерна на всех конечных интервалах. *Указание.* Разбейте ось абсцисс на интервалы, на каждом из которых колебание u меньше ε .

6. Пусть F_n сосредоточено в точке n^{-1} и $u(x) = \sin(x^2)$. Тогда $F_n \star u \rightarrow u$ поточечно, но неравномерно.

7. а) Если совместное распределение (X_n, Y_n) сходится к совместному распределению (X, Y) , то распределение $X_n + Y_n$ сходится к распределению $X + Y$.

б) Покажите, что теорема 2 из § 3 — частный случай этого утверждения.

в) Сделанное утверждение, вообще говоря, неверно, если известна только сходимость частных распределений X_n и Y_n .

8. Пусть $F_n \rightarrow F$, где F — несобственное распределение. Если $u \in C_0(-\infty, \infty)$, то

$$F_n \star u \rightarrow F \star u$$

равномерно в каждом конечном интервале (это — обобщение теоремы 1 из § 3).

9. Если сходимость $F_n \rightarrow F$ несобственная, то отсюда не следует с необходимостью, что $F_n \star F_n \rightarrow F \star F$. *Пример.* Пусть F_n имеет атомы веса $1/3$ в точках $-n$, 0 и n .

10. На плоскости каждая непрерывная исчезающая на бесконечности функция может быть равномерно аппроксимирована конечными линейными комбинациями вида $\sum c_k \varphi_k(x) \psi_k(y)$, где φ_k и ψ_k — бесконечно дифференцируемые функции. *Указание.* Используйте пример 3, а), полагая $G_k(x, y) = \mathfrak{N}_k(x) \mathfrak{N}_k(y)$, где $\mathfrak{N}$ — нормальное распределение.

Метрики. Функция ρ называется *расстоянием* между распределениями вероятностей, если $\rho(F, G)$ определено для любой пары распределений F, G и обладает следующими тремя свойствами: $\rho(F, G) \geq 0$ и $\rho(F, G) = 0$ тогда и только тогда, когда $F = G$; далее, $\rho(F, G) = \rho(G, F)$; наконец, ρ удовлетворяет неравенству треугольника:

$$\rho(F_1, F_2) \leq \rho(F_1, G) + \rho(G, F_2).$$

11. *Метрика П. Леви.* Пусть F и G — два собственных распределения вероятностей. Определим $\rho(F, G)$ как точную нижнюю грань всех $h > 0$, для которых при всех x

$$F(x-h) - h \leq G(x) \leq F(x+h) + h. \quad (10.2)$$

Проверьте, что ρ является расстоянием. Докажите, что $F_n \rightarrow F$ в собственном смысле тогда и только тогда, когда $\rho(F_n, F) \rightarrow 0$.

12. *Расстояние «по саринации».* Положим $\rho(F, G) = \sup \|\mathfrak{N}u - \mathfrak{G}u\|$, где $u \in C_0$ и $\|u\| = 1$. Покажите, что ρ — расстояние¹⁾. Если F и G — атомические распределения, приписывающие точке a_k веса p_k и q_k соответственно, то

$$\rho(F, G) = \sum |p_k - q_k|. \quad (10.3)$$

Если F и G имеют плотности f и g , то

$$\rho(F, G) = \int_{-\infty}^{+\infty} |f(x) - g(x)| dx. \quad (10.4)$$

Указание. Достаточно доказать (10.4) для непрерывных f и g . Общий случай изучается с помощью надлежащей аппроксимации.

¹⁾ Это определение может быть распространено на разности любых конечных мер и определяет для мер «топологию, порождаемую нормой». Задача 13 показывает, что соответствующее понятие сходимости оказывается несовместимым с точки зрения теории вероятностей.

13. *Продолжение.* Покажите, что из $\rho(F_n, G) \rightarrow 0$ вытекает собственная сходимость $F_n \rightarrow G$. Обратное неверно (рассмотрите нормальные функции распределения $\mathfrak{N}(nx)$ и распределения F_n , сосредоточенные в n^{-1}).

14. *Продолжение.* Если $U = F_1 \star \dots \star F_n$ и $V = G_1 \star \dots \star G_n$, то

$$\rho(U, V) \leq \sum_{k=1}^n \rho(F_k, G_k). \quad (10.5)$$

Последнее неравенство обобщает основное неравенство (3.9). *Указание.* Используйте (3.9) и функцию u , такую, что $\|\mathfrak{N}u - \mathfrak{N}u\|$ близко к $\rho(U, V)$.

15. *Аппроксимация пуассоновским распределением¹⁾.* Пусть F приписывает вес p точке 1 и вес $q=1-p$ точке 0. Если G — распределение Пуассона с математическим ожиданием p , то покажите, что $\rho(F, G) \leq \frac{1}{4}p^2$, где p — расстояние, определенное в (10.3). Выведите отсюда утверждение: если F — распределение числа успехов в n испытаниях Бернулли с вероятностями успеха $p_1, p_2, \dots, p_n$ и если G — распределение Пуассона с математическим ожиданием $p_1 + \dots + p_n$, то $\rho(F, G) \leq \frac{1}{4}(p_1^2 + \dots + p_n^2)$.

16. Закон больших чисел, установленный в гл. VII, 7, утверждает, что для независимых одинаково распределенных величин $X_k \in E(X_k)=0$,

$$(X_1 + \dots + X_n)/n \xrightarrow{P} 0.$$

Докажите это методом, использованным в теореме 1 § 4.

17. Условие Линдеберга (4.15) выполняется, если при некотором $\delta > 0$ существуют $\alpha_k = E(|X_k|^{2+\delta})$ и $\alpha_1 + \dots + \alpha_n = o(s_n^{2+\delta})$ (условие Ляпунова).

18. Пусть F_k — симметричное распределение и $1 - F_k(x) = x^{-2-1/k/2}$ для $x > 1$. Покажите, что условие Линдеберга (4.15) не выполняется.

19. Пусть $X_k = \pm 1$ с вероятностью $1/2(1-k^{-2})$ и $X_k = \pm k$ с вероятностью $1/2k^{-2}$. Покажите с помощью урезания, что величина $S_n/\sqrt{n}$ ведет себя асимптотически так, как если бы $X_k = \pm 1$ с вероятностью $1/2$. Таким образом, распределение $S_n/\sqrt{n}$ сходится к $\mathfrak{N}$, но $\text{Var}(S_n/\sqrt{n}) \rightarrow 2$.

20. Видоизмените условие предыдущей задачи так, чтобы $E(X_k^2) = \infty$, но распределение $S_n/\sqrt{n}$ сходилась к $\mathfrak{N}$.

21²⁾. *Центральная предельная теорема для симметрично зависимых величин.* Пусть каждому значению θ соответствует распределение F_θ с нулевым математическим ожиданием и дисперсией $\sigma^2(\theta)$. Значение θ выбирается в соответствии с распределением вероятностей G . Затем рассматриваются взаимно независимые случайные величины X_n с одним и тем же распределением F_θ . Пусть a^2 — математическое ожидание σ^2 , взятое по отношению к G . Покажите, что распределение $S_n/(a\sqrt{n})$ сходится к распределению

$$\int_{-\infty}^{+\infty} \mathfrak{N}\left(\frac{ax}{\sigma(\theta)}\right) G\{d\theta\}.$$

Это распределение будет нормальным только в случае, когда G сосредоточено в одной точке.

22. *Шумы в электронных лампах и подобные процессы.* Рассмотрим стохастический процесс из примера 3, з) гл. VI с «дискретизированным» временем. Предполагая, что вероятность появления электрона в момент kh равна ah , покажите, что сила тока в дискретной модели определяется бесконечной сверт-

¹⁾ Задача подсказана неравенствами работы Le Cam L., An approximation theorem for the Poisson binomial distribution, Pacific J. Math., 10 (1960), 1181—1197.

²⁾ Blum J. R., Chernoff H., Rosenblatt M., and Teicher H., Central limit theorems for interchangeable processes, Canadian J. Math., 10 (1958), 222—229.

кой. Переход к пределу при $h \rightarrow 0$ приводит к теореме Кемпбелла [(3.5) гл. VI].

Прodelайте то же самое в примере с занятыми линиями [см. пример 3, и гл. VI]. Обобщите модель на случай, когда последствие (after-effect) в момент kh есть случайная величина, принимающая значения 1, 2, ... с вероятностями $p_1, p_2, \dots$.

23. Последовательность случайных величин с распределениями $\{F_n^*\}$ никогда не бывает стохастически ограниченной. Указание. Достаточно рассмотреть симметричные распределения. Кроме того, можно допустить, что F имеет бесконечные хвосты, в ином случае $F_n^* \rightarrow 0$ по центральной предельной теореме. Используйте (5.10) гл. V.

24. Продолжение. Тем не менее возможно, что при каждом x

$$\limsup_{n \rightarrow \infty} F_n^*(x) = 1, \quad \liminf_{n \rightarrow \infty} F_n^*(x) = 0.$$

В самом деле, можно выбрать две крайне быстро растущие последовательности целых чисел a_k и n_k так, что

$$(-1)^k \frac{1}{2ka_k} S_{n_k} \xrightarrow{p} 1.$$

Указание. Возьмите распределение $P\{X = (-1)^k a_k\} = p_k$. При подходящем выборе констант с большой вероятностью почти $2k$ членов последовательности $X_1, \dots, X_{n_k}$ будут равны $(-1)^k a_k$, и ни один не превзойдет по абсолютной величине a_k . Тогда для четного k $S_{n_k} > a_k - n_k a_{k-1}$. Покажите, что для этой цели пригодны значения

$$n_k = (2k)!, \quad p_k \sim \frac{1}{(2k-1)!}, \quad a_k \sim (n_k)^k.$$

25. В доказательстве леммы 8.1 достаточно предположить, что множество A плотно в некотором открытом интервале.

26. Распределение максимума. Пусть $X_1, \dots, X_n$ — независимые случайные величины с одним и тем же распределением F и $X_n^* = \max(X_1, \dots, X_n)$. Пусть G_n — распределение величины $a_n^{-1} X_n^*$.

а) Если $F(x) = 1 - e^{-x}$ и $a_n = n$, то G_n сходится к распределению, сосредоточенному в точке 0. Покажите прямым подсчетом, что ни при каком выборе a_n не может получиться распределение, имеющее более одной точки роста.

б) Если F — распределение Коши с плотностью $\frac{1}{\pi(1+x^2)}$ и $a_n = n/\pi$, то $G_n(x) \rightarrow e^{-x^{-1}}$ при $x > 0$.

27. Если X и Y имеют одно и то же распределение F , такое, что $1 - F(x) \sim x^{-p} L(x)$, где L — медленно меняющаяся функция, то

$$P\{X > t | X + Y > t\} \rightarrow \frac{1}{2}$$

при $t \rightarrow \infty$. Грубо говоря, большие значения суммы, как правило, возникают за счет больших значений одного из двух слагаемых¹⁾.

28. Пусть $v > 0$ и $a > 0$ на $0, \infty$. Допустим, что предел

$$\lim_{t \rightarrow \infty} [a(t)v(tx) + b(t)x] = z(x)$$

существует и непрерывно зависит от x . Докажите, что при фиксированном $x_0 > 0$ функция $\frac{v(x_0 x)}{x_0 x} - \frac{v(x)}{x}$ правильно меняется. Установите, что или $z(x) =$

¹⁾ Это явление было впервые подмечено, кажется, Б. Мандельбротом.

$= cx^\alpha$, или $z(x) = cx + c_1 x \log x$ при том только условии, что сама функция u не является правильно меняющейся [в этом случае $z(x) = cx^\alpha + c_1 x$].

29. Пусть G — симметричное устойчивое распределение, т. е. $G^{r*}(c_r x) = G(x)$ (см. гл. VI, 1). Выведите из последнего следствия § 8, что $1 - G(x) \sim x^{-\alpha} L(x)$, где $\alpha < 2$. При этом исключается случай $r[1 - G(c_r x)] \rightarrow 0$, приводящий к нормальному распределению.

Указание. Неравенства симметризации гл. V, (5.10) показывают, что последовательность $r[1 - G(c_r x)]$ ограничена. Дальнейшие рассуждения просты.

30. Обобщите эти утверждения на случай несимметричных устойчивых законов.

31. Пусть $\{X_n\}$ — последовательность взаимно независимых положительных случайных величин с одинаковым распределением F , которое сосредоточено на $0, \infty$. Пусть N — пуассоновская случайная величина. Случайная сумма $S_N = X_1 + \dots + X_N$ имеет обобщенное распределение Пуассона

$$U = e^{-c} \sum \frac{c^n}{n!} F^{n*},$$

Пусть функция L медленно изменяется на бесконечности. Докажите, что

$$\text{если } 1 - F(x) \sim x^{-\rho} L(x),$$

$$\text{то } 1 - U(x) \sim cx^{-\rho} L(x).$$

Указание. Совершенно очевидно, что вероятность $P\{S_N > x\}$ превышает вероятность того, что ровно одна из компонент X_j больше x , иначе

$$1 - U(x) \geq c[1 - F(t)]e^{-c[1 - F(t)]},$$

С другой стороны, для достаточно больших x из события $S_N > x$ вытекает, что или какая-либо одна из величин X_j больше $(1-\varepsilon)x$, или по крайней мере две превышают $x^{2/3}$, или, наконец, $N > x^{1/3}$. Вероятность второго по счету события есть $o(1 - F(x))$, тогда как вероятность того, что $N > \log x$, стремится к 0 быстрее, чем любая степень x .

32. Пусть F — атомическое распределение, приписывающее точке 2^n вес, пропорциональный $n^{-1}2^{-2^n}$. Покажите, что функция U_2 [см. (9.10)] медленно меняется и $U_2(\infty) = \infty$, но $1 - F$ не является правильно меняющейся функцией. Указание. Для доказательства последнего утверждения достаточно рассмотреть величину скачков.

Замечание. Оставшиеся задачи имеют отношение к обобщению понятия правильно изменяющейся функций¹⁾. Удобную исходную точку зрения задает следующее

Определение. Монотонная функция u называется изменяющейся доминируемо (varies dominantly) на бесконечности, если отношения $u(2x)/u(x)$ лежат между двумя конечными положительными постоянными.

33. Покажите, что неубывающая функция u изменяется доминируемо тогда и только тогда, когда существуют такие постоянные A, p и t_0 , что

$$\frac{u(tx)}{u(t)} < Ax^p, \quad t > t_0, \quad x > 1. \quad (10.5)$$

Для невозрастающей функции u тот же критерий применим с $x < 1$ вместо

¹⁾ О последующих результатах и деталях см.: Feller W., One-sided analogues of Karamata's regular variation, в мемориальном томе, посвященном Карамате, l'Enseignement Mathématique, vol. 15 (1969), 107—121. См. также: Feller W., On regular variation and local limit theorems, Proc. Fifth. Berkeley Symposium Math. Stat. and Probability, vol. 2, part 1, 373—388 (1965—66),

$x > 1$.

34. Обобщение теоремы 2 из § 9. Определим U_ξ и V_η так же, как в (9.10) (где $-\infty < \eta < \xi$). Положим $R(t) = t^{\xi-\eta} V_\eta(t) / U_\xi(t)$.

Покажите, что U_ξ меняется доминируемо в том и только том случае, когда $\limsup R(t) < \infty$. Аналогично V_η меняется доминируемо в том и только том случае, когда $\liminf R(t) > 0$.

35. Продолжение. Более точно: если $R(t) \leq M$ при $t > t_0$, то

$$\frac{U_\xi(tx)}{U_\xi(t)} \leq (M+1)x^p, \quad x > 1, \quad t > t_0, \quad (10,6)$$

где $p = (\xi - \eta) M / (M + 1)$. Обратно, из (10,6) при $p < \xi - \eta$ следует, что

$$R(t) \leq \frac{M(\xi - \eta) + p}{\xi - \eta - p}, \quad (10,7)$$

Эти утверждения остаются справедливыми, если R заменить обратной величиной R^{-1} и в то же время отношение $U_\xi(tx)/U_\xi(t)$ заменить отношением $V_\eta(t)/V_\eta(tx)$.

36. Докажите следующий критерий. Если существует число $s > 1$, такое, что $\liminf U_\xi(st)/U_\xi(t) > 1$, то V_η меняется доминируемо. Аналогично если $\liminf V_\eta(t/s)/V_\eta(t) > 1$, то U_ξ меняется доминируемо.

Цель этой главы — показать, что основные теоремы, касающиеся безгранично делимых распределений, процессов с независимыми приращениями, устойчивых распределений и их областей притяжения, могут быть получены с помощью естественного обобщения рассуждений, использованных при доказательстве центральной предельной теоремы. Заметим, что *излагаемая теория будет развита заново (и в большем объеме) на основе анализа Фурье*. По этой причине здесь мы ограничимся лишь основными фактами. Настоящая глава представляет в значительной степени методологический интерес, связывая разбираемые вопросы с общей теорией марковских процессов. Метод Фурье, там, где он применим, приводит к более сильным результатам. Чтобы облегчить понимание важных фактов, некоторые теоремы доказываются дважды. Так, общая теорема о строении рассматриваемых полугрупп распределений доказывается сначала для случая конечных дисперсий и т. п. В этом смысле § 1—4 дают замкнутое изложение основных фактов.

Полугрупповая операция, изучаемая в этой главе, — свертка. Другие полугруппы будут изучены независимо новыми методами в следующей главе.

§ 1. ОБЩЕЕ ЗНАКОМСТВО С ТЕМОЙ

Предельные теоремы этой главы служат естественным обобщением центральной предельной теоремы, а безгранично делимые распределения тесно связаны с нормальным распределением. Чтобы понять это, стоит повторить доказательство теоремы 1, гл. VIII, 4 в несколько иной обстановке.

Мы рассмотрим на этот раз схему серий $\{X_{k,n}\}$, где при каждом n случайные величины ¹⁾ $X_{1,n}, \dots, X_{n,n}$ взаимно независимы

¹⁾ Схема серий была определена в гл. VI, 3. Следует помнить, что в действительности мы имеем дело с функциями распределения $F_{k,n}$; случайные величины $X_{k,n}$ используются лишь для того, чтобы упростить обозначения. Соответственно случайные величины из различных строк не обязаны находиться в каких бы то ни было отношениях друг к другу (в частности, они могут быть заданы на разных вероятностных пространствах).

вероятности того, что до момента t произойдет бесконечно много скачков.

г) Применение к флуктуационным шумам, телефонным вызовам и тому подобному см. в задаче 22. ►

§ 6. ТЕОРЕМЫ О ВЫБОРЕ

Один из стандартных методов доказательства сходимости числовой последовательности состоит в том, что сначала устанавливается существование хотя бы одной предельной точки, а затем—ее единственность. Аналогичный способ применим и к распределениям. Аналог утверждения о существовании предельной точки дается следующей важной теоремой, приписываемой обычно Хелли. Как и все теоремы этого параграфа, она верна в пространстве любого числа измерений (специальный случай был использован в I; гл. XI,6).

Теорема 1. (i) Каждая последовательность $\{F_k\}$ распределений вероятностей в $\mathcal{R}^r$ содержит подпоследовательность $F_{n_1}, F_{n_2}, \dots$, сходящуюся (в собственном или несобственном смысле) к некоторому пределу F .

(ii) Для того чтобы все такие пределы были собственными, необходимо и достаточно, чтобы последовательность $\{F_n\}$ была стохастически ограниченной (см. определение 2 в § 2).

(iii) Для того чтобы $F_n \rightarrow F$, необходимо и достаточно, чтобы предел любой сходящейся подпоследовательности был равен F .

Доказательство основано на следующей лемме.

Лемма. Пусть $a_1, a_2, \dots$ —произвольная последовательность точек. Каждая последовательность $\{u_n\}$ числовых функций содержит подпоследовательность $u_{n_1}, u_{n_2}, \dots$, сходящуюся во всех точках a_j (может быть, к $\pm\infty$).

Доказательство. Мы используем «диагональный метод» Кантора. Найдется последовательность номеров $v_1, v_2, \dots$, такая, что последовательность значений $u_{v_k}(a_1)$ сходится. Чтобы избежать сложных индексов, положим $u_k^{(1)} = u_{v_k}$, так что $\{u_k^{(1)}\}$ —подпоследовательность $\{u_n\}$, сходящаяся в точке a_1 . Из этой подпоследовательности выделим другую $u_1^{(2)}, u_2^{(2)}, \dots$, сходящуюся в точке a_2 . Продолжая по индукции, мы построим при каждом n последовательность $u_1^{(n)}, u_2^{(n)}, \dots$, сходящуюся в точке a_n и содержащуюся в предыдущей последовательности. Рассмотрим теперь «диагональную» последовательность $u_1^{(1)}, u_2^{(2)}, u_3^{(3)}, \dots$. Эта последовательность, за исключением ее $n-1$ первых членов, содержится в n -й последовательности $u_1^{(n)}, u_2^{(n)}, \dots$ и потому сходится в a_n . Так как это верно при каждом n , то диагональная последовательность $\{u_n^{(n)}\}$ сходится во всех точках $a_1, a_2, \dots$. Лемма доказана. ►

удовлетворяющих (1.6), и прийти к заключению, что их суммы по строкам ведут себя асимптотически одинаково. Если для одной из этих схем распределения S_n сходятся к пределу G , то это же будет верно и для всех других наших схем. Мы докажем это позже.

Пример. б) *Распределение Пуассона.* Допустим, что $X_{t,n}$ равно единице с вероятностью p_n и нулю с вероятностью $1-p_n$. Если $np_n \rightarrow \alpha$, то

$$n[\mathfrak{F}_n u(x) - u(x)] = np_n[u(x-1) - u(x)] \rightarrow \alpha[u(x-1) - u(x)]. \quad (1.7)$$

На этот раз мы выберем в качестве $\mathfrak{G}_n$ оператор, связанный с распределением Пуассона с математическим ожиданием α/n . Простой подсчет показывает, что $n[\mathfrak{E}_n u - u]$ также сходится к правой части (1.7). Отсюда выводим, как и раньше, что $\mathfrak{F}_n u \rightarrow \mathfrak{G}_1$. Таким образом, распределение S_n сходится к распределению Пуассона с математическим ожиданием α . [Правая часть (1.7) дает нам один из возможных примеров оператора $\mathfrak{U}$ в (1.6). Другой пример простой схемы серий см. в задаче 2.] ▶

В двух указанных выше примерах нам посчастливилось знать форму предельного распределения заранее. В общем случае схема серий может служить для определения предельного распределения, и таким путем мы можем получить новые функции распределения. Эта процедура была использована в I; гл. VI, для определения распределения Пуассона как предела биномиальных распределений.

Напомним, что предельные распределения для сумм S_n были названы (см. гл. VI, 3) *безгранично делимыми*. Мы покажем, что такое предельное распределение существует всякий раз, когда выполняется соотношение типа (1.6), а также покажем, что это условие необходимо. Другой подход к рассматриваемой задаче основан на изучении мер $ny^2 F_n \{dy\}$. В обоих примерах существовала предельная мера; в примере а) она была сосредоточена в нуле, в примере б) — в точке единичица. В общем случае соотношение (1.6) тесно связано с существованием такой меры Ω , что $ny^2 F_n \{dy\} \rightarrow \Omega \{dy\}$. При этом безгранично делимые распределения могут быть описаны либо с помощью оператора $\mathfrak{U}$, либо с помощью меры Ω (которая может быть неограниченной).

Третий подход к разбираемой задаче использует уравнение свертки

$$Q_s \star Q_t = Q_{s+t}, \quad (1.8)$$

в котором Q_t — распределение вероятностей, зависящее от параметра $t > 0$.

Пример. в) Нормальное распределение и распределение Пуассона удовлетворяют (1.8) с t , пропорциональным дисперсии. Выраженное формулой (2.3) гл. II свойство свертки гамма-распре-

делений из (2.2) гл. II представляет собой частный случай (1.8). То же самое справедливо и для аналогичных свойств свертки, выведенных для распределения Коши (см. (4.5) гл. II) и одно-сторонних устойчивых распределений (см. (4.7) гл. II). ►

В схеме серий с $F_n = Q_{1/n}$ соотношение (1.6) выражает свойство (1.8) при t , пробегающем последовательность $1/2, 1/3, \dots$. Можно ожидать, что (1.6) будет выполняться при произвольном стремлении $t \rightarrow 0+$.

Отметим, что (1.8) есть не что иное, как основное уравнение для процессов со *стационарными независимыми приращениями* (гл. VI, 4). Оно тесно связано с теорией полугрупп. В этом контексте $\mathfrak{U}$ выступает в роли «производящего оператора». Оказывается, что именно эта теория обеспечивает самый легкий подход к предельным теоремам и безгранично делимым распределениям. Поэтому мы с нее и начнем.

§ 2. ПОЛУГРУППЫ СО СВЕРТКОЙ

Пусть Q_t при $t > 0$ — семейство распределений в $\mathcal{R}^1$, удовлетворяющих (1.8), и $\mathfrak{Q}(t)$ — соответствующие операторы, т. е.

$$\mathfrak{Q}(t)u(x) = \int_{-\infty}^{+\infty} u(x-y) Q_t\{dy\}. \quad (2.1)$$

Тогда (1.8) эквивалентно

$$\mathfrak{Q}(s+t) = \mathfrak{Q}(s)\mathfrak{Q}(t). \quad (2.2)$$

Семейство операторов, удовлетворяющих (2.2), называется *полугруппой*. [Оно не будет, как правило, группой, так как оператор $\mathfrak{Q}(t)$ в общем случае не имеет обратного.] Операторы полугруппы могут иметь произвольную природу, и поэтому удобно ввести термин, подчеркивающий наше допущение о связи $\mathfrak{Q}(t)$ и вероятностных распределений.

Определение 1. *Полугруппа со сверткой $\{\mathfrak{Q}(t)\}$ (где $t > 0$) — это семейство операторов, связанных с распределениями вероятностей и удовлетворяющих (2.2).*

В качестве области определения операторов мы возьмем $C_0[-\infty, \infty]$. Операторы $\mathfrak{Q}(t)$ являются операторами перехода, т. е. из $0 \leq u \leq 1$ вытекает $0 \leq \mathfrak{Q}(t)u \leq 1$ и, кроме того, $\mathfrak{Q}(t)1 = 1$.

Нам придется также иметь дело и с операторами [подобными $\partial^2/\partial x^2$ в (1.3)], которые определены не для всех непрерывных функций. Наши теперешние цели позволяют, к счастью, избежать утомительного обсуждения точной области определения этих операторов, так как нам достаточно рассмотреть только класс таких функций u , что $u \in C[-\infty, \infty]$ и u имеет производные всех поряд-

ков, также принадлежащие $C[-\infty, \infty]$. Эти функции будут называться *бесконечно дифференцируемыми*¹⁾, и их совокупность будет обозначаться C^∞ . Пока мы ограничимся только операторами $\mathfrak{A}$, определенными для всех $u \in C^\infty$ и таким, что $\mathfrak{A}u \in C^\infty$, так что все встречающиеся операторы можно рассматривать как *операторы, отображающие C^∞ в C^∞* . Как мы видели в гл. VIII, 3, для операторов, связанных с распределениями вероятностей, сходимость $\mathfrak{F}_n \rightarrow \mathfrak{F}$ равносильна сходимости $\mathfrak{F}_n u \rightarrow \mathfrak{F}u$ для всех $u \in C^\infty$. Мы распространим теперь это определение сходимости на произвольные операторы.

Определение 2. Пусть $\mathfrak{A}_n$ и $\mathfrak{A}$ — операторы, отображающие C^∞ в C^∞ . Мы скажем, что $\mathfrak{A}_n$ сходится к $\mathfrak{A}$, в обозначениях $\mathfrak{A}_n \rightarrow \mathfrak{A}$, если при каждом $u \in C^\infty$

$$\|\mathfrak{A}_n u - \mathfrak{A}u\| \rightarrow 0. \quad (2.3)$$

Соотношение (2.3) показывает, что $\mathfrak{A}_n u \rightarrow \mathfrak{A}u$ равномерно. Обратно, если для каждого $u \in C^\infty$ последовательность $\{\mathfrak{A}_n u\}$ сходится равномерно к пределу $v \in C^\infty$, то равенство $\mathfrak{A}u = v$ определяет некоторый оператор $\mathfrak{A}$ и, очевидно, $\mathfrak{A}_n \rightarrow \mathfrak{A}$.

Определение 3. Полугруппа со сверткой $\{\mathfrak{Q}(t)\}$ называется непрерывной, если

$$\mathfrak{Q}(h) \rightarrow 1, \quad h \rightarrow 0+, \quad (2.4)$$

где 1 — тождественный оператор. В этом случае мы полагаем $\mathfrak{Q}(0) = 1$.

Из неравенства $\|\mathfrak{Q}(t)u\| \leq \|u\|$ и из определения (2.2) получаем при $h > 0$

$$\|\mathfrak{Q}(t+h)u - \mathfrak{Q}(t)u\| \leq \|\mathfrak{Q}(h)u - u\|. \quad (2.5)$$

При достаточно малом h левая часть будет $< \varepsilon$, каково бы ни было t . В этом смысле полугруппа со сверткой является *равномерно непрерывной*.

Определение 4. Говорят, что оператор $\mathfrak{A}$, отображающий C^∞ в C^∞ , порождает полугруппу со сверткой $\{\mathfrak{Q}(t)\}$, если при $h \rightarrow 0+$

$$\frac{\mathfrak{Q}(h) - 1}{h} \rightarrow \mathfrak{A}. \quad (2.6)$$

Оператор $\mathfrak{A}$ будет называться производящим оператором²⁾.

¹⁾ Класс C^∞ введен только для того, чтобы избежать появления нового термина. Он мог бы быть заменен, скажем, классом функций с четырьмя ограниченными производными или (что еще проще) классом всех линейных комбинаций нормальных функций распределения с произвольными математическими ожиданиями и дисперсиями.

²⁾ Мы ограничили область определения $\mathfrak{A}$ классом C^∞ . Поэтому наша терминология слегка отличается от общепринятой, ср., например, книгу Хилле и Филлипса (1962).

Более определено, всякий раз, когда предел существует, оператор $\mathfrak{U}$ определяется соотношением

$$t^{-1} \int_{-\infty}^{+\infty} [u(x-y) - u(x)] Q_t \{dy\} \rightarrow \mathfrak{U}u(x). \quad (2.7)$$

Очевидно, что полугруппа, обладающая производящим оператором, автоматически непрерывна. Мы покажем в дальнейшем, что любая непрерывная полугруппа со сверткой имеет производящий оператор, но это утверждение ни в коей мере не очевидно.

Формально (2.6) определяет $\mathfrak{U}$ как производную $\mathfrak{Q}(t)$ при $t=0$. Существование $\mathfrak{U}$ влечет дифференцируемость и при $t>0$, так как

$$\frac{\mathfrak{Q}(t+h) - \mathfrak{Q}(t)}{h} = \frac{\mathfrak{Q}(h) - 1}{h} \mathfrak{Q}(t) \rightarrow \mathfrak{U} \mathfrak{Q}(t) \quad (2.8)$$

при $h \rightarrow 0+$ и аналогично при $h \rightarrow 0-$.

Приводимые ниже примеры будут использованы в дальнейшем.

Примеры. а) *Обобщенная пуассоновская полугруппа.* Пусть

$$Q_t = e^{-\alpha t} \sum_{k=0}^{\infty} \frac{(\alpha t)^k}{k!} F^k \star \quad (2.9)$$

— обобщенное пуассоновское распределение. Здесь

$$\mathfrak{Q}(h)u - u = (e^{-\alpha h} - 1)u + \alpha h e^{-\alpha h} \left[\mathfrak{F}u + \frac{\alpha h}{2!} \mathfrak{F}^2 u + \dots \right]. \quad (2.10)$$

Деля обе части на h , видим, что (2.6) выполняется с $\mathfrak{U} = \alpha(\mathfrak{F} - 1)$. Таким образом, обобщенная пуассоновская полугруппа (2.9) порождается оператором $\alpha(\mathfrak{F} - 1)$ и ее элементы мы будем записывать сокращенно в виде $\mathfrak{Q}(t) = e^{\alpha t(\mathfrak{F} - 1)t}$.

б) *Сдвиги.* Обозначим через T_a распределение, сосредоточенное в точке a , и через $\mathfrak{F}_a$ — соответствующий оператор. При фиксированном $\beta > 0$ выполняется полугрупповое свойство $T_{\beta s} \star T_{\beta t} = T_{\beta(t+s)}$. Кроме того, $\mathfrak{F}(\beta t)u(x) = u(x - \beta t)$. График функции $\mathfrak{F}(\beta t)u$ получается сдвигом графика u . Поэтому мы будем говорить о полугруппе сдвигов. Производящий оператор равен $-\beta \frac{d}{dx}$. Заметим, что он может быть представлен как предел при $h \rightarrow 0$ производящих операторов $\alpha(\mathfrak{F} - 1)$, где $\alpha = \beta/h$ и F сосредоточено в h . Оператор $\alpha(\mathfrak{F} - 1)$ — разностный оператор, и переход к пределу рассматривался в гл. VII, 5. Разумно обозначить эту полугруппу как $\mathfrak{F}(t) = \exp\left(-\beta t \frac{d}{dx}\right)$.

в) *Сложение производящих операторов.* Пусть $\mathfrak{A}_1$ и $\mathfrak{A}_2$ порождают полугруппы со сверткой $\{\mathfrak{D}_1(t)\}$ и $\{\mathfrak{D}_2(t)\}$. Тогда оператор $\mathfrak{A}_1 + \mathfrak{A}_2$ порождает полугруппу операторов со сверткой $\mathfrak{D}(t) = \mathfrak{D}_1(t) \mathfrak{D}_2(t)$. [Такая полугруппа связана со сверткой распределений, соответствующих $\mathfrak{D}_1(t)$ и $\mathfrak{D}_2(t)$; см. теорему 2 гл. VIII, 3.] Высказанное утверждение очевидным образом следует из формулы

$$\frac{\mathfrak{D}_1(h) \mathfrak{D}_2(h) - 1}{h} = \frac{\mathfrak{D}_1(h) - 1}{h} + \mathfrak{D}_1(h) \frac{\mathfrak{D}_2(h) - 1}{h}. \quad (2.11)$$

г) *Сдвинутые полугруппы.* Как частный случай предыдущего примера, отметим правило: если $\mathfrak{A}$ порождает полугруппу операторов $\mathfrak{D}(t)$, связанных с распределениями Q_t , то оператор $\mathfrak{A} - \beta d/dx$ порождает полугруппу $\{\mathfrak{D}^*(t)\}$, такую, что $Q_t^*(x) = Q_t(x - \beta t)$.

д) *Нормальные полугруппы.* Пусть Q_t — нормальное распределение с нулевым математическим ожиданием и дисперсией ct . Как уже говорилось, это распределение определяет полугруппу, и мы воспользуемся для отыскания соответствующего производящего оператора определением (2.7). По формуле Тейлора

$$u(x-y) - u(x) = -yu'(x) + \frac{1}{2}y^2u''(x) - \frac{1}{6}y^3u'''(x - \theta y). \quad (2.12)$$

Третий абсолютный момент распределения Q_t пропорционален $t^{3/2}$, и потому для функций u с тремя ограниченными производными предел в (2.7) существует и равен $\frac{1}{2}cu''(x)$. Таким образом, производящий оператор $\mathfrak{A}$ есть $\frac{1}{2}c \frac{d^2}{dx^2}$. ▶

(Дальнейшие примеры см. в задачах 3—5.)

Замечание об упрощении Фоккера—Планка. Рассмотрим семейство функций, определенных формулой $v(t, x) = \mathfrak{D}(t)f(x)$. Соотношение (2.8) показывает, что для гладких f

$$\frac{\partial v}{\partial t} = \mathfrak{A}v. \quad (2.13)$$

Это — уравнение Фоккера—Планка для рассматриваемого процесса, и v — его единственное решение, удовлетворяющее начальному условию $v(0, x) = f(x)$. Уравнение (2.13) описывает процесс. Заметим, что традиционные попытки заменить (2.13) уравнением для самих переходных вероятностей Q_t приводят к ненужным усложнениям. Возьмем, например, сдвинутую обобщенную пуассоновскую полугруппу, порожденную оператором $\mathfrak{A} = \alpha(\mathfrak{F} - 1) - \beta \frac{d}{dx}$. Уравнение Фоккера—Планка (2.13) выполняется при любой начальной функции $f(x) = v(0, x)$, имеющей непрерывную производную. Его формальный аналог для переходных вероятностей имеет вид

$$\frac{\partial Q_t}{\partial t} = -\beta \frac{\partial Q_t}{\partial x} - \alpha Q_t + \alpha F \star Q_t. \quad (2.14)$$

Это уравнение имеет смысл, только если Q имеет плотность, и потому неприменимо к дискретным процессам. Обычные ссылки на (2.14) вместо (2.13) порождают лишь затруднения.

§ 3. ПОДГОТОВИТЕЛЬНЫЕ ЛЕММЫ

В этом параграфе собрано вместе несколько простых лемм, на которые опирается вся теория. Несмотря на всю свою простоту, основную роль играет следующее неравенство.

Лемма 1. Если операторы $\mathfrak{A}$ и $\mathfrak{A}^*$ порождают полугруппы со сверткой $\{\mathfrak{D}(t)\}$ и $\{\mathfrak{D}^*(t)\}$ соответственно, то при всех $t > 0$

$$\|\mathfrak{D}(t)u - \mathfrak{D}^*(t)u\| \leq t \|\mathfrak{A}u - \mathfrak{A}^*u\|. \quad (3.1)$$

Доказательство. Из полугруппового свойства и основного неравенства (1.5) при $r=1, 2, \dots$ получаем

$$\begin{aligned} \|\mathfrak{D}(t)u - \mathfrak{D}^*(t)u\| &\leq r \left\| \mathfrak{D}\left(\frac{t}{r}\right)u - \mathfrak{D}^*\left(\frac{t}{r}\right)u \right\| = \\ &= t \left\| \frac{\mathfrak{D}(t/r) - 1}{t/r} u - \frac{\mathfrak{D}^*(t/r) - 1}{t/r} u \right\|. \end{aligned} \quad (3.2)$$

При $r \rightarrow \infty$ правая часть стремится к правой части (3.1), так что это неравенство верно. $\blacktriangleright$

Следствие. Различные полугруппы со сверткой не могут иметь один и тот же производящий оператор.

Лемма 2. (Сходимость.) Пусть при каждом n оператор $\mathfrak{A}_n$ порождает полугруппу со сверткой $\{\mathfrak{D}_n(t)\}$.

Если $\mathfrak{A}_n \rightarrow \mathfrak{A}$, то $\mathfrak{A}$ порождает полугруппу со сверткой $\{\mathfrak{D}(t)\}$ и $\mathfrak{D}_n(t) \rightarrow \mathfrak{D}(t)$ при каждом $t > 0$.

Доказательство. При каждом $t > 0$ последовательность $\{\mathfrak{D}_n(t)u\}$ сходится равномерно, так как по (3.1)

$$\|\mathfrak{D}_n(t)u - \mathfrak{D}_m(t)u\| \leq t \|\mathfrak{A}_n u - \mathfrak{A}_m u\|. \quad (3.3)$$

По критерию 1 гл. VIII, 3, существует оператор $\mathfrak{D}(t)$, связанный с некоторым распределением вероятностей, такой, что $\mathfrak{D}_n(t) \rightarrow \mathfrak{D}(t)$. Тогда

$$\mathfrak{D}_n(s+t) = \mathfrak{D}_n(s) \mathfrak{D}_n(t) \rightarrow \mathfrak{D}(s) \mathfrak{D}(t) \quad (3.4)$$

(см. теорему 2 гл. VIII, 3). Поэтому $\{\mathfrak{D}(t)\}$ — полугруппа со сверткой. Покажем, что она порождается оператором $\mathfrak{A}$. Отметим с этой целью, что

$$\left\| \frac{\mathfrak{D}(t) - 1}{t} u - \mathfrak{A}u \right\| \leq \left\| \frac{\mathfrak{D}_n(t) - 1}{t} u - \mathfrak{A}u \right\| + \left\| \frac{\mathfrak{D}(t)u - \mathfrak{D}_n(t)u}{t} \right\|.$$

Первое слагаемое в правой части стремится к $\|\mathfrak{A}_n u - \mathfrak{A}u\|$ при $t \rightarrow 0$. Полагая в (3.3) $m \rightarrow \infty$, видим, что второе слагаемое меньше $\|\mathfrak{A}u - \mathfrak{A}_n u\|$.

Следовательно, при фиксированном n верхний предел левой части меньше величины $2\|\mathfrak{A}u - \mathfrak{A}_n u\|$, которая может быть сделана произвольно малой при достаточно больших n .

Последующая лемма делает по крайней мере правдоподобным то, что каждая непрерывная полугруппа имеет производящий оператор.

Лемма 3. Пусть $\{\Omega(t)\}$ — непрерывная полугруппа со сверткой. Если для некоторой последовательности $t_1, t_2, \dots \rightarrow 0$

$$\frac{\Omega(t_k) - 1}{t_k} \rightarrow \mathfrak{A}, \quad (3.5)$$

то $\mathfrak{A}$ порождает эту полугруппу.

Доказательство. Обозначим левую часть $\mathfrak{A}_k$. Как было показано в примере 2, а), оператор $\mathfrak{A}_k$ порождает обобщенную пуассоновскую полугруппу и по последней лемме существует полугруппа $\{\Omega^*(t)\}$, порождаемая $\mathfrak{A}$. Докажем, что $\Omega^*(t) = \Omega(t)$. Поступая, как в (3.2), получим

$$\|\Omega(rt_k)u - \Omega^*(rt_k)u\| \leq rt_k \left\| \mathfrak{A}_k u - \frac{\Omega^*(t_k) - 1}{t_k} u \right\|. \quad (3.6)$$

Пусть $k \rightarrow \infty$ и $r \rightarrow \infty$ так, что $rt_k \rightarrow t$. Правая часть стремится к нулю, а левая в силу (2.5) — к $\|\Omega(t)u - \Omega^*(t)u\|$. ►

Эти леммы мы используем немедленно. Следующая лемма приведена здесь потому, что она представляет собой вариант леммы 2 и доказательство ее почти такое же. Мы воспользуемся лишь частным случаем $v_n = n$ и $t = 1$, который будет служить связующим звеном между схемами серий и полугруппами со сверткой.

Лемма 4. Пусть при каждом n определен оператор $\mathfrak{F}_n$, соответствующий распределению вероятностей F_n . Если

$$n(\mathfrak{F}_n - 1) \rightarrow \mathfrak{A}, \quad (3.7)$$

то $\mathfrak{A}$ порождает полугруппу со сверткой $\{\Omega(t)\}$. Если $n \rightarrow \infty$ и $\frac{v_n}{n} \rightarrow t$, то

$$\mathfrak{F}_n^{v_n} \rightarrow \Omega(t). \quad (3.8)$$

В частности, $\mathfrak{F}_n^n \rightarrow \Omega(1)$. Лемма остается верной, если n пробегает подпоследовательность $n_1, n_2, \dots$.

Доказательство. Левая часть (3.7) порождает обобщенную пуассоновскую полугруппу [пример 2, а)], так что по лемме 2 $\mathfrak{A}$ будет производящим оператором. По основному неравенству (1.5) имеем

$$\left\| \mathfrak{F}_n^{v_n} u - \Omega\left(\frac{v_n}{n}\right) u \right\| \leq \frac{v_n}{n} \left\| n[\mathfrak{F}_n u - u] - n \left[\Omega\left(\frac{1}{n}\right) u - u \right] \right\|. \quad (3.9)$$

Если $u \in C^\infty$, то каждый из членов разности, стоящей под знаком нормы справа в (3.9), равномерно сходится к $\mathfrak{A}u$. ►

§ 4. СЛУЧАЙ КОНЕЧНЫХ ДИСПЕРСИЙ

Полугруппы распределений с конечными дисперсиями особенно важны, а их теория столь проста, что заслуживает отдельного изложения. Многим читателям не будут интересны более общие полугруппы, а другим этот параграф даст полезный предварительный пример.

Рассмотрим полугруппу со сверткой $\{\Omega(t)\}$ и обозначим соответствующие распределения вероятностей через Q_t . Допустим, что Q_t имеет конечную дисперсию $\sigma^2(t)$. В силу полугруппового свойства $\sigma^2(s+t) = \sigma^2(s) + \sigma^2(t)$. Единственное решение этого уравнения¹⁾ имеет вид $\sigma^2(t) = ct$.

Предположим, что Q_t имеет нулевое математическое ожидание. Тогда второй момент Q_t индуцирует *распределение вероятностей* Ω_t , определенное равенством

$$\Omega_t\{dy\} = \frac{1}{ct} y^2 Q_t\{dy\}. \quad (4.1)$$

По теореме о выборе существует последовательность $\{t_n\}$, стремящаяся к 0 и такая, что если t пробегает ее, то Ω_t стремится к (возможно, несобственному) распределению вероятностей Ω .

Поскольку Q_t имеет нулевое математическое ожидание, то имеет место тождество

$$\frac{\Omega(t) - 1}{t} u(x) = c \int_{-\infty}^{+\infty} \frac{u(x-y) - u(x) + y u'(x)}{y^2} \Omega_t\{dy\}, \quad (4.2)$$

где подынтегральное выражение при фиксированном x есть непрерывная функция от y , принимающая в нуле значение $\frac{1}{2} u''(x)$. На бесконечности эта функция и ее производная обращаются в нуль. Отсюда вытекает, что если t пробегает последовательность $\{t_n\}$ и как следствие $\Omega_t \rightarrow \Omega$, то интеграл в (4.2) равномерно сходится к аналогичному интегралу относительно распределения Ω . В силу леммы 3 из § 3 это означает, что наша полугруппа имеет производящий оператор $\mathfrak{A}$, определяемый равенством

$$\mathfrak{A}u(x) = c \int_{-\infty}^{+\infty} \frac{u(x-y) - u(x) + y u'(x)}{y^2} \Omega\{dy\}. \quad (4.3)$$

Это представление $\mathfrak{A}$ *единственно*, поскольку для функций вида

$$u(x) = 1 + \frac{x^2}{1+x^2} f(-x) \quad (4.4)$$

¹⁾ Уравнение $\sigma(s+t) = \sigma(s) + \sigma(t)$ называют уравнением Гамеля. Полагая $u(t) = e^{\sigma(t)}$, получаем $u(s+t) = u(s)u(t)$. В этой форме уравнение встречалось нам несколько раз и изучалось в I, гл. XVII, 6. Математическое ожидание Q_t также удовлетворяет уравнению Гамеля; следовательно, или оно имеет вид mt , или крайне причудливо; см. § 5а.

имеем

$$\mathfrak{U}u(0) = c \int_{-\infty}^{+\infty} \frac{f(y)}{1+y^2} \Omega\{dy\}. \quad (4.5)$$

Поэтому знание $\mathfrak{U}u$ для всех $u \in C^\infty$ определяет однозначно меру $(1+y^2)^{-1} \Omega\{dy\}$, а вместе с тем и меру Ω .

Как следствие доказанной единственности представления, получаем, что предельное распределение Ω не зависит от выбора последовательности $\{t_k\}$, т. е. $\Omega_t\{dy\} \rightarrow \Omega\{dy\}$ при любом стремлении $t \rightarrow 0$.

Мы покажем, что Ω — собственное распределение вероятностей и что каждый оператор вида (4.3) является производящим оператором. Доказательство использует анализ двух частных случаев, которые содержатся в следующих примерах.

Примеры. а) *Нормальные полугруппы.* Если распределение вероятностей Ω сосредоточено в нуле, то (4.3) сводится к $\mathfrak{U}u(x) = \frac{1}{2}cu''(x)$. В примере 2, д) мы видели, что такой оператор $\mathfrak{U}$ порождает полугруппу нормальных распределений с нулевыми математическими ожиданиями и дисперсиями ct . Легко проверить, что распределения Ω_t стремятся к распределению вероятностей, сосредоточенному в нуле.

б) *Обобщенные пуассоновские полугруппы.* Пусть распределение вероятностей F сосредоточено на интервалах $|x| > \eta$ и имеет математическое ожидание m_1 и дисперсию m_2 . Распределения Q_t обобщенной пуассоновской полугруппы из примера 2, а) имеют математические ожидания $\alpha m_1 t$ и дисперсии $\alpha m_2 t$. Полугруппа порождается оператором $\alpha(\mathfrak{F}-1)$. Согласно примеру 2, г), те же распределения Ω_t , но с нулевыми математическими ожиданиями образуют полугруппу, порождаемую оператором вида

$$\alpha[\mathfrak{F}-1-m_1 d/dx]$$

или

$$\mathfrak{U}u(x) = \int_{|y|>\eta} [u(x-y) - u(x) - yu'(x)] F\{dy\}. \quad (4.6)$$

С точностью до обозначений $\Omega\{dy\} = y^2 F\{dy\}/m_2$ и $\alpha m_2 = c$ это сводится к (4.3). Обратно, если распределение вероятностей Ω сосредоточено на $|y| > 0$, то (4.3) можно переписать в виде (4.6) и, следовательно, оператор $\mathfrak{U}$ такого вида порождает обобщенную пуассоновскую полугруппу, распределения Q_t которой имеют нулевые математические ожидания и дисперсии $m_2 t = ct$.

Теперь мы в состоянии сформулировать основную теорему.

Теорема. Пусть Q_t имеет нулевое математическое ожидание и дисперсию ct . Полугруппа со сверткой $\{\Omega(t)\}$ имеет производящий оператор $\mathfrak{U}$ вида (4.3), где Ω — собственное распределение вероят-

ностей. Представление (4.3) единственно. Обратно, каждый оператор такого вида порождает полуруппу распределений со сверткой, причем эти распределения имеют нулевое математическое ожидание и дисперсию ct .

Доказательство. Мы установили существование производящего оператора вида (4.3), но относительно Ω знаем лишь, что соответствующая полная масса $\omega \leq 1$. Остается доказать, что если Ω имеет полную массу ω , то оператор $\mathfrak{A}$ из (4.3) порождает такую полуруппу, что Q_t имеет нулевое математическое ожидание и дисперсию $\leq c\omega t$.

Пусть $\mathfrak{A}_\eta$ — оператор, получаемый из (4.3) исключением интервалов $0 < |y| \leq \eta$ из области интегрирования. Обозначим через m и ω_η массы, которые распределение Ω приписывает соответственно точке нуль и интервалам $|y| > \eta$. Как следует из вышеприведенных примеров, оператор $\mathfrak{A}_\eta$ представляет собой сумму двух операторов, из которых один порождает нормальную полуруппу с дисперсиями $cm t$, а второй порождает обобщенную пуассоновскую полуруппу с дисперсиями $c\omega_\eta t$. По правилу сложения производящих операторов из примера 2, в) оператор $\mathfrak{A}_\eta$ сам порождает полуруппу с дисперсиями $c(m + \omega_\eta)t$. Оператор $\mathfrak{A}$ как предел $\mathfrak{A}_\eta$ при $\eta \rightarrow 0$ сам служит порождающим оператором некоторой полуруппы. Дисперсии соответствующих этой полуруппе распределений лежат между дисперсиями $c(m + \omega_\eta)t$ распределений, соответствующих $\mathfrak{A}_\eta$, и их предельными значениями $c\omega t$. Это доказывает, что оператор $\mathfrak{A}$ действительно порождает полуруппу с дисперсиями $c\omega t$. ▶

§ 5. ОСНОВНЫЕ ТЕОРЕМЫ

В этом параграфе $\{\Omega(t)\}$ обозначает произвольную непрерывную полуруппу со сверткой. Соответствующие функции распределения будут снова обозначаться через Q_t . По аналогии с (4.1) определим новую меру Ω_t равенством

$$\Omega_t\{dy\} = t^{-1}y^2 Q_t\{dy\}. \quad (5.1)$$

Новая особенность состоит, однако, в том, что при отсутствии второго момента у распределения Q_t мера Ω_t не обязана быть конечной на всей прямой. Но Ω_t принимает конечные значения $\Omega_t\{I\}$ на любом конечном интервале I . Кроме того, поскольку $Q_t\{dy\} = ty^{-2}\Omega_t\{dy\}$, то y^{-2} интегрируема относительно Ω_t в любой области, содержащей некоторую окрестность нуля. Мы увидим, что при $t \rightarrow 0$ меры Ω_t будут сходиться к мере Ω с аналогичными свойствами. Эта мера Ω будет определять производящий генератор некоторой полуруппы. Удобно поэтому ввести следующее

Определение. Мера Ω на действительной прямой называется канонической, если $\Omega\{I\} < \infty$ для всех конечных интервалов I .

и если интегралы

$$\psi^+(x) = \int_x^\infty y^{-2} \Omega \{dy\}, \quad \psi^-(-x) = \int_{-\infty}^{-x} y^{-2} \Omega \{dy\} \quad (5.2)$$

сходятся при каждом $x > 0$.

(Для определенности будем считать интервалы интегрирования замкнутыми.)

Мы собираемся показать, что здесь применима теория предыдущего параграфа, с тем лишь исключением, что мы должны иметь дело не с распределениями вероятностей, а с каноническими мерами и что при отсутствии математических ожиданий мы должны выбрать способ центрирования. Определим *урезающую функцию* τ_s , как такую непрерывную монотонную функцию, что

$$\tau_s(x) = x \text{ при } |x| \leq s, \quad \tau_s(x) = \pm s \text{ при } |x| \geq s, \quad (5.3)$$

где $s > 0$ произвольно, но фиксировано.

По аналогии с (4.2) мы теперь имеем тождество

$$\frac{\Omega_t(t-1)}{t} u(x) = \int_{-\infty}^{+\infty} \frac{u(x-y) - u(x) - \tau_s(y) u'(x)}{y^2} \Omega_t \{dy\} + b_t u'(x), \quad (5.4)$$

где

$$b_t = \int_{-\infty}^{+\infty} \tau_s(y) y^{-2} \Omega_t \{dy\} = t^{-1} \int_{-\infty}^{+\infty} \tau_s(y) Q_t \{dy\}. \quad (5.5)$$

Подынтегральное выражение в (5.4) снова является (при фиксированном x) ограниченной непрерывной функцией y , принимающей в нуле значение $\frac{1}{2} u''(x)$. Следует отметить, что наш выбор (5.3) в качестве урезающей функции не имеет исключительного значения: мы могли бы выбрать в качестве τ любую ограниченную непрерывную функцию, которая вблизи нуля дважды непрерывно дифференцируема, при этом $\tau(0) = \tau''(0) = 0$ и $\tau'(0) = 1$.

Мы имеем теперь ситуацию, аналогичную той, с которой мы встретились в предыдущем параграфе. Докажем поэтому аналогичную теорему. Интеграл в (5.4) сохраняет смысл, когда мера Ω_t заменена любой канонической мерой. Определим оператор $\mathfrak{A}^{(\tau)}$ равенством

$$\mathfrak{A}^{(\tau)} u(x) = \int_{-\infty}^{+\infty} \frac{u(x-y) - u(x) - \tau_s(y) u'(x)}{y^2} \Omega \{dy\}. \quad (5.6)$$

Индекс τ служит для указания на зависимость от урезающей функции τ . Изменение τ_s [или точки s в определении (5.3)] сводится к добавлению слагаемого bd/dx к $\mathfrak{A}^{(\tau)}$, и поэтому семейство опе-

вероятности того, что до момента t произойдет бесконечно много скачков.

г) Применение к флуктуационным шумам, телефонным вызовам и тому подобному см. в задаче 22. ►

§ 6. ТЕОРЕМЫ О ВЫБОРЕ

Один из стандартных методов доказательства сходимости числовой последовательности состоит в том, что сначала устанавливается существование хотя бы одной предельной точки, а затем—ее единственность. Аналогичный способ применим и к распределениям. Аналог утверждения о существовании предельной точки дается следующей важной теоремой, приписываемой обычно Хелли. Как и все теоремы этого параграфа, она верна в пространстве любого числа измерений (специальный случай был использован в 1; гл. XI,6).

Теорема 1. (i) Каждая последовательность $\{F_k\}$ распределений вероятностей в $\mathcal{R}^r$ содержит подпоследовательность $F_{n_1}, F_{n_2}, \dots$, сходящуюся (в собственном или несобственном смысле) к некоторому пределу F .

(ii) Для того чтобы все такие пределы были собственными, необходимо и достаточно, чтобы последовательность $\{F_n\}$ была стохастически ограниченной (см. определение 2 в § 2).

(iii) Для того чтобы $F_n \rightarrow F$, необходимо и достаточно, чтобы предел любой сходящейся подпоследовательности был равен F .

Доказательство основано на следующей лемме.

Лемма. Пусть $a_1, a_2, \dots$ —произвольная последовательность точек. Каждая последовательность $\{u_n\}$ числовых функций содержит подпоследовательность $u_{n_1}, u_{n_2}, \dots$, сходящуюся во всех точках a_j (может быть, к $\pm\infty$).

Доказательство. Мы используем «диагональный метод» Кантора. Найдется последовательность номеров $v_1, v_2, \dots$, такая, что последовательность значений $u_{v_k}(a_1)$ сходится. Чтобы избежать сложных индексов, положим $u_k^{(1)} = u_{v_k}$, так что $\{u_k^{(1)}\}$ —подпоследовательность $\{u_n\}$, сходящаяся в точке a_1 . Из этой подпоследовательности выделим другую $u_1^{(2)}, u_2^{(2)}, \dots$, сходящуюся в точке a_2 . Продолжая по индукции, мы построим при каждом n последовательность $u_1^{(n)}, u_2^{(n)}, \dots$, сходящуюся в точке a_n и содержащуюся в предыдущей последовательности. Рассмотрим теперь «диагональную» последовательность $u_1^{(1)}, u_2^{(2)}, u_3^{(3)}, \dots$. Эта последовательность, за исключением ее $n-1$ первых членов, содержится в n -й последовательности $u_1^{(n)}, u_2^{(n)}, \dots$ и потому сходится в a_n . Так как это верно при каждом n , то диагональная последовательность $\{u_n^{(n)}\}$ сходится во всех точках $a_1, a_2, \dots$. Лемма доказана. ►

бегают так выбранную последовательность $\{n_k\}$, то

$$t^{-1}[\Sigma(t) - 1]u(x) \rightarrow \mathfrak{U}u(x), \quad (5.11)$$

причем сходимости здесь равномерная. В силу леммы 3 из § 3 это означает, что полугруппа $\{\Sigma(t)\}$ порождается оператором $\mathfrak{U}$ и поэтому (5.11) выполняется при любом стремлении $t \rightarrow 0$.

Таким образом, мы доказали, что производящий оператор $\mathfrak{U}$ существует и может быть представлен в форме (5.6)—(5.7) в терминах канонической меры Ω . Как и в предыдущем параграфе, единственность меры Ω в этом представлении следует из того факта, что для функций вида (4.4) величина $\mathfrak{U}u(0)$ задается формулой (4.5) [где c включено в Ω].

Из единственности Ω следует, что (5.8) выполняется при любом стремлении $t \rightarrow 0$. Более того, (5.10) гарантирует, что величины в (5.9) равномерно малы при достаточно больших x . Эти величины были определены равенствами (5.2) и аналогичными соотношениями с Ω_t вместо Ω . Отсюда вытекает, что (5.9) есть следствие (5.8).

Остается показать, что мера Ω может быть выбрана произвольным образом. Доказательство проводится так же, как в случае конечных дисперсий. Следуя примеру 4, 6), можно убедиться в том, что если мера Ω сосредоточена на множестве $|y| > \eta > 0$, то оператор $\mathfrak{U}$, определенный в (5.6)—(5.7), порождает обобщенную пуассоновскую полугруппу, распределения которой централизованы теперь иначе в отсутствие конечных математических ожиданий. Таким образом, оператор $\mathfrak{U}$ снова может быть представлен как предел производящих операторов и поэтому сам является производящим оператором. ►

Пример. Полугруппы Коши. Распределения Q_t с плотностью $\pi^{-1}t(t^2 + x^2)^{-1}$ образуют полугруппу. Легко проверить, что пределы в соотношениях (5.9) равны $\psi^+(x) = \psi^-(-x) = \pi x^{-1}$ и мера $\pi\Omega$ совпадает с мерой Лебега или обычной длиной.

Следующая ниже теорема заключает в себе различные важнейшие характеристические свойства безгранично делимых распределений. Доказательство части (v) откладывается до § 7 (иное доказательство — непосредственное — см. в задаче 11). Эта часть допускает дальнейшее обобщение на схемы серий с различными распределениями (см. § 9; общая теория будет развита в гл. XVII). Исторические замечания по поводу этой теории см. гл. VI, 3.

Теорема 2. Следующие классы вероятностных распределений совпадают:

- (i) Безгранично делимые распределения.
- (ii) Распределения, соответствующие непрерывным полугруппам со сверткой (т. е. распределения приращений в процессах со стационарными независимыми приращениями).
- (iii) Пределы последовательностей обобщенных пуассоновских распределений.

(iv) *Пределы последовательностей безгранично делимых распределений.*

(v) *Предельные распределения для сумм в схемах серий $\{X_{k,n}\}$, где случайные величины $X_{k,n}$ n -й серии имеют одинаковое распределение.*

Доказательство. Пусть $\{\mathfrak{D}(t)\}$ — непрерывная полугруппа со сверткой. Как было показано в примере 2, а), оператор $\mathfrak{A}_h = [\mathfrak{D}(h) - 1]/h$ при фиксированном $h > 0$ порождает обобщенную пуассоновскую полугруппу операторов $\mathfrak{D}_h(t)$. При $h \rightarrow 0$ производящие операторы $\mathfrak{A}_h$ сходятся к производящему оператору $\mathfrak{A}$ и, следовательно, по лемме 2 из § 3 $\mathfrak{D}_h(t) \rightarrow \mathfrak{D}(t)$. Таким образом, Q_t является пределом обобщенных пуассоновских распределений, и поэтому класс (ii) содержится в классе (iii). Очевидно, что класс (iii) содержится в (iv).

Пусть при каждом n распределение $\mathfrak{G}^{(n)}$ безгранично делимо. По определению $G^{(n)}$ есть распределение суммы n независимых одинаково распределенных случайных величин, и таким образом последовательность $\{G^{(n)}\}$ приводит к описанной в п. (v) схеме серий. Следовательно, класс (iv) содержится в классе (v). В § 7 будет показано, что класс (v) содержится в (ii), и поэтому классы (ii) — (v) совпадают. Наконец, класс всех безгранично делимых распределений (i) является подклассом класса (iv) и содержит класс (ii). ▴

Согласно теореме 2, каждое безгранично делимое распределение оказывается распределением Q_t из соответствующей полугруппы со сверткой. Значение параметра t может быть зафиксировано произвольным образом посредством изменения масштаба на t -оси. Однако *существует только одна полугруппа $\{\mathfrak{D}(t)\}$, к которой принадлежит данное безгранично делимое распределение F .* Это означает, что представление $F = F_n^n *$ распределения F в виде n -кратной свертки *единственно*. Это утверждение правдоподобно, но требует доказательства. На самом деле единственность становится очевидной при использовании соответствующего анализа Фурье, тогда как в настоящем контексте доказательство увело бы нас в сторону от основного содержания и в то же время было бы лишено поучительности. По этой причине мы отказываемся на этот раз от доказательства указанного утверждения в рамках того и другого подходов.

Применение к случайным процессам. Пусть случайные величины $X(t)$ описывают случайный процесс со *стационарными независимыми приращениями* (гл. VI, 4). Условимся интерпретировать Q_t как распределение приращения $X(t+s) - X(s)$. Рассмотрим временной интервал $s, s+1$ единичной длины. Разделим его точками $s = s_0 < s_1 < \dots < s_n = s+1$ на подынтервалы длины n^{-1} . Тогда $P\{X(s_k) - X(s_{k-1}) > x\} = 1 - Q_{1/n}(x)$, так что $n[1 - Q_{1/n}(x)]$ равно мате-

математическому ожиданию числа интервалов s_{k-1}, s_k , приращения на которых $> x$. При $n \rightarrow \infty$ это математическое ожидание сходится к $\psi^+(x)$. Чтобы упростить дальнейшее обсуждение, предположим, что пределы $X(t+)$ и $X(t-)$ существуют при всех t и что $X(t)$ лежит между ними. Пусть s_{k-1}, s_k — тот интервал нашего разбиения, который содержит точку t . Для достаточно больших n приращение $X(s_k) - X(s_{k-1})$ будет близко к скачку $X(t+) - X(t-)$ и интуитивно ясно, что предел $\psi^+(x)$ представляет собой математическое ожидание числа моментов t (в единицу времени), для которых $X(t+) - X(t-) > x$. Это рассуждение можно сделать строгим, но мы не будем входить в детали. Из сказанного следует, что математическое ожидание числа разрывов будет нулем, только если $\psi^+(x) = 0$ и $\psi^-(x) = 0$ для всех $x > 0$. В последнем случае Ω сосредоточено в нуле, т. е. приращения $X(t+s) - X(s)$ будут нормально распределены. Для такого процесса траектории непрерывны с вероятностью единица (теорема П. Леви и Н. Винера). Итак, траектории непрерывны с вероятностью единица тогда и только тогда, когда процесс нормален.

В качестве второй иллюстрации мы рассмотрим обобщенный пуассоновский процесс (2.8). Математическое ожидание числа скачков за единицу времени равно α , а вероятность получить скачок, превосходящий $x > 0$, равна $1 - F(x)$. Таким образом, $\alpha[1 - F(x)]$ есть математическое ожидание числа скачков, которые больше x , в полном согласии с нашими интуитивными доводами.

Ба. Разрывные полугруппы

Естественно поставить вопрос о том, существуют ли разрывные полугруппы. Задача не имеет никакой практической ценности, но ответ до некоторой степени любопытен: каждая полугруппа со сверткой $\{\Omega(t)\}$ лишь центрированием отличается от некоторой непрерывной полугруппы $\{\Omega^*(t)\}$. В частности, если распределения Q_t симметричны, то полугруппа обязательно непрерывна. В общем случае существует такая функция ϕ , что распределения Q_t^* , определенные как $Q_t(x + \phi(t))$, связаны с некоторой непрерывной полугруппой. Функция ϕ необходимо должна удовлетворять уравнению

$$\phi(t+s) = \phi(t) + \phi(s). \quad (5.12)$$

Это — известное уравнение Гамеля, непрерывными решениями которого могут быть только функции вида ct (см. сноску на стр. 341). Более того, единственная бэровская функция, которая удовлетворяет (5.12), — это линейная функция. Другие решения слишком причудливы. Например, нелинейное решение принимает в любом интервале сколь угодно большие и сколь угодно малые значения. Его невозможно получить аналитически последовательными предельными переходами. Короче говоря, уместно спросить, что в точности означает его «существование».

Вернемся на землю и рассмотрим произвольную полугруппу со сверткой $\{\Omega(t)\}$ и схему серий $\{X_k, n\}$, связанную с распределениями $Q_{1/n}$. Суммы по строкам имеют одно и то же распределение Q_1 . Поэтому мы можем воспользоваться последней леммой и извлечь такую подпоследовательность $n_1, n_2, \dots$, что когда n пробегает ее, $n[\Omega(1/n) - 1] \rightarrow \mathfrak{A}^*$, где $\mathfrak{A}^*$ — производящий опис-

ратор некоторой непрерывной полугруппы $\{\Sigma^*(t)\}$. Мы можем выбрать n_k вида 2^v . Неравенство (3.2) тогда показывает, что $\Sigma(t) = \Sigma^*(t)$ при всех t , которые при каком-либо k кратны $1/n_k$, т. е. при всех t вида $t = a2^{-v}$, где a и v — целые. Таким образом, *всегда существует непрерывная полугруппа $\{\Sigma^*(t)\}$, такая, что $\Sigma(t) = \Sigma^*(t)$ при всех t из некоторого всюду плотного множества Σ .*

Мы можем теперь доказать исходное утверждение. Выберем $\varepsilon_n > 0$ так, что $t + \varepsilon_n$ лежит в Σ . Тогда

$$\Sigma^*(t + \varepsilon_n) = \Sigma(t + \varepsilon_n) = \Sigma(t) \Sigma(\varepsilon_n). \quad (5.13)$$

При $\varepsilon_n \rightarrow 0$ левая часть сходится к $\Sigma^*(t)$ и потому достаточно доказать, что если $\Sigma(\varepsilon_n) \rightarrow \mathcal{F}$, то распределение F сосредоточено в одной точке. Выберем точки h_n в Σ так, что $0 < \varepsilon_n < h_n$ и $h_n \rightarrow 0$. Тогда $\Sigma^*(h_n) = \Sigma(h_n - \varepsilon_n) \Sigma(\varepsilon_n)$. Левая часть сходится к тождественному оператору, так что F в самом деле может иметь только одну точку роста.

§ 6. ПРИМЕР: УСТОЙЧИВЫЕ ПОЛУГРУППЫ

Полугруппа $\{\Sigma(t)\}$ называется *устойчивой*, если соответствующие распределения имеют вид

$$Q_t(x) = G(\lambda_t(x - \beta_t)), \quad (6.1)$$

где $\lambda_t > 0$ и β_t — постоянные, непрерывно зависящие от t , а G — фиксированное распределение. Очевидно, что G является устойчивым распределением в смысле гл. VI, 1. Теория устойчивых полугрупп излагается здесь главным образом как иллюстрация результатов предыдущего параграфа. Кроме того, мы выпишем их производящие операторы. Результаты этого параграфа независимо выводятся (обходным путем) в § 8.

В силу предположенной непрерывности λ_t и β_t полугруппа (если она существует) будет непрерывна. При $t \rightarrow 0$ распределение Q_t стремится к распределению, сосредоточенному в нуле, и, следовательно, $\lambda_t \rightarrow \infty$ и $\beta_t \rightarrow 0$. Первое из соотношений (5.14) принимает форму

$$\frac{1 - G(\lambda_t(x - \beta_t))}{t} \rightarrow \psi^+(x), \quad x > 0. \quad (6.2)$$

Так как функция $\beta_t \rightarrow 0$ и G монотонна, то это соотношение остается верным, даже если выбросить β_t и тогда (6.2) можно переписать в виде

$$\frac{1 - G(\lambda_t x)}{1 - G(\lambda_t)} \rightarrow \psi^+(x). \quad (6.3)$$

(Здесь мы предполагаем, что 1 является точкой непрерывности для ψ^+ ; этого можно достигнуть изменением масштаба.) Теперь (6.3) является частным случаем соотношения (8.1) из гл. VIII, в котором определяется понятие правильного изменения. заключаем, что либо ψ^+ тождественно равно нулю, либо хвост $1 - G$ правильно меняется на бесконечности и

$$\psi^+(x) = c^+ x^{-\alpha}, \quad x > 0, \quad c^+ > 0. \quad (6.4)$$

Следовательно, на положительной полуоси мера Ω имеет плотность $\alpha c^+ x^{-\alpha-1}$. Приходим к заключению, что $0 < \alpha < 2$, потому что Ω приписывает конечные массы конечным окрестностям нуля и $\psi^+(x) \rightarrow 0$ при $x \rightarrow \infty$. В силу подобных же причин либо ψ^- тождественно равно нулю, либо $\psi^-(x) = c^- |x|^{-\alpha}$ при $x < 0$. Показатель α одинаков для обоих хвостов, потому что сумма правого и левого хвостов $1 - G(x) + G(-x)$ также правильно меняется.

Функции ψ^+ и ψ^- определяют меру Ω с точностью до атома в нуле. Мы увидим, что такой атом может существовать, только если обе функции ψ^+ и ψ^- тождественно равны нулю, а мера Ω сосредоточена в нуле.

Производящий оператор $\mathfrak{A}$ задан формулой (5.7). В нашем случае удобно записать его в виде

$$\mathfrak{A} = c^+ \mathfrak{A}_\alpha^+ + c^- \mathfrak{A}_\alpha^- + bd/dx, \quad (6.5)$$

где операторы $\mathfrak{A}^+$ и $\mathfrak{A}^-$ соответствуют вкладам обеих полуосей и определяются следующим образом: если $0 < \alpha < 1$, то

$$\mathfrak{A}_\alpha^+ u(x) = \int_0^\infty [u(x-y) - u(x)] y^{-\alpha-1} dy; \quad (6.6)$$

если $1 < \alpha < 2$, то

$$\mathfrak{A}_\alpha^+ u(x) = \int_0^\infty [u(x-y) - u(x) - yu'(x)] y^{-\alpha-1} dy, \quad (6.7)$$

и, наконец, если $\alpha = 1$, то

$$\mathfrak{A}_1^+ u(x) = \int_0^\infty [u(x-y) - u(x) - \tau_s(y) u'(x)] y^{-2} dy. \quad (6.8)$$

Оператор $\mathfrak{A}_\alpha^-$ определяется аналогичными интегралами на $-\infty, 0$, причем $y^{-\alpha-1}$ заменяется на $|y|^{-\alpha-1}$. Центрирование в (6.6) и (6.7) отличается от центрирования в (5.6), но эта разница поглощается членом bd/dx в общей записи производящего оператора (6.5). Атом меры Ω в нуле добавил бы член $\gamma d^2/dx^2$ к производящему оператору $\mathfrak{A}$. В примере 4, а) показано, что этот член сам порождает полугруппу нормальных распределений.

Теорема. а) При $b=0$ и $0 < \alpha < 1$ или $1 < \alpha < 2$ оператор (6.5) порождает строго устойчивую полугруппу распределений вида

$$Q_t(x) = G(xt^{-1/\alpha}); \quad (6.9)$$

при $\alpha=1$ и $b=0$ он порождает устойчивую полугруппу распределений вида

$$Q_t(x) = G(xt^{-1} - (c^+ - c^-) \log t). \quad (6.10)$$

б) Устойчивая полугруппа либо порождается оператором (6.5), либо является полугруппой нормальных распределений.

[В примере 2, б) мы видели, что bd/dx порождает полугруппу операторов сдвига, и для получения полугруппы, порожденной оператором (6.5), при $b \neq 0$ достаточно заменить в (6.9) и (6.10) x на $x + bt$.]

Доказательство. а) Изменяя масштаб, можно заменить распределение Q_t полугруппы на распределения, определенные равенством $Q_t^\#(x) = Q_t(x/\rho)$. Они образуют новую полугруппу $\{\Omega^\#(t)\}$. Если положить $v(x) = u(\rho x)$, то из определения свертки видно, что $\Omega^\#(t)u(x) = \Omega(t)v(x/\rho)$. Это означает для производящих операторов, что для нахождения $\mathfrak{U}^\#u(x)$ достаточно просто вычислить $\mathfrak{U}v(x)$ и заменить x на x/ρ . Подстановка $y = z/\rho$ в (6.7) и (6.8) показывает, что для соответствующих производящих операторов $\mathfrak{U}_\alpha^\# = \rho^\alpha \mathfrak{U}_\alpha$. Оператор $\rho^\alpha \mathfrak{U}_\alpha$, очевидно, порождает полугруппу $\{\Omega(\rho^\alpha t)\}$, и, таким образом, из единственности производящих операторов мы заключаем, что $Q_t(x/\rho) = Q_{t\rho^\alpha}(x)$. Если положить $G = Q_1$ и $\rho = t^{-1/\alpha}$, то получим (6.9).

Подобные же рассуждения можно применить в случае $\alpha = 1$ с той разницей, что при использовании подстановки $y = z/\rho$ в (6.8) центрирующая функция дает дополнительный член вида $(c^+ - c^-) \times (\rho \log \rho) u'(x)$, и это приводит к (6.10).

б) Мере Ω , сосредоточенной в нуле, соответствует нормальная полугруппа. Мы видели, что производящий оператор любой другой устойчивой полугруппы имеет вид $\mathfrak{U}_\alpha + \gamma d^2/dx^2$. Как показано в примере 2, в), распределения соответствующей полугруппы будут свертками нашего устойчивого распределения Q_t и нормальных распределений с дисперсией $2\gamma t$, и ясно, что такая полугруппа не может быть устойчивой. ►

В доказанной теореме утверждается, что функция λ_t в определении устойчивой полугруппы имеет вид $\lambda_t = t^{-1/\alpha}$. Рассматривая (6.2) и его аналог для отрицательной полуоси, получаем

Следствие. Если $0 < \alpha < 2$, $c^+ \geq 0$, $c^- \geq 0$ (но $c^+ + c^- > 0$), то существует ровно одна устойчивая функция распределения G , такая, что при $x \rightarrow \infty$

$$x^\alpha [1 - G(x)] \rightarrow c^+, \quad x^\alpha G(-x) \rightarrow c^-. \quad (6.11)$$

За пределами этого остается единственное устойчивое распределение — нормальное распределение [оно удовлетворяет условию (6.11) при $\alpha = 2$ и $c^+ = c^- = 0$].

Утверждение в квадратных скобках будет доказано в § 8.

§ 7. СХЕМЫ СЕРИЙ С ОДИНАКОВЫМИ РАСПРЕДЕЛЕНИЯМИ

Пусть при каждом n величины $X_{1,n}, \dots, X_{n,n}$ взаимно независимы и имеют одно и то же распределение F_n . Нас интересуют возможные предельные распределения суммы $S_n = X_{1,n} + \dots + X_{n,n}$, но полезно начать с анализа необходимого условия существования предельного распределения, которое состоит в том, что последовательность $\{S_n\}$ *стохастически ограничена*. Напомним определение гл. VIII,2 по которому последовательность $\{S_n\}$ называется стохастически ограниченной, если каждому $\varepsilon > 0$ соответствует такое a , что $P\{|S_n| > a\} < \varepsilon$ при всех n . Грубо говоря, это означает, что никакая вероятностная масса не «утекает» в бесконечность. Очевидно, это условие является необходимым для существования собственного предельного распределения.

Обратимся за помощью к операции урезания. Очень удобно еще раз использовать урезающую функцию τ_s , введенную в (5.3), чтобы избежать разрывов; τ_s — непрерывная монотонная функция, равная $\tau_s(x) = x$ при $|x| \leq s$ и $\tau_s(x) = \pm s$ при $|x| \geq s$. Положим при таком урезании

$$X'_{k,n} = \tau_s(X_{k,n}), \quad X_{k,n} = X'_{k,n} + X''_{k,n}. \quad (7.1)$$

Эти новые величины зависят от параметра s , хотя в наших обозначениях это не подчеркивается. Суммы по строкам в каждой из схем серий $\{X'_{k,n}\}$ и $\{X''_{k,n}\}$ обозначим через S'_n и S''_n соответственно, так что $S_n = S'_n + S''_n$. Величины $X'_{k,n}$ ограничены; их математическое ожидание обозначим

$$\beta_n = E(X'_{k,n}) \quad (7.2)$$

(разумеется, β_n не зависит от k). Наконец, введем, как аналог мер Ω_t из § 5, меру Φ_n , определенную формулой

$$\Phi_n\{dx\} = nx^2 F_n\{dx\}. \quad (7.3)$$

Значения $\Phi_n(I)$ конечны на конечных интервалах I , значение Φ_n на всей числовой прямой может быть бесконечно.

Весьма правдоподобно, что $\{S_n\}$ не может оставаться стохастически ограниченной без того, чтобы отдельные компоненты S_n были малы в том смысле, что для всякого $\varepsilon > 0$

$$P\{|X_{k,n}| > \varepsilon\} \rightarrow 0, \quad n \rightarrow \infty \quad (7.4)$$

(левая часть не зависит от k). Схемы серий, обладающие свойством (7.4), называются *нулевыми схемами*. Мы увидим, что только в нулевых схемах суммы по строкам стохастически ограничены, но пока что введем (7.4) как исходное предположение.

«Необходимость» условия следующей леммы используется в доказательстве теоремы 1 § 5: условие (7.4) выполняется, так как рассматриваемая там полугруппа непрерывна.

Лемма. (Компактность). Для того чтобы суммы по строкам S_n нулевой схемы серий $\{X_{k,n}\}$ оставались стохастически ограниченными, необходимо и достаточно, чтобы

- (i) $\Phi_n(I)$ было ограничено для каждого конечного интервала I и
(ii) для больших x хвосты

$$T_n(x) = n[1 - F_n(x) + F_n(-x)] \quad (7.5)$$

были равномерно малы.

Другими словами, требуется, чтобы каждому $\varepsilon > 0$ соответствовало такое t , что $T_n(x) < \varepsilon$ при $x > t$. (Заметим, что T_n — убывающая функция.)

Доказательство. В частном случае симметричных распределений F_n необходимость условия (ii) вытекает из неравенства

$$P\{|S_n| > a\} \geq \frac{1}{2}(1 - \exp(-T_n(a))) \quad (7.6)$$

(см. (5.10) гл. V). Для произвольного распределения F_n применим уже знакомую нам симметризацию. Вместе с S_n симметризованные величины ${}^0S_n$ будут стохастически ограниченными, и, таким образом, условие (ii) применимо к хвостам ${}^0T_n$ симметризованных распределений. Но в случае нулевой схемы ясно, что для каждого $\delta > 0$ в конечном счете ${}^0T_n(a) \geq \frac{1}{2}T_n(a + \delta)$ и, следовательно, условие (ii) необходимо во всех случаях.

В предположении, что условие (ii) выполнено, можно выбрать точку урезания s такой большой, что $T_n(s) < 1$ при всех n . Тогда число отличных от нуля членов среди $X_{1,n}^*, \dots, X_{n,n}^*$ является биномиально распределенной случайной величиной с математическим ожиданием и дисперсией, меньшими 1. Следовательно, можно так подобрать числа N и c , что с вероятностью, сколь угодно близкой к единице, среди величин $X_{k,n}^*$ менее чем N будут отличны от 0, и все они будут $\leq c$. Это означает, что суммы S_n^* остаются стохастически ограниченными, и при этих обстоятельствах $\{S_n^*\}$ стохастически ограничена тогда и только тогда, когда стохастически ограничена $\{S_n'\}$.

Остается доказать, что условие (i) необходимо и достаточно для стохастической ограниченности последовательности $\{S_n'\}$.

Положим $\sigma_n^2 = \text{Var}(S_n)$. Если $\sigma_n \rightarrow \infty$, то применение к величинам $(X_{k,n} - \beta_n)/\sigma_n$ центральной предельной теоремы из примера 1, а) показывает, что для большого n распределение сумм S_n' будет приближенно нормальным с дисперсией $\sigma_n^2 \rightarrow \infty$ и, следовательно, $P\{S_n' \in I\} \rightarrow 0$ для любого конечного интервала I . Те же самые доводы применимы к последовательностям и доказывают, что $\{S_n'\}$ стохастически ограничена, если только $\text{Var}(S_n')$ остается стохастически ограниченной. Но в этом случае с помощью неравенства Чебышева можно показать, что $\{S_n' - n\beta_n\}$ стохастически ограни-

чена и что для S'_n то же самое верно в том и только том случае, когда $\{n\beta_n\}$ ограничено. Так как $\{X_{k,n}\}$ является нулевой схемой серий, то $\beta_n \rightarrow 0$ и, таким образом, из ограниченности $n\beta_n$ следует, что $\text{Var}(S'_n) \sim E(S_n'^2)$.

Таким образом, мы показали, что ограниченность $E(S_n'^2)$ является необходимым условием стохастической ограниченности $\{S'_n\}$, а в силу неравенства Чебышева это условие является и достаточным. Но

$$E(S_n'^2) = \Phi_n \{ -s, s \} + s^2 T_n(s), \quad (7.7)$$

и, следовательно, при этих обстоятельствах условие (i) эквивалентно тому, что $E(S_n'^2)$ ограничено. ►

Предположение о том, что $\{X_{k,n}\}$ является нулевой схемой серий, использовалось лишь в связи с симметризацией и может быть опущено для схем серий с симметричными распределениями F_n . Однако из ограниченности $\Phi_n(I)$ следует, что $E(X_{k,n}^2) = O(n^{-1})$, и легко заключить из леммы, что схема серий со стохастически ограниченными суммами по строкам и симметричными распределениями необходимо является нулевой схемой серий. В силу симметризации отсюда следует, что в общем случае существуют такие числа μ_n (например, медианы случайных величин $X_{k,n}$), что $\{X_{k,n} - \mu_n\}$ является нулевой схемой. Другими словами, при подходящем центрировании получается нулевая схема серий, и в этом смысле только нулевые схемы представляют интерес.

Пример. Пусть $X_{k,n}$ имеют нормальное распределение с математическим ожиданием β_n и дисперсией n^{-1} . Тогда $S_n - n\beta_n$ имеет стандартное нормальное распределение, но, поскольку β_n произвольны, $\{S_n\}$ не обязательно стохастически ограничена. Это иллюстрирует важность центрирования. ►

Теоретически возможно такое центрирование схемы серий, что $\beta_n = 0$ для всех n , но соответствующий такому центрированию критерий было бы трудно применять в конкретных ситуациях. При произвольном центрировании критерии включают нелинейные члены и становятся громоздкими. Мы изучим этот случай во всей его общности в гл. XVII, 7. А здесь мы примем компромиссное решение: потребуем только, чтобы

$$\beta_n^2 = o(E(X_{k,n}^2)), \quad n \rightarrow \infty. \quad (7.8)$$

По-видимому, это условие выполняется во всех ситуациях, встречающихся на практике. Во всяком случае, оно настолько умеренно, что его легко удовлетворить с помощью подходящего центрирования, в то время как более ограничительное условие $\beta_n = 0$ потребовало бы сложных вычислений,

Теорема. Пусть $\{X_{k,n}\}$ — нулевая схема серий, удовлетворяющая условию (7.8).

Для существования постоянных b_n , таких, что распределение $S_n - b_n$ сходится к собственному предельному распределению, необходимо и достаточно, чтобы существовала такая каноническая мера Ω , что

$$\Phi_n \{I\} \rightarrow \Omega \{I\} \quad (7.9)$$

для каждого конечного интервала и при $x > 0$

$$n[1 - F_n(x)] \rightarrow \psi^+(x), \quad nF_n(-x) \rightarrow \psi^-(-x). \quad (7.10)$$

Если эти условия выполнены, то распределение $S_n - n\beta_n$ сходится к распределению Q_1 , соответствующему полугруппе со сверткой, порожденной оператором $\mathfrak{A}$, который определяется равенством

$$\mathfrak{A}u(x) = \int_{-\infty}^{+\infty} \frac{u(x-y) - u(x) + \tau_s(y) u'(x)}{y^2} \Omega \{dy\}. \quad (7.11)$$

Доказательство. Прежде всего заметим, что при произвольном b_n условия (i) и (ii) леммы необходимы для того, чтобы последовательность $\{S_n - b_n\}$ была стохастически ограниченной. Доказательство сохраняется с тем упрощением, что соотношение $E(S_n^2) \sim \sim \text{Var}(S_n^2)$ теперь является следствием (7.8), тогда как раньше нам приходилось его выводить из ограниченности S_n' .

Теперь предположим, что условия леммы удовлетворены. В силу теоремы о выборе существует такая последовательность $\{n_k\}$, что когда n пробегает ее, то (7.9) выполняется для конечных интервалов. Для конечного интервала $0 < a < x \leq b$ имеем

$$n[F_n(b) - F_n(a)] = \int_a^b y^{-2} \Phi_n \{dy\}, \quad (7.12)$$

а из (7.9) следует, что величина (7.12) тоже стремится к пределу. Условие (ii) утверждает, что $n[1 - F_n(b)]$ будет меньше произвольного числа $\varepsilon > 0$, если только b достаточно велико. Отсюда следует, что при $0 < a < b \leq \infty$ интегралы в (7.12) сходятся к аналогичным интегралам относительно Ω . Таким образом, Ω является канонической мерой и (7.10) справедливо, когда n пробегает последовательность $\{n_k\}$.

Нам известно, что оператор $\mathfrak{A}$ из (7.11) определяет полугруппу операторов со сверткой $\{\Omega(t)\}$. Пусть Ω_n — оператор, индуцированный распределением G_n величин $X_{k,n} - \beta_n$, т. е. $G_n(x) = F_n(x + \beta_n)$. Как известно из § 1, для того чтобы доказать, что распределение разности $S_{n_k} - n_k \beta_{n_k}$ стремится к распределению Q_1 , связанному с оператором $\Omega(1)$, достаточно показать, что

$$n[\Omega_n - 1] \rightarrow \mathfrak{A}, \quad (7.13)$$

когда n пробегает последовательность $\{n_k\}$. Теперь

$$n[\mathfrak{D}_n - 1]u(x) = n \int_{-\infty}^{+\infty} [u(x + \beta_n - y) - u(x - y)] F_n\{dy\}. \quad (7.14)$$

Мы выражаем $u(x + \beta_n - y)$, используя формулу Тейлора вплоть до члена второй степени. Поскольку $\beta_n \rightarrow 0$, из (7.8) и ограниченности $\Phi_n\{I\}$ следует, что $n\beta_n^2 \rightarrow 0$. Так как к тому же u'' ограничено, то

$$n[\mathfrak{D}_n - 1]u(x) = \int_{-\infty}^{+\infty} [u(x - y) - u(x) + \tau_s(y) u'(x)] nF_n\{dy\} + \varepsilon_n(x), \quad (7.15)$$

где ε_n равномерно стремится к нулю. Интеграл можно переписать в виде (7.11), но с той разницей, что интегрирование будет производиться относительно Φ_n (вместо Ω). Как мы неоднократно доказывали, из предельных соотношений (7.9) и (7.10) следует, что интеграл в (7.15) сходится к интегралу в (7.11) и, таким образом, (7.13) выполняется, когда n пробегает последовательность $\{n_k\}$. Наконец, из единственности полугруппы, содержащей Q_1 , следует, что наши предельные соотношения должны выполняться при произвольном стремлении $n \rightarrow \infty$, что завершает доказательство. ►

§ 8. ОБЛАСТИ ПРИТЯЖЕНИЯ

В этом параграфе $X_1, X_2, \dots$ предполагаются независимыми случайными величинами с одним и тем же распределением F . В соответствии с определением 2 гл. VI,1, распределение F принадлежит области притяжения G , если существуют постоянные $a_n > 0$ и b_n , такие, что распределение $\frac{1}{a_n}(X_1 + \dots + X_n) - b_n$ сходится к G , где G — собственное распределение, не сосредоточенное в одной точке. Несмотря на уже известные предварительные результаты гл. VI,1 и § 6, мы будем строить здесь теорию с самого начала. (В гл. XVII,5 эта теория будет изложена независимым образом и с большими деталями.)

Всюду в этом параграфе мы будем использовать обозначение

$$U(x) = \int_{-x}^x y^2 F\{dy\}, \quad x > 0. \quad (8.1)$$

Напомним, что в соответствии с теорией правильного изменения функций из гл. VIII,8, определенная на $\overline{0, \infty}$ положительная функция L называется *медленно меняющейся* (на ∞), если при

всех $x > 0$

$$\frac{L(sx)}{L(s)} \rightarrow 1, \quad s \rightarrow \infty. \quad (8.2)$$

Теорема 1. *Распределение F принадлежит области притяжения некоторого распределения G тогда и только тогда, когда существует медленно меняющаяся функция L , такая, что*

$$U(x) \sim x^{2-\alpha} L(x), \quad x \rightarrow \infty, \quad (8.3)$$

где $0 < \alpha \leq 2$, и при $\alpha < 2$

$$\frac{1 - F(x)}{1 - F(x) + F(-x)} \rightarrow p, \quad \frac{F(-x)}{1 - F(x) + F(-x)} \rightarrow q. \quad (8.4)$$

При $\alpha = 2$ достаточно одного условия (8.3), если только F не сосредоточена в одной точке¹⁾.

Мы увидим, что (8.3) при $\alpha = 2$ приводит к сходимости к нормальному распределению. Этот случай охватывает распределения с конечной дисперсией, а также многие распределения с неограниченной медленно меняющейся функцией U [см. пример 4, а) из гл. VIII].

По теореме 2 гл. VIII,9, при $\zeta = 2$ и $\eta = 0$ соотношение (8.3) эквивалентно²⁾

$$\frac{x^2 [1 - F(x) + F(-x)]}{U(x)} \rightarrow \frac{2 - \alpha}{\alpha} \quad (8.5)$$

в том смысле, что каждое из этих соотношений влечет за собой другое.

При $0 < \alpha < 2$ мы можем переписать (8.5) в виде

$$1 - F(x) + F(-x) \sim \frac{2 - \alpha}{\alpha} x^{-\alpha} L(x), \quad (8.6)$$

и из (8.6) в свою очередь следуют (8.3) и (8.5). Мы приходим к новой формулировке теоремы, интуитивно более понятной, так как она описывает поведение отдельных хвостов (другие варианты см. в задаче 17).

¹⁾ Для распределений с конечной дисперсией U является медленно меняющейся функцией, кроме того случая, когда F сосредоточено в нуле. Во всех других случаях (8.3) и (8.4) остаются неизменными, если $F(x)$ заменить на $F(x+b)$.

²⁾ Условие (8.4) налагает аналогичные ограничения на каждый хвост по отдельности:

$$\frac{x^2 [1 - F(x)]}{U(x)} \rightarrow p \frac{2 - \alpha}{\alpha}, \quad \frac{x^2 F(-x)}{U(x)} \rightarrow q \frac{2 - \alpha}{\alpha}. \quad (*)$$

При $\alpha = 2$ эти соотношения вытекают из (8.5), чем и объясняется отсутствие второго условия при $\alpha = 2$. Теорема 1 могла бы быть сформулирована более сжато (но более искусственно) следующим образом: *F принадлежит некоторой области притяжения тогда и только тогда, когда выполняется (*) и $0 < \alpha \leq 2$, $p \geq 0$, $q \geq 0$, $p + q = 1$.*

Теорема 1а. (Другая форма.) (i) Распределение F принадлежит области притяжения нормального распределения тогда и только тогда, когда U медленно меняется.

(ii) Оно принадлежит области притяжения, отличной от нормальной, тогда и только тогда, когда для некоторого $0 < \alpha < 2$ выполняются (8.4) и (8.6).

Доказательство. Мы применим теорему § 7 к схеме серий, образованной величинами $X_{k,n} = X_k/a_n$ с распределениями $F_n(x) = F(a_n x)$. Суммы по строкам для схемы $\{X_{k,n}\}$ равны

$$S_n = \frac{X_1 + \dots + X_n}{a_n}.$$

С очевидностью $a_n \rightarrow \infty$, и, следовательно, $\{X_{k,n}\}$ является нулевой схемой серий. Чтобы показать, что условие (7.8) выполнено, положим

$$v(x) = \int_{-x}^x y F(dy) \quad (8.7)$$

и заметим, что (7.8) наверняка выполняется, если

$$v^2(x) = o(U(x)). \quad (8.8)$$

Тогда, если $U(x) \rightarrow \infty$ при $x \rightarrow \infty$, ясно, что $v(x) = o(xU(x))$ и $U(x) = o(x^{-2})$, и, следовательно, (8.8) выполнено. Случай ограниченной функции U не представляет интереса, поскольку мы знаем, что центральная предельная теорема применима к величинам с конечными дисперсиями. Однако соотношение (8.8) выполняется, даже если функция U ограничена, при условии, что распределение F имеет нулевое математическое ожидание.

Условие (i) последней теоремы требует¹⁾, чтобы при $x > 0$

$$na_n^{-2}U(a_n x) \rightarrow \Omega\{-x, x\}, \quad n \rightarrow \infty, \quad (8.9)$$

в то время как условие (ii) сводится к

$$n[1 - F(a_n x)] \rightarrow \psi^+(x), \quad nF(-a_n x) \rightarrow \psi^-(x) \quad (8.10)$$

[обозначения см. в (5.2)]. Легко видеть²⁾, что $a_{n+1}/a_n \rightarrow 1$.

Таким образом, в соответствии с леммой 3 из гл. VIII, 8. из (8.9) следует, что U правильно меняется, и предел справа в (8.9) пропорционален степени x . Следуя обычаю, установленному П. Леви, обозначим показатель степени $2 - \alpha$. Таким образом,

$$\Omega\{-x, x\} = Cx^{2-\alpha}, \quad x > 0. \quad (8.11)$$

¹⁾ Как обычно, молчаливо предполагается, что сходимость нужна лишь в точках непрерывности.

²⁾ Для симметричных распределений это следует из того факта, что $(X_1 + \dots + X_n)/a_n$ и $(X_1 + \dots + X_n)/a_{n+1}$ имеют одно и то же предельное распределение. Для произвольных F утверждение становится верным после симметризации.

Поскольку левая часть является неубывающей функцией от x и ограничена в окрестности нуля, то $0 < \alpha \leq 2$. Отсюда следует, что U в самом деле имеет вид, указанный в формуле (8.3).

Кроме того, та же лемма 3 из гл. VIII, 8, утверждает, что пределы в (8.10) либо тождественно равны нулю, либо пропорциональны степени x . Тогда из (8.5) видно, что единственно возможной степенью является $x^{-\alpha}$; действительно, при $\alpha = 2$ оба предела — тождественные нули в то время, как при $\alpha < 2$ пределы обязательно имеют вид $Ax^{-\alpha}$ и $Bx^{-\alpha}$, где $A \geq 0$ и $B \geq 0$, но $A + B > 0$. Отсюда следует, что условия теоремы являются необходимыми.

Если предположить, что (8.3) выполняется, то возможно построить такую последовательность $\{a_n\}$, что

$$na_n^{-2}U(a_n) \rightarrow 1. \quad (8.12)$$

Например, в качестве a_n можно взять нижнюю грань всех t , для которых $nt^{-2}U(t) \leq 1$. Тогда (8.3) гарантирует, что при $x > 0$

$$na_n^{-2}U(a_n x) \rightarrow x^{2-\alpha}. \quad (8.13)$$

Таким образом, условие (7.9) удовлетворяется для интервалов вида $I = \{-x, x\}$. В случае $\alpha = 2$ предельная мера Ω сосредоточена в нуле и, следовательно, (7.9) автоматически удовлетворяется для всех конечных интервалов; в этом случае из (8.5) следует, что выполняется также и условие (7.10) при тождественно равных нулю ψ^+ и ψ^- . Если $\alpha < 2$, то из соотношений (8.3) — (8.5) вытекает, что при $x \rightarrow \infty$

$$1 - F(x) \sim p \frac{2-\alpha}{\alpha} x^{-\alpha} L(x), \quad F(-x) \sim q \frac{2-\alpha}{\alpha} x^{-\alpha} L(x),$$

если только $p > 0$ и $q > 0$ (в противном случае символ $\sim$ надо заменить «о» малым, а это — несущественное изменение). Следовательно, условие (7.10) выполнено, а отсюда в свою очередь вытекает, что (7.9) применимо к произвольным интервалам, находящимся на положительном расстоянии от нуля. ▀

Замечательно, что все результаты § 7 содержатся в настоящей теореме и ее доказательстве. Это доказательство дает также другую важную информацию. Прежде всего имеет место очевидное

Следствие. Если $\alpha = 2$, то предельное распределение является нормальным, а в противном случае оно является устойчивым распределением, удовлетворяющим условию (6.11). В любом случае оно определено с точностью до произвольных масштабных параметров.

Мы видели также, что (8.12) приводит к возможной последовательности нормирующих постоянных a_n . Легко видеть, что другую последовательность $\{a'_n\}$ можно взять в том и только том случае, если отношения a'_n/a_n стремятся к положительному пределу.

В условиях теоремы 1 мы установили существование предельного распределения для $S_n - n\beta_n$, где $[v]$ определено в (8.7)]

$$\beta_n = E(X_{k,n}) = a_n^{-1}v(sa_n) + s[1 - F(sa_n) - F(-sa_n)]. \quad (8.15)$$

Теперь мы приступаем к доказательству того приятного факта, что центрирующие постоянные β_n несущественны, за исключением того случая, когда $\alpha = 1$.

Если $\alpha < 1$, то мы применяем теорему 2 из гл. VIII, 9, при $\xi = 1$ и $\eta = 0$ отдельно к каждой из двух полуосей и находим, что при $x \rightarrow \infty$

$$v(x) \sim \frac{\alpha}{1-\alpha} x[1 - F(x) - F(-x)]. \quad (8.16)$$

Отсюда и из (8.10) следует, что $n\beta_n$ стремится к конечному пределу и, следовательно, играет несущественную роль.

При $\alpha > 1$ та же самая теорема 2 из гл. VIII, 9, с $\xi = 2$ и $\eta = 1$ показывает, что F имеет математическое ожидание, и мы естественно считаем математическое ожидание нулевым. Тогда область интегрирования в интеграле (8.7) можно заменить на $|y| > x$, и можно установить, что (8.16) остается без изменения. Таким образом, распределения сумм S_n стремятся к пределу, который снова имеет нулевое математическое ожидание. Аналогично, если $\alpha < 1$, то предельное распределение центрируется так, чтобы быть строго устойчивым. Таким образом, мы доказали теорему

Теорема 2. *Предположим, что F удовлетворяет условиям теоремы 1. Если $\alpha < 1$, то $F^{n*}(a_n x) \rightarrow G(x)$, где G — строго устойчивое распределение, удовлетворяющее (6.11). Если $\alpha > 1$, то же самое верно при условии, что F имеет нулевое математическое ожидание.*

[О центрировании при $\alpha = 1$ см. гл. XVIII, 5. Относительно моментов F см. задачу 16.]

§ 9. РАЗЛИЧНЫЕ РАСПРЕДЕЛЕНИЯ.

ТЕОРЕМА О ТРЕХ РЯДАХ

На короткое время мы перейдем к общим схемам серий $\{X_{k,n}\}$, в которых случайные величины¹⁾ $X_{1,n}, \dots, X_{n,n}$, входящие в n -ю строку, взаимно независимы, но могут иметь произвольные распределения $F_{k,n}$. Для того чтобы сохранить характер наших предельных теорем, мы рассмотрим только нулевые серии: потребуем, чтобы для любых $\eta > 0$ и $\varepsilon > 0$ и всех достаточно больших n

$$P\{|X_{k,n}| > \eta\} < \varepsilon, \quad k = 1, \dots, n. \quad (9.1)$$

Теория, развитая в § 7, переносится на этот случай с единственным изменением, что выражения типа $n \text{Var}(X'_{1,n})$ заменяются соответствующими суммами. В частности, в качестве предельных

¹⁾ Относительно числа слагаемых в n -й строке см. задачу 10 в гл. VII, 13.

распределений сумм по строкам нулевых схем серий появляются только безгранично делимые распределения. Проверка требует стандартных приемов и может быть предоставлена читателю.

Обсудим несколько интересных частных случаев. Мы сохраним обозначения § 7, но в дальнейшем не имеет значения, какой тип урезания мы использовали; быть может, проще всего определить урезанные величины так: $X'_{k,n} = X_{k,n}$ при $|X_{k,n}| < s$ и $X'_{k,n} = 0$ в противном случае. Уровень урезания s здесь произвольный.

Первая теорема является вариантом леммы о компактности и эквивалентна ей.

Теорема 1. (Закон больших чисел.) Пусть S_n обозначает сумму по строкам нулевой схемы серий. Для того чтобы существовали такие постоянные b_n , что¹⁾ $S_n - b_n \xrightarrow{P} 0$, необходимо и достаточно, чтобы для каждого $\eta > 0$ и каждого уровня урезания s выполнялись соотношения

$$\sum_{k=1}^n P\{|X_{k,n}| > \eta\} \rightarrow 0, \quad \sum_{k=1}^n \text{Var}(X'_{k,n}) \rightarrow 0. \quad (9.2)$$

В этом случае можно взять $b_n = \sum_k E(X'_{k,n})$.

В качестве приложения докажем следующую теорему, которая уже обсуждалась в гл. VIII, 5.

Теорема 2. (Бесконечные свертки.) Пусть $Y_1, Y_2, \dots$ — независимые случайные величины с распределениями $G_1, G_2, \dots$. Для того чтобы распределения $G_1 \star G_2 \star \dots \star G_n$ сумм $T_n = Y_1 + \dots + Y_n$ сходились к собственному предельному распределению G , необходимо и достаточно, чтобы при каждом $s > 0$

$$\sum P\{|Y_k| > s\} < \infty, \quad \sum \text{Var}(Y'_k) < \infty \quad (9.3)$$

и

$$\sum_{k=1}^n E(Y'_k) \rightarrow m. \quad (9.4)$$

Доказательство. Для данной возрастающей последовательности $v_1, v_2, \dots$ целых чисел и $k = 1, \dots, n$ положим $X_{k,n} = Y_{v_n+k}$. Распределения $G_1 \star \dots \star G_n$ сходятся тогда и только тогда, когда все полученные таким образом схемы серий подчиняются закону больших чисел с центрирующими постоянными $b_n = 0$. Из теоремы 1 ясно видно, что для этого необходимы и достаточны условия (9.3) и (9.4) ►

Теореме 2 можно придать следующую более яркую формулировку.

¹⁾ Напомним из гл. VIII, 2, что последовательность Z_n сходится по вероятности к 0, если $P\{|Z_n| > \varepsilon\} \rightarrow 0$ для всякого $\varepsilon > 0$.

Теорема 3. (Теорема Колмогорова о трех рядах.) Ряд $\sum Y_k$ сходится с вероятностью единица, если выполняются соотношения (9.3) и (9.4), и сходится с вероятностью нуль в противном случае.

Доказательство. Допустим, что имеют место (9.3) и (9.4). По теореме 2 гл. VII, 8, второе из условий (9.3) гарантирует сходимость с вероятностью единица ряда $\sum [Y'_k - E(Y'_k)]$, и тогда из (9.4) следует то же самое для ряда $\sum Y'_k$. По лемме Бореля — Кантелли (см. 1; гл. VIII, 3) из первого условия (9.3) следует, что с вероятностью единица лишь конечное число среди величин Y_k отличается от Y'_k . Поэтому ряд $\sum Y_k$ сходится с вероятностью единица.

Для доказательства необходимости наших условий напомним (см. гл. IV, 6), что вероятность сходимости равна или нулю, или единице. В противном случае распределения частных сумм должны сходиться, и поэтому будут выполнены (9.3) и (9.4). ►

Процессы с нестационарными приращениями

Развитая в этой главе теория полугрупп представляет собой орудие, особенно пригодное для изучения процессов со стационарными независимыми приращениями. При отказе от условия стационарности распределения приращений $X(t) - X(\tau)$ будут зависеть от двух параметров t и τ , и нам приходится иметь дело с двухпараметрическим семейством операторов $\Omega(\tau, t)$, $0 < \tau < t$, удовлетворяющих уравнению свертки

$$\Omega(\tau, s) \Omega(s, t) = \Omega(\tau, t), \quad \tau < s < t. \quad (9.5)$$

Будут ли распределения, связанные с такими операторами, безгранично делимы? Мы можем разбить t на n интервалов t_{k-1}, t_k и рассмотреть случайные величины $X(t_k) - X(t_{k-1})$. Однако для применения теории, развитой для схем серий, мы должны принять ограничение (9.1), сводящееся к требованию равномерной непрерывности рассматриваемых распределений по двум временным параметрам. Но $\Omega(\tau, t)$ не обязано зависеть от t непрерывно. В самом деле, частные суммы последовательности $X_1, X_2, \dots$ независимых случайных величин образуют процесс с независимыми приращениями, в котором все изменения происходят только в целочисленные моменты времени. Этот процесс разрывен. Однако в некотором смысле это есть единственный тип существенно разрывного процесса. Термин «существенно» необходим. В самом деле, мы показали в § 5а, что даже в обычных полугруппах искусственное центрирование может породить беспорядок. Он не имеет особенных последствий, но требует осторожности в формулировках. Для простоты мы ограничимся симметричными распределениями и докажем следующую лемму.

Лемма. Если распределения, связанные с $\Omega(\tau, t)$, симметричны, то при каждом t существует односторонний предел $\Omega(\tau, t-)$.

Доказательство. Пусть $\tau < t_1 < t_2 < \dots$ и $t_n \rightarrow t$. Последовательность распределений, связанных с $\Omega(\tau, t_n)$, стохастически ограничена, так что она содержит сходящуюся подпоследовательность. Чтобы избежать двойных индексов, допустим, что $\Omega(\tau, t_n) \rightarrow \mathcal{U}$, где $\mathcal{U}$ соответствует собственному распределению U . Легко установить, что $\Omega(t_n, t_{n+1}) \rightarrow 1$ и потому $\Omega(t_n, s_n) \rightarrow 1$ для любой последовательности моментов времени, такой, что $t_n < s_n < t_{n+1}$. В силу (9.5) это означает, что $\Omega(\tau, s_n) \rightarrow \mathcal{U}$, поэтому предел $\mathcal{U}$ не зависит от выбора последовательности $\{t_n\}$. Лемма доказана. ►

Следуя П. Леви, мы назовем момент t *фиксированным разрывом*, если пределы $\Omega(\tau, t+)$ и $\Omega(\tau, t-)$ различны. Из теоремы 2 сразу следует, что множество фиксированных разрывов счетно. Используя симметризацию, можно показать, что в общем случае разрывы обусловлены только центрированием (исключая не более чем счетное множество моментов времени) и что их можно устранить подходящим центрированием. Вклад $\Omega_{\alpha}(\tau, t)$ всех фиксированных разрывов в $\Omega(\tau, t)$ представляет собой бесконечную свертку. Поэтому процесс можно разложить на дискретную и непрерывную части. Для схем серий, возникающих из непрерывных процессов, нетрудно показать (используя теорему 2), что условие равномерной малости (9.1) автоматически выполнено. Мы приходим к заключению, что *распределения, связанные с непрерывными процессами, безгранично делимы*. П. Леви показал, что выборочные функции таких процессов «достаточно хороши» в том смысле, что с вероятностью единица они имеют при каждом t правые и левые предельные значения.

§ 10. ЗАДАЧИ

1. Покажите, что в примере 1, а) $\sum X_{k,n}^2 \xrightarrow{P} 1$ при $n \rightarrow \infty$. Указание. Используйте дисперсии.

2. Пусть T — момент первого прохождения через точку $+1$ в обычном симметричном случайном блуждании. Другими словами, T — такая случайная величина, что

$$P\{T=2r-1\} = \frac{1}{2r-1} \binom{2r-1}{r} 2^{-2r+1}.$$

Рассмотрите схему серий, в которой $X_{k,n}$ распределено как T/n^2 . Используя элементарные методы § 1, покажите прямым вычислением, что

$$n \{F_{n^2} u(x) - u(x)\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \frac{u(x-y) - u(x)}{\sqrt{y^3}} dy, \quad (*)$$

Установите, что распределения сумм по строкам сходятся к устойчивому распределению F_1 , определенному в гл. II, (4.7), и обладающему свойством (4.9) гл. II. Дайте интерпретацию полученного результата в терминах случайного блуждания, в котором скачки равны $\pm 1/n$, а промежутки между последовательными скачками равны $1/n^2$. Указание: Ряд, определяющий $F_{n^2} u(x)$, можно приблизить интегралом.

3. Рассмотрим гамма-распределения (2.2) из гл. II при $\alpha=1$. Покажите, что из свойства свертки (2.3) гл. II следует, что они образуют полугруппу с производящим оператором вида

$$\mathfrak{A}u(x) = \int_0^{\infty} \frac{u(x-y) - u(x)}{y} e^{-y} dy.$$

Обсудите отсутствие центрирующего члена.

4. Односторонние устойчивые распределения (4.7) из гл. II удовлетворяют свойству свертки (4.9) из гл. II и, следовательно, образуют полугруппу. Покажите, что производящий оператор задается правой частью соотношения (*) из задачи 2.

5. Пусть распределения Q_k некоторой полугруппы сосредоточены на множестве целых чисел. Обозначим вес k через $q_k(t)$. Покажите, что

$$\mathfrak{A}u(x) = -q'_0(0)u(x) + \sum_{k \neq 0} q'_k(0)u(x-k).$$

Сравните с канонической формой (5.9). Рассмотрите в связи с этим производящие функции безгранично делимых распределений, найденные в I; гл. XII, § 3.

6. Распространите понятия § 2 на полугруппы с несобственными распределениями. Покажите, что если $\mathfrak{U}$ порождает $\{\Omega(t)\}$, то $\mathfrak{U}-c1$ порождает $\{e^{-ct}\Omega(t)\}$.

7. Обозначение $e^{t(\mathfrak{F}-1)}$ для обобщенной пуассоновской полугруппы наталкивает на мысль написать в общем случае $\Omega(t) = e^{t\mathfrak{U}}$. Для нормальной полугруппы это приводит к формальному операторному соотношению

$$\exp\left(\frac{1}{2}t\frac{d^2}{dx^2}\right)u(x) = \sum \frac{1}{n!}\left(\frac{t}{2}\right)^n u^{(2n)}(x).$$

Покажите, что оно верно, если ряд Тейлора для u сходится при всех x и ряд в правой части сходится при $t > 0$. *Указание.* Начните с ряда Тейлора для свертки u и нормального распределения. Используйте моменты нормального распределения.

8. Распределения полугруппы имеют конечные математические ожидания тогда и только тогда, когда функция $\frac{1}{1+|x|}$ интегрируема относительно меры Ω , входящей в формулу для производящего оператора.

9. Покажите непосредственно, что если $n[\mathfrak{F}_n-1] \rightarrow \mathfrak{U}$, то оператор $\mathfrak{U}$ обязательно имеет форму производящего оператора. *Указание.* Используйте метод § 4, рассматривая функции вида (4.4), но не привлекайте теорию полугрупп. Ваша цель — вывести общий вид производящего оператора, не доказывая заранее его существование.

10. Пусть F_k приписывает вероятности $\frac{1}{2}$ двум точкам $\pm \mu^k$. Тогда

$$\sum_{k=-\infty}^{+\infty} 2^{-k} (\mathfrak{F}_k - 1)$$

порождает полугруппу, такую, что $Q_{2t}(x) = Q_t(x\mu)$, но Q_t не являются устойчивыми распределениями. (П. Леви.)

11. *Прямое доказательство того, что предельные распределения сумм по строкам схем серий безгранично делимы.* Пусть $\{X_{k,n}\}$ — схема серий, составленная одинаково распределенными величинами. Стохастическая ограниченность S_n влечет то же самое для частных сумм $X_{1,n} + \dots + X_{m,n}$, где m обозначает наибольшее целое число $\leq n/r$. Используя теорему о выборе, покажите, что отсюда следует, что предельное распределение G сумм S_n есть r -кратная свертка распределения G^r .

12. Для любого распределения F и гладкой функции u

$$\|(\mathfrak{F}-1)u\| \leq 100 \cdot (\|u\| + \|u''\|) \int_{-\infty}^{+\infty} \frac{x^2}{1+x^2} F(dx) + \|u'\|.$$

13. Схеме серий $\{X_{k,n}\}$ с распределениями F_n соответствует другая схема серий $\{X_{k,n}^\#\}$ с обобщенными пуассоновскими распределениями

$$\mathfrak{F}_n^\# = e^{\mathfrak{F}_n - 1}.$$

Покажите, что если $\{S_n\}$ стохастически ограничена, то $n[\mathfrak{F}_n - \mathfrak{F}_n^\#] \rightarrow 0$. Отсюда видно, что суммы по строкам S_n и $S_n^\#$ асимптотически эквивалентны. Так как распределение $S_n^\#$ связано с $e^{n[\mathfrak{F}_n - 1]}$, мы получаем второй метод, пригодный для вывода основных теорем § 7. Этот метод применим и к схемам серий с различно распределенными величинами. *Указание.* Используйте задачу 12.

14. Положим в обозначениях § 5 $M_n = \max\{X_{1,n}, \dots, X_{n,n}\}$. Если S_n имеет предельное распределение, то $\psi^+(x) = -\lim \log P\{M_n < x\}$.

15. Если S_n имеет предельное распределение, то этим же свойством обладают суммы по строкам в схеме серий, образованной квадратами $X_{k,n}^2$.

16. *Области притяжения.* Пусть F принадлежит области притяжения устойчивого распределения с показателем α . Используя теорему 2 из гл. VIII, 9, покажите, что F обладает абсолютными моментами всех порядков $< \alpha$. Если $\alpha < 2$, то не существует моментов порядка $> \alpha$. Последнее утверждение неверно при $\alpha = 2$.

17. *Продолжение.* Теория § 8 основана на урезанных вторых моментах (функция U). Причина этого — традиция. Теорема 2 гл. VIII, 9, позволяет нам заменить y^2 в (8.1) на $|y|^p$ с другим показателем p и при каждом p — заменить (8.3) и (8.5) равносильными соотношениями.

18. Пусть $X_1, X_2, \dots$ — независимые случайные величины с одним и тем же распределением F . Пусть $1 - F(x) + F(-x)$ медленно меняется. Выведите из леммы о компактности, что в этом случае последовательность $(S_{n_k}/a_k) + b_k$ не может иметь собственное предельное распределение, не сосредоточенное в одной точке. (Это можно выразить также, говоря, что F не принадлежит никакой области частичного притяжения. См. гл. XVII, 9.) *Указание:* Используйте симметризацию,

Эта глава открывается элементарным обзором наиболее распространенных типов марковских процессов, или, точнее, основных уравнений, которым подчиняются их переходные вероятности. Затем мы перейдем к введенному Бохнером понятию подчиненности процессов и к полугрупповой трактовке марковских процессов. Связующим звеном между этими вопросами служит так называемая показательная формула теории полугрупп. Существование производящих операторов будет доказано только в гл. XIII методом, использующим теорию резольвент. Теоретически настоящее изложение могло бы охватывать в качестве частных случаев процессы и полугруппы предыдущей главы. Однако методы и применения столь различны, что последующая теория излагается замкнуто и независимо от гл. IX. Результаты будут усилены в гл. XIII, но теория марковских процессов не используется в оставшейся части книги.

Эта глава по своему характеру — обзор, в котором не делается попытки достичь общности или полноты¹⁾. В частности, мы не будем обсуждать свойства выборочных функций, и всюду в этой главе существование рассматриваемых процессов считается заранее данным фактом. Наши интересы полностью концентрируются вокруг аналитических свойств переходных вероятностей и операторов.

Теория индуцированных полугрупп будет с достаточной степенью общности рассмотрена в § 8—9. В предшествующих им параграфах основное пространство есть интервал на прямой или вся прямая, хотя часть теории годится и в более общей обстановке. Чтобы избежать введения специальных обозначений, мы условимся считать, что если пределы не указаны, то интеграл берется по фиксированному множеству Ω , служащему основным пространством.

¹⁾ Более детально полугрупповой подход к марковским процессам описан в книгах Е. Дынкина (1966) и М. Лозва (1963). Книга К. Иосады (1967) дает сжатое и ясное введение в аналитическую теорию полугрупп и ее применение к теории диффузии и эргодической теории.

§ 1. ПСЕВДОПУАССОНОВСКИЙ ТИП

Всюду в этой главе мы ограничиваемся марковскими процессами со *стационарными переходными вероятностями* Q_t , определяемыми формулой

$$Q_t(x, \Gamma) = P\{X(t + \tau) \in \Gamma \mid X(\tau) = x\} \quad (1.1)$$

в предположении, что правая часть не зависит от τ (см. гл. VI, 11).

Простое расширение понятия обобщенного пуассоновского процесса приводит к некоторому важному классу процессов, из которых все другие можно получить с помощью аппроксимации. Теория полугрупп опирается на аналитический аналог этого факта (§ 10).

Пусть $N(t)$ обозначает величины, образующие пуассоновский процесс. Обобщенный пуассоновский процесс был введен в гл. VI, 4, посредством случайных сумм $S_{N(t)}$, где $S_0, S_1, \dots$ — частные суммы последовательности независимых и одинаково распределенных случайных величин. Псевдопуассоновский процесс определяется тем же способом, но теперь случайные величины $S_0, S_1, \dots$ образуют марковскую цепь, переходные вероятности которой задаются стохастическим ядром K (см. гл. VI, 11). Случайные величины $X(t) = S_{N(t)}$ образуют новый случайный процесс, который формально можно описать следующим образом.

В промежутке между скачками пуассоновского процесса типичная выборочная функция остается постоянной. Переход из x в Γ может произойти за 0, 1, 2, ... шагов, и, следовательно,

$$Q_t(x, \Gamma) = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} K^{(n)}(x, \Gamma), \quad t > 0. \quad (1.2)$$

Обобщенное пуассоновское распределение гл. VI, (4.2), получается из этой формулы как частный случай, когда Ω совпадает со всей прямой и S_n представляет собой сумму независимых случайных величин с одним и тем же распределением F .

Правило композиции

$$Q_{t+\tau}(x, \Gamma) = \int Q_t(x, dy) Q_\tau(y, \Gamma) \quad (1.3)$$

($t, \tau > 0$), аналогичное гл. VI, (4.1), легко проверить аналитически¹). Оно называется уравнением *Чепмена—Колмогорова* и утверждает, что переход от состояния x в момент 0 во множество Γ в момент $t + \tau$ происходит через некоторую точку y в момент t и что последующее движение не зависит от прошлого²) [см. 1; гл. XVII, 9, а также гл. VI, (11.3)].

¹) Аппроксимация Q_t и Q_τ их частными суммами с n членами показывает, что правая часть в (1.3) $\leq$ левой части, но $\geq n$ -й частной суммы $Q_{t+\tau}$.

²) Иногда объявляют (1.3) либо законом природы, либо формулой полной вероятности. Но (1.3) не выполняется для немарковских процессов; см. 1, гл. XV, 13.

Примеры. а) *Частицы, движущиеся со столкновениями.* Рассмотрим частицу, движущуюся с постоянной скоростью в однородной среде и время от времени подвергающуюся случайным соударениям. Каждое из них приводит к изменению энергии, описываемому стохастическим ядром K . Если число соударений образует пуассоновский процесс, то переходные вероятности для энергии $X(t)$ имеют вид (1.2). Этот случай имеет место при хорошо знакомых нам предположениях пространственной однородности и отсутствия последействия.

Обычно предполагают, что *доля энергии*, теряемая при каждом соударении, не зависит от ее начальной величины. Это означает, что $K(x, dy) = V \{dy/x\}$, где V — распределение вероятностей, сосредоточенное на $\overline{0, 1}$. Имея в виду последующие применения, рассмотрим частный случай $V(x) = x^\lambda$. Тогда K имеет плотность

$$k(x, y) = \lambda x^{-\lambda} y^{\lambda-1}, \quad 0 < y < x. \quad (1.4)$$

При $\lambda=1$ наше предположение означает, что доля теряемой энергии *распределена равномерно*¹⁾. Итерированные ядра $K^{(n)}$ были вычислены в гл. VI, (11.5). Подставляя их значение в (1.2), видим, что Q_t имеет атом веса $e^{-\alpha t}$ в нуле (соответствующий отсутствию соударений), а при $0 < y < x$ имеет плотность

$$q_t(x, y) = e^{-\alpha t} \sqrt{\lambda \alpha t} \frac{y^{\lambda-1}}{x^\lambda \sqrt{\log(x/y)}} I_1(2\sqrt{\alpha t \lambda \log(x/y)}), \quad (1.5)$$

где I_1 — функция Бесселя, введенная в гл. II, (7.1) [см. примеры 2, а) и 2, б)].

б) *Потеря энергии быстрыми частицами при ионизации*²⁾. Поучительный вариант предыдущего примера получается в предельном случае, когда энергию частицы предполагают бесконечно большой. Потери энергии при последовательных соударениях будут представлять собой независимые случайные величины с одним и тем же распределением V , сосредоточенным на $\overline{0, \infty}$. Пусть $X(t)$ обозначает полную величину потерянной за время $\overline{0, t}$ энергии. Тогда случайные величины $X(t)$ образуют *обобщенный пуассоновский процесс*. Переходные вероятности для него определяются по (1.2) с заменой $K^{(n)}$ на свертку V^{n*} .

в) *Изменение направления.* Вместо энергии частицы мы можем рассмотреть направление ее движения и построить модель, по-

¹⁾ Это предположение использовано Гейтлером и Яноши (Heitler W., Janossy L., Absorption of meson producing nucleons, Proc. Phys. Soc., Ser. A, 62 (1949), 374—385. В этой статье введено (но не решено) уравнение Фоккера—Планка (1.8).

²⁾ См. статью: Ландау Л. Д., J. Phys., USSR, 8 (1944), 201—205. Ландау использует другую терминологию, но его предположения равносильны нашим. Ландау вывел прямое уравнение (1.6).

добную приведенной в примере а). Основное различие будет в том, что направление в $\mathcal{R}^3$ определяется двумя величинами, так что плотность ядра k будет зависеть от четырех вещественных переменных.

г) Рандомизированное случайное блуждание примера гл. II, 7, в) представляет собой псевдопуассоновский процесс, состояниями которого служат целые числа. При фиксированном x ядро K приписывает вес $1/2$ каждой из двух точек $x \pm 1$. ►

Из (1.2) легко получаем

$$\frac{\partial Q_t(x, \Gamma)}{\partial t} = -\alpha Q_t(x, \Gamma) + \alpha \int Q_t(x, dz) K(z, \Gamma). \quad (1.6)$$

Это — прямое уравнение Колмогорова¹⁾, которое (в более общей постановке) будет обсуждаться в § 3. Там будет показано, что (1.2) является единственным решением, удовлетворяющим естественным вероятностным требованиям. Если K имеет плотность k , то уравнение (1.6) принимает более привычную форму. В точке x распределение Q_t имеет атом веса $e^{-\alpha t}$, равный вероятности отсутствия изменений. Если исключить этот атом, то Q_t будет иметь плотность q_t , удовлетворяющую уравнению

$$\frac{\partial q_t(x, \xi)}{\partial t} = -\alpha q_t(x, \xi) + \alpha \int q_t(x, z) k(z, \xi) dz. \quad (1.6a)$$

Если μ_0 — распределение вероятностей в момент 0, то распределение в момент t определяется по формуле

$$\mu_t\{\Gamma\} = \int \mu_0\{dx\} Q_t(x, \Gamma) \quad (1.7)$$

и в силу (1.6)

$$\frac{\partial \mu_t\{\Gamma\}}{\partial t} = -\alpha \mu_t\{\Gamma\} + \alpha \int \mu_t\{dz\} K(z, \Gamma). \quad (1.8)$$

Этот вариант (1.6) известен физикам под названием уравнения Фоккера — Планка (или уравнения неразрывности).

Это уравнение будет проанализировано в § 3. В случае когда K и начальное распределение μ_0 имеют плотности, μ_t также имеет плотность m_t и уравнение Фоккера — Планка принимает вид

$$\frac{\partial m_t(\xi)}{\partial t} = -\alpha m_t(\xi) + \alpha \int m_t(z) k(z, \xi) dz. \quad (1.8a)$$

§ 2. ВАРИАНТ: ЛИНЕЙНЫЕ ПРИРАЩЕНИЯ

В физике, теории очередей и других применениях встречается один вариант рассмотренного нами процесса. Предположения, касающиеся скачков, остаются теми же самыми, но между скач-

¹⁾ Относительно терминологии см. сноску в конце 1; гл. XVII, 8. — Прим. перев.

ками $X(t)$ изменяется линейно со скоростью c . Это означает, что величины $X(t) - ct$ образуют уже описанный псевдопуассоновский процесс; обозначим переходные вероятности нового процесса через Q_t . Тогда вероятности $Q_t(x, \Gamma + ct)$ должны удовлетворять (1.6). Получающееся отсюда уравнение для Q_t имеет необычную форму, однако если существуют дифференцируемые плотности, то они удовлетворяют привычным уравнениям. Так, для m_t мы должны уменьшить ξ в (1.8а) на $\xi + ct$. После замены $y = \xi + ct$ получим уравнение Фоккера—Планка ¹⁾

$$\frac{\partial m_t(y)}{\partial t} = -c \frac{\partial m_t(y)}{\partial y} - \alpha m_t(y) + \alpha \int m_t(z) k(z - ct, y - ct) dz. \quad (2.1)$$

Уравнение, аналогичное (1.6а), получается простым добавлением члена $-c \partial q_t / \partial y$ в правой части.

В связи с теорией полугрупп мы придадим уравнению Фоккера—Планка более гибкую форму, не зависящую от неестественных условий дифференцируемости [см. пример 10, б)].

Примеры. а) *Частицы, испытывающие соударения.* Пример 1, а) встречается в физической литературе обычно в другой постановке. Предполагается, что между соударениями энергия рассеивается с постоянной скоростью благодаря поглощению или трению. Модель (2.1) подходит к этой ситуации, если потеря энергии пропорциональна плотности вероятности m_t энергии в момент t . В других ситуациях физики предполагают, что между соударениями энергия рассеивается со скоростью, пропорциональной ее мгновенному значению. В этом случае логарифм энергии убывает с постоянной скоростью.

б) *Звездное излучение* ²⁾. В этой модели переменная t обозначает *расстояние*, а $X(t)$ — интенсивность светового луча, проходящего в пространстве. Предполагается, что (в экваториальной плоскости) каждый элемент объема порождает излучение постоян-

¹⁾ Многие частные случаи уравнения Фоккера—Планка (1.8) были открыты независимо друг от друга. Много энергии было затрачено на обобщение (2.1). Общая теория уравнений Фоккера—Планка была развита Колмогоровым в его известной работе «Об аналитических методах в теории вероятностей» (Math. Ann., 104 (1931), 415—458, перевод в Успехах Матем. Наук, 5 (1938), 5—41). В этой работе Колмогоров отмечает возможность добавить к правой части произвольный «диффузионный член»

$$\gamma \frac{\partial^2 m_t}{\partial y^2} - c \frac{\partial m_t}{\partial y}.$$

Формула (2.1) представляет собой лишь частный случай этой более общей формулы. Уже первые теоремы существования охватывали случай нестационарного общего уравнения [Feller W., Math. Ann., 113 (1936)].

²⁾ Физические предпосылки взяты из статьи: В. А. Амбарцумян. О флуктуациях яркости Млечного пути, ДАН СССР, 44 (1944), 223—226. В этой статье косвенным путем получен один вариант (2,1),

ной интенсивности и, следовательно, $X(t)$ возрастает линейно. Но в пространстве также размещены поглощающие черные облака, которые мы будем представлять себе в виде пуассоновского множества точек. Встречая облако, каждый луч теряет (случайную) часть интенсивности, так что мы оказываемся в точности в тех условиях, которые привели нас к (2.1). Кажется правдоподобным (и на самом деле это может быть доказано), что плотность m_t величины $X(t)$ приближается к *стационарной плотности* m , которая не зависит от t и удовлетворяет (2.1) с левой частью, замененной нулем.

Амбарцумян предполагает, в частности, что потеря интенсивности при прохождении отдельного облака регулируется переходным ядром (1.4). В этом случае явный вид решения содержится в (1.5), но он представляет вторичный интерес. Более важен (и легко проверяется) тот факт, что (2.1) имеет не зависящее от времени (или стационарное) решение, задаваемое гамма-плотностью

$$m(y) = \left(\frac{\alpha}{c}\right)^{\lambda+1} \frac{1}{\Gamma(\lambda+1)} y^{\lambda} e^{-(\alpha/c)y}, \quad y > 0. \quad (2.2)$$

Этот результат показывает, что относящаяся к делу информация может быть получена из (2.1), даже минуя отыскание явного решения. Например, легко найти прямым интегрированием, что стационарное решение имеет математическое ожидание $c[\alpha(1-\mu)]^{-1}$, где μ — математическое ожидание для распределения поглощаемой энергии V .

в) *Задача о разорении* из гл. VI, 5, представляет собой частный случай, в котором ядро k зависит от разности аргументов. Значения этого процесса получаются добавлением $-ct$ к значениям некоторого обобщенного пуассоновского процесса. Аналогичные «задачи о разорении» могут быть сформулированы для любого псевдопуассоновского процесса. Они приводят к уравнению (2.1).

§ 3. СКАЧКООБРАЗНЫЕ ПРОЦЕССЫ

В псевдопуассоновском процессе время ожидания следующего скачка имеет фиксированное показательное распределение с математическим ожиданием $1/\alpha$. Мы приходим к естественному обобщению, допуская, что это распределение может зависеть от значения $X(t)$ (в данный момент (в примере 1, а) это сводится к предположению, что вероятность столкновения зависит от энергии частицы). Марковский характер процесса требует, чтобы распределение времени до следующего скачка было показательным, но его математическое ожидание может зависеть от значения $X(t)$ в данный момент. В соответствии со сказанным мы введем следующие

Основные постулаты. При условии $X(t) = x$ время ожидания следующего скачка имеет показательное распределение с математическим ожиданием $1/\alpha(x)$ и не зависит от прошлого. Вероятность того, что следующий скачок приводит к значению из некоторого множества Γ , равна $K(x, \Gamma)$.

В аналитических терминах эти постулаты сводятся к интегральным уравнениям для вероятностей перехода $Q_t(x, \Gamma)$ рассматриваемого процесса (в предположении, что такой процесс на самом деле существует). Рассмотрим фиксированную точку x и фиксированное множество Γ , не содержащее x . Событие $\{X(t) \in \Gamma\}$ может произойти только при условии, что первый скачок из x произошел в некоторый момент $s < t$. При этом условии условная вероятность события $\{X(t) \in \Gamma\}$ получается интегрированием $K(x, dy) Q_{t-s}(y, \Gamma)$ по множеству Ω всех возможных значений y . Теперь момент первого скачка — это случайная величина с показательной плотностью $\alpha(x) e^{-\alpha(x)s}$. Интегрируя по отношению к этой плотности, мы получаем вероятность события $\{X(t) \in \Gamma\}$ в форме

$$Q_t(x, \Gamma) = \alpha(x) \int_0^t e^{-\alpha(x)s} ds \int_{\Omega} K(x, dy) Q_{t-s}(y, \Gamma). \quad (3.1a)$$

Для множества Γ , содержащего x , мы должны добавить вероятность того, что до момента t не произойдет ни одного скачка. Мы получим

$$Q_t(x, \Gamma) = e^{-\alpha(x)t} + \alpha(x) \int_0^t e^{-\alpha(x)s} ds \int_{\Omega} K(x, dy) Q_{t-s}(y, \Gamma). \quad (3.1b)$$

Эти два уравнения, справедливые при $x \in \bar{\Gamma}$ и $x \in \Gamma$ соответственно, являются аналитическими эквивалентами основных постулатов. Они упрощаются при переходе к новой переменной интегрирования $\tau = t - s$. Дифференцирование по t приводит оба уравнения к одинаковому виду, и тогда пара уравнений заменяется одним интегро-дифференциальным уравнением

$$\frac{\partial Q_t(x, \Gamma)}{\partial t} = -\alpha(x) Q_t(x, \Gamma) + \alpha(x) \int_{\Omega} K(x, dy) Q_t(y, \Gamma). \quad (3.2)$$

Это не что иное, как *обратное уравнение Колмогорова*, которое служит отправной точкой для аналитического подхода к задаче, так как позволяет избежать раздражающего различия между двумя указанными случаями.

Обратное уравнение (3.2) допускает простую интуитивную интерпретацию, которая позволяет представить основные постулаты в более близких к практике терминах. В терминах разно-

стных отношений (3.2) равносильно соотношению

$$Q_{t+h}(x, \Gamma) = [1 - \alpha(x)h] Q_t(x, \Gamma) + \alpha(x)h \int_{\Omega} K(x, dy) Q_t(y, \Gamma) + o(h). \quad (3.3)$$

Для интуитивной интерпретации последнего рассмотрим изменение состояния на временном интервале $\overline{0, t+h}$ как результат изменения за короткий интервал $\overline{0, h}$ и последующего изменения в интервале $\overline{h, t+h}$ длины t . Тогда, очевидно, (3.3) утверждает, что если $X(0) = x$, то вероятность одного скачка в интервале $\overline{0, h}$ равна $\alpha(x)h + o(h)$, а вероятность большего числа скачков равна $o(h)$. Наконец, если на $\overline{0, h}$ скачок происходит, то условные вероятности тех или иных переходов определяются мерой $K(x, dy)$. Эти три постулата приводят к (3.3), а тем самым и к (3.2). По существу, они повторяют основные постулаты¹⁾.

С вероятностной точки зрения обратное уравнение представляется несколько искусственным, так как в нем окончательное состояние Γ играет роль параметра и (3.2) описывает зависимость $Q_t(x, \Gamma)$ от начального положения x . Казалось бы более естественным получить уравнение для Q_{t+h} , разбивая интервал $\overline{0, t+h}$ на длинный начальный интервал $\overline{0, t}$ и короткий конечный интервал $\overline{t, t+h}$. Вместо (3.3) мы получим формально

$$Q_{t+h}(x, \Gamma) = \int_{\Gamma} Q_t(x, dz) [1 - \alpha(z)h] + \int_{\Omega} Q_t(x, dz) \alpha(z)h K(z, \Gamma) + o(h), \quad (3.4)$$

и, следовательно,

$$\frac{\partial Q_t(x, \Gamma)}{\partial t} = - \int_{\Gamma} Q_t(x, dz) \alpha(z) + \int_{\Omega} Q_t(x, dz) \alpha(z) K(z, \Gamma). \quad (3.5)$$

Это — *прямое уравнение Колмогорова* (в частном случае известное физикам, как *уравнение неразрывности или уравнение Фоккера — Планка*). Оно сводится к (1.6), когда α не зависит от z .

Формальный характер приведенного вывода мы подчеркнули потому, что на самом деле прямое уравнение не вытекает из наших основных постулатов. Это объясняется тем, что величина $o(h)$ в (3.3) зависит от z , и так как z служит в (3.4) переменной интегрирования, то эта величина должна была бы стоять под

¹⁾ Дифференцируемость Q_t по отношению к t и тот факт, что вероятность более чем одного скачка равна $o(h)$, теперь принимаются в качестве новых постулатов; в то же время они вытекают из прежней более сложной формулировки.

знаком интеграла. Но тогда возникает вопрос о том, сходятся ли интегралы в (3.5) и законен ли предельный переход при $h \rightarrow 0$. [Ни одно из этих затруднений не возникает в связи с обратным уравнением, так как начальное состояние x при этом фиксировано и интеграл в (3.2) существует в силу ограниченности Q_t .]

Прямое уравнение можно обосновать, добавляя к нашим основным постулатам необходимое ограничение на остаточный член в (3.3), но такой вывод перестал бы быть интуитивно привлекательным. Кроме того, кажется невозможным сформулировать условия, которые охватывали бы все типичные случаи, встречающиеся на практике. Если допустить, что функция α может быть неограниченной, то существование интегралов в (3.5) сомнительно и уравнение нельзя обосновать априори. С другой стороны, обратное уравнение есть необходимое следствие основных предположений. Поэтому в качестве отправной точки лучше использовать его, а затем исследовать, в каких условиях из него может быть выведено прямое уравнение.

Решение обратного уравнения легко построить, используя последовательные приближения, имеющие простой вероятностный смысл. Обозначим через $Q_t^{(n)}(x, \Gamma)$ вероятность перехода из $X(0) = x$ в $X(t) \in \Gamma$ с не более чем n скачками. Переход без скачков возможен, только если $x \in \Gamma$, и, так как время пребывания в x имеет показательное распределение, мы имеем

$$Q_t^{(0)}(x, \Gamma) = e^{-\alpha(x)t} K^{(0)}(x, \Gamma) \quad (3.6)$$

(где $K^{(0)}(x, \Gamma)$ равно 1 или 0 в зависимости от того, входит x в Γ или нет). Предположим затем, что первый скачок происходит в момент $s < t$ и приводит из x в y . Суммируя по всем возможным s и y , мы получаем [как и в (3.1)] рекуррентную формулу

$$Q_t^{(n+1)}(x, \Gamma) = Q_t^{(n)}(x, \Gamma) + \int_0^t e^{-\alpha(x)s} \alpha(x) ds \int K(x, dy) Q_{t-s}^{(n)}(y, \Gamma), \quad (3.7)$$

верную при $n = 0, 1, \dots$. Очевидно, что $Q_t^{(0)} \leq Q_t^{(1)}$, и по индукции $Q_t^{(1)} \leq Q_t^{(2)} \leq \dots$.

Мы видим, что при каждом x и Γ существует предел

$$Q_t^{(\infty)}(x, \Gamma) = \lim_{n \rightarrow \infty} Q_t^{(n)}(x, \Gamma), \quad (3.8)$$

но он может быть в принципе бесконечным.

Мы покажем, что

$$Q_t^{(\infty)}(x, \Omega) \leq 1, \quad (3.9)$$

откуда будет вытекать $Q_t^{(\infty)}(x, \Gamma) \leq 1$ для всех множеств из Ω . Для этого достаточно доказать, что при всех n

$$Q_t^{(n)}(x, \Omega) \leq 1. \quad (3.10)$$

При $n=0$ неравенство тривиально, и мы используем индукцию. Предполагая (3.10) справедливым при некотором фиксированном n , мы получим из (3.7) [вспоминая, что K —стохастическое ядро]

$$Q_t^{(n+1)}(x, \Omega) \leq e^{-\alpha(x)t} + \int_0^t e^{-\alpha(x)s} \alpha(x) ds = 1. \quad (3.11)$$

Следовательно, (3.10) выполняется при всех n .

Из (3.7) в силу монотонной сходимости следует, что $Q_t^{(\infty)}$ удовлетворяет обратному уравнению в исходной интегральной форме (3.1) [и, следовательно, в интегро-дифференциальной форме (3.2)]. Очевидно, что для любого другого положительного решения Q_t уравнения (3.1) справедливо $Q_t \geq Q_t^{(0)}$; сравнивая (3.1) с (3.7), заключаем, что $Q_t \geq Q_t^{(n)}$ при всех n и поэтому $Q_t \geq Q_t^{(\infty)}$. По этой причине $Q_t^{(\infty)}$ называется *минимальным решением* обратного уравнения; из (3.9) следует, что $Q_t^{(\infty)}$ является стохастическим или субстохастическим ядром.

Из (3.8) вытекает, что $Q_t^{(\infty)}(x, \Gamma)$ есть вероятность перехода из x в Γ с конечным числом скачков. Соответственно для субстохастического решения $Q_t^{(\infty)}$ дефект $1 - Q_t^{(\infty)}(x, \Omega)$ представляет собой вероятность того, что при переходе из x как начальной точки за время t может произойти бесконечно много скачков. Из 1; XVII.4, и примера 5, в) гл. VIII нам известно, что это свойственно некоторым процессам чистого размножения, и, следовательно, субстохастические ядра существуют, но они являются скорее исключением, чем правилом. В частности, если коэффициент $\alpha(x)$ ограничен, то минимальное решение—строго стохастическое, т. е.

$$Q_t^{(\infty)}(x, \Omega) = 1, \quad t > 0. \quad (3.12)$$

Действительно, если $\alpha(x) < a < \infty$ при всех x , то мы покажем, используя индукцию, что

$$Q_t^{(n)}(x, \Omega) \geq 1 - (1 - e^{-at})^n \quad (3.13)$$

при всех n и $t > 0$. При $n=0$ это неравенство тривиально. Предположим, что оно справедливо при некотором n . Заметим, что правая часть (3.12) есть убывающая функция, которую мы обозначим $f(t)$, и рассмотрим (3.7) с $\Gamma = \Omega$. В силу (3.13) внутренний интеграл в (3.7) $\geq f(t)$. Так как $f(t)$ не зависит от переменной интегрирования, то, интегрируя $\alpha(x)e^{-\alpha(x)s}$, приходим к выводу, что (3.13) выполняется и при замене n на $n+1$.

Наконец, отметим, что в строго стохастическом случае (3.12) минимальное решение единственно. В самом деле, в силу минимальности $Q_t^{(\infty)}$ любое другое приемлемое решение должно удовлетворять цепочке соотношений

$$\begin{aligned} 1 &\geq Q_t(x, \Omega) = Q_t(x, \Gamma) + Q_t(x, \Omega - \Gamma) \geq \\ &\geq Q_t^{(\infty)}(x, \Gamma) + Q_t^{(\infty)}(x, \Omega - \Gamma) = Q_t^{(\infty)}(x, \Omega) = 1, \end{aligned} \quad (3.14)$$

что невозможно, если только знаки неравенства в двух местах не заменить равенствами. Таким образом, мы доказали теорему.

Теорема. Обратное уравнение допускает минимальное решение $Q_t^{(\infty)}$, определяемое формулой (3.8) и соответствующее процессу, в котором переходы из x в Γ происходят только с конечным числом скачков. Ядро $Q_t^{(\infty)}$ является стохастическим или субстохастическим.

В субстохастическом случае дефект $1 - Q_t^{(\infty)}(x, \Omega)$ представляет вероятность того, что за конечное время t произойдет бесконечно много скачков, и в этом случае минимальный процесс обрывается.

В строго стохастическом случае (3.12) минимальное решение есть единственное вероятностное решение обратного уравнения. Эта ситуация возникает, когда коэффициент $\alpha(x)$ ограничен.

Обнаружение несобственных решений произвело шок на ранней стадии развития теории в 1930-х годах, но и стимулировало исследования, приведшие к целиной теории марковских процессов. Процессы, в которых переходы из x в Γ возможны после бесконечного числа скачков, допускают аналогию с диффузионными процессами с граничными условиями и потому не так патологичны, какими казались вначале. Возможность появления бесконечного числа скачков объясняет также трудности, возникающие при непосредственном¹⁾ выводе прямых уравнений. Обратные уравнения были выведены на основе предположения о том, что при данном положении x в настоящий момент время ожидания следующего скачка имеет показательное распределение с математическим ожиданием $1/\alpha(x)$. Прямые уравнения зависят от положения именно до момента t и поэтому зависят от пространства Ω в целом. В частности, нелегко непосредственно выразить условие существования последнего скачка до момента t .

Однако в приложении в конце этого параграфа будет показано, что минимальное решение $Q_t^{(\infty)}$ обратного уравнения автоматически удовлетворяет прямому уравнению и является минимальным также и для него. Отсюда, в частности, следует, что если α ограничено, то $Q_t^{(\infty)}$ представляет собой единственное решение прямого уравнения. Если $Q_t^{(\infty)}$ — субстохастическое ядро, то существуют различные процессы, включающие переходы с бесконечным числом скачков и удовлетворяющие обратным уравнениям. Переходные вероятности таких процессов могут, но не обязаны удовлетворять прямым уравнениям. Этот неожиданный факт показывает, что прямые уравнения могут удовлетворяться в ситуациях, когда вывод уравнения из (3.4) терпит неудачу.

В заключение снова отметим, что (по контрасту с процессами § 2 и диффузионными процессами, в которых используется диф-

¹⁾ Ср. с аналогичным рассуждением в I; XVII, 9.

ференцирование по x) чисто скачкообразные процессы *никоим образом не зависят от природы пространства состояний* и наши формулы применимы для любого множества Ω , на котором определено стохастическое ядро K .

Пример. *Счетные пространства состояний.* Если случайные величины $X(t)$ положительны и целочисленны, то основное выборочное пространство (пространство состояний) состоит из чисел $1, 2, \dots$. Теперь достаточно знать вероятности $P_{ik}(t)$ перехода от одного целого числа к другому. Все прочие переходные вероятности получаются суммированием по k . Теория подобных марковских процессов была намечена в 1; гл. XVII, 9, где рассматривались и нестационарные переходные вероятности. Чтобы выделить внутри этой теории стационарный случай, следует считать коэффициенты c_i и вероятности p_{ik} постоянными (не зависящими от t). Тогда прежние предположения становятся равносильными принятым здесь и, как легко видеть, две системы уравнений Колмогорова, полученные в 1; гл. XVII, 9, становятся частными случаями уравнений (3.2) и (3.5) [с заменой $\alpha(i)$ на c_i и $K(i, j)$ на p_{ij}]. Расходящийся процесс размножения из 1; гл. XVII, 4, служит примером процесса с бесконечным числом скачков за конечный промежуток времени. Мы вернемся к этому процессу в гл. XIV, 8, чтобы продемонстрировать возможность перехода из состояния i в состояние j , включающего бесконечное число скачков.

Приложение ¹⁾. *Минимальное решение прямого уравнения.* Конструкция минимального решения $Q_i^{(\infty)}$ обратного уравнения может быть приспособлена для прямого уравнения. Мы в краткой форме укажем, как это может быть сделано и каким образом можно проверить, что оба решения в действительности идентичны. Детали доказательства предоставляются читателю.

Пусть

$$K_i^{\#}(x, \Gamma) = \int_{\Gamma} e^{-\alpha(y)t} K(x, dy). \quad (3.15)$$

Для построения решения прямого уравнения определим $Q_i^{(0)}$ равенством (3.6) и положим

$$Q_i^{(n+1)}(x, \Gamma) = Q_i^{(0)}(x, \Gamma) + \int_0^t \int Q_i^{(n-s)}(x, dy) \alpha(y) K_s^{\#}(y, \Gamma). \quad (3.16)$$

Это равенство задает вероятности перехода за не более чем $n+1$ шагов в терминах *последнего* скачка точно так же, как (3.7) имеет отношение к первому

¹⁾ Та же теория, использующая преобразования Лапласа, дана в гл. XIV, 7. (Только для упрощения рассматриваются счетные пространства, хотя рассуждения применимы и в общем случае без существенных изменений.) Непосредственный метод, данный в тексте, менее элегантен, но обладает тем преимуществом, что он применим и к нестационарным процессам, переходные вероятности которых зависят от временного параметра. Общая теория была развита Феллером, см.: Feller W., Trans. Amer. Math. Soc., vol. 48 (1940), 488—515 (поправка vol. 58, 474).

скачку. Повторяя доказательство последней теоремы, выводим, что $Q_i^{(\infty)} = \lim Q_i^{(n)}$ есть минимальное решение прямого уравнения.

Хотя мы и использовали те же самые буквы, но обе рекуррентные формулы (3.7) и (3.16) получены независимо одна от другой, и пока ниоткуда не следует, что полученные ядра идентичны. Чтобы показать, что это на самом деле так, положим $P_i^{(n)} = Q_i^{(n)} - Q_i^{(n-1)}$, что соответствует переходам ровно за n шагов. Тогда (3.16) сводится к

$$P_i^{(n+1)}(x, \Gamma) = \int \int P_{i-s}^{(n)}(x, dy) \alpha(y) K_s^{\#}(y, \Gamma). \quad (3.17)$$

Мы запишем (3.17) с помощью сокращенного обозначения: $P_i^{(n+1)} = P_i^{(n)} \mathfrak{A}$. Рекуррентную формулу (3.7) мы запишем подобным образом: $P_i^{(n+1)} = \mathfrak{B} P_i^{(n)}$. Величина $P_i^{(0)}$ одна и та же в любом случае и определена в (3.6). Докажем теперь, что обе рекуррентные формулы приводят к одному результату. Более точно, мы докажем, что $P_i^{(n)}$, определенные соотношением $P_i^{(n+1)} = P_i^{(n)} \mathfrak{A}$, удовлетворяют соотношению $P_i^{(n+1)} = \mathfrak{B} P_i^{(n)}$. Доказательство проведем по индукции. Предположим, что утверждение справедливо при всех $n \leq r$. Тогда

$$P_i^{(r+1)} = P_i^{(r)} \mathfrak{A} = (\mathfrak{B} P_i^{(r-1)}) \mathfrak{A} = \mathfrak{B} (P_i^{(r-1)} \mathfrak{A}) = \mathfrak{B} P_i^{(r)},$$

и таким образом предположение индукции справедливо и при $n = r+1$.

Мы доказали, следовательно, что минимальное решение является общим для обратного и прямого уравнений.

§ 4. ДИФФУЗИОННЫЕ ПРОЦЕССЫ В $\mathcal{R}^1$

Покончив с процессами, в которых все изменения происходят скачками, мы обратимся к другому крайнему случаю, к процессам, у которых (с вероятностью единица) выборочные функции непрерывны. Их теория похожа на развитую в предыдущем параграфе, однако основные уравнения требуют более сложного анализа. Мы удовлетворимся поэтому выводом обратного уравнения и кратким резюме результатов, относящихся к минимальным решениям и другим проблемам. Прототипом диффузионных процессов служит броуновское движение (или винеровский процесс). Это процесс с независимыми нормально распределенными приращениями. Его переходные вероятности имеют плотности $q_t(x, y)$, которые нормальны с математическим ожиданием x и дисперсией at , где $a > 0$ — некоторая постоянная. Эти плотности удовлетворяют обычному уравнению диффузии

$$\frac{\partial q_t(x, y)}{\partial t} = \frac{1}{2} a \frac{\partial^2 q_t(x, y)}{\partial x^2}. \quad (4.1)$$

Мы покажем теперь, что и другие переходные вероятности удовлетворяют соответствующим уравнениям в частных производных. Цель этого вывода — дать представление о типах процессов и о возникающих задачах, т. е. он должен служить начальным шагом. Вследствие этого мы не будем тратить усилий на достижение общности и полноты изложения.

Из предположенной нормальности распределений очевидно, что в броуновском движении приращения за короткий промежуток времени длины t обладают следующими свойствами: (i) при фиксированном $\delta > 0$ вероятность смещения, превосходящего δ , равна $o(t)$; (ii) математическое ожидание смещения равно нулю; (iii) его дисперсия равна at . Мы сохраним первое требование, а два других приспособим к условиям неоднородной среды. Иначе говоря, мы допустим, что a зависит от x , и допустим ненулевое среднее смещение. В этой обстановке математическое ожидание и дисперсия смещения уже не будут в точности пропорциональны t , и мы можем лишь постулировать, что при данном $X(\tau) = x$ смещение $X(t + \tau) - X(\tau)$ имеет математическое ожидание $b(x)t + o(t)$ и дисперсию $a(x)t + o(t)$. Моменты не обязаны существовать, но, имея в виду первое требование, естественно рассмотреть урезанные моменты. Мы приходим к следующим постулатам.

Постулаты¹⁾ относительно переходных вероятностей Q_t . При каждом $\delta > 0$ при $t \rightarrow 0$

$$\frac{1}{t} \int_{|y-x| \geq \delta} Q_t(x, dy) \rightarrow 0, \quad (4.2)$$

$$\frac{1}{t} \int_{|y-x| < \delta} (y-x) Q_t(x, dy) \rightarrow b(x), \quad (4.3)$$

$$\frac{1}{t} \int_{|y-x| < \delta} (y-x)^2 Q_t(x, dy) \rightarrow a(x). \quad (4.4)$$

Заметим, что если (4.2) верно при *всех* $\delta > 0$, то асимптотическое поведение величин в левой части (4.3) и (4.4) не зависит от δ . Возможно поэтому в последних двух соотношениях заменить δ на 1.

Первое условие делает мало вероятными большие смещения и было введено в 1936 г. в надежде, что оно необходимо и достаточно для непрерывности выборочных функций²⁾. Оно было названо именем Линдеберга ввиду сходства с его условием применимости центральной предельной теоремы. Можно показать, что при довольно слабых предположениях о регулярности переходных вероятностей существование пределов в (4.3) и (4.4) на самом деле вытекает из условия Линдеберга (4.2). Мы не будем обсуждать подобные детали, так как сейчас мы не интересуемся систе-

¹⁾ Первый вывод (4.1) из вероятностных предположений принадлежит Эйнштейну. Первый вывод обратного уравнения (4.6) и прямого уравнения (5.2) был дан в известной работе А. Н. Колмогорова 1931 г. (см. § 2). Усовершенствованный вариант постулатов, данный здесь, принадлежит Феллеру (1936). Феллер же доказал первые теоремы существования и исследовал связь между двумя уравнениями.

²⁾ Это предположение было подтверждено Д. Рэем (D. Ray).

матическим развитием теории¹⁾. Наша цель очень умеренна — объяснить природу и эмпирический смысл диффузионных уравнений в простейшей ситуации. На пути к ней мы покажем формально, как из (4.2) — (4.4) можно вывести некоторые дифференциальные уравнения. Мы, однако, не будем обсуждать, когда существует решение этих уравнений²⁾. Поэтому мы допустим, что коэффициенты a и b суть ограниченные непрерывные функции и $a(x) > 0$.

В качестве пространства состояний мы возьмем конечный или бесконечный интервал I на прямой и, как и ранее, будем придерживаться следующего соглашения: если пределы не указаны, то интегрирование производится по всем I . Для упрощения записи и в качестве подготовительного шага для применения теории полугрупп мы введем преобразования

$$u(t, x) = \int Q_t(x, dy) u_0(y), \quad (4.5)$$

переводящие (при фиксированном t) ограниченную непрерывную «начальную функцию» u_0 в функцию³⁾ со значениями $u(t, x)$.

Ясно, что, зная правую часть (4.5) для всех начальных функций u_0 , мы можем однозначно определить Q_t . Теперь мы покажем, что при очень слабых условиях регулярности функция u должна удовлетворять обратному уравнению

$$\frac{\partial u}{\partial t} = \frac{1}{2} a \frac{\partial^2 u}{\partial x^2} + b \frac{\partial u}{\partial x}, \quad (4.6)$$

обобщающему стандартное уравнение диффузии (4.1). Мы ищем функцию u , удовлетворяющую (4.6) и такую, что $u(t, x) \rightarrow u_0(x)$ при $t \rightarrow 0$. В случае единственности это решение необходимо имеет вид (4.5) и Q_t называется функцией Грина данного уравнения. Случаи, когда единственности нет, будут рассмотрены в следующем параграфе.

Чтобы получить обратное уравнение (4.6), мы отправимся от тождества

$$u(s+t, x) = \int Q_s(x, dy) u(t, y), \quad s, t > 0, \quad (4.7)$$

которое очевидным образом следует из уравнения Чепмена—Кол-

¹⁾ Современная теория полугрупп позволила автору получить наиболее общую форму обратного уравнения (или производящего оператора) для марковских процессов, удовлетворяющих требованию типа условия Линдеберга. Классические дифференциальные операторы заменяются при этом их модернизированными вариантами: роль коэффициента b играет «естественный масштаб» (natural scale), а роль a играет «мера скорости» (speed measure). Такие процессы были объектом плодотворных исследований Дынкина и его школы, с одной стороны, и К. Ито и Г. Маккина — с другой. Теория изложена в книгах авторов.

²⁾ Исследование диффузионных уравнений с помощью преобразования Лапласа в гл. XIV, 5.

³⁾ В терминах теории случайных процессов $u(t, x)$ есть условное математическое ожидание $u_0(X(t))$ при гипотезе $X(0) = x$.

могорова (1.3). При $h > 0$ из (4.7) выводим

$$\frac{u(t+h, x) - u(t, x)}{h} = \frac{1}{h} \int Q_h(x, dy) [u(t, y) - u(t, x)]. \quad (4.8)$$

Допустим теперь, что вероятности перехода Q_t достаточно регулярны, во всяком случае, что функция u в (4.5) имеет две ограниченные непрерывные производные по x (хотя бы для бесконечно дифференцируемой u_0). При фиксированных x и $\varepsilon > 0$ найдется в силу формулы Тейлора такое $\delta > 0$, что

$$\left| u(t, y) - u(t, x) - (y-x) \frac{\partial u(t, x)}{\partial x} - \frac{1}{2} (y-x)^2 \frac{\partial^2 u(t, x)}{\partial x^2} \right| < \varepsilon |y-x|^2 \quad (4.9)$$

при всех $|y-x| \leq \delta$. Выбрав таким образом δ , рассмотрим в (4.8) отдельно интегралы по области $|y-x| > \delta$ и области $|y-x| \leq \delta$. Первый интеграл стремится к нулю в силу условия Линдберга (4.2) и ограниченности u . Благодаря условиям (4.3) и (4.4) ясно, что при достаточно малых h второй интеграл отличается от правой части (4.6) менее чем на $\varepsilon \cdot a(x)$. Так как ε произвольно, это означает, что при $h \rightarrow 0$ правая часть (4.8) сходится к правой части (4.6). В соответствии с этим существует по крайней мере правая производная $\frac{\partial u}{\partial t}$, и она удовлетворяет (4.6). Главный вывод изложенной теории состоит, грубо говоря, в следующем. Если переходные вероятности марковского процесса удовлетворяют условию непрерывности (4.2), то процесс определяется коэффициентами b и a . Это утверждение звучит как чисто теоретическое, но в практических ситуациях именно коэффициенты b и a задаются априори, исходя из их эмпирического смысла и характера процесса.

Для объяснения смысла b и a рассмотрим приращение $X(t+\tau) - X(\tau)$ за короткий промежуток времени, предполагая, что $X(\tau) = x$. Если бы моменты в (4.3) и (4.4) не были урезаны, это приращение имело бы условное математическое ожидание $b(x)t + o(t)$ и условную дисперсию $a(x)t - b^2(x)t^2 + o(t) = a(x)t + o(t)$. Таким образом, $b(x)$ служит мерой локальной средней скорости смещения (которая может быть нулем по причинам симметрии), а $a(x)$ служит мерой дисперсии. Мы будем называть b *инфинитезимальной скоростью* (сносом), а a — *инфинитезимальной дисперсией*.

Следующие примеры показывают, как определяются эти коэффициенты в конкретных ситуациях.

Примеры. а) *Броуновское движение.* Предположим, что пространство (ось x) однородно и симметрично. Тогда $a(x)$ не должно зависеть от x , а $b(x)$ должно быть нулем. Мы приходим к классическому уравнению диффузии (4.1).

б) *Процесс Орнштейна—Уленбека* получается при воздействии упругой силы на совершающую броуновское движение частицу. Аналитически это означает наличие сноса по направлению к

началу координат, причем величина сноса пропорциональна расстоянию от начала, т. е. $b(x) = -\rho x$. Так как это не изменяет инфинитезимальной дисперсии, то функция $a(x)$ остается постоянной, скажем равной единице. Обратное уравнение принимает вид

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \frac{\partial^2 u(t, x)}{\partial x^2} - \rho x \frac{\partial u(t, x)}{\partial x}. \quad (4.10)$$

Это уравнение решается удивительно легко. В самом деле, замена

$$v(t, x) = u(t, xe^{\rho t})$$

приводит его к виду

$$e^{2\rho t} \frac{\partial v}{\partial t} = \frac{1}{2} \frac{\partial^2 v}{\partial x^2}, \quad (4.11)$$

и последующая замена

$$\tau = \frac{1 - e^{-2\rho t}}{2\rho} \quad (4.12)$$

превращает (4.11) в стандартное уравнение диффузии (4.1). Отсюда следует, что *переходные плотности $q_t(x, y)$ процесса Орнштейна—Уленбека нормальны с математическим ожиданием $xe^{-\rho t}$ и дисперсией τ , определяемой по (4.12).*

В примере гл. III, 8, д) было показано, что процесс Орнштейна—Уленбека, описываемый уравнением (4.10) и нормальным начальным распределением, является единственным *стационарным нормальным марковским процессом* со стационарными переходными вероятностями. (Броуновское движение включается сюда как частный случай при $\rho = 0$.)

в) *Диффузия в генетике.* Рассмотрим популяцию, включающую различные поколения и имеющую постоянный размер N (типичный пример — поле зерновых культур). Имеется $2N$ генов, каждый из них принадлежит одному из генотипов. Обозначим X_n долю генов типа A . Если нет преимуществ при отборе и не рассматривается возможность мутаций, то гены $(n+1)$ -го поколения могут рассматриваться как случайная выборка объема $2N$ из совокупности генов n -го поколения. Тогда X_n будет марковским процессом, $0 \leq X_n \leq 1$, и при условии, что $X_n = x$, распределение величины $2NX_{n+1}$ будет биномиальным с математическим ожиданием $2Nx$ и дисперсией $2Nx(1-x)$. Изменение за одно поколение будет иметь математическое ожидание 0 и дисперсию, пропорциональную $x(1-x)$.

Допустим теперь, что мы прослеживаем огромное число поколений и вводим такую шкалу, что изменения представляются непрерывными. При такой аппроксимации мы будем иметь дело с марковским процессом, переходные вероятности которого удовлетворяют нашим основным условиям с $b(x) = 0$ и $a(x)$, пропорциональной $x(1-x)$. Множитель пропорциональности зависит от выбора единицы измерения времени. Его можно считать равным 1.

Тогда (4.6) принимает вид

$$\frac{\partial u(t, x)}{\partial t} = x(1-x) \frac{\partial^2 u(t, x)}{\partial x^2}, \quad (4.13)$$

и на этот раз процесс протекает на конечном интервале $\overline{0, 1}$. Влияние селекции и мутаций породило бы снос и привело бы к появлению в уравнении (4.13) члена первого порядка. Генетическую модель, математически равносильную нашей, предложили Р. Фишер и С. Райт, однако их рассуждения были другими. Генетические выводы в какой-то степени сомнительны в силу предположенной неизменности размера популяции. Эффект этого допущения не всегда правильно оценивается. Правильная модель¹⁾ приводит к уравнению с двумя пространственными переменными (частота генов и размер популяции).

г) *Рост популяции.* Мы желаем описать рост большой популяции, в которой индивидуумы статистически независимы и скорость размножения не зависит от размера популяции. Для очень большой популяции процесс приближенно непрерывен, т. е. описывается диффузионным уравнением. Из независимости индивидуумов следует, что инфинитезимальная скорость и дисперсия пропорциональны размеру популяции. Таким образом, процесс описывается обратным уравнением (4.6) с $a = \alpha x$ и $b = \beta x$. Постоянные α и β зависят от выбора единиц измерения времени и размера популяции, и при подходящем выборе можно считать $\alpha = 1$ и $\beta = 1, -1$ или 0 (в зависимости от знака скорости).

В 1; гл. XVII, (5.7), этот же самый рост популяции описывался дискретной моделью. Предполагалось, что при условии $X(\tau) = n$ вероятности событий $X(t + \tau) = n + 1$, $n - 1$ и n отличаются от $\lambda n t$, $\mu n t$ и $1 - (\lambda + \mu) n t$ соответственно членами вида $o(t)$, так что инфинитезимальная скорость и дисперсия равны $(\lambda - \mu)n$ и $(\lambda + \mu)n$. Диффузионный процесс получается простым переходом к пределу. Можно показать, что его переходные вероятности получаются предельным переходом из вероятностей в дискретной модели.

Подобная аппроксимация дискретных процессов диффузионными процессами часто бывает весьма практичной. Типичным примером может служить переход от обычных случайных блужданий к диффузионным процессам, описанный в 1; XIV, 6 [продолжение см. в примере 5, а)].

§ 5. ПРЯМОЕ УРАВНЕНИЕ. ГРАНИЧНЫЕ УСЛОВИЯ

В этом параграфе мы для простоты предположим, что переходные вероятности Q_t обладают плотностью вероятности q_t , которая определяется плотностью стохастического ядра $q_t(x, y)$.

¹⁾ Feller W., Second Berkeley Symposium on Math. Statist. and Probability, 1951, 227—246.

Преобразование (4.5) и последующее обратное уравнение (4.6) описывают переходные вероятности как функции от начального положения x . С вероятностной точки зрения представляется более естественным считать начальную точку x фиксированной и рассмотреть $q_t(x, y)$ как функцию конечной точки y . С этой точки зрения преобразование (4.5) должно быть заменено преобразованием

$$v(s, y) = \int v_0(x) q_s(x, y) dx. \quad (5.1)$$

Здесь v_0 — произвольная плотность вероятности. Из стохастического свойства q_t вытекает, что при любом фиксированном $s > 0$ преобразованная величина v снова является плотностью вероятности. Иными словами, тогда как преобразование (4.5) переводит непрерывные функции в непрерывные, новое преобразование превращает плотности вероятности в новые плотности.

В предыдущем параграфе мы смогли показать с помощью вероятностных рассуждений, что преобразованная функция (4.5) удовлетворяет обратному уравнению (4.6). Даже при том, что новое преобразование более естественно с вероятностной точки зрения, аналогичный непосредственный вывод прямого уравнения невозможен. Однако общая теория дифференциальных уравнений в частных производных делает правдоподобным заключение о том, что v должно удовлетворять уравнению¹⁾

$$\frac{\partial v(s, y)}{\partial s} = \frac{1}{2} \frac{\partial^2}{\partial y^2} [a(y) v(s, y)] - \frac{\partial}{\partial y} [b(y) v(s, y)]. \quad (5.2)$$

В теории вероятностей последнее известно под названием *прямого уравнения* или уравнения *Фоккера — Планка*.

¹⁾ Приводим неформальный набросок доказательства (5.2). Из уравнения Чепмена — Колмогорова (1.3) для переходных вероятностей следует, что интеграл

$$\int v(s, y) u(t, y) dy$$

зависит только от суммы $s+t$. Тогда

$$\int \frac{\partial v(s, y)}{\partial s} u(t, y) dy = \int v(s, y) \frac{\partial u(t, y)}{\partial t} dy. \quad (*)$$

Выразим теперь $\partial u / \partial t$ в соответствии с обратным уравнением (4.6) и применим к полученному в правой части (*) интегралу интегрирование по частям. Если обозначить правую часть уравнения (5.2) через $R(s, y)$, то можно сделать вывод, что (*) равно

$$\int R(s, y) u(t, y) dy$$

плюс некоторое выражение, зависящее только от величин u, v и их производных, вычисленных на границах (или на бесконечности). При подходящих условиях эти граничные слагаемые могут быть отброшены и переход к пределу при $t \rightarrow 0$ приводит в таком случае к равенству

$$\int \left[\frac{\partial v(s, y)}{\partial s} - R(s, y) \right] u_0(y) dy = 0.$$

Сейчас мы покажем, какого рода информацию можно получить из (5.2) более легким путем, чем из обратного уравнения.

Пример. а) Рост популяции. Пример 4, г) приводит к прямому уравнению

$$\frac{\partial v(s, y)}{\partial s} = \alpha \frac{\partial^2 yv(s, y)}{\partial y^2} - \beta \frac{\partial yv(s, y)}{\partial y}. \quad (5.3)$$

Можно доказать, что при данной начальной плотности v_0 существует лишь одно решение. Хотя для него трудно получить явные формулы, можно получить массу информации о нем прямо из уравнения. Например, чтобы вычислить математическое ожидание размера популяции $M(s)$, можно умножить (5.3) на y и проинтегрировать обе части по y от 0 до ∞ . Слева мы получим при этом производную $M'(s)$. Предполагая, что v стремится к нулю на бесконечности быстрее, чем $1/y^2$, и производя интегрирование по частям, мы увидим, что правая часть равна $\beta M(s)$. Таким образом,

$$M'(s) = \beta M(s),$$

так что $M(s)$ пропорционально $e^{\beta s}$. Аналогичные формальные выкладки показывают, что дисперсия пропорциональна $2\alpha\beta^{-1}e^{\beta s} \times (e^{\beta s} - 1)$ [сравните с аналогичным результатом для дискретного случая в 1; гл. XVII, формулы (5.10) и (10.9)]. Конечно, эти выкладки требовали бы обоснования, но и на эвристическом уровне их нельзя было бы получить из обратного уравнения, минуя явное вычисление q_t . ►

Связь между прямым и обратным уравнениями подобна той, которая описана в § 3 для скачкообразных процессов. Мы дадим здесь без доказательства краткое перечисление результатов.

Рассмотрим обратное уравнение (4.6) на открытом интервале x_1, x_2 , который может быть конечным или бесконечным. Мы предполагаем, конечно, что $a > 0$ и что коэффициенты a и b достаточно регулярны для того, чтобы (5.2) имело смысл. При этих условиях существует *единственное минимальное решение* Q_t , такое,

Если это справедливо для произвольной функции u_0 , то выражение внутри квадратных скобок должно быть равно 0, т. е. (5.2) должно быть справедливо.

Это рассуждение оправдывается во многих имеющих практический интерес ситуациях, и тогда прямое уравнение (5.2) справедливо. Однако в так называемых процессах с возвращением [см. 1; XVII, 9] граничные члены, которыми мы пренебрегли, играют роль. Поэтому переходные вероятности таких процессов удовлетворяют обратным уравнениям (4.6), но не удовлетворяют (5.2); прямое уравнение имеет в этом случае другой вид.

Заслуживает внимания также и то, что (5.6) бессмысленно без предположения о дифференцируемости a и b , тогда как никаких подобных ограничений по отношению к обратному уравнению нет. Справедливое прямое уравнение может быть выписано также и тогда, когда a и b не дифференцируемы, но тогда оно включает обобщенные дифференциальные операторы, о которых было упомянуто в сноске 1 на стр. 380.

что (4.5) дает решение обратного уравнения (4.6). При фиксированных t и x ядро $Q_t(x, \Gamma)$ может определять *несобственное распределение*. В любых обстоятельствах распределения Q_t имеют плотности, и функция v , определяемая в (5.1), удовлетворяет прямому уравнению. На самом деле это решение будет минимальным в очевидном смысле слова (который был уточнен в § 3). В таком понимании прямое уравнение является следствием обратного уравнения. Однако эти уравнения определяют процесс единственным образом только в случае, когда минимальное решение собственно. Во всех других случаях природа процесса определяется дополнительными граничными условиями.

Характер граничных условий лучше всего может быть понят по аналогии с простым случайным блужданием на $0, \infty$, обсуждавшимся в 1; гл. XIV. Можно принять различные соглашения относительно течения процесса после его возвращения в нуль. В задаче о разорении процесс на этом останавливается. В этом случае говорят, что нуль является *поглощающим экраном*. С другой стороны, если нуль представляет собой *отражающий экран*, то частица мгновенно возвращается в точку 1 и процесс, таким образом, продолжается бесконечно. Следует подчеркнуть, что граничные условия нужны тогда и только тогда, когда граничная точка достижима. Событие «граничная точка x_2 достигается ранее момента t » вполне определено для диффузионных процессов ввиду непрерывности их траекторий. Оно тесно связано по своему характеру с событием «до момента t происходит бесконечно много скачков» в скачкообразных процессах.

В некоторых диффузионных процессах с вероятностью единица не достигается ни одна из граничных точек. Таково, например, броуновское движение [пример 4, а)]. В подобных случаях минимальное решение является собственным распределением вероятностей и не существует иных решений. В других ситуациях процесс описывается минимальным решением лишь до момента достижения границы. Оно соответствует поглощающим границам, т. е. описывает процесс, который останавливается по достижении границ. Это наиболее важный тип процессов, и не только потому, что все остальные получаются из него, но и потому, что все вероятности *первого прохождения* состояний могут быть вычислены искусственным введением поглощающих барьеров. Этот метод широко применим, но мы объясним его на простейшем примере (он уже был нами неявно использован при изучении случайных блужданий и в других местах, см., например, задачу 18 в 1; гл. XVII, 10).

В последующих примерах мы сосредоточим наше внимание на простом уравнении (5.4). Более общие уравнения диффузии будут рассмотрены в гл. XIV, 5, с помощью методов, использующих преобразования Лапласа.

Примеры. б) Один поглощающий барьер. Время первого прохождения. Рассмотрим броуновское движение на $\overline{0, \infty}$ с поглощающим барьером в нуле. Более точно, броуновское движение, начинающееся в точке $x > 0$ в момент 0, останавливается в момент первого попадания в нуль. По причинам симметрии и обратное, и прямое уравнения принимают вид классического уравнения диффузии

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}. \quad (5.4)$$

Нужное граничное условие таково: $q_t(0, y) = 0$ при всех t (так же как и в примере случайных блужданий). Проверить это можно либо предельным переходом из формул 1; гл. XIV, 6, или исходя из минимального характера соответствующего решения.

Найдем решение (5.4), определенное при $t \geq 0$, $x \geq 0$, и такое, что $u(0, x) = u_0(x)$ и $u(t, 0) = 0$, где u_0 — заданная функция. Его построение опирается на метод отражений, введенный Кельвином¹⁾. Доопределим u_0 на левой полуоси равенством $u_0(-x) = -u_0(x)$ и решим (5.4) с этим начальным условием на $-\infty, \infty$. По причинам симметрии решение должно удовлетворять условию $u(t, 0) = 0$. Ограничиваясь теперь лишь положительными x , мы найдем требуемое решение. Оно равно интегралу по $\overline{0, \infty}$ от функции $u_0(y) q_t(x, y)$, где

$$q_t(x, y) = \frac{1}{\sqrt{2\pi t}} \left[\exp\left(-\frac{(y-x)^2}{2t}\right) - \exp\left(-\frac{(y+x)^2}{2t}\right) \right]. \quad (5.5)$$

Таким образом, функции q_t совпадают с плотностями переходных вероятностей нашего процесса ($t > 0$, $x > 0$, $y > 0$). Легко видеть, что при фиксированном y q_t будет решением (5.4) с граничным условием $q_t(0, y) = 0$ [другой вывод, использующий общий прием, см. в примере 5, а) гл. XIV].

Интегрированием по y находим вероятность находиться в момент t в состоянии, отличном от нуля

$$\int_0^{\infty} q_t(x, y) dy = \mathfrak{N}\left(\frac{x}{\sqrt{t}}\right) - 1, \quad (5.6)$$

где $\mathfrak{N}$ обозначает стандартное нормальное распределение. Иными словами, (5.6) дает вероятность того, что процесс, начинающийся в точке $x > 0$, не достигнет нуля до момента t . Поэтому (5.6) можно истолковать и как распределение времен первого прохождения в «свободном» броуновском движении (т. е. при отсутствии барьеров). Отметим, что функцию (5.6) можно охарактеризовать как решение дифференциального уравнения (5.4), определенное

¹⁾ См. задачи 15—18 в 1; гл. XIV, 9, где этот же метод применяется к разностным уравнениям,

при $x > 0$ и удовлетворяющее начальному условию $u(0, x) = 1$ и граничному условию $u(t, 0) = 0$.

[Читатель может узнать в (5.6) устойчивое распределение с показателем $\alpha = 1/2$; этот же результат был получен в гл. VI, 2, предельным переходом от случайных блужданий к броуновскому движению.]

в) *Два поглощающих барьера.* Рассмотрим теперь броуновское движение при наличии двух поглощающих барьеров — в точках 0 и $a > 0$. Это означает, что при фиксированном $0 < y < a$ переходные плотности q_t должны удовлетворять дифференциальному уравнению (5.4) и граничным условиям $q_t(0, y) = q_t(a, y) = 0$.

Легко видеть, что решение имеет вид¹⁾

$$q_t(x, y) = \frac{1}{\sqrt{2\pi t}} \sum_{k=-\infty}^{+\infty} \left\{ \exp\left(-\frac{(y-x+2ka)^2}{2t}\right) - \exp\left(-\frac{(y+x+2ka)^2}{2t}\right) \right\}, \quad (5.7)$$

где $0 < x, y < a$. В самом деле, ряд (5.7) очевидным образом сходится и приведение подобных членов показывает, что $q_t(0, y) = q_t(a, y) = 0$ при всех $t > 0$ и $0 < y < a$. [Преобразование Лапласа для (5.7) см. в гл. XIV, (5.17).]

Интегрируя (5.7) по $0 < y < a$, найдем, что «непоглощенная» вероятностная масса в момент t равна

$$\lambda_a(t, x) = \sum_{k=-\infty}^{\infty} \left\{ \mathfrak{N}\left(\frac{2ka+a-x}{\sqrt{t}}\right) - \mathfrak{N}\left(\frac{2ka-x}{\sqrt{t}}\right) - \mathfrak{N}\left(\frac{2ka+a+x}{\sqrt{t}}\right) + \mathfrak{N}\left(\frac{2ka+x}{\sqrt{t}}\right) \right\}. \quad (5.8)$$

Эта формула дает вероятность того, что *частица, выходящая из точки x , не будет поглощена до момента t .*

Функция λ_a является решением дифференциального уравнения (5.4), стремящимся к 1 при $t \rightarrow 0$ и удовлетворяющим граничным условиям $\lambda_a(t, 0) = \lambda_a(t, a) = 0$. Это решение можно получить также стандартным применением метода Фурье. При этом оно имеет вид²⁾

$$\lambda_a(t, x) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \exp\left(-\frac{(2n+1)^2 \pi^2}{2a^2} t\right) \sin \frac{(2n+1) \pi x}{a}. \quad (5.9)$$

¹⁾ Формула получается последовательными приближениями с использованием повторных отражений. В (5.5) указано решение, удовлетворяющее граничным условиям в точке 0 (но не в точке a). Отражение в a приводит к четырехчленному решению, удовлетворяющему граничному условию в a (но не в нуле). Чередующиеся отражения в 0 и в a приводят к пределу (5.7). Аналогичное решение для случайных блужданий дано в 1; гл. XIV, (9.1); откуда (5.7) может быть получено предельным переходом (см. 1; гл. XIV, 6).

²⁾ Аналогичная формула для случайных блужданий получена в 1; гл. XIV, 5, где, однако, граничные условия имели вид $\lambda_a(t, 0) = 1$ и $\lambda_a(t, a) = 0$.

Таким образом, мы получили два совершенно различных представления¹⁾ для одной и той же функции λ_a . Это благоприятно для вычислений, так как ряд (5.8) удовлетворительно сходится лишь при малых t , а (5.9) применимо при больших t .

Можно дать другую интерпретацию λ_a . Рассмотрим положение $X(t)$ частицы, находящейся в свободном броуновском движении, начинающемся в нуле. Событие: частица за время $0, t$ не выходит из интервала $-a/2, a/2$, равносильно событию: в процессе с поглощающими границами $\pm a/2$, начинающемся в 0, поглощение не происходит до момента t . Таким образом, величина $\lambda_a(t, a/2)$ равна вероятности того, что в свободном (неограниченном) броуновском движении, начинающемся в нуле, $|X(s)| < a/2$ при всех s из интервала $0 < s < t$.

г) Применение к предельным теоремам и критериям Колмогорова—Смирнова. Пусть $Y_1, Y_2, \dots$ — независимые случайные величины с одним и тем же распределением вероятностей, $E(Y_i) = 0$, $E(Y_i^2) = 1$. Положим $S_n = Y_1 + \dots + Y_n$. Пусть $T_n = \max[|S_1|, \dots, |S_n|]$. Центральная предельная теорема делает правдоподобным заключение, что асимптотическое поведение T_n будет таким же, как если бы Y_i были нормальны. В последнем же случае нормированную сумму $S_n/\sqrt{n}$ можно интерпретировать как значение процесса броуновского движения в момент времени k/n ($k=0, 1, \dots, n$). Вероятность того, что это броуновское движение остается в интервале $-a/2, a/2$, равна, как следует из сказанного ранее, $\lambda_a(1, a/2)$. Наши правдоподобные рассуждения поэтому приводят к гипотезе, что при $n \rightarrow \infty$

$$P\{T_n < z\} \rightarrow L(z), \quad (5.10)$$

где $L(z) = \lambda_{az}(1, z)$ находится с помощью (5.8) и (5.9):

$$\begin{aligned} L(z) &= 2 \sum_{k=-\infty}^{\infty} \{ \mathfrak{N}((4k+1)z) - \mathfrak{N}((4k-1)z) \} = \\ &= \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \exp\left(-\frac{(2n+1)^2 \pi^2}{8z^2}\right). \end{aligned} \quad (5.11)$$

Эта гипотеза была доказана в 1946 г. Эрдешем и Кацем. Лежащая в основе доказательства идея получила с тех пор название принципа инвариантности. Этот принцип устанавливает, грубо говоря, что предельное распределение некоторых функций от случайных величин нечувствительно к изменениям распреде-

¹⁾ Равенство выражений (5.8) и (5.9) может служить стандартным примером формулы суммирования Пуассона [см. гл. XIX, (5.8)]. Открытое первоначально в связи с теорией Якоби преобразования η -эта-функций, это равенство приобрело некоторую историческую известность. См. теорему 277 в книге Ландау (Landau E., Verteilung, der Primzahlen, 1909).

ления этих величин и что оно может быть найдено построением надлежащего аппроксимирующего случайного процесса. Принцип инвариантности был усовершенствован М. Донскером, П. Биллингсли, Ю. Прохоровым и другими и стал мощным орудием доказательства предельных теорем.

По сходным причинам распределение (5.11) играет важную роль в обширной литературе о непараметрических критериях описанного в гл. I, 12, типа¹⁾.

д) *Отражающие экраны.* По аналогии с обычным случайным блужданием при отражающем экране в нуле следует ввести граничное условие $\frac{\partial q_1(0, y)}{\partial y} = 0$. Легко проверить, что решение для интервала $\overline{0, \infty}$ дается формулой (5.5), где знак минус заменен на плюс. Формальный вывод с применением метода отражений остается таким же, как и раньше, с той разницей, что теперь мы полагаем $u_0(-x) = u_0(x)$. Решение для интервала $\overline{0, a}$ при отражающих экранах в 0 и в a получается заменой знака минус на знак плюс в (5.7) (другое выражение для решения, получаемое применением метода Фурье или формулы суммирования Пуассона, дано в задаче 11 гл. XIX, 9).

Нужно отметить, что в случае отражающих экранов q_1 будет плотностью собственного распределения вероятностей.

§ 6. ДИФФУЗИЯ В МНОГОМЕРНОМ СЛУЧАЕ

Предыдущая теория легко переносится на двумерный случай. Чтобы не загромождать формулы индексами, мы будем обозначать координатные величины через $(X(t), Y(t))$, а *плотности* перехода — через $q_1(x, y; \xi, \eta)$; здесь (x, y) — начальная точка и q_1 — плотность в точке (ξ, η) . Сохраняются постулаты § 4 с той разницей, что инфинитезимальная скорость $b(x)$ заменяется вектором, а дисперсия $a(x)$ — матрицей ковариаций. Вместо (4.6) мы получим тогда *обратное уравнение диффузии*

$$\frac{du}{dt} = a_{11} \frac{\partial^2 u}{\partial x^2} + 2a_{12} \frac{\partial^2 u}{\partial x \partial y} + a_{22} \frac{\partial^2 u}{\partial y^2} + b_1 \frac{\partial u}{\partial x} + b_2 \frac{\partial u}{\partial y} \quad (6.1)$$

с коэффициентами, зависящими от x и y . В случае двумерного броуновского движения мы налагаем условие круговой симметрии, что приводит (с точностью до несущественной константы) к урав-

¹⁾ Эта тема сравнительно нова, и тем не менее происхождение часто используемого тождества (5.11) уже успели, как кажется, забыть. [См.: Renyi A., On the distribution function, *L(z)*. Selected Translation in Math. Statist. and Probability, 4 (1963), 219—224. Новое доказательство Реньи опирается на классические рассуждения с привлечением зета-функций и тем самым затмевает простой вероятностный смысл (5.11).]

нению

$$\frac{\partial u}{\partial t} = \frac{1}{2} \left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right]. \quad (6.2)$$

Соответствующие переходные плотности нормальны с дисперсией t и центром в (x, y) . Очевидное разложение этих плотностей на множители показывает, что $X(t)$ и $Y(t)$ статистически независимы.

Наиболее интересной из величин, связанных с этим процессом, является расстояние $R(t)$ от начала координат ($R^2 = X^2 + Y^2$). Интуитивно ясно, что величины $R(t)$ образуют одномерный диффузионный процесс. Интересно сравнить различные способы вывода соответствующего диффузионного уравнения. Нормальные переходные плотности, отвечающие (6.2), в полярных координатах имеют вид

$$\frac{\rho}{2\pi t} \exp \left(-\frac{\rho^2 + r^2 - 2\rho r \cos(\theta - \alpha)}{2t} \right) \quad (6.3)$$

(где $x = r \cos \alpha$ и т. д.). При данных r и α в момент 0 частная плотность $R(t)$ получается интегрированием (6.3) по θ . Параметр α исчезает, и мы получаем¹⁾ переходные плотности процесса $R(t)$:

$$\omega_t(r, \rho) = \frac{1}{t} \exp \left(-\frac{r^2 + \rho^2}{2t} \right) I_0 \left(\frac{r\rho}{t} \right), \quad (6.4)$$

где I_0 — функция Бесселя, определенная в гл. II, (7.1). Здесь r обозначает начальное положение в момент 0. Из самого вывода ясно, что при фиксированном ρ переходные вероятности ω_t должны удовлетворять (6.2) в полярных координатах; иными словами, должно быть

$$\frac{\partial \omega_t}{\partial t} = \frac{1}{2} \left(\frac{\partial^2 \omega_t}{\partial r^2} + \frac{1}{r} \frac{\partial \omega_t}{\partial r} \right). \quad (6.5)$$

Это — обратное уравнение для процесса $R(t)$. Оно просто выводится из (6.2) и требования круговой симметрии. Уравнение (6.5) показывает, что инфинитезимальная скорость процесса $R(t)$ равна $1/2 r$. Существование сноса от начальной точки можно объяснить следующим образом. Рассмотрим двумерное броуновское движение, начавшееся от точки r на оси x . По причинам симметрии ее абсцисса в момент $h > 0$ с одинаковой вероятностью будет $> r$ или $< r$. В первом случае, конечно, $R(h) > r$. Однако это соотношение может выполняться и во втором случае. Таким образом, соотношение $R(h) > r$ имеет вероятность $> 1/2$ и в среднем R обязано возрастать.

Наш вывод формул для переходных вероятностей применим и в случае трех измерений, причем с одним существенным упрощением: якобиан ρ в (6.3) заменяется теперь на $\rho^2 \sin \theta$ и интег-

¹⁾ Интеграл хорошо известен. Стандартная проверка результата состоит в разложении $e^{\cos \theta}$ в степенной ряд по $\cos \theta$.

рал элементарно вычисляется. Вместо (6.4) мы получаем следующую *переходную плотность для процесса $R(t)$ в трех измерениях*:

$$\omega_t(r, \rho) = \frac{1}{\sqrt{2\pi t}} \frac{\rho}{r} \left[\exp\left(-\frac{(\rho-r)^2}{2t}\right) - \exp\left(-\frac{(\rho+r)^2}{2t}\right) \right]. \quad (6.6)$$

(Снова r обозначает начальное положение в момент 0.)

§ 7. ПОДЧИНЕННЫЕ ПРОЦЕССЫ

Пусть $\{X(t)\}$ — марковский процесс со стационарными вероятностями перехода $Q_t(x, \Gamma)$. Исходя из $\{X(t)\}$, можно построить множество новых процессов, вводя то, что называют *рандомизированным операционным временем*.

Предположим, что каждому значению $t > 0$ соответствует случайная величина $T(t)$ с распределением U_t . Новое стохастическое ядро P_t может быть определено в таком случае равенством

$$P_t(x, \Gamma) = \int_0^\infty Q_s(x, \Gamma) U_t\{ds\}. \quad (7.1)$$

Это есть распределение случайной величины $X(T(t))$ при условии, что $X(0) = 0$.

Пример. а) Если $T(t)$ имеет распределение Пуассона с математическим ожиданием αt , то

$$P_t(x, \Gamma) = \sum_{n=0}^{\infty} e^{-\alpha t} \frac{(\alpha t)^n}{n!} Q_n(x, \Gamma). \quad (7.2)$$

Эти P_t являются переходными вероятностями некоторого псевдопуассоновского процесса. В § 9 будет показано, что рандомизация посредством пуассоновских распределений приводит к так называемой показательной формуле теории марковских полугрупп. Теперь мы увидим, что удобные результаты могут быть получены и для множества других распределений, каждое из которых приводит к аналогу показательной формулы. ►

Случайные величины $X(T(t))$ образуют новый случайный процесс, который, вообще говоря, уже не обязан быть марковским. Для того чтобы процесс был марковским, очевидно, необходимо, чтобы P_t удовлетворяли уравнению Чепмена — Колмогорова

$$P_{s+t}(x, \Gamma) = \int_{-\infty}^{+\infty} P_s(x, dy) P_t(y, \Gamma). \quad (7.3)$$

Это означает, что распределение $X(T(t+s))$ получается интегрированием $P_t(y, \Gamma)$ по распределению $X(T(s))$, так что по определению условных вероятностей

$$P_t(y, \Gamma) = P\{X(T(t+s)) \in \Gamma \mid X(T(s)) = y\}. \quad (7.4)$$

Аналогичное вычисление переходных вероятностей более высокого порядка показывает, что условия (7.3) достаточно, чтобы гарантировать марковский характер полученного процесса $\{X(T(t))\}$.

Мы хотим описать распределения U_t , которые приводят к решениям P_t уравнения (7.3). Непосредственный подход к решению этой задачи затруднен, но эти трудности можно обойти, если для начала рассмотреть простой частный случай. Пусть случайные величины $T(t)$ принимают только значения, кратные фиксированному числу $h > 0$. Для распределения $T(t)$ введем обозначение

$$P\{T(t) = nh\} = a_n(t). \quad (7.5)$$

При условии $X(0) = x$ величина $X(T(t))$ имеет распределение

$$P_t(x, \Gamma) = \sum_{k=0}^{\infty} a_k(t) Q_{kh}(x, \Gamma). \quad (7.6)$$

Так как ядра $\{Q_t\}$ удовлетворяют уравнению Чепмена—Колмогорова, то

$$\int_{-\infty}^{+\infty} P_s(x, dy) P_t(y, \Gamma) = \sum_{i,k} a_j(s) a_k(t) \cdot Q_{(j+k)h}(x, \Gamma). \quad (7.7)$$

Отсюда видно, что ядра P_t удовлетворяют уравнению Чепмена—Колмогорова (7.3) тогда и только тогда, когда

$$a_0(s) a_n(t) + a_1(s) a_{n-1}(t) + \dots + a_n(s) a_0(t) = a_n(s+t) \quad (7.8)$$

при всех $s > 0$ и $t > 0$. Последнее соотношение выполняется для процесса $\{T(t)\}$ со стационарными независимыми приращениями. Общее решение (7.8) было изучено в 1; гл. XII, 2.

Полученный результат приводит к предположению, что и в общем случае уравнение (7.3) будет удовлетворяться всякий раз, когда случайные величины $T(t)$ образуют процесс со стационарными независимыми приращениями, т. е. тогда, когда распределения U_t удовлетворяют¹⁾ соотношению

$$U_{t+s}(x) = \int_0^x U_s(x-y) U_t\{dy\}. \quad (7.9)$$

Мы проверим это предположение переходом к пределу²⁾. Представим U_t как предел последовательности арифметических распределений $U_t^{(n)}$ только что рассмотренного типа: $U_t^{(n)}$ сосредоточены на множестве чисел, кратных числу h_n , причем массы, кото-

¹⁾ Наиболее общее решение (7.9) будет найдено с помощью преобразования Лапласа в гл. XIII, 7. То же самое можно получить также из общей теории безгранично делимых распределений.

²⁾ Непосредственная проверка требует аналитического мастерства. Предлагаемая процедура еще раз показывает, что наивный подход иногда может быть более действенным.

рые $U_i^{(v)}$ приписывают точкам nh_v , удовлетворяют соотношению вида (7.8). Таким образом, при каждом v мы получим ядро $P_i^{(v)}$, соответствующее (7.6), которое удовлетворяет уравнению Чепмена—Колмогорова (7.3). Для того чтобы показать, что ядро P_i из (7.1) также удовлетворяет этому уравнению, достаточно доказать, что $P_i^{(v)} \rightarrow P_i$ или, что означает то же самое, что

$$\int_{-\infty}^{+\infty} P_i^{(v)}(x, dy) f(y) \rightarrow \int_{-\infty}^{+\infty} P_i(x, dy) f(y) \quad (7.10)$$

для каждой непрерывной функции $f(y)$, обращающейся в нуль на бесконечности. Если мы положим

$$F(t, x) = \int_{-\infty}^{+\infty} Q_t(x, dy) f(y), \quad (7.11)$$

то (7.10) можно переписать иначе:

$$\int_0^{\infty} F(s, x) U_i^{(v)} \{ds\} \rightarrow \int_0^{\infty} F(s, x) U_i \{ds\}. \quad (7.12)$$

Это соотношение естественно выполняется, когда F непрерывно, и это накладывает на Q_t очень умеренное условие регулярности. Таким образом, мы получаем следующий основной результат.

Пусть $\{X(t)\}$ —марковский процесс с непрерывными переходными вероятностями Q_t и $\{T(t)\}$ —процесс с неотрицательными независимыми приращениями. Тогда $\{X(T(t))\}$ —также марковский процесс с переходными вероятностями P_t , определяемыми по (7.7). Мы будем говорить, что этот процесс подчинен¹⁾ $\{X(t)\}$ с использованием операционного времени $T(t)$. Процесс $\{T(t)\}$ называется направляющим процессом.

Наиболее интересен частный случай, когда процесс $X(t)$ также имеет независимые приращения. В этом случае переходные вероятности зависят только от разностей $\Gamma-x$ и могут быть заменены соответствующими функциями распределения. Тогда (7.7) принимает более простой вид:

$$P_t(x) = \int_0^{\infty} Q_s(x) U_t \{ds\}. \quad (7.13)$$

Все наши примеры будут именно этого типа.

Примеры. б) Процесс Коши подчинен броуновскому движению. Пусть $\{X(t)\}$ —броуновское движение (винеровский процесс) с переходными плотностями $q_t(x) = (2\pi t)^{-\frac{1}{2}} e^{-\frac{1}{2} x^2/t}$. В качестве $\{T(t)\}$

¹⁾ Понятие подчиненных полугрупп введено С. Бохнером в 1949. Систематическое изложение на «высшем уровне» дано в статье: Nelson E., A functional calculus using singular Laplace integrals, Trans Amer. Math. Soc., 88 (1958), 400—413.

мы возьмем устойчивый процесс с показателем $1/2$ и переходными плотностями

$$u_t(x) = \frac{t}{\sqrt{2\pi} \sqrt{x^3}} e^{-\frac{1}{2} t^2/x}.$$

Распределение (7.13) имеет при этом плотность

$$p_t(x) = \frac{t}{2\pi} \int_{-\infty}^{+\infty} s^{-2} e^{-(x^2+t^2)s^2} e^{(2xs)} ds = \frac{t}{\pi(t^2+x^2)}, \quad (7.14)$$

и таким образом наша процедура «подчинения» приводит к процессу Коши.

Этот результат можно следующим образом интерпретировать в терминах двух независимых броуновских движений $X(t)$ и $Y(t)$. В примере 2, д) гл. VI было показано, что U_t можно интерпретировать как распределение времени ожидания момента, в который процесс $Y(s)$ впервые достигнет значения $t > 0$. Соответственно процесс Коши $Z(t)$ может быть реализован так: случайная величина $Z(t)$ — это значение процесса X в момент $T(t)$, когда $Y(s)$ впервые достигает значения t . [О другой связи процесса Коши с временами достижения в броуновском движении см. в примере 2, е) гл. VI.]

в) *Устойчивые процессы.* Последний пример легко распространяется на произвольные строго устойчивые процессы $\{X(t)\}$ и $\{T(t)\}$ с показателями α и β соответственно. Здесь $\alpha \leq 2$, но так как величина $T(t)$ должна быть положительна, то необходимо $\beta < 1$. Переходные вероятности Q_t и U_t имеют вид $Q_t(x) = Q(xt^{-1/\alpha})$ и $U_t(x) = U(xt^{-1/\beta})$, где Q и U — фиксированные устойчивые распределения. Мы покажем, что подчиненный процесс $X(T(t))$ устойчив с показателем $\alpha\beta$. Это утверждение эквивалентно соотношению $P_{\lambda, t}(x) = P_t(x^{-1/\alpha\beta})$. Это соотношение тривиально выводится из определения Q_s и U_t и формулы (7.13) с применением подстановки $s = y\lambda^{1/\beta}$.

[Наш результат по существу эквивалентен формуле для произведений, установленной в примере 2, з) гл. VI. При $X(t) > 0$ эта формула может быть переписана в терминах преобразований Лапласа, см. гл. XIII, 7, д). Вариант, использующий преобразование Фурье, см. в задаче 9 гл. XVII, 12.]

г) *Обобщенный пуассоновский процесс, направляемый гамма-процессом.* Пусть Q_t — обобщенное пуассоновское распределение, порожаемое распределением вероятностей F , и пусть U_t имеет гамма-плотность $e^{-x}x^{t-1}/\Gamma(t)$. Тогда (7.13) принимает вид

$$P_t = \sum_{n=0}^{\infty} a_n(t) F^{n*}, \quad (7.15)$$

где

$$a_n(t) = \int_0^{\infty} e^{-s} \frac{s^n}{n!} \cdot e^{-s} \frac{s^{t-1}}{\Gamma(t)} ds = \frac{\Gamma(n+t)}{n! \Gamma(t)} \cdot 2^{-n-t}. \quad (7.16)$$

Легко проверить, что вероятностям $a_n(t)$ соответствует безгранично делимая производящая функция

$$\sum a_n(t) \xi^n = (2 - \xi)^{-t}.$$

д) *Гамма-процесс, направляемый законом Пуассона.* Рассмотрим теперь те же самые распределения, поменяв их ролями. Тогда операционное время будет целочисленным, и нуль имеет вес e^{-t} . Отсюда следует, что результирующее распределение также имеет атом веса e^{-t} в нуле. Непрерывная часть имеет плотность

$$\sum_{n=0}^{\infty} e^{-x} \frac{x^{n-1}}{(n-1)!} e^{-t} \frac{t^n}{n!} = e^{-t-x} \sqrt{\frac{t}{x}} I_1(2\sqrt{xt}), \quad (7.17)$$

где I_1 — функция Бесселя (см. гл. II, (7.1)). Из сказанного следует, что это распределение безгранично делимо. Прямая проверка этого утверждения нелегка. ►

§ 8. МАРКОВСКИЕ ПРОЦЕССЫ И ПОЛУГРУППЫ

В гл. VIII были отмечены преимущества трактовки вероятностных распределений как операторов, действующих на непрерывные функции. Преимущества операторного подхода к стохастическим ядрам еще больше, и теория полугрупп приводит к цельной теории марковских процессов, что недостижимо при других методах. Пусть даны стохастическое ядро K в $\mathcal{N}$ и ограниченная непрерывная функция u . Тогда соотношение

$$U(x) = \int_{-\infty}^{+\infty} K(x, dy) u(y) \quad (8.1)$$

определяет нам новую функцию. Мы почти не теряем общности рассуждений, предположив, что функция U также непрерывна. Тогда мы могли бы перейти к изучению свойства ядра K в терминах порожденного им преобразования $u \rightarrow U$ семейства непрерывных функций. По двум приводимым ниже причинам целесообразно рассмотреть более общую ситуацию. Во-первых, преобразования вида (8.1) имеют смысл в произвольных пространствах, и было бы крайне неэкономно развивать теорию, которая не охватывает простейший и очень важный частный случай, а именно процессы со счетным множеством состояний [когда (8.1) сводится к матричному преобразованию]. Во-вторых, даже в теории, ограничивающейся непрерывными функциями на прямой, различные типы граничных условий вынуждают нас вводить специальные

классы непрерывных функций. С другой стороны, большая общность достигается без всяких затрат. Читатели в зависимости от настроения могут пренебречь общностью и рассматривать все теоремы в одной (или нескольких) из следующих типичных ситуаций: (i) пространство состояний Σ есть вещественная прямая и $\mathcal{L}$ есть класс ограниченных непрерывных функций, исчезающих на бесконечности; (ii) пространство Σ есть конечный замкнутый интервал I в $\mathcal{R}^1$ или $\mathcal{R}^2$ и $\mathcal{L}$ — класс непрерывных функций на нем; (iii) Σ состоит из целых чисел и $\mathcal{L}$ — из ограниченных последовательностей. В последнем случае последовательности лучше всего представлять себе в виде векторов-столбцов, а преобразования — в виде матриц.

Как и в гл. VIII, норма ограниченной действительной функции u определяется как $\|u\| = \sup |u(x)|$. Последовательность функций u_n сходится равномерно к u тогда и только тогда, когда $\|u_n - u\| \rightarrow 0$.

Начиная с этого места, $\mathcal{L}$ будет обозначать семейство действительных функций, определенных на некотором множестве Σ и обладающих следующими свойствами: (i) если u_1 и u_2 принадлежат $\mathcal{L}$, то каждая линейная комбинация $c_1 u_1 + c_2 u_2 \in \mathcal{L}$; (ii) если $u_n \in \mathcal{L}$ и $\|u_n - u\| \rightarrow 0$, то $u \in \mathcal{L}$; (iii) если $u \in \mathcal{L}$, то и u^+ , и u^- принадлежат $\mathcal{L}$ (здесь $u = u^+ - u^-$ — обычное разложение u на положительную и отрицательную части). Другими словами, $\mathcal{L}$ замкнуто относительно образования линейных комбинаций, равномерного перехода к пределу и взятия абсолютной величины. Первые два свойства означают, что $\mathcal{L}$ — банахово пространство, последнее — что $\mathcal{L}$ есть структура.

Ниже приводятся стандартные определения. Линейное преобразование T называют эндоморфизмом на $\mathcal{L}$, если каждое $u \in \mathcal{L}$ имеет образ $Tu \in \mathcal{L}$, такой, что $\|Tu\| \leq m\|u\|$, где m — постоянная, не зависящая от u . Наименьшая постоянная с этим свойством называется нормой $\|T\|$ преобразования T . Преобразование T положительно, если из $u \geq 0$ вытекает $Tu \geq 0$. В этом случае $-Tu^- \leq Tu \leq Tu^+$. Сжатием называется положительный оператор T , такой, что $\|T\| \leq 1$. Если постоянная функция 1 принадлежит $\mathcal{L}$ и T — положительный оператор, такой, что $T1 = 1$, то T называется оператором перехода (он автоматически будет сжатием).

Пусть даны два эндоморфизма S и T на $\mathcal{L}$. Тогда их произведение ST есть эндоморфизм, отображающий u в $S(Tu)$. Очевидно, $\|ST\| \leq \|S\| \cdot \|T\|$. В общем случае $ST \neq TS$ в отличие от частного случая операторов свертки из гл. VIII, 3, которые перестановочны один с другим.

Мы всерьез интересуемся лишь преобразованиями вида (8.1), где K — стохастическое или по крайней мере субстохастическое ядро. Операторы такого вида суть сжатия или операторы перехода. Им присуще также

Свойство монотонной сходимости. Если $u_n \geq 0$ и $u_n \uparrow u$ ($u_n \in \mathcal{L}$, $u \in \mathcal{L}$), то $Tu_n \rightarrow Tu$ в смысле поточечной сходимости.

Практически во всех ситуациях сжатия, обладающие этим свойством, имеют вид (8.1). Два примера иллюстрируют это замечание.

Примеры. а) Пусть Σ — действительная прямая и $\mathcal{L} = C$, где C — семейство всех ограниченных непрерывных функций на ней. Пусть $C_0 \subset \mathcal{L}$ — подкласс функций, обращающихся в нуль на $\pm \infty$. Если T — сжатие на $\mathcal{L}$, то при $u \in C_0$ значение $Tu(x)$ функции Tu в фиксированной точке x является положительным линейным функционалом на $\mathcal{L}$. По теореме Рисса о представлении таких функционалов существует (возможно, несобственное) распределение вероятностей F , такое, что $Tu(x)$ равно математическому ожиданию u относительно F . Так как F зависит от x , то $F(\Gamma)$ мы обозначим $K(x, \Gamma)$. Тогда для $u \in C_0$

$$Tu(x) = \int_{-\infty}^{+\infty} K(x, dy) u(y). \quad (8.2)$$

Если оператор T обладает свойством монотонной сходимости, то это соотношение можно распространить на все ограниченные непрерывные функции.

При фиксированном x ядро K как функция от Γ является мерой. Если Γ — открытый интервал и $\{u_n\}$ — возрастающая последовательность непрерывных функций, такая, что $u_n(x) \rightarrow 1$ при $x \in \Gamma$ и $u_n(x) \rightarrow 0$ в других случаях, то $K(x, \Gamma) = \lim Tu_n(x)$ (в силу основных свойств интегралов). Так как функции Tu_n непрерывны, мы заключаем, что при фиксированном Γ ядро K есть бэровская функция от x . Поэтому K обладает всеми свойствами стохастических или субстохастических ядер. То же самое распространяется на случай интервалов или $\mathcal{R}^n$.

б) Пусть Σ — множество целых чисел, а $\mathcal{L}$ — множество числовых последовательностей $u = \{u_n\}$ с $\|u\| = \sup |u_n|$. Если p_{ij} — элементы стохастической или субстохастической матрицы, то определим преобразование T так, что i -я компонента Tu равна

$$(Tu)_i = \sum p_{ik} u_k. \quad (8.3)$$

Очевидно, что T — оператор сжатия, обладающий свойством монотонной сходимости. Если матрица (p_{ij}) строго стохастическая, то T — оператор перехода. ▶

Эти примеры типичны, и найти сжатие, не порожденное стохастическим ядром, в самом деле трудно. Как бы то ни было, мы оправдали дальнейшее развитие общей теории сжимающих операторов и убедились в том, что применения к вероятностным задачам будут носить очевидный характер. (В действительности мы никогда не выйдем за рамки указанных примеров.)

Переходные вероятности марковского процесса образуют одно-

параметрическое семейство ядер, удовлетворяющих уравнению Чепмена — Колмогорова

$$Q_{s+t}(x, \Gamma) = \int Q_s(x, dy) Q_t(y, \Gamma) \quad (8.4)$$

($s > 0, t > 0$); интегрирование производится по всему пространству состояний. Каждое из этих ядер порождает оператор перехода $\Omega(t)$, определяемый соотношением

$$\Omega(t)u(x) = \int Q_t(x, dy) u(y). \quad (8.5)$$

Очевидно, что (8.4) эквивалентно

$$\Omega(s+t) = \Omega(s)\Omega(t), \quad s > 0, t > 0. \quad (8.6)$$

Семейство эндоморфизмов с этим свойством называют *полугруппой*. Ясно, что $\Omega(s)\Omega(t) = \Omega(t)\Omega(s)$, т. е. *каждые два элемента полугруппы перестановочны*.

Последовательность $\{T_n\}$ эндоморфизмов на $\mathcal{L}$ называется сходящейся¹⁾ к эндоморфизму T тогда и только тогда, когда $\|T_n u - Tu\| \rightarrow 0$ при каждом $u \in \mathcal{L}$. Если последнее соотношение выполнено, мы пишем $T_n \rightarrow T$.

Начиная с этого момента, мы сосредоточимся на полугруппах сжимающих операторов. При этом мы введем некоторые условия регулярности. Обозначим опять через 1 тождественный оператор: $1u = u$.

Определение. Мы будем называть полугруппу сжимающих операторов $\Omega(t)$ *непрерывной²⁾*, если $\Omega(0) = 1$ и $\Omega(h) \rightarrow 1$ при $h \rightarrow 0+$.

При $0 \leq t' < t''$ мы имеем

$$\|\Omega(t'')u - \Omega(t')u\| \leq \|\Omega(t'' - t')u - u\|. \quad (8.7)$$

Для непрерывной полугруппы существует такое $\delta > 0$, что правая часть $< \varepsilon$ при $t'' - t' < \delta$. Таким образом, верно не только то, что $\Omega(t) \rightarrow \Omega(t_0)$ при $t \rightarrow t_0$, но и большее: формула (8.7) показывает, что $\Omega(t)u$ при любом фиксированном u есть равномерно непрерывная функция t ³⁾.

¹⁾ Такая сходимость называется *сильной*. Она введена в гл. VIII, 3. Из нее не вытекает сходимость $\|T_n - T\| \rightarrow 0$ (называемая *равномерной*). Более слабый тип сходимости определяется требованием $T_n u(x) \rightarrow Tu(x)$ при каждом x (но не обязательно равномерно); см. задачу 6 в гл. VIII, 10.

²⁾ Мы используем этот термин вместо стандартного «сильно непрерывна в нуле».

³⁾ Существуют полугруппы, такие, что $\Omega(h)$ стремится к оператору $T \neq 1$, но они патологичны. Определим, например, эндоморфизм T равенством $Tu(x) = \frac{1}{2}u(0)[1 + \cos x] + \frac{1}{2}u(\pi)[1 - \cos x]$ и положим $\Omega(t) = T$ при всех $t \geq 0$.

Преобразование (8.1)—это, конечно, то же самое, что и (4.5); оно служило исходным пунктом при выводе обратного уравнения для диффузионных процессов. Теперь семейство марковских переходных вероятностей порождает также некоторую полугруппу преобразований мер. При этом мера μ переходит в меру $T(t)\mu$, приписывающую множеству Γ массу

$$T(t)\mu(\Gamma) = \int_{\Gamma} \{dx\} Q_t(x, \Gamma). \quad (8.8)$$

Если Q_t имеет плотность q_t , то это преобразование совпадает с (5.1) и было использовано для вывода прямого уравнения. Так как теория вероятностей имеет в первую очередь дело с мерами, а не с функциями, то возникает вопрос: почему мы начали с полугруппы $\{\Omega(t)\}$, а не с $\{T(t)\}$? Ответ интересен и проливает новый свет на сложную связь прямого и обратного уравнений.

Причина состоит в том (и это подтверждают приведенные выше примеры), что в обычной обстановке непрерывные полугруппы сжимающих операторов на функциональном пространстве $\mathcal{L}$ порождаются переходными вероятностями: изучение наших полугрупп $\{\Omega(t)\}$ —это практически то же самое, что изучение марковских переходных вероятностей. Для полугрупп на пространстве мер это неверно. Существуют аналитически очень естественные полугруппы сжимающих операторов, не порождаемые марковскими процессами. Чтобы построить пример, рассмотрим любую марковскую полугруппу вида (8.8) на прямой, предполагая лишь, что абсолютно непрерывная мера μ переходит в абсолютно непрерывную меру $T(t)\mu$ [например, $T(t)$ может быть сверткой с нормальным распределением с дисперсией t]. Пусть $\mu = \mu_c + \mu_s$ —разложение μ на абсолютно непрерывную и сингулярную компоненты. Определим новую полугруппу $\{S(t)\}$, положив

$$S(t)\mu = T(t)\mu_c + \mu_s.$$

Эта полугруппа непрерывна, и $S(0) = 1$, но легко показать, что она не связана ни с какой системой переходных вероятностей и что она с вероятностной точки зрения бессмысленна.

§ 9. «ПОКАЗАТЕЛЬНАЯ ФОРМУЛА» В ТЕОРИИ ПОЛУГРУПП

Псевдопуассоновские процессы § 1 являются, конечно, простейшими марковскими процессами. Теперь мы докажем, что практически все марковские процессы представимы как пределы псевдопуассоновских процессов¹⁾. Абстрактный вариант этого утверждения играет фундаментальную роль в теории полугрупп. Мы увидим сейчас, что он на самом деле есть следствие закона больших чисел.

Пусть T —оператор, порождаемый стохастическим ядром K . Тогда оператор $\Omega(t)$, индуцированный псевдопуассоновским распределением (1.2), принимает вид

$$\Omega(t) = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} T^n, \quad (9.1)$$

где сумма ряда определяется как предел частных сумм. Из уравнения Чепмена—Колмогорова (1.3) следует, что эти операторы

¹⁾ Специальный случай, когда $\Omega(t)$ суть операторы свертки, был разобран в гл. IX,

образуют полугруппу. Предпочтительнее, однако, доказать этот факт заново для произвольных сжимающих операторов T .

Теорема 1. Если T — сжатие на $\mathcal{L}$, то операторы (9.1) образуют непрерывную полугруппу сжатий. Если T — оператор перехода, то это же верно и для $\mathfrak{D}(t)$.

Доказательство. Очевидно, что $\mathfrak{D}(t)$ — положительный оператор и $\|\mathfrak{D}(t)\| \leq e^{-\alpha t + \alpha t \|T\|} \leq 1$. Полугрупповое свойство легко проверяется формальным перемножением рядов для $\mathfrak{D}(s)$ и $\mathfrak{D}(t)$ (см. сноску 1 на стр. 366). Соотношение $\mathfrak{D}(h) \rightarrow 1$ очевидно из (9.1). ►

Мы будем записывать (9.1) сокращенно:

$$\mathfrak{D}(t) = e^{\alpha t (T-1)}. \quad (9.2)$$

Полугруппы сжатий такого вида будут называться псевдопуассоновскими¹⁾. Мы скажем, что $\{\mathfrak{D}_h(t)\}$ порождается оператором $\alpha(T-1)$.

Рассмотрим теперь произвольную непрерывную полугруппу сжатий $\mathfrak{D}(t)$. Она ведет себя во многих отношениях так же, как действительная непрерывная функция. В частности, применима без каких-либо серьезных изменений теория приближения функций, развитая в гл. VII на основе закона больших чисел. Мы покажем, что процедура примера 1, б) из гл. VII приводит к важнейшей формуле общей теории полугрупп. Введем при фиксированном $h > 0$ операторы

$$\mathfrak{D}_h(t) = e^{-th^{-1}} \sum_{n=0}^{\infty} \frac{(th^{-1})^n}{n!} \mathfrak{D}(nh), \quad (9.3)$$

которые могут быть получены рандомизацией параметра t в $\mathfrak{D}(t)$. Сравнение с (9.1) показывает, что операторы $\mathfrak{D}_h(t)$ образуют псевдопуассоновскую полугруппу, порождаемую оператором $[\mathfrak{D}(h) - 1]/h$. Докажем теперь, что

$$\mathfrak{D}_h(t) \rightarrow \mathfrak{D}(t), \quad h \rightarrow 0. \quad (9.4)$$

Сформулируем этот важный результат в виде теоремы.

Теорема 2. Каждая непрерывная полугруппа сжатий $\mathfrak{D}(t)$ является пределом (9.4) псевдопуассоновских полугрупп $\{\mathfrak{D}_h(t)\}$, порождаемых эндоморфизмами $h^{-1}[\mathfrak{D}(h) - 1]$.

Доказательство. Отправным пунктом будет тождество

$$\mathfrak{D}_h(t)u - \mathfrak{D}(t)u = e^{-th^{-1}} \sum_{n=0}^{\infty} \frac{(th^{-1})^n}{n!} [\mathfrak{D}(nh)u - \mathfrak{D}(t)u]. \quad (9.5)$$

¹⁾ Существуют полугруппы сжатий вида e^{tS} , где S — эндоморфизм, не представимый в форме $\alpha(T-1)$. Таковы, например, полугруппы, связанные со скачкообразными процессами § 3, если $\alpha(x)$ ограничена.

Выберем δ так, что $\|\mathfrak{Q}(s)u - u\| < \varepsilon$ при $0 < s < \delta$. Тогда ввиду (8.7) имеем при $|nh - t| < \eta t$

$$\|\mathfrak{Q}(nh)u - \mathfrak{Q}(t)u\| \leq \varepsilon. \quad (9.6)$$

Распределение Пуассона в (9.5) имеет математическое ожидание и дисперсию t/h . Поэтому вклад слагаемых с $|nh - t| \geq \delta$ в сумму (9.5) можно оценить с помощью неравенства Чебышева, и мы получаем

$$\|\mathfrak{Q}_h(t)u - \mathfrak{Q}(t)u\| < \varepsilon + 2\|u\|th\delta^{-2}.$$

Отсюда следует, что сходимость (9.4) равномерна на всех конечных интервалах изменения t . ▶

Другие варианты можно получить, ставя на место распределения Пуассона другие безгранично делимые распределения. Из 1; гл. XII, 2, мы знаем, что производящая функция безгранично делимого распределения $\{u_n(t)\}$, сосредоточенного на множестве целых чисел $n \geq 0$, имеет вид

$$\sum_{n=0}^{\infty} u_n(t) \zeta^n = \exp(t\alpha[p(\zeta) - 1]), \quad (9.7)$$

где

$$p(\zeta) = p_0 + p_1\zeta + \dots, \quad p_j \geq 0, \quad \sum p_j = 1. \quad (9.8)$$

Допустим, что распределение $\{u_n(t)\}$ имеет математическое ожидание bt и конечную дисперсию ct . Заменяя в (9.3) распределение Пуассона на $\{u_n(t/b)\}$, получаем оператор

$$\mathfrak{Q}_h(t) = \sum u_n(t/bn) \mathfrak{Q}(nh) = \sum u_n(t/bh) \mathfrak{Q}^n(h). \quad (9.9)$$

Как и прежде, простое применение закона больших чисел показывает, что $\mathfrak{Q}_h(t) \rightarrow \mathfrak{Q}(t)$ при $h \rightarrow 0$. Таким путем мы получаем аналог «показательной формулы», в котором распределение $\{u_n(t)\}$ играет роль распределения Пуассона¹⁾.

Чтобы понять вероятностное содержание и пути обобщения этих рассуждений, обозначим через $\{X(t)\}$ марковский процесс, соответствующий полугруппе $\{\mathfrak{Q}(t)\}$, и через $\{T(t)\}$ — процесс с независимыми приращениями, описываемый вероятностями перехода $\{u_n(t)\}$. Операторы (9.9) соответствуют переходным вероятностям процесса $X(hT(t/bh))$. Другими словами, мы ввели специальный процесс, подчиненный $X(t)$. Закон больших чисел, примененный к T -процессу, делает правдоподобным предположение о том, что при $h \rightarrow 0$ распределения нового процесса сходятся к распределениям исходного марковского процесса. Эта процедура аппроксимации никоим образом не связана с целочисленностью случайных величин $T(t)$. В самом деле, мы можем взять в качестве $T(t)$ любой процесс с положительными независимыми приращениями, такой, что $E((T(t))) = bt$ и что дисперсия $T(t)$ конечна.

Таким образом, данный марковский процесс $\{X(t)\}$ выступает как предел подчиненных марковских процессов вида $X(hT(t/bh))$.

Дело в том, что полугруппы аппроксимирующих процессов могут иметь значительно более простую структуру, чем исходная полугруппа. Действительно, полугруппа операторов $\mathfrak{Q}_h(t)$, определенных в (9.9), принадлежит псевдопуассоновскому типу. Для проверки положим

$$\mathfrak{Q}^* = p(\mathfrak{Q}(h)) = \sum_{n=0}^{\infty} p_n \mathfrak{Q}^n(h) = \sum_{n=0}^{\infty} p_n \mathfrak{Q}(nh). \quad (9.10)$$

¹⁾ Это отметил Чжун К. Л. (см. гл. VII, 5).

Оператор $\Omega^\#$ — это смесь операторов перехода и потому тоже оператор перехода. Сравнение соотношений (9.7) и (9.9) приводит к формальному равенству

$$\Omega_h(t) = \exp \left\{ \frac{\alpha}{bh} (\Omega^\# - 1) \right\}, \quad (9.11)$$

что несомненно является разновидностью (9.2). Равенство (9.11) нетрудно проверить элементарными методами, но мы увидим, что (9.11) на самом деле только частный случай формулы для производящих операторов подчиненных полугрупп (см. пример 9, б) гл. XIII).

§ 10. ПРОИЗВОДЯЩИЕ ОПЕРАТОРЫ. ОБРАТНОЕ УРАВНЕНИЕ

Рассмотрим псевдопуассоновскую полугруппу $\{\Omega(t)\}$ сжимающих операторов, порождаемую оператором $\mathfrak{A} = \alpha(T-1)$. Так как это эндоморфизм, то $v = \mathfrak{A}u$ определено для всех $u \in \mathcal{S}$ и

$$\frac{\Omega(h) - 1}{h} u \rightarrow v, \quad h \rightarrow 0+. \quad (10.1)$$

Было бы хорошо, если бы это свойство имело место для всех полугрупп, однако этого невозможно ожидать. Возьмем, например, полугруппу, связанную с броуновским движением. Уравнение диффузии (4.1) показывает, что для дважды непрерывно дифференцируемой функции u левая часть (10.1) сходится к $u''/2$. Однако не существует никакого предела, если u не дифференцируема. Тем не менее уравнение диффузии однозначно описывает процесс, так как полугруппа однозначно определяется своим действием на дважды дифференцируемые функции. Поэтому мы не можем ожидать, что (10.1) будет выполняться для *всех* функций u , но для всех практических целей хватает требования, чтобы оно выполнялось для «достаточно многих» функций. Имея это в виду, примем

Определение. Если для некоторых u и v из $\mathcal{S}$ выполняется соотношение (10.1) (в смысле равномерной сходимости), то мы полагаем $v = \mathfrak{A}u$. Так определенный оператор называется производящим оператором¹⁾ полугруппы $\{\Omega(t)\}$.

Умножая (10.1) слева на $\Omega(t)$, получаем

$$\frac{\Omega(t+h) - \Omega(t)}{h} u \rightarrow \Omega(t)v. \quad (10.2)$$

Таким образом, если $\mathfrak{A}u$ существует, то все функции $\Omega(t)u$ лежат в области определения $\mathfrak{A}$ и

$$\frac{\Omega(t+h) - \Omega(t)}{h} u \rightarrow \Omega(t)\mathfrak{A}u = \mathfrak{A}\Omega(t)u. \quad (10.3)$$

¹⁾ При анализе полугрупп со сверткой в гл. IX мы ограничивались бесконечно дифференцируемыми функциями, благодаря чему все производящие операторы имели одну и ту же область определения. Столь удобного средства нет в общей теории полугрупп.

Это соотношение, по существу, то же самое, что обратное уравнение для марковских процессов. В самом деле, в обозначениях § 4 мы можем положить

$$u(t, x) = \mathfrak{D}(t) u_0(x),$$

где u_0 — начальная функция. Тогда (10.3) превращается в

$$\frac{\partial u(t, x)}{\partial t} = \mathfrak{A}u(t, x). \quad (10.4)$$

Это — знакомое нам обратное уравнение, но оно нуждается в правильной интерпретации. Вероятности перехода диффузионных процессов § 4 так гладки, что обратное уравнение верно для *всех* непрерывных начальных функций u_0 . В общем случае этого может не быть.

Примеры. а) *Сдвиги*. Пусть $\mathcal{L}$ состоит из всех непрерывных функций на прямой, обращающихся в нуль на бесконечности. Положим $\mathfrak{D}(t)u(x) = u(x+t)$. Очевидно, что (10.1) выполняется тогда и только тогда, когда u имеет непрерывную производную u' , обращающуюся в нуль на бесконечности. В этом случае $\mathfrak{A}u = u'$.

Формально обратное уравнение (10.4) сводится к

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}. \quad (10.5)$$

Формальное решение, совпадающее при $t=0$ с данной начальной функцией u_0 , было бы $u(t, x) = u_0(t+x)$. Но u в самом деле будет решением только в том случае, когда функция u_0 дифференцируема.

б) Как и в § 2, рассмотрим псевдопуассоновский процесс $X(t)$ и другой процесс $X^*(t) = X(t) - ct$. Соответствующие полугруппы связаны очевидным образом: значение $\mathfrak{D}^*(t)u$ в точке x равно значению $\mathfrak{D}(t)$ в точке $x+ct$. Для производящих операторов имеем связь

$$\mathfrak{A}^* = \mathfrak{A} - c \frac{d}{dx}, \quad (10.6)$$

так что область определения $\mathfrak{A}^*$ ограничена дифференцируемыми функциями. Обратное уравнение выполняется, если только начальная функция u_0 имеет непрерывную производную, но не для произвольных функций. В частности, сами переходные вероятности не обязаны удовлетворять обратному уравнению. Этим объясняются как трудности устаревших теорий [обсуждавшихся в связи с гл. IX, (2.14)], так и необходимость принятия неестественных условий регулярности при выводе прямого уравнения (2.1). ►

Польза понятия производящего оператора проистекает из следующего факта: для любой непрерывной полугруппы сжатий про-

изводящий оператор однозначно описывает полугруппу. Простое доказательство этой теоремы будет дано в гл. XIII, 9.

Эта теорема дает нам возможность обращаться с обратными уравнениями, не прибегая к излишним ограничениям, и значительно упрощает вывод этих уравнений. Так, наиболее общая форма диффузионных операторов, упомянутая в сноске 1 на стр. 380 (§ 4), не могла бы быть получена без априорной уверенности в том, что производящий оператор на самом деле существует.

Процессы восстановления были введены в гл. VI, 6, и были иллюстрированы примерами в гл. VI, 7. Теперь мы начнем с общей теории так называемого уравнения восстановления, с которым часто приходится сталкиваться в различных вопросах. Ярким примером применения общей теоремы восстановления служит предельная теорема из § 8. Параграфы 6 и 7 содержат улучшенные и обобщенные варианты некоторых асимптотических оценок, первоначальный вывод которых был трудоемким и требовал применения глубоких аналитических методов. Здесь еще раз демонстрируется та экономия в работе мысли и количестве аналитических средств, которая достигается за счет общетеоретического подхода к трудным индивидуальным проблемам. Анализ задач теории восстановления с помощью преобразований Лапласа см. в гл. XIV, 1—3.

Немало работ было написано и много сил затрачено с целью освободить теорему восстановления от условия положительности случайных величин. Ввиду солидной истории вопроса мы включим в § 9 новое и сильно упрощенное доказательство общей теоремы.

§ 1. ТЕОРЕМА ВОССТАНОВЛЕНИЯ

Пусть F — распределение, сосредоточенное¹⁾ на $\overline{0, \infty}$. Иными словами, мы предполагаем, что $F(0) = 0$. Мы не требуем существования математического ожидания. Однако в силу условия положительности мы можем принять

$$\mu = \int_0^{\infty} yF\{dy\} = \int_0^{\infty} [1 - F(y)] dy, \quad (1.1)$$

где $\mu \leq \infty$. При $\mu = \infty$ мы условимся интерпретировать символ μ^{-1} как 0.

В этом параграфе мы исследуем асимптотическое поведение при $x \rightarrow \infty$ функции

$$U = \sum_{n=0}^{\infty} F_n^*. \quad (1.2)$$

¹⁾ Если допустить атом с массой $p < 1$ в нуле, то никаких существенных изменений не требуется (см. задачу 1).

Вскоре мы увидим, что эта задача тесно связана с асимптотическим поведением решения Z уравнения восстановления

$$Z(x) = z(x) + \int_0^x Z(x-y) F\{dy\}, \quad x > 0. \quad (1.3)$$

Для определенности мы считаем интервал интегрирования замкнутым, но из данного контекста должно быть понятно, что функции z и Z равны нулю на отрицательной полуоси. Поэтому пределы интегрирования можно заменить без ущерба на $-\infty$ и ∞ и переписать уравнение восстановления в форме уравнения свертки

$$Z = z + F \star Z \quad (1.4)$$

(это замечание относится ко всем встречающимся в дальнейшем интегралам этого типа).

Вероятностный смысл U и вероятностные применения уравнения восстановления довольно подробно обсуждались в гл. VI, 6, 7. Поэтому сейчас мы приступим к чисто аналитическому исследованию. Здесь уместно отметить, что в процессе восстановления значение $U(x)$ равно математическому ожиданию числа моментов восстановления на интервале $\overline{0, x}$, при этом нуль считается моментом восстановления. В соответствии с этим U следует интерпретировать как меру, сосредоточенную на $\overline{0, \infty}$. Интервалу $I = \overline{a, b}$ отвечает мера $U\{I\} = U(b) - U(a)$. Отметим, что нуль является атомом этой меры, который имеет массу 1 и вносится нулевым членом ряда (1.2).

Следующая лемма просто повторяет теорему 1 из гл. VI, 6, однако новое доказательство, приводимое здесь, способствует независимости настоящего параграфа.

Лемма. $U(x) < \infty$ для всех x . Если z ограничена, то функция Z , определенная равенством

$$Z(x) = \int_0^x z(x-y) U\{dy\}, \quad x > 0, \quad (1.5)$$

есть единственное решение уравнения восстановления (1.3), которое ограничено на конечных интервалах.

[При соглашении о том, что $z(x) = Z(x) = 0$ при $x < 0$, можно записать (1.5) в форме $Z = U \star z$.]

Доказательство. Положим $U_n = F^0 \star \dots \star F^n$ и выберем такие положительные числа τ и η , что $1 - F(\tau) > \eta$. Тогда

$$\int_0^x [1 - F(x-y)] U_n\{dy\} = 1 - F^{(n+1)\star}(x), \quad x > 0, \quad (1.6)$$

и, следовательно, $\eta[U_n(x) - U_n(x-\tau)] < 1$. Устремляя $n \rightarrow \infty$, мы делаем вывод, что $U\{I\} \leq \eta^{-1}$ для любого интервала I , длина которого $< \tau$. Поскольку произвольный интервал длины a можно представить как объединение самое большее $1 + a/\tau$ интервалов, длины τ , то отсюда следует, что

$$U(x) - U(x-a) \leq C_a, \quad (1.7)$$

где $C_a = (a + \tau)/(\tau\eta)$. Таким образом, $U\{I\}$ равномерно ограничено на всех интервалах I данной длины.

Далее, $Z_n = U_n \star z$ удовлетворяет уравнению $Z_{n+1} = z + F \star Z_n$. Устремляя $n \rightarrow \infty$, убеждаемся в том, что интеграл в (1.5) имеет смысл и что Z есть решение уравнения (1.3).

Для доказательства единственности заметим, что разность двух решений должна удовлетворять уравнению $V = F \star V$, а следовательно, и соотношению

$$V(x) = \int_0^x V(x-y) F^r \star \{dy\}, \quad x > 0, \quad (1.8)$$

при $r = 1, 2, \dots$. Но $F^r \star(x) \rightarrow 0$ при $r \rightarrow \infty$ и поскольку предполагается, что V ограничено на $0, \infty$, то отсюда вытекает, что $V(x) = 0$ при всех $x > 0$. ▶

Формулировка теоремы восстановления затрудняется из-за специальной роли, которую играют распределения, сосредоточенные на множестве чисел, кратных числу λ . Согласно определению 3 из гл. V, 2, такое распределение называется *арифметическим*, а наибольшее λ с тем свойством, что F сосредоточено на множестве точек $\lambda, 2\lambda, \dots$, называется *шагом* F . В этом случае мера U чисто атомическая, и мы обозначим через u_n массу, соответствующую точке $n\lambda$. Теорема восстановления из 1; XIII, 11, утверждает, что $u_n \rightarrow \lambda/\mu$. Приводимая ниже теорема¹⁾ обобщает этот результат на случай произвольных распределений, сосредоточенных на $0, \infty$. Случай арифметического распределения F ради полноты формулировки упомянут снова. (Мы напоминаем о соглашении: $\mu^{-1} = 0$, если $\mu = \infty$.)

Теорема восстановления. (Основная форма.) Если распределение F не является арифметическим, то при любом фиксированном $h > 0$ и $t \rightarrow \infty$

$$U(t) - U(t-h) \rightarrow \frac{h}{\mu}. \quad (1.9)$$

¹⁾ Случай арифметических распределений был рассмотрен П. Эрдешем, В. Феллером, Г. Пшардом (1949). Их доказательство было незамедлительно обобщено Д. Блекуэллом. В тексте дано новое доказательство. Распространение на распределения, не сосредоточенные на $0, \infty$, см. в § 9. Важнейшие обобщения в другом направлении содержатся в работе Чжоу и Роббинса (Chow Y. S., Robbins H. E., A renewal theorem for random variables which are dependent distributed, Ann. Math. Stat., vol. 34 (1963), 390–401.

Если же F — арифметическое распределение, то (1.9) верно для h , кратных шагу λ .

Прежде чем доказывать теорему, мы переформулируем ее в терминах асимптотических свойств решений (1.5) уравнения восстановления. Поскольку любая данная функция z может быть разложена на свои положительную и отрицательную части, то мы можем допустить, что $z \geq 0$. Для определенности мы предположим, что распределение F не является арифметическим и имеет математическое ожидание $\mu < \infty$.

Если $z(x) = 1$ для $0 \leq a \leq x < b < \infty$ и $z(x) = 0$ для всех других x , то мы получим из (1.5)

$$Z(t) = U(t-a) - U(t-b) \rightarrow (b-a)/\mu, \quad t \rightarrow \infty. \quad (1.10)$$

Этот результат немедленно обобщается на конечные ступенчатые функции. Пусть $I_1, \dots, I_r$ — непересекающиеся интервалы на положительной полуоси с длинами $L_1, \dots, L_r$. Если z принимает значение a_k на интервале I_k и обращается в нуль за пределами объединения интервалов I_k , $k = 1, \dots, r$, то очевидно, что

$$Z(t) \rightarrow \mu^{-1} \sum_{k=1}^r a_k L_k = \mu^{-1} \int_0^{\infty} z(x) dx. \quad (1.11)$$

Классический интеграл Римана функции z определяется в терминах конечных ступенчатых функций, аппроксимирующих z , и потому правдоподобно, что предельное соотношение (1.11) выполняется, когда z интегрируема по Риману. Чтобы прояснить это место, вспомним определение интеграла Римана функции z по конечному интервалу $0 \leq x \leq a$. Достаточно рассмотреть разбиения на подынтервалы равной длины $h = a/n$. Пусть $\underline{m}_k$ и $\overline{m}_k$ соответственно такие наибольшее и наименьшее число, что

$$\underline{m}_k \leq z(x) \leq \overline{m}_k \text{ при } (k-1)h \leq x < kh. \quad (1.12)$$

Очевидная зависимость $\underline{m}_k$ и $\overline{m}_k$ от h должна быть учтена. Нижняя и верхняя римановы суммы при заданном шаге h определяются равенствами

$$\underline{\sigma} = h \sum \underline{m}_k, \quad \overline{\sigma} = h \sum \overline{m}_k. \quad (1.13)$$

При $h \rightarrow 0$ и $\underline{\sigma}$, и $\overline{\sigma}$ стремятся к конечным пределам. Если $\overline{\sigma} - \underline{\sigma} \rightarrow 0$, то эти пределы совпадают и интеграл Римана функции z определяется как значение этого общего предела. Любая ограниченная функция, которая имеет только разрывы первого рода, интегрируема в этом смысле.

Когда мы переходим к интегралам по области $\overline{0, \infty}$, то классическое определение приводит к осложнениям, которые, впрочем, можно обойти. Чтобы сделать класс интегрируемых функций возможно более широким, интеграл по $\overline{0, \infty}$ уславливаются понимать

как предел интегралов по областям $\overline{0, a}$. Непрерывная неотрицательная функция $z(x)$ интегрируема в этом смысле тогда и только тогда, когда площадь между ее графиком и x -осью конечна. К сожалению, это не устраняет сильных колебаний функции $z(x)$ на бесконечности (см. пример а)). Очевидно, что нет причин предполагать, что решение Z стремится к конечному пределу, если данная функция z колеблется беспорядочным образом. Другими словами, усложненное стандартное определение делает слишком много функций интегрируемыми, и для наших целей предпочтительнее поступать проще, распространяя первоначальное определение на бесконечные интервалы. Из-за отсутствия устоявшегося термина мы будем говорить о *непосредственном* интегрировании в противоположность косвенной процедуре, содержащей предельный переход от конечных интервалов.

Определение. Функция $z \geq 0$ называется непосредственно интегрируемой по Риману, если верхняя и нижняя римановы суммы, определенные в (1.12), (1.13), конечны и стремятся при $h \rightarrow 0$ к одному и тому же пределу.

Это определение не делает различия между конечными и бесконечными интервалами. Легко убедиться в том, что z непосредственно интегрируема на $\overline{0, \infty}$, если она интегрируема на каждом конечном интервале $\overline{0, a}$ и если $\bar{\sigma} < \infty$ при некотором h (тогда $\bar{\sigma} < \infty$ автоматически для всех h). Второе условие как раз и исключает большие колебания функции.

Мы можем переосмыслить это определение в терминах аппроксимирующих ступенчатых функций. При фиксированном $h > 0$ положим $z_k(x) = 1$ в случае $(k-1)h \leq x < kh$ и $z_k(x) = 0$ в других случаях. Тогда две функции

$$\underline{z} = \sum \underline{m}_k z_k \quad \text{и} \quad \bar{z} = \sum \bar{m}_k z_k \quad (1.14)$$

являются конечными ступенчатыми функциями и $\underline{z} \leq z \leq \bar{z}$. Интеграл для z есть общий предел при $h \rightarrow 0$ интегралов от этих ступенчатых функций. Обозначим через Z_k решение уравнения восстановления, соответствующее z_k . Тогда решения, соответствующие $\underline{z}$ и $\bar{z}$, определяются равенствами

$$\underline{Z} = \sum Z_k \underline{m}_k \quad \text{и} \quad \bar{Z} = \sum \bar{m}_k Z_k. \quad (1.15)$$

При каждом фиксированном k по теореме восстановления $Z_k(x) \rightarrow h/\mu$. Кроме того, (1.7) гарантирует нам, что $Z_k(x) \leq C_h$ при всех k и x . Остаточные члены рядов в (1.15) сходятся поэтому равномерно к 0, и мы заключаем, что при $x \rightarrow \infty$

$$\underline{Z}(x) \rightarrow \underline{\sigma}/\mu, \quad \bar{Z}(x) \rightarrow \bar{\sigma}/\mu. \quad (1.16)$$

Но $\underline{Z} \leq Z \leq \bar{Z}$, и поэтому все предельные значения $Z(x)$ лежат

между $\underline{\sigma}/\mu$ и $\bar{\sigma}/\mu$. Если z непосредственно интегрируема по Риману, то из последних рассуждений вытекает, что

$$Z(x) = \mu^{-1} \int_0^{\infty} z(y) dy, \quad x \rightarrow \infty. \quad (1.17)$$

До сих пор мы предполагали, что F — не арифметическое распределение и $\mu < \infty$. Наши рассуждения применимы без изменений в случае $\mu = \infty$, если μ^{-1} считать нулем.

Если распределение F арифметическое с шагом λ , то решение Z уравнения (1.5) имеет вид

$$Z(x) = \sum z(x - k\lambda) u_k, \quad (1.18)$$

где $u_k \rightarrow \lambda/\mu$. Легко сделать вывод, что при фиксированном x

$$Z(x + n\lambda) \rightarrow \lambda \mu^{-1} \sum_{j=1}^{\infty} z(x + j\lambda), \quad n \rightarrow \infty, \quad (1.19)$$

при условии сходимости ряда, что, конечно, имеет место, если z непосредственно интегрируема.

Соотношения (1.17) и (1.19) выведены нами из теоремы восстановления, но эти соотношения сами содержат теорему восстановления как частный случай, когда z является индикатором интервала $\overline{0, h}$. Таким образом, нами доказана следующая

Теорема восстановления. (Альтернативная форма.) ¹⁾ Если z непосредственно интегрируема по Риману, то решение Z уравнения восстановления удовлетворяет соотношению (1.17), если F неарифметическое, и соотношению (1.19), если F арифметическое с шагом λ .

Возникает вопрос: может ли быть опущено условие непосредственной интегрируемости, по крайней мере для непрерывных функций z , стремящихся к 0 на бесконечности? Следующие ниже примеры показывают, что это невозможно. Пример 3, б) покажет подобным же образом, что теорема восстановления может быть

¹⁾ Можно надеяться, что эта форма теоремы и предшествующее ей обсуждение положат конец достойной сожаления путанице, распространившейся ныне в литературе. Наиболее широко распространены ссылки на статью Смита: Smith W. L., Renewal theory and its ramifications, J. Roy. Stat. Soc. ser. B, vol. 20 (1958), 243—302. (Есть русский перевод: Смит В., Теория восстановления, сб. Математика, 5:3 (1961), 95—150.) Со времени своего появления «ключевая теорема восстановления», доказанная Смитом, практически вытеснила все применявшиеся до нее варианты (которые не всегда были корректны). Ключевая теорема доказывает (1.17) при излишнем предположении, что z монотонна. Доказательство Смита 1954 г. основано на глубоких тауберовых теоремах, принадлежащих Винеру, а работа Смита 1958 г. производит впечатление, что этот обходной путь доказательства проще, чем прямое сведение к основной форме теоремы восстановления. Кроме того, условие ограниченности z пропущено в работе. [Относительно его необходимости см. пример 3, б).]

неверна для неограниченной функции z даже при том, что она обращается в нуль вне некоторого конечного интервала. Поэтому несобственные интегралы Римана не находят применения в теории восстановления, а непосредственная интегрируемость появляется как естественная основа теории.

Примеры. а) *Непрерывная функция $z(x)$ может быть неограниченной и тем не менее интегрируемой по Риману на $0, \infty$ функцией.* Чтобы убедиться в этом, положим $z(n) = a_n$ при $n = 1, 2, \dots$, и пусть z тождественно равна нулю вне объединения интервалов $|x - n| < h_n < 1/2$; пусть между n и $n \pm h_n$ функция z есть линейная функция x . График функции z при этом образуется последовательностью треугольников с площадями $a_n h_n$, и, следовательно, z интегрируема по Риману тогда и только тогда, когда $\sum a_n h_n < \infty$. Это не мешает тому, что $a_n \rightarrow \infty$.

б) Чтобы исследовать роль условия непосредственной интегрируемости в теореме восстановления, достаточно рассмотреть арифметические распределения F . Таким образом, можно предположить, что мера U сосредоточена на множестве целых чисел и что масса u_n , приписанная точке n , стремится к пределу $\mu^{-1} > 0$. Тогда для любого целого положительного числа n мы имеем

$$Z(n) = u_n z(0) + u_{n-1} z(1) + \dots + u_0 z(n).$$

Выберем теперь в качестве z функцию из предыдущего примера с $a_n = 1$. Тогда $Z(n) \sim n\mu^{-1}$ и поэтому функция Z даже не ограничена. Очевидно, что то же самое справедливо, если a_n стремится к 0 достаточно медленно. Таким образом, мы получаем пример непрерывной интегрируемой функции z , такой, что $z(x) \rightarrow 0$, но $Z(x)$ неограничена. ►

§ 2. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ ВОССТАНОВЛЕНИЯ

Для арифметических распределений теорема восстановления была доказана в I; XIII, II, и поэтому мы предположим, что распределение F неарифметическое. Для доказательства нам нужны будут две леммы (первая из них вновь появится в усиленной форме как следствие из теоремы в § 9).

Лемма 1. Пусть ограниченная равномерно непрерывная функция ξ такова, что $\xi(x) \leq \xi(0)$ при $-\infty < x < \infty$. Если

$$\xi(x) = \int_0^{\infty} \xi(x-y) F\{dy\}, \quad (2.1)$$

то имеет место тождество $\xi(x) = \xi(0)$.

Доказательство. Вычисляя свертки с F и используя индукцию,

получаем из (2.1), что

$$\zeta(x) = \int_0^{\infty} \zeta(x-y) F^{r*} \{dy\}, \quad r=1, 2, \dots \quad (2.2)$$

Подынтегральное выражение в (2.2) не превосходит $\zeta(0)$, и поэтому для $x=0$ равенство возможно только в том случае, если $\zeta(-y) = \zeta(0)$ для каждого значения y , которое служит точкой роста F^{r*} . По лемме 2 из гл. V, 4а, образованное такими точками множество Σ асимптотически плотно на бесконечности, и ввиду равномерной непрерывности ζ отсюда следует, что $\zeta(-y) \rightarrow \zeta(0)$ при $y \rightarrow \infty$. При $r \rightarrow \infty$ масса распределения F^{r*} стремится сосредоточиться «вблизи бесконечности». Поэтому при больших r интеграл в (2.2) по существу зависит только от больших значений y , и для таких значений $\zeta(x-y)$ близко к $\zeta(0)$. Устремляя r к ∞ , мы выводим таким образом из (2.2), что $\zeta(x) = \zeta(0)$, что и утверждалось. ►

Лемма 2. Пусть z — непрерывная функция, обращающаяся в нуль вне интервала $0, h$. Тогда соответствующее решение Z уравнения восстановления равномерно непрерывно и для любого a

$$Z(x+a) - Z(x) \rightarrow 0, \quad x \rightarrow \infty. \quad (2.3)$$

Доказательство. Разности $z(x+\delta) - z(x)$ обращаются в нуль на интервале длины $h+2\delta$, и поэтому в силу (1.5) и (1.7)

$$|Z(x+\delta) - Z(x)| \leq C_{h+2\delta} \max |z(x+\delta) - z(x)|. \quad (2.4)$$

Отсюда вытекает, что если z равномерно непрерывна, то это же самое справедливо для Z .

Допустим теперь, что z имеет непрерывную производную z' . Тогда Z' существует и удовлетворяет уравнению восстановления

$$Z'(x) = z'(x) + \int_0^x Z'(x-y) F \{dy\}. \quad (2.5)$$

Таким образом, производная Z' ограничена и равномерно непрерывна. Пусть

$$\limsup Z'(x) = \eta. \quad (2.6)$$

Выберем такую последовательность t_n , что $Z'(t_n) \rightarrow \eta$. Семейство функций ζ_n , определенных равенством

$$\zeta_n(x) = Z'(t_n + x), \quad (2.7)$$

равностепенно непрерывно и

$$\zeta_n(x) = z'(t_n + x) + \int_0^{x+t} \zeta_n(x-y) F \{dy\}. \quad (2.8)$$

Следовательно, существует такая подпоследовательность t_{n_r} , что ξ_{n_r} сходится к пределу ξ . Из (2.8) следует, что этот предел удовлетворяет условиям леммы 1 и поэтому $\xi'(x) = \xi'(0) = \eta$ для всех x .

Таким образом, $Z'(t_{n_r} + x) \rightarrow \eta$ или

$$Z(t_{n_r} + a) - Z(t_{n_r}) \rightarrow \eta a. \quad (2.9)$$

Поскольку это верно при любом a и Z ограничена, то $\eta = 0$. Те же самые аргументы применимы при рассмотрении нижнего предела, и в итоге получаем, что $Z'(x) \rightarrow 0$.

Таким образом, мы доказали лемму для непрерывно дифференцируемой функции z . Но произвольная непрерывная функция z может быть приближена непрерывно дифференцируемой функцией z_i , которая обращается в нуль вне интервала $\bar{0, h}$. Пусть Z_i — соответствующее z_i решение уравнения восстановления. Тогда из $|z - z_i| < \varepsilon$ следует $|Z - Z_i| < C_h \varepsilon$, и поэтому $|Z(x + a) - Z(x)| < (2C_h + 1)\varepsilon$ для всех достаточно больших x . Таким образом, (2.3) справедливо для произвольной непрерывной функции x . ▀

Завершение доказательства теперь облегчено. Если I — это интервал $\alpha \leq x \leq \beta$, то пусть $I + t$ обозначает интервал $\alpha + t \leq x \leq \beta + t$. По теореме о выборе (теорема 2 из гл. VIII, 6) существует последовательность $t_k \rightarrow \infty$ и мера V , такие, что

$$U\{t_k + dy\} \rightarrow V\{dy\}. \quad (2.10)$$

Мера V конечна на конечных интервалах, но не сосредоточена на $\overline{0, \infty}$.

Пусть теперь z — непрерывная функция, равная нулю вне конечного интервала $\bar{0, a}$. Соответствующее решение Z уравнения восстановления удовлетворяет тогда соотношению

$$Z(t_k + x) = \int_0^a z(-s) U\{t_k + x + ds\} \rightarrow \int_0^a z(-s) V\{x + ds\}. \quad (2.11)$$

Из леммы 2 следует, что семейство мер $V\{x + ds\}$ не зависит от x , и, следовательно, значение $V\{I\}$ должно быть пропорционально длине интервала I . Таким образом, (2.10) можно переписать в виде

$$U(t_k) - U(t_k - h) \rightarrow \gamma h. \quad (2.12)$$

Это соотношение есть не что иное, как утверждение (1.9) теоремы восстановления, с той лишь разницей, что множитель η^{-1} заменен неизвестной величиной γ и что t пробегает последовательность $\{t_k\}$. Однако наш вывод теоремы восстановления в ее альтернативной форме остается справедливым, и, таким образом,

$$Z(t_k) \rightarrow \gamma \int_0^\infty z(y) dy, \quad (2.13)$$

если только z непосредственно интегрируема.

Функция $z = 1 - F$ монотонна, и ее интеграл равен μ . Соответствующее z решение Z уравнения восстановления становится равным 1. Если $\mu < \infty$, то функция z непосредственно интегрируема и из (2.13) следует, что $\gamma\mu = 1$. Если $\mu = \infty$, то мы урезаем z и выводим из (2.13), что γ^{-1} превосходит интеграл от z по произвольному интервалу $0, a$. Таким образом, из $\mu = \infty$ следует $\gamma = 0$. Следовательно, предел в (2.12) не зависит от последовательности $\{t_k\}$, и (2.12) совпадает с утверждением (1.9) теоремы восстановления. ►

§ 3*). УТОЧНЕНИЯ

В этом параграфе мы покажем, как свойства регулярности распределения F могут привести к усиленным вариантам теоремы восстановления.

Теорема 1. Если F — неарифметическое распределение с математическим ожиданием μ и дисперсией σ^2 , то

$$0 \leq U(t) - \frac{t}{\mu} \rightarrow \frac{\sigma^2 + \mu^2}{2\mu^2}. \quad (3.1)$$

Сама теорема восстановления лишь утверждает, что $U(t) \sim t/\mu$, а оценка (3.1) значительно сильнее. Даже тогда, когда дисперсия не существует, (3.1) выполняется с заменой правой части на ∞ . [Аналог теоремы для арифметических распределений дан в (12.2) в I, XIII.]

Доказательство. Положим

$$Z(t) = U(t) - \frac{t}{\mu}, \quad (3.2)$$

Легко проверить, что это есть решение уравнения восстановления, которому соответствует

$$z(t) = \frac{1}{\mu} \int_t^\infty [1 - F(y)] dy. \quad (3.3)$$

Интегрируя по частям, получаем

$$\int_0^\infty z(t) dt = \frac{1}{2\mu} \int_0^\infty y^2 F\{dy\} = \frac{\sigma^2 + \mu^2}{2\mu}. \quad (3.4)$$

Так как функция z монотонна, она непосредственно интегрируема, и по теореме восстановления (альтернативная форма) заключаем, что (3.1) справедливо. ►

Перейдем теперь к свойствам гладкости функции U . Если F

*) Этот параграф может быть опущен при первом чтении.

имеет плотность f , то уравнение восстановления для U принимает вид

$$U(x) = 1 + \int_0^x U(x-y)f(y)dy. \quad (3.5)$$

Если бы f была непрерывна, то формально выполненное дифференцирование показало бы, что U имеет производную u , удовлетворяющую уравнению

$$u(x) = f(x) + \int_0^x u(x-y)f(y)dy. \quad (3.6)$$

Это уравнение восстановления стандартного типа, а мы знаем, что его решение единственно, если функция f ограничена (но обязательно непрерывна). Легко проверить, что функция U , определенная равенством

$$U(t) = 1 + \int_0^t u(y)dy, \quad t > 0, \quad (3.7)$$

удовлетворяет (3.5), и, следовательно, решение u уравнения (3.6) действительно является плотностью U . Как следствие из теоремы восстановления в ее альтернативной форме мы получили следующий результат.

Теорема 2. Если F имеет непосредственно интегрируемую плотность f , то U имеет плотность u , такую, что $u(t) \rightarrow \mu^{-1}$.

Плотности, не являющиеся непосредственно интегрируемыми, вряд ли могут встретиться в практических ситуациях. Однако и относительно них мы можем сделать некоторые заключения. В самом деле, рассмотрим плотность

$$f_2(t) = \int_0^t f(t-y)f(y)dy \quad (3.8)$$

распределения $F \star F$. Вообще говоря, f_2 «много лучше», чем f . Например, если $f < M$, то из соображений симметрии получаем

$$f_2(t) < 2M \left[1 - F\left(\frac{1}{2}t\right) \right]. \quad (3.9)$$

Правая часть — монотонная интегрируемая функция при $\mu < \infty$ и отсюда следует, что f_2 непосредственно интегрируема. Разность $u - f$ есть решение уравнения восстановления с $z = f_2$. Таким образом, имеет место

Теорема 2а. Если F имеет ограниченную плотность f и конечное математическое ожидание μ , то

$$u(t) - f(t) \rightarrow \mu^{-1}. \quad (3.10)$$

Полученный результат любопытен, так как означает, что если колебания f велики, то u изменяется так, что компенсирует их. (Близкие результаты см. в задачах 7, 8.)

Условие ограниченности f является существенным. Мы покажем это на примере, который, кроме того, проливает новый свет на условие непосредственной интегрируемости в теореме восстановления.

Примеры. а) Пусть распределение вероятностей G , сосредоточенное на $\overline{0, 1}$, определено формулой

$$G(x) = \frac{1}{\log(e/x)}, \quad 0 < x \leq 1. \quad (3.11)$$

Это распределение обладает плотностью, которая непрерывна в открытом интервале, но не ограничена вблизи нуля, поскольку $x^{-1}G(x) \rightarrow \infty$ при $x \rightarrow 0$. Сумма n независимых случайных величин с распределением G , очевидно, меньше x , если каждая компонента $< x/n$. Следовательно,

$$G^{n*}(x) \geq (G(x/n))^n. \quad (3.12)$$

Отсюда следует, что при каждом n плотность распределения G^{n*} не ограничена вблизи нуля.

Положим теперь $F(x) = G(x-1)$. Тогда $F^{n*}(x) = G^{n*}(x-n)$ и, следовательно, F^{n*} имеет плотность, которая обращается в нуль при $x < n$, непрерывна при $x > n$, но не ограничена вблизи числа n . Поэтому плотность u функции восстановления $U = \sum F^{n*}$ не ограничена в окрестности любого целого числа $n > 0$.

б) Плотность u предыдущего примера удовлетворяет уравнению (3.6), которое совпадает при $z=f$ с обычным уравнением восстановления (1.3). Функция z интегрируема, обращается в нуль при $x > 2$ и непрерывна всюду, за исключением точки 1. Тот факт, что решение $Z=u$ не ограничено вблизи любого целого числа, показывает, что теорема восстановления неверна, даже если z сосредоточена на конечном интервале, не являясь интегрируемой по Риману в собственном смысле (последнее вытекает из неограниченности z). ►

§ 4. УСТОЙЧИВЫЕ (ВОЗВРАТНЫЕ) ПРОЦЕССЫ ВОССТАНОВЛЕНИЯ

Мы используем теперь теорему восстановления для вывода различных предельных теорем для процессов восстановления, описанных в гл. VI, 6. Мы рассмотрим последовательность взаимно независимых случайных величин $T_1, T_2, \dots$ с одним и тем же распределением F , которые назовем *промежутками между восстановлениями*. В этом параграфе мы предположим, что F — собственное распределение и $F(0) = 0$. В дополнение к T_k может быть задана неотрицательная случайная величина S_0 , имеющая собственное распределение F_0 . Положим

$$S_n = S_0 + T_1 + \dots + T_n. \quad (4.1)$$

Случайные величины S_n называются *моментами восстановления*. Процесс восстановления $\{S_n\}$ называется *чистым*, если $S_0 = 0$, и называется процессом с *запаздыванием* в других случаях.

Воспользуемся обозначением $U = \sum F^{n*}$, введенным в (1.2).

Математическое ожидание числа восстановлений на $\overline{0, t}$ равно

$$V(t) = \sum_{n=0}^{\infty} P\{S_n \leq t\} = F_0 \star U. \quad (4.2)$$

Таким образом¹⁾, при $h > 0$

$$V(t+h) - V(t) = \int_0^{t+h} [U(t+h-y) - U(t-y)] F_0\{dy\}. \quad (4.3)$$

Если F — неарифметическое распределение, то при $t \rightarrow \infty$ подынтегральное выражение стремится к $\mu^{-1}h$. Поэтому основная теорема может быть распространена на случай процессов с запаздыванием: если распределение F — неарифметическое, то математическое ожидание числа восстановлений в $t, t+h$ стремится к $\mu^{-1}h$. В этом утверждении содержатся две теоремы о равномерной распределенности: во-первых, скорость восстановления стремится к постоянной и, во-вторых, эта постоянная не зависит от начального распределения. В этом смысле мы имеем здесь аналог эргодических теорем для цепей Маркова [см. 1; гл. XV].

Если $\mu < \infty$, то $V(t) \sim \mu^{-1}t$ при $t \rightarrow \infty$. Естественно поставить вопрос о том, возможно ли выбрать F_0 так, чтобы получить $V(t) = \mu^{-1}t$, т. е. постоянную скорость восстановления. Но V удовлетворяет уравнению восстановления

$$V = F_0 + F \star V, \quad (4.4)$$

и потому $V(t) = \mu^{-1}t$ тогда и только тогда, когда

$$F_0(t) = \frac{t}{\mu} - \frac{1}{\mu} \int_0^t (t-y) F\{dy\}. \quad (4.5)$$

Интегрирование по частям показывает, что это эквивалентно равенству

$$F_0(t) = \frac{1}{\mu} \int_0^t [1 - F(y)] dy. \quad (4.6)$$

Функция F_0 является функцией распределения, так что ответ утвердителен: при начальном распределении (4.6) скорость восстановления постоянна, $V(t) = \mu^{-1}t$.

Распределение (4.6) появляется и как предельное распределение времени ожидания восстановления (или как предельное распределение величины перескока через некоторый уровень). Каж-

¹⁾ Вспомним о договоренности в § 1 — интервалы интегрирования предполагаются замкнутыми, пределы интегрирования могут быть заменены на $-\infty$ и ∞ .

дому $t > 0$ соответствует случайный индекс N_t , такой, что

$$S_{N_t} \leq t < S_{N_t+1}. \quad (4.7)$$

По терминологии гл. VI, 7, случайная величина $S_{N_t+1} - t$ называется *остающимся временем ожидания*, соответствующим моменту t . Обозначим через $H(t, \xi)$ вероятность того, что это время $\leq \xi$. Другими словами, $H(t, \xi)$ — это вероятность того, что *первый* после t момент восстановления лежит в интервале $t, t + \xi$ или что перескок через уровень t не превосходит ξ . Это событие происходит тогда и только тогда, когда некоторый момент восстановления S_n равен $x \leq t$, а время до следующего момента восстановления лежит между $t - x$ и $t - x + \xi$. В случае чистого процесса восстановления получаем, суммируя по всем x и n ,

$$H(t, \xi) = \int_0^t U\{dx\} [F(t - x + \xi) - F(t - x)]. \quad (4.8)$$

Этот интеграл содержит ξ как свободный параметр и имеет стандартный вид $U \star z$, где $z(t) = F(t + \xi) - F(t)$. Заметим, что последняя функция является непосредственно интегрируемой¹⁾. Предположим, что распределение F — неарифметическое. Из равенства

$$\int_0^\infty z(t) dt = \int_0^\infty ([1 - F(t)] - [1 - F(t + \xi)]) dt = \int_0^\xi (1 - F(s)) ds \quad (4.9)$$

получаем *предельную теорему*

$$\lim_{t \rightarrow 0} H(t, \xi) = \frac{1}{\mu} \int_0^\xi [1 - F(s)] ds \quad (4.10)$$

(легко проверить, что это соотношение верно и для процессов с запаздыванием при любом начальном распределении F_0). Эта предельная теорема замечательна в нескольких отношениях. Как показывает дальнейшее обсуждение, она тесно связана с «парадоксом контроля» из гл. VI, 7, и «парадоксом времени ожидания» из гл. I, 4.

Если $\mu < \infty$, то предельное распределение (4.10) совпадает с распределением (4.6). Таким образом, при $\mu < \infty$ *остающееся время ожидания имеет собственное предельное распределение, которое совпадает с распределением, приводящим к постоянной скорости*

¹⁾ При $n\xi < x < (n+1)\xi$ имеем $z(x) \leq F((n+2)\xi) - F(n\xi)$, и ряд из этих членов, очевидно, сходится.

восстановления. В этом примере мы сталкиваемся еще раз со сходимостью к «стационарному режиму».

Предельное распределение в (4.10) имеет конечное математическое ожидание в том и только том случае, когда F имеет дисперсию. Это указывает, грубо говоря, на то, что «вероятности вхождения» ведут себя хуже, чем F . В случае $\mu = \infty$ имеем при всех ξ

$$H(t, \xi) \rightarrow 0, \quad (4.11)$$

т. е. вероятность того, что *перескок через уровень t превзойдет сколь угодно большое число ξ* , стремится к единице (для случая правильно меняющихся хвостов распределения более точная информация дается в гл. XIV, 3).

Примеры. а) *Суперпозиция процессов восстановления.* Пусть даны n процессов восстановления. Соединяя все моменты восстановления вместе, мы получим некоторый новый процесс. Вообще говоря, новый процесс не будет процессом восстановления, однако для него легко подсчитать время ожидания W первого после нуля восстановления. Мы покажем, что при весьма общих условиях *распределение W приближается к показательному закону*, а составной процесс — к процессу Пуассона. Этот результат объясняет, почему многие процессы (такие, как поток вызовов, поступающих на телефонную станцию) похожи на процесс Пуассона.

Рассмотрим n взаимно независимых процессов восстановления, у которых промежутки между восстановлениями имеют распределения $F_1, \dots, F_n$ с математическими ожиданиями $\mu_1, \dots, \mu_n$ соответственно. Положим

$$\frac{1}{\mu_1} + \dots + \frac{1}{\mu_n} = \frac{1}{\alpha}. \quad (4.12)$$

Мы потребуем, грубо говоря, чтобы моменты восстановления в каждом из объединяемых процессов были расположены крайне редко, так что суммарный эффект складывается из многих малых частей. Чтобы удовлетворить это требование, мы допустим, что при фиксированных k и y вероятности $F_k(y)$ малы, а μ_k велики. Это предположение оказывается осмысленным в форме предельной теоремы.

Рассмотрим «стационарную» ситуацию, предполагая, что рассматриваемые процессы протекают уже длительное время. Тогда для времени ожидания W_k ближайшего момента восстановления в k -м процессе имеем приближенное равенство

$$P\{W_k \leq t\} \approx \frac{1}{\mu_k} \int_0^t (1 - F_k(y)) dy \approx \frac{t}{\mu_k}. \quad (4.13)$$

[Последняя аппроксимация основана на малости $F_k(y)$.] Время ожидания W в объединенном процессе равно наименьшему из W_k ,

и потому

$$P\{W > t\} \approx \left(1 - \frac{t}{\mu_1}\right) \dots \left(1 - \frac{t}{\mu_n}\right) \approx e^{-t\alpha}. \quad (4.14)$$

Эти рассуждения легко сделать строгими, и при указанных выше условиях показательное распределение будет предельным при $n \rightarrow \infty$.

б) *Вероятности достижения областей в случайных блужданиях.* Рассмотрим последовательность независимых случайных величин $X_1, X_2, \dots$ и их частные суммы

$$Y_n = X_1 + \dots + X_n.$$

Для положительных X_k случайное блуждание $\{Y_n\}$ представляет собой процесс восстановления, однако сейчас мы рассмотрим произвольные X_k . Допустим, что случайное блуждание устойчиво, так что для любого $t > 0$ с вероятностью единица найдется такое n , что $Y_n > t$. Если N — наименьший индекс, для которого это верно, то Y_N называется точкой *первого вхождения* в $\overline{t, \infty}$. Случайная величина $Y_N - t$ равна перескоку через уровень t при первом вхождении и соответствует остающемуся времени ожидания в процессах восстановления. Положим снова $P\{Y_N \leq t + \xi\} = H(t, \xi)$ и покажем, каким образом предельная теорема для остающегося времени ожидания может быть применена к этому распределению.

Определим S_1 как точку первого вхождения в $\overline{0, \infty}$ и по индукции определим S_{n+1} как точку первого вхождения в $\overline{S_n, \infty}$. Последовательность $S_1, S_2, \dots$ совпадает с *последовательностью лестничных высот*, введенных в гл. VI, 8, и образует процесс восстановления, так как разности $S_{n+1} - S_n$ очевидным образом независимы и имеют такое же распределение, как S_1 . Разность $Y_N - t$ есть остающееся время ожидания в процессе восстановления $\{S_n\}$, так что можно применить (4.10). ►

Методом, использованным при доказательстве (4.10), можно показать, что *прошедшее время ожидания* $t - S_{N_t}$ имеет то же самое предельное распределение. Для длины промежутка $L_t = S_{N_t+1} - S_{N_t}$ между восстановлениями, содержащего t , имеем

$$P\{L_t < \xi\} = \int_{t-\xi}^t U\{dx\} [F(\xi) - F(t-x)]. \quad (4.15)$$

Следовательно,

$$\lim_{t \rightarrow \infty} P\{L_t < \xi\} = \frac{1}{\mu} \int_0^{\xi} [F(\xi) - F(y)] dy = \frac{1}{\mu} \int_0^{\xi} x F\{dx\}. \quad (4.16)$$

Любопытные следствия этой формулы обсуждались в связи с «парадоксом контроля» в гл. VI, 7 и «парадоксом времени ожидания» в гл. I, 4.

Легко видеть, что три семейства случайных величин $t - S_{N_t}$, $S_{N_t+1} - t$, L_t суть однородные марковские процессы. Доказанные нами три предельные теоремы следуют из эргодических теорем для марковских процессов (см. также гл. XIV, 3).

§ 5. ЧИСЛО N_t МОМЕНТОВ ВОССТАНОВЛЕНИЯ

Рассмотрим для простоты чистый процесс восстановления, так что r -й момент восстановления является суммой r независимых случайных величин с одинаковым распределением F . Нуль считается моментом восстановления. Обозначим через N_t число моментов восстановления на $\overline{0, t}$. Событие $\{N_t > r\}$ равносильно тому, что r -й момент восстановления находится на $\overline{0, t}$. Поэтому

$$P\{N_t > r\} = F^{r*}(t). \quad (5.1)$$

Очевидно, что $N_t \geq 1$. Тогда

$$E(N_t) = \sum_{r=0}^{\infty} P\{N_t > r\} = U(t). \quad (5.2)$$

(О моментах N_t второго порядка см. в задаче 13.) Случайные величины N_t встречаются также в задачах последовательного анализа. Предположим, что последовательные испытания (наблюдения) с результатами $T_1, T_2, \dots, T_n$ продолжаются до тех пор, пока сумма результатов испытаний не превзойдет в первый раз t . Тогда N_t равно полному числу испытаний. Можно было бы избежать многих утомительных вычислений, если применить оценку (3.1), предлагаемую уточненной теоремой восстановления.

Если F имеет математическое ожидание μ и дисперсию σ^2 , то асимптотическое поведение распределения величины N_t определяется тем, что распределение F^{r*} асимптотически нормально. Соответствующие расчеты можно найти в 1, XIII, 6, и они не связаны с арифметическим характером F . Итак, имеет место следующий результат.

Центральная предельная теорема для случайной величины N_t . Если F имеет математическое ожидание μ и дисперсию σ^2 , то для больших t число N_t моментов восстановления распределено приблизительно нормально с математическим ожиданием $t\mu^{-1}$ и дисперсией $t\sigma^2\mu^{-3}$.

Пример. а) Счетчики первого типа. Поступающие частицы образуют пуассоновский поток. Частица, попадающая в счетчик в момент, когда он работает, регистрируется, но «закрывает» счетчик на фиксированное время ξ . Частицы, достигающие счетчика в то время, когда он «закрывает», не регистрируются. Для простоты рассмотрим процесс с того момента, когда новая частица попадает в работающий счетчик. Мы имеем два процесса восста-

новления. Первый процесс — входной поток. Это процесс Пуассона, т. е. промежутки между появлениями частиц имеют показательное распределение $1 - e^{-ct}$ с математическим ожиданием c^{-1} и дисперсией c^{-2} . Последовательные регистрации образуют вторичный процесс восстановления, в котором промежутки между последовательными регистрациями представляют собой суммы ξ и показательно распределенных случайных величин. Поэтому их математическое ожидание равно $\xi + c^{-1}$, а дисперсия равна c^{-2} . Таким образом, число регистраций на временном интервале $0, t$ распределено приблизительно нормально с математическим ожиданием $tc(1 + c\xi)^{-1}$ и дисперсией $tc(1 + c\xi)^{-3}$.

Из того что эти величины различны, следует, что число регистраций не подчиняется распределению Пуассона. В свое время не было отчетливого понимания того факта, что процесс регистрации существенно отличается от первоначального процесса. Результаты наблюдений привели некоторых физиков к ошибочному заключению, что ливни космических частиц не имеют пуассоновского характера «чистой случайности».

Предельное распределение для N_t существует тогда и только тогда, когда F принадлежит области притяжения некоторого закона. Области притяжения описаны в гл. IX, 8, и XVII, 5, откуда можно вывести, что N_t имеет собственное предельное распределение тогда и только тогда, когда

$$1 - F(x) \sim x^{-\alpha} L(x), \quad x \rightarrow \infty, \quad (5.3)$$

где L — медленно меняющаяся функция и $0 < \alpha < 2$. Предельное распределение для N_t легко вычислить. При этом обнаруживаются парадоксальные свойства флуктуаций. Их поведение существенно различно при $\alpha < 1$ и $\alpha > 1$.

Рассмотрим случай $0 < \alpha < 1$. Если a_r выбраны так, что

$$r[1 - F(a_r)] \rightarrow \frac{2 - \alpha}{\alpha}, \quad (5.4)$$

то $F^{r*}(a_r x) \rightarrow G_\alpha(x)$, где G_α — одностороннее устойчивое распределение, удовлетворяющее условию $x^\alpha [1 - G_\alpha(x)] \rightarrow (2 - \alpha)/\alpha$ при $x \rightarrow \infty$. (Ср. с гл. IX, 6, и гл. XVII, 5, а также гл. XIII, 6.) Пусть r и t возрастают таким образом, что $t \sim a_r x$. Тогда из (5.3) и (5.4) с учетом того, что функция L медленно меняющаяся, получаем

$$r \sim \frac{2 - \alpha}{\alpha} \frac{x^{-\alpha}}{1 - F(t)}, \quad (5.5)$$

откуда в силу (5.1)

$$P \left\{ [1 - F(t)] N_t \geq \frac{2 - \alpha}{\alpha} x^{-\alpha} \right\} \rightarrow G_\alpha(x). \quad (5.6)$$

Это соотношение — аналог центральной предельной теоремы. Частный случай для $\alpha = 1/2$ содержится в 1; гл. XIII, 6. Неожидаанные черты поведения N_t связаны с нормирующим множителем $1 - F(t)$ в (5.6). Грубо говоря, $1 - F(t)$ имеет порядок $t^{-\alpha}$, так что величина N_t имеет порядок t^α ; плотность моментов восстановления должна неограниченно убывать (что согласуется с асимптотическим поведением вероятностей достижения).

При $1 < \alpha < 2$ распределение F имеет математическое ожидание $\mu < \infty$ и

вычисления, аналогичные проделанным выше, показывают, что

$$P \left\{ N_t \geq \frac{t - \lambda(t)x}{\mu} \right\} \rightarrow G_\alpha(x), \quad (5.7)$$

где $\lambda(t)$ удовлетворяет соотношению

$$t[1 - F(\lambda(t))] \rightarrow \frac{2 - \alpha}{\alpha} \mu. \quad (5.8)$$

В этом случае среднее число моментов восстановления растет линейно, однако наличие нормирующего множителя $\lambda(t)$ показывает, что случайные отклонения от среднего могут быть весьма велики.

§ 6. ОБРЫВАЮЩИЕСЯ (НЕВОЗВРАТНЫЕ) ПРОЦЕССЫ

Общая теория процессов восстановления с несобственным распределением F почти тривиальна. Однако соответствующее уравнение восстановления часто встречается в различных «масках», когда случайные черты затемняют общее основание. Ясное понимание этого основного факта позволит избежать громоздких рассуждений в отдельных задачах. В частности, асимптотическая формула теоремы 2 влечет результаты, для доказательства которых ранее использовался в специально преобразованном виде аппарат уравнений Винера—Хопфа.

Чтобы избежать путаницы в обозначениях, мы будем использовать для исходного распределения букву L (вместо F). Иными словами, в этом параграфе L обозначает *несобственное распределение вероятностей* с $L(0) = 0$ и $L(\infty) = L_\infty < 1$. Оно играет роль распределения (несобственных) промежутков T_k между последовательными восстановлениями, причем «дефект» $1 - L_\infty$ равен вероятности обрыва процесса. Начало координат на оси времени считается моментом восстановления с нулевым номером, а $S_n = T_1 + \dots + T_n$ представляет собой n -й момент восстановления; это несобственная случайная величина с распределением L^{n*} , которому соответствует полная масса $L^{n*}(\infty) = L_\infty^n$. «Дефект» $1 - L_\infty^n$ равен вероятности обрыва процесса до n -го момента восстановления. Положим снова

$$U = \sum_{n=0}^{\infty} L^{n*}. \quad (6.1)$$

Как и в случае возвратных процессов, $U(t)$ равно *математическому ожиданию числа моментов восстановления* в интервале $\overline{0, t}$. На этот раз, однако, математическое ожидание полного числа восстановлений является конечным и равно

$$U(\infty) = \frac{1}{1 - L_\infty}. \quad (6.2)$$

Вероятность того, что n -й момент восстановления S_n будет последним и при этом $\leq x$, равна $(1 - L_\infty) L^{n*}(x)$. Таким образом, верна

Теорема 1. *Обрывающийся процесс восстановления, начинающийся в начале координат, обрывается с вероятностью единица. Момент обрыва M (т. е. максимум последовательности $0, S_1, S_2, \dots$) имеет собственное распределение*

$$P\{M \leq x\} = (1 - L_\infty) U(x). \quad (6.3)$$

Вероятность того, что n -й момент восстановления будет последним, равна $(1 - L_\infty) L_\infty^n$, т. е. число моментов восстановления имеет геометрическое распределение.

Эти результаты возможно изложить и в терминах несобственного уравнения восстановления

$$Z(t) = z(t) + \int_0^t Z(t-y) L\{dy\}, \quad (6.4)$$

но при несобственном распределении L соответствующая теория тривиальна. В предположении, что $z(x) = 0$ при $x \leq 0$, единственное решение задается формулой

$$Z(t) = \int_0^t z(t-y) U\{dy\} \quad (6.5)$$

и, как очевидно,

$$Z(t) \rightarrow \frac{z(\infty)}{1 - L_\infty}, \quad (6.6)$$

если только $z(t) \rightarrow z(\infty)$ при $t \rightarrow \infty$.

Примеры. а) Событие $\{M \leq t\}$ происходит, если процесс заканчивается на S_0 или если T_1 принимает некоторое положительное значение $y \leq t$ и остающийся процесс «доживает» до момента, который $\leq t - y$. Таким образом, $Z(t) = P\{M \leq t\}$ удовлетворяет уравнению восстановления

$$Z(t) = 1 - L_\infty + \int_0^t Z(t-y) L\{dy\}. \quad (6.7)$$

Это уравнение эквивалентно (6.3).

б) *Вычисление моментов.* В последнем соотношении собственное распределение Z представляется в виде суммы двух несобственных распределений: одно определено, как свертка, а другое имеет единственный атом в нуле. Чтобы вычислить математическое ожидание Z , положим

$$E_L = \int_0^\infty x L\{dx\}. \quad (6.8)$$

Аналогичное обозначение примем и для других распределений (как собственных, так и несобственных). Так как L — несобственное распределение, то свертка в (6.7) имеет математическое ожи-

дание $L_{\infty} \cdot E_Z + E_L$. Таким образом, $E_Z = E_L / (1 - L_{\infty})$. Для более общего уравнения (6.4) подобным же путем получаем

$$E_Z = \frac{E_Z + E_L}{1 - L_{\infty}}. \quad (6.9)$$

Моменты более высокого порядка могут быть подсчитаны этим же методом. ▶

Асимптотические оценки

В применениях обычно $z(t)$ стремится к пределу $z(\infty)$, и в таком случае $Z(t)$ стремится к пределу $Z(\infty)$, который дан в (6.6). Поэтому важно установить асимптотические оценки для разности $Z(\infty) - Z(t)$. Это можно сделать методом, который широко применим в теории случайных блужданий [см. сопряженные случайные блуждания в примере 4, б) гл. XII]. Этот метод связан с (как правило, безобидным) предположением, что существует такое число κ , что

$$\int_0^{\infty} e^{\kappa y} L \{dy\} = 1. \quad (6.10)$$

Такой корень κ уравнения (6.10), очевидно, единственный, и так как распределение L несобственное, то $\kappa > 0$. Введем теперь собственное распределение вероятностей L^* равенством

$$L^* \{dy\} = e^{\kappa y} L \{dy\} \quad (6.11)$$

и сопоставим каждой функции f новую функцию f^* , определенную соотношением

$$f^*(x) = e^{\kappa x} f(x).$$

Достаточно взглянуть на (6.4), чтобы понять, что выполняется уравнение восстановления

$$Z^*(t) = z^*(t) + \int_0^t Z^*(t-y) L^* \{dy\}. \quad (6.12)$$

Если $Z^*(t) \rightarrow a \neq 0$, то $Z(t) \sim ae^{-\kappa t}$. Таким образом, если z^* непосредственно интегрируема [и в этом случае $z(\infty) = 0$], то из теоремы восстановления следует, что

$$e^{-\kappa t} Z(t) \rightarrow \frac{1}{\mu^*} \int_0^{\infty} e^{\kappa x} z(x) dx, \quad (6.13)$$

где

$$\mu^* = \int_0^{\infty} e^{\kappa y} y L \{dy\}. \quad (6.14)$$

В (6.13) содержится хорошая оценка для $Z(t)$ при больших t . С незначительным изменением эта процедура применима и

тогда, когда $z(\infty) \neq 0$. Положим

$$z_1(t) = z(\infty) - z(t) + z(\infty) \frac{L - L(t)}{1 - L_\infty}.$$

Легко проверить, что разность $Z(\infty) - Z(t)$ удовлетворяет обычному уравнению восстановления (6.4) с заменой z на z_1 . Интегрируя по частям, находим, что интеграл от $z_1^{\#}(x) = z_1(x) e^{x\kappa}$ равен правой части (6.15). Применяя (6.13) к Z_1 , мы приходим к следующему результату.

Теорема 2. Если выполняется (6.10), то решение уравнения восстановления удовлетворяет условию

$$\mu^* e^{x\kappa} [Z(\infty) - Z(t)] \rightarrow \frac{z(\infty)}{\kappa} + \int_0^\infty e^{x\kappa} [z(\infty) - z(x)] dx \quad (6.15)$$

при дополнительном предположении, что $\mu^* \neq \infty$ и что z_1 непосредственно интегрируема.

Для частного случая (6.7) получаем

$$P\{M > t\} \sim \frac{1 - L_\infty}{\kappa \mu^*} e^{-\kappa t}. \quad (6.16)$$

В следующем параграфе мы покажем, как поразительно сильна эта оценка. В частности, в примере б) будет показано, как с помощью нашего простого метода можно иногда быстро получить результаты, требующие обычно более глубоких и трудоемких методов.

Случай $L_\infty > 1$. Если $L_\infty > 1$, то существует постоянная $\kappa < 0$, такая, что (6.10) верно, и тогда описанное выше преобразование сводит интегральное уравнение (6.4) к (6.12). Теорема восстановления приводит, таким образом, к точным оценкам асимптотического поведения $Z(t) e^{t\kappa}$. [Дискретный случай охватывается теоремой 1 из 1; гл. XIII, 10. О применениях в демографии см. пример 10, д) из 1; гл. XII.]

§ 7. РАЗЛИЧНЫЕ ПРИМЕНЕНИЯ

Как уже указывалось, теория предыдущего параграфа может быть применена к задачам, которые не связаны непосредственно с процессами восстановления. В этом параграфе мы дадим два не зависящих друг от друга примера.

а) *Крамеровская оценка для риска.* В гл. VI, 5, было показано, что в решении задачи о разорении в обобщенном пуассоновском процессе, а также в решении проблем, связанных с размером хранилищ, назначением пациентов и т. п., используется распределение R , сосредоточенное на $\overline{0, \infty}$ и удовлетворяющее интегро-дифференциальному уравнению

$$R'(z) = \frac{\alpha}{c} R(z) - \frac{\alpha}{c} \int_0^z R(z-x) F\{dx\}, \quad (7.1)$$

где F — собственное распределение¹⁾. Интегрируя (7.1) на $\overline{0, t}$ и производя очевидное интегрирование по частям, получаем

$$R(t) - R(0) = \frac{\alpha}{c} \int_0^t R(t-x) [1 - F(x)] dx. \quad (7.2)$$

Здесь $R(0)$ — неизвестная постоянная, но в остальном (7.2) представляет собой уравнение восстановления с *несобственным* распределением L , имеющим плотность $(\alpha/c)[1 - F(x)]$. Обозначим математическое ожидание F через μ . Тогда полная масса L равна $L_\infty = \alpha\mu/c$. [Рассматриваемый процесс имеет смысл только при $L_\infty < 1$, так как в противном случае $R(t) = 0$ при всех t .] Заметим, что (7.2) является частным случаем (6.4) и что $R(\infty) = 1$. Вспоминая (6.6), получаем

$$R(0) = 1 - \frac{\alpha\mu}{c}. \quad (7.3)$$

Подставляя это значение в уравнение (7.2), мы приведем его к уравнению вида (6.7) для распределения времени жизни M обрывающегося процесса, в котором промежуток между соседними восстановлениями имеет распределение L . Из (6.16) следует, что *если существует постоянная κ , такая, что*

$$\frac{\alpha}{c} \int_0^\infty e^{\kappa x} [1 - F(x)] dx = 1 \quad (7.4)$$

и

$$\mu^* = \frac{\alpha}{c} \int_0^\infty e^{\kappa x} x [1 - F(x)] dx < \infty, \quad (7.5)$$

то при $t \rightarrow \infty$

$$1 - R(t) \sim \frac{1}{\kappa\mu^*} \left(1 - \frac{\alpha\mu}{c}\right) e^{-\kappa t}. \quad (7.6)$$

Это известная оценка, принадлежащая Крамеру, была первоначально получена при помощи глубоких методов теории комплексной переменной. Моменты R могут быть вычислены, как указано в примере 6, б).

б) *Пробелы в процессе Пуассона*. В гл. VI, 7, мы вывели уравнение восстановления для распределения V времени ожидания первого пробела длины $\geq \xi$ в процессе восстановления. Когда этот процесс является процессом Пуассона, то промежутки между последовательными восстановлениями распределены показательно, и уравнение восстановления (7.2) гл. VI имеет стандартную

¹⁾ Это специальный случай (5.4) гл. VI с распределением F , сосредоточенным на $\overline{0, \infty}$. Он будет рассмотрен с помощью преобразований Лапласа в гл. XIV, 2, б). Общую ситуацию см. в гл. XII, 5, г).

форму $V = z + V \star L$ с

$$L(x) = \begin{cases} 1 - e^{-cx}, & z(x) = 0 \quad \text{при } x < \xi, \\ 1 - e^{-c\xi}, & z(x) = e^{-c\xi} \quad \text{при } x \geq \xi. \end{cases} \quad (7.7)$$

Так как $z(\infty) = 1 - L_\infty$, то решение V будет собственным распределением вероятностей, что соответствует смыслу задачи.

Моменты рассматриваемого нами времени ожидания W легко вычислить методом, описанным в примере 6, б). Мы получаем

$$\begin{aligned} E(W) &= \frac{ec\xi - 1}{c}, \\ \text{Var}(W) &= \frac{e^{2c\xi} - 1 - 2c\xi e^{c\xi}}{c^2}. \end{aligned} \quad (7.8)$$

Условимся интерпретировать W как *время ожидания пешеходом возможности пересечь поток машин*. Тогда указанные формулы показывают, как влияет на ответ возрастание плотности потока. Математическое ожидание числа машин, проходящих за время, необходимое для пересечения потока, равно $c\xi$. Полагая $c\xi = 1$ и 2, видим, что $E(W) \approx 1,72\xi$ и $E(W) \approx 3,2\xi$ соответственно. Дисперсия возрастает от (примерно) ξ^2 до $6\xi^2$. [Явное решение и связь с теоремой о покрытии указаны в примере 2, а) гл. XIV.] Можно применить асимптотическую оценку (6.14). При $c\xi > 1$ основное уравнение (6.10) сводится к

$$ce^{(\kappa - c)\xi} = \kappa, \quad 0 < \kappa < c, \quad (7.9)$$

и стандартными вычислениями мы выводим из (6.14), что

$$1 - V(t) \sim \frac{1 - \kappa/c}{1 - \kappa\xi} e^{-\kappa t}. \quad (7.10) \quad \blacktriangleright$$

§ 8. СУЩЕСТВОВАНИЕ ПРЕДЕЛОВ В СЛУЧАЙНЫХ ПРОЦЕССАХ

Сила теоремы восстановления наиболее ярко демонстрируется, пожалуй, тем, что она позволяет без всяких усилий доказать существование «стационарного режима» в обширном классе случайных процессов. От самого процесса мы потребуем лишь, чтобы рассматриваемые вероятности были определены. В противном случае теорема имеет чисто аналитический характер¹⁾.

Рассмотрим случайный процесс со счетным множеством состояний $E_0, E_1, \dots$ и обозначим $P_k(t)$ вероятность состояния E_k в момент $t > 0$. Нижеследующая теорема опирается на существование «рекуррентных событий», т. е. моментов, в которые процесс как бы начинается заново. Более точно, мы предполагаем, что существует момент S_1 , такой, что течение процесса после S_1

¹⁾ Более сложные результаты см. в статье: Beneš V. E., A «renewal» limit theorem for general stochastic processes, Ann. Math. Statist., 33 (1962), 98—113, или в книге того же автора (1963),

является точной вероятностной копией всего процесса, начинающегося в момент 0. Отсюда вытекает существование целой последовательности моментов $S_2, S_3, \dots$ с тем же самым свойством. Последовательность $\{S_n\}$ представляет собой возвратный процесс восстановления. Мы допустим, что среднее время между восстановлениями $\mu = E(S_1)$ конечно. Обозначим $P_k(t)$ условную вероятность E_k в момент $t + s$ при условии, что $S_1 = s$. Предполагается, что эти вероятности не зависят от s . При этих условиях мы докажем следующее важное утверждение.

Теорема. Предел

$$\lim_{t \rightarrow \infty} P_k(t) = p_k \quad (8.1)$$

существует, причем $p_k \geq 0$ и $\sum p_k = 1$.

Доказательство. Пусть $q_k(t)$ — вероятность объединения событий: $S_1 > t$ и «в момент t система находится в состоянии E_k ». Тогда

$$\sum_{k=0}^{\infty} q_k(t) = 1 - F(t), \quad (8.2)$$

где F — распределение «времен восстановления» $S_{n+1} - S_n$. По предположению

$$P_k(t) = q_k(t) + \int_0^t P_k(t-y) F(dy). \quad (8.3)$$

Функция q_k непосредственно интегрируема, поскольку она мажорируется монотонной интегрируемой функцией $1 - F$. Поэтому по второй теореме восстановления

$$\lim_{t \rightarrow \infty} P_k(t) = \frac{1}{\mu} \int_0^{\infty} q_k(t) dt. \quad (8.4)$$

Интегрируя (8.2), видим, что сумма этих пределов равна единице. Теорема доказана. ►

Примечательно, что существование предела (8.1) установлено безотносительно к способу его вычисления.

Замечание. Если F_0 — собственное распределение, то (8.1) влечет

$$\int_0^t P_k(t-y) F_0(dy) \rightarrow p_k, \quad t \rightarrow \infty. \quad (8.5)$$

Таким образом, теорема охватывает и случай процесса $\{S_n\}$ восстановления с запаздыванием, в котором S_0 имеет распределение F_0 .

Примеры. а) *Теория очередей.* Рассмотрим систему (типа телефонной станции, почтового отделения или части вычислительной машины), состоящую из одного или более «обслуживающих устройств», и пусть E_k обозначает состояние: в системе имеется k ожидающих «клиентов». В большинстве моделей процесс начинается заново, когда прибывающие «клиенты» заставляют систему в состоянии E_0 . В этом случае наша предельная теорема применима тогда и только тогда, когда с вероятностью единица наступает такой момент и когда математическое ожидание времени между последовательными наступлениями конечно.

б) *Двухстепенный процесс восстановления.* Допустим, что имеются два возможных состояния E_1 и E_2 . Первоначально система находится в E_1 . Последовательные промежутки пребывания в E_1 суть случайные величины X_j с одним и тем же распределением F_1 . Они перемежаются с промежутками пребывания в E_2 , имеющими одно и то же распределение F_2 . Предполагая, как обычно, что все рассматриваемые случайные величины независимы, мы получаем «вложенный» процесс восстановления с распределением промежутков между восстановлениями, равным $F = F_1 \star F_2$. Допустим, что $E(X_j) = \mu_1 < \infty$ и $E(Y_j) = \mu_2 < \infty$. Очевидно, $q_1(t) = 1 - F_1(t)$, и потому при $t \rightarrow \infty$ вероятности состояний E_k стремятся к пределам

$$P_1(t) \rightarrow \frac{\mu_1}{\mu_1 + \mu_2}, \quad P_2(t) \rightarrow \frac{\mu_2}{\mu_1 + \mu_2}. \quad (8.6)$$

Эти рассуждения легко переносятся на случай многостепенных систем.

в) Дифференциальные уравнения из 1; гл. XVII, соответствуют случайным процессам, в которых последовательные возвращения в *любое* состояние образуют процесс восстановления требуемого типа. Поэтому наша теорема гарантирует существование предельных вероятностей в этом случае. Их явный вид можно легко найти из дифференциальных уравнений, в которых производные заменены нулем. [См., например, 1; гл. XVII, (7.3). Мы вернемся к более систематическому изложению этого вопроса в гл. XIV, 9. Точно такие же рассуждения применимы к полумарковским процессам, описанным в задаче 14 гл. XIV, 10.] ►

§ 9*). ТЕОРИЯ ВОССТАНОВЛЕНИЯ НА ВСЕЙ ПРЯМОЙ

В этом параграфе теория восстановления будет распространена на распределения, не сосредоточенные на полупрямой. Чтобы избежать тривиальностей, мы предположим, что $F\{-\infty, 0\} > 0$, $F\{0, \infty\} > 0$ и что F — неарифметическое распределение. Случай

*) Материал не используется в дальнейшем.

арифметических распределений требует видоизменений, которые производятся очевидным образом по аналогии с § 1.

Мы напомним (см. гл. VI, 10), что распределение F *невозвратно* тогда и только тогда, когда

$$U\{I\} = \sum_{n=0}^{\infty} F^n \star \{I\} \quad (9.1)$$

принимает конечные значения для всех конечных интервалов. В противном случае $U\{I\} = \infty$ для *каждого* интервала и F называется *возвратным*. Для невозвратных распределений естественно возникает вопрос: переносятся ли на них теоремы восстановления § 1? Эта задача интересовала многих математиков, и, пожалуй, не из-за ее важности, а в основном по причине связанных с ней неожиданных трудностей. Так, теорема восстановления переносилась шаг за шагом на различные специальные классы невозвратных распределений Блекуэллом, Чжуном, Чжуном и Полардом, Чжуном и Вольфовицем, Карлином и Смитом. Однако общая теорема была доказана только в 1961 г. Феллером и Ореем с использованием как вероятностных приемов, так и анализа Фурье. Приводимое ниже доказательство значительно проще и более элементарно. Заметим, что если F имеет конечное математическое ожидание, то применимо без изменений доказательство § 2. (О теории восстановления на плоскости см. в задаче 20.)

Для дальнейшего нам нужно вспомнить, что распределение с математическим ожиданием $\mu \neq 0$ является невозвратным (теорема 4 из гл. VI.10). Как обычно, $I+t$ обозначает интервал, полученный сдвигом I на t .

Общая теорема восстановления. а) Если F имеет математическое ожидание $\mu > 0$, то для всякого конечного интервала I длины $h > 0$

$$U\{I+t\} \rightarrow \frac{h}{\mu}, \quad t \rightarrow \infty, \quad (9.2)$$

$$U\{I+t\} \rightarrow 0, \quad t \rightarrow -\infty. \quad (9.3)$$

б) Если F *невозвратно* и не имеет математического ожидания, то $U\{I+t\} \rightarrow 0$ при $t \rightarrow \pm \infty$ для *каждого* конечного интервала I .

В дальнейшем предполагается, что F *невозвратно* и что z — непрерывная функция, которая неотрицательна при $-h < x < h$ и обращается в нуль вне этого интервала.

Прежде чем приступить к доказательству теоремы, напомним несколько фактов из гл. VI, 10.

Свертка $Z = U \star z$ вполне определена равенством

$$Z(x) = \int_{-\infty}^{+\infty} z(x-y) U\{dy\}, \quad (9.4)$$

так как в действительности область интегрирования конечна. По

теореме 2 гл. VI, 10, функция Z непрерывна, удовлетворяет уравнению восстановления

$$Z = z + F \star Z \quad (9.5)$$

и имеет максимум в точке ξ , где $z(\xi) > 0$.

Положим $U_n = F^0 \star + \dots + F^n \star$. Любое неотрицательное решение Z уравнения (9.5) удовлетворяет неравенству $Z \geq z = U_0 \star z$ и, следовательно, удовлетворяет (по индукции) $Z \geq U_n \star z$. Отсюда следует, что решение (9.4) минимально в том смысле, что $Z_1 \geq Z$ для любого другого неотрицательного решения Z_1 . Поскольку $Z_1 = Z + \text{const}$ есть снова решение, то

$$\liminf Z(x) = 0, \quad x \rightarrow \pm \infty. \quad (9.6)$$

Лемма 1. Для любой постоянной a

$$Z(x+a) - Z(x) \rightarrow 0, \quad x \rightarrow \pm \infty. \quad (9.7)$$

Доказательство совпадает с доказательством леммы 2 в § 2. Там мы использовали тот факт, что ограниченное равномерно непрерывное решение уравнения свертки

$$\xi = F \star \xi, \quad (9.8)$$

достигающее своего максимума при $x=0$, равно постоянной. Это остается верным и для распределений, не сосредоточенных на $0, \infty$, а соответствующее доказательство в действительности проще, так как в этом случае множество Σ , образованное точками роста $F, F^2 \star, \dots$, является всюду плотным. ►

Приведем ниже, хотя и не будем использовать, интересное следствие из леммы 1.

Следствие 1¹⁾. Любое ограниченное непрерывное решение уравнения (9.8) равно постоянной.

Доказательство. В случае когда само решение (9.8) ξ равномерно непрерывно, то применимо без изменений доказательство леммы 1. Если же G есть произвольное распределение вероятностей, то $\xi_1 = G \star \xi$ снова есть решение уравнения (9.8). Можно выбрать такое распределение G , что ξ_1 имеет ограниченную производную и поэтому ξ_1 равномерно непрерывна. В частности, свертка ξ с произвольным нормальным распределением дает постоянную. Если теперь уменьшать дисперсию G до нуля, то мы видим, что функция ξ сама сводится к постоянной. ►

Доказательство теоремы восстановления в случае, когда математическое ожидание μ существует. При $0 < \mu < \infty$ применимо доказательство из § 2 с одним тривиальным изменением. В итоговом соотношении (2.13) мы использовали функцию $z=1-F$, и для нее решение Z было тождественно равно 1. Вместо этой

¹⁾ Это утверждение верно для распределений на произвольных группах, см.; Choquet G., Deny C. R., Acad. Sci. Paris, vol. 250 (1960), 799–801.

функции мы используем здесь

$$z = F^0 \star -F. \quad (9.9)$$

При этом $U_n \star z = F^0 \star -F^{(n+1)} \star$, и, поскольку $\mu > 0$, ясно, что $Z = F^0 \star$. Следует отметить, что это доказательство применимо и тогда, когда $\mu = +\infty$ (в том очевидном смысле, что интеграл от $x F\{dx\}$ расходится на $0, \infty$, но сходится на $-\infty, 0$). ►

Если математическое ожидание не существует, то доказательство теоремы требует более тонкого анализа. Следующая лемма показывает, что достаточно доказать основное утверждение для одного хвоста.

Лемма 2. *Предположим, что F не имеет математического ожидания и что*

$$U\{I-t\} \rightarrow 0, \quad t \rightarrow +\infty. \quad (9.10)$$

Тогда

$$U\{I+t\} \rightarrow 0, \quad t \rightarrow +\infty. \quad (9.11)$$

Доказательство. Мы воспользуемся результатом примера 4, б), имеющего прямое отношение к вероятностям достижения состояний в случайном блуждании, которое управляется распределением F . Обозначим через $H(t, \xi)$ вероятность того, что первое вхождение в t, ∞ имеет место в промежутке от t до $t+\xi$. Интервал $I+t$ по отношению к $t+x$ занимает такое же положение, как интервал $I-x$ по отношению к t . Поэтому

$$U\{I+t\} = \int_0^\infty H(t, d\xi) U\{I-\xi\}. \quad (9.12)$$

Рассматривая первый шаг в случайном блуждании, получим

$$1 - H(0, \xi) \geq 1 - F(\xi). \quad (9.13)$$

Мы уже знаем, что утверждение леммы справедливо, если $\mu < \infty$ или $\mu = -\infty$, т. е. если правая часть (9.13) интегрируема на $0, \infty$. В ином случае H имеет бесконечное математическое ожидание, и поэтому $H(t, \xi) \rightarrow 0$ при $t \rightarrow \infty$ и любом ξ . Следовательно, при больших t только большие значения ξ играют главную роль и при этих значениях $U\{I-\xi\}$ мало. Таким образом, соотношение (9.11) есть прямое следствие (9.10) и (9.12). ►

Лемма 3. *Предположим, что $Z(x) \leq t$, и выберем $p > 0$, такое, что $p' = 1 - pt > 0$. При заданном $\varepsilon > 0$ существует такое s_ε , что при $s > s_\varepsilon$ или*

$$Z(s) < \varepsilon, \quad (9.14)$$

или же

$$Z(s+x) \geq pZ(s)Z(x) \quad \text{для всех } x. \quad (9.15)$$

Доказательство. Исходя из равномерной непрерывности Z и из леммы 1, мы можем выбрать s_e таким, что

$$Z(s+x) - Z(x) > -\varepsilon p' \text{ при } s > s_e \text{ и } |x| < h. \quad (9.16)$$

Положим

$$V_s(x) = Z(s+x) - pZ(s)Z(x), \quad v_s(x) = z(s+x) - pZ(s)z(x). \quad (9.17)$$

Функция V_s удовлетворяет уравнению восстановления $V_s = v_s + F \star V_s$, и из замечания, предваряющего лемму 1, следует, что если функция V_s принимает отрицательные значения, то она достигает минимума в точке ξ , где $v(\xi) < 0$ и, следовательно, $|\xi| < h$. Тогда ввиду (9.16) мы имеем при $s > s_e$

$$V_s(\xi) > -\varepsilon p' + Z(s)[1 - pZ(\xi)] \geq p'[Z(s) - \varepsilon]. \quad (9.18)$$

Таким образом, или выполняется (9.14), или же V_s принимает неотрицательные значения, и в этом случае выполняется (9.15). ►

Лемма 4. Пусть

$$\limsup Z(x) = \eta, \quad x \rightarrow \pm \infty. \quad (9.19)$$

Тогда

$$\limsup [Z(x) + Z(-x)] = \eta, \quad x \rightarrow \infty. \quad (9.20)$$

Доказательство. Выберем такое a , что $Z(a) < \delta$ (это возможно вследствие (9.6)). По предыдущей лемме при достаточно больших s или

$$pZ(s)Z(a-s) \leq Z(a) < \delta, \quad (9.21)$$

или же $Z(s) < \varepsilon$. Поскольку ε произвольно, то неравенство (9.21) будет справедливо при всех достаточно больших s в любом случае. Ввиду леммы 1 отсюда вытекает¹⁾, что

$$Z(s)Z(-s) \rightarrow 0. \quad (9.22)$$

Таким образом, если x велико, то или $Z(x)$ мало, или $Z(-x)$ мало, а так как $Z \geq 0$, то ясно, что (9.19) влечет (9.20). ►

Доказательство теоремы. Допустим, что $\eta > 0$, поскольку в противном случае нечего доказывать. Рассмотрим свертки функций Z и z с равномерным распределением на $\overline{0, t}$, именно

$$W_t(x) = \frac{1}{t} \int_{x-t}^x Z(y) dy, \quad w_t(x) = \frac{1}{t} \int_{x-t}^x z(y) dy. \quad (9.23)$$

¹⁾ Легко видеть, что (9.22) эквивалентно соотношению $U\{I+t\}U\{I-t\} \rightarrow 0$. Обозначим $p\{I\}$ вероятность вхождения и интервал I при случайном блуждании $\{S_n\}$, которое управляется распределением F . Тогда (9.22) эквивалентно также соотношению $p\{I+t\}p\{I-t\} \rightarrow 0$. Если бы это не имело места, то вероятность попадания в окрестность нуля после захода в интервал $I+t$ не стремилась бы к нулю, и распределение F не было бы невозвратным,

Наша ближайшая цель — показать, что при $t \rightarrow \infty$ должно иметь место одно из соотношений

$$W_t(t) = \frac{1}{t} \int_0^t Z(y) dy \rightarrow \eta \text{ или } W_t(0) = \frac{1}{t} \int_{-t}^0 Z(y) dy \rightarrow \eta. \quad (9.24)$$

В силу (9.7) верхние границы для $Z(x)$ и $W_t(x)$ (с фиксированным t) одни и те же. Поэтому максимум W_t будет $\geq \eta$. С другой стороны, W_t удовлетворяет уравнению восстановления (9.5) с z , замененным на w_t . Как отмечалось раньше, отсюда вытекает, что максимум W_t достигается в точке, где w_t положительно, т. е. между $-h$ и $t+h$. Далее для $\frac{1}{2}t \leq x < t$

$$W_t(x) = \frac{1}{t} \int_{t-x}^x Z(y) dy + \frac{1}{t} \int_0^{t-x} [Z(y) + Z(-y)] dy. \quad (9.25)$$

Суммарная длина обоих интервалов интегрирования равна x . Из (9.20) поэтому следует, что при достаточно больших t $W_t(x) < \eta(\frac{x}{t}) + \varepsilon$. Таким образом, если $W_t(x) \geq \eta$, то точка x должна быть близка к t и $W_t(t)$ должно быть близко к η . Если максимум W_t достигается в точке $x \leq t/2$, то, как показывают аналогичные рассуждения, x должно быть близко к нулю, а $W_t(0)$ — к η .

Мы доказали, что при больших t или $W_t(t)$, или $W_t(0)$ должно быть близко к η . Но из (9.23) и (9.20) следует, что

$$\limsup [W_t(t) + W_t(0)] = \limsup t^{-1} \int_0^1 [Z(y) + Z(-y)] dy \leq \eta. \quad (9.26)$$

В силу непрерывности обеих функций должно быть или $W_t t \rightarrow \eta$ и $W_t(0) \rightarrow 0$, или же $W_t(t) \rightarrow 0$ и $W_t(0) \rightarrow \eta$.

По причине симметрии мы можем допустить, что $W_t(t) \rightarrow \eta$, т. е.

$$W_t(t) = t^{-1} \int_0^t Z(y) dy = t^{-1} \int_0^{t/2} \left[Z\left(\frac{1}{2}t + y\right) + Z\left(\frac{1}{2}t - y\right) \right] dy \rightarrow \eta. \quad (9.27)$$

Отсюда следует, что для произвольно больших t существуют такие значения x , что как $Z(x)$, так и $Z(t-x)$ должны быть близки к η . Из леммы 3 вытекает, что при больших t значения $Z(t) > c > 0$ и поэтому $Z(-t) \rightarrow 0$ вследствие соотношения (9.22). Таким образом, $U\{I-t\} \rightarrow 0$ при $t \rightarrow \infty$, что ввиду леммы 2 завершает доказательство. ►

§ 10. ЗАДАЧИ

(См. также задачи 12—20 в гл. VI, 13.)

1. Отбрасывая предположение $F(0)=0$, мы заменим F распределением $F^* = pH_0 + qF$, где H_0 сосредоточено в нуле и $p+q=1$. Тогда U заменяется на $U^* = U/q$. Покажите, что с вероятностной точки зрения это очевидно. Проверьте утверждение, формально а) вычисляя свертки, б) из уравнения восстановления.

2. Пусть F — равномерное распределение на $\overline{0, 1}$. Покажите, что

$$U(t) = \sum_{k=0}^n (-1)^k e^{t-k} \frac{(t-k)^k}{k!}$$

при $n \leq t \leq n+1$. Эта формула неоднократно переоткрывалась в теории очередей, но она мало что говорит о природе U . Асимптотическая формула $0 \leq U(t) - 2t \rightarrow 2/3$ гораздо более интересна. Она является прямым следствием (3.1).

3. Предположим, что $z \geq 0$ и что $|z'|$ интегрируема на $\overline{0, \infty}$. Покажите, что z непосредственно интегрируема.

Замечание. В следующих ниже трех задачах подразумевается, что Z и Z' суть решения стандартного уравнения восстановления (1.3), соответствующие z и z' .

4. Пусть $z \rightarrow \infty$ и $z_1 \sim z$ при $x \rightarrow \infty$. Покажите, что $Z_1 \sim Z$.

5. Если z_1 есть интеграл от z и $z_1(0)=0$, то Z_1 есть интеграл от Z . Докажите, что если $z = x^{n-1}$, то $Z \sim x^n / (n\mu)$ при $\mu < \infty$.

6. **Обобщение.** Если $z_1 = G \star z$ (где мера G конечна на конечных интервалах), то $Z_1 = G \star Z$. Полагая $G(x) = x^a$ при $a \geq 0$, выведите:

$$\text{если } z(x) \sim x^{n-1}, \text{ то } Z(x) \sim x^a / (a\mu).$$

7. К теореме 2 из § 3. Обозначим f_r плотность распределения $F_r \star$ и положим $v = u - f - \dots - f_r$. Покажите, что если f_r ограничена, то $v(x) \rightarrow 1/\mu$. В частности, если $f \rightarrow 0$, то $u \rightarrow 1/\mu$.

8. Если F^{2*} имеет непосредственно интегрируемую плотность, то распределение $V = U - 1 - F$ имеет плотность v , которая стремится к $1/\mu$.

9. Выведите из (4.4), что $Z(t) = V(t) - V(t-h)$ удовлетворяет стандартному уравнению восстановления с $z'(t) = F_0(t) - F_0(t-h)$. Выведите непосредственно из теоремы восстановления, что $V(t) - V(t-h) \rightarrow h/\mu$.

10. **Совместное распределение для прошедшего и остающегося времени ожидания.** В обозначениях (4.7) докажите, что при $t \rightarrow \infty$

$$P\{t - S_{N_t} > x, S_{N_t+1} - t > y\} \rightarrow \frac{1}{\mu} \int_{x+y}^{\infty} [1 - F(s)] ds,$$

Указание. Получите уравнение восстановления для левой части.

11. **Стационарные свойства.** Рассмотрим процесс восстановления с задержкой, с начальным распределением F_0 , задаваемым (4.6). Вероятность $H(t, \xi)$ того, что n -й момент восстановления S_n лежит между t и $t+\xi$, удовлетворяет уравнению восстановления $H(t, \xi) = F_0(t+\xi) - F_0(t) + F_0 \star H(t, \xi)$. Докажите без вычислений, что $H(t, \xi) = F_0(\xi)$ тождественно.

12. **Максимальное наблюдаемое время жизни.** Пусть $V(t, \xi)$ — вероятность того, что максимальный промежуток между восстановлениями, наблюдаемый до момента t в стандартном возвратном процессе восстановления, имеет дли-

тельность $> \xi$. Покажите, что

$$V(t, \xi) = 1 - F(\xi) + \int_0^{\xi} V(t-y, \xi) F(dy).$$

Изучите характер решения.

13. Пусть N_t — число моментов восстановления, определенное в § 5. Покажите, что

$$E(N_t^2) = \sum_{k=0}^{\infty} (2k+1) F^{k*}(t) = 2U \star U(t) - U(t).$$

Используя интегрирование по частям, выведите из этого равенства и теоремы восстановления, что

$$E(N_t^2) = \frac{2}{\mu} \int_0^t U(x) dx + \frac{\sigma^2}{\mu^3} t + o(t)$$

и далее $\text{Var}(N_t) \sim (\sigma^2/\mu^3) t$, что соответствует оценке в центральной предельной теореме. (Замечание: этот метод применим также к арифметическим распределениям, и он предпочтительнее для вывода того же самого результата, который описан в задаче 23 в I; XIII, 12.)

14. Если F — собственное распределение и a — постоянная, то интегро-дифференциальное уравнение $Z' = aZ - aZ \star F$ сводится к

$$Z(t) = Z(0) + a \int_0^t Z(t-x) [1 - F(x)] dx,$$

15. *Обобщенные счетчики второго типа.* Поступающие частицы образуют пуассоновский поток. Очередная j -я частица «закрывает» счетчик на время T_j и аннулирует последствие (если оно было) предыдущих частиц. Предполагается, что случайные величины T_j независимы между собой, не зависят от поступающего потока и имеют одно и то же распределение G . Пусть Y — длина промежутка времени, в течение которого счетчик закрыт, и $Z(t) = P\{Y > t\}$. Покажите, что Y — собственная случайная величина и

$$Z(t) = [1 - G(t)] e^{\alpha t} + \int_0^t Z(t-x) [1 - G(x)] \alpha e^{-\alpha x} dx,$$

Покажите, что этот процесс восстановления будет обрывающимся в том и только том случае, когда G имеет математическое ожидание $\mu < \alpha^{-1}$. Обсудите применимость асимптотических оценок § 6.

16. *Эффект «островков безопасности».* [Пример 7, б).] Два независимых пуассоновских потока автомашин с одинаковыми плотностями движутся по двум сторонам магистрали. Время, необходимое для пересечения магистрали, равно 2ξ . Формулы (7.10) применимы и в этом случае. Наличие «островка безопасности» приводит к тому, что время, необходимое для пересечения магистрали, становится суммой двух независимых случайных величин с математическими ожиданиями и дисперсиями, определяемыми (7.10). Обсудите практические следствия этого факта.

17. Частицы, попадающие на счетчик, образуют возвратный процесс восстановления с распределением F . После каждой регистрации счетчик закрывается на фиксированное время ξ , в течение которого поступающие частицы не оказывают никакого воздействия на счетчик. Покажите, что распределение для промежутка времени между концом «периода нечувствительности»

и моментом появления очередной частицы задается формулой

$$\int_0^{\xi} [F(\xi + t - y) - F(\xi - y)] U(dy).$$

Если F — показательное распределение, то и это распределение будет показательным.

18. Нелинейное восстановление. Частица имеет показательное время жизни, в конце которого она с вероятностью p_k производит k точно таких же частиц, существующих независимо друг от друга ($k=0, 1, \dots$). Вероятность $F(t)$ того, что весь процесс оборвется до момента t , удовлетворяет уравнению

$$F(t) = p_0(1 - e^{-\alpha t}) + \sum_{n=1}^{\infty} p_n \int_0^t \alpha e^{-\alpha(t-x)} F^n(x) dx.$$

(Неизвестно никакого общего метода обращения с такими уравнениями.)

19. Пусть F — произвольное распределение в $\mathcal{R}^1$ с математическим ожиданием $\mu > 0$ и конечным вторым моментом m_2 . Покажите, что

$$\sum_{n=0}^{\infty} F^{n*}(x) - \frac{x_+}{\mu} \rightarrow \frac{m_2}{2\mu^2},$$

где, как обычно, x_+ обозначает положительную часть x . *Указание.* Обозначим $Z(t)$ левую часть. Тогда Z удовлетворяет уравнению восстановления

$$z(x) = \begin{cases} \frac{1}{\mu} \int_{-\infty}^t F(x) dx, & t < 0, \\ \frac{1}{\mu} \int_t^{\infty} (1 - F(x)) dx, & t > 0. \end{cases}$$

20. Теорема восстановления в $\mathcal{R}^2$. Пусть распределение пары (X, Y) сосредоточено в первом квадранте. Пусть I — интервал $0 \leq x, y \leq 1$. Обозначим $I+a$ (где a — произвольный вектор) интервал, получаемый сдвигом I на a . Тогда лемма 1 из § 9 может быть обобщена следующим образом. Для любых фиксированных векторов a и b

$$U\{I+a+tb\} - U\{I+tb\} \rightarrow 0$$

при $t \rightarrow \infty$.

а) Считая это соотношение выполненным, покажите, что из теоремы восстановления для маргинальных распределений вытекает $U\{I+tb\} \rightarrow 0$.

б) Покажите, что доказательство леммы тривиальным образом применимо ¹⁾.

¹⁾ Более подходящая формулировка задач теории восстановления для плоскости была недавно предложена в работе: Bickel P. J., Yahav J. A. Renewal theory in the plane, Ann. Math. Statist., 36 (1965), 946—955. Эти авторы рассматривали математическое ожидание числа попаданий в область, заключенную между окружностями радиусов r и $r+a$ при $r \rightarrow \infty$.

В этой главе рассматриваются задачи теории случайных блужданий с упором на комбинаторные методы и систематическое использование «лестничных величин». Некоторые из результатов будут выведены заново и дополнены в гл. XVIII методами анализа Фурье. (В другом аспекте случайные блуждания изучались в гл. VI, 10.) Мы ограничимся в основном двумя центральными вопросами. Во-первых, будет показано, что полученные в I; гл. III, любопытные результаты, относящиеся к бросанию монеты, переносятся на очень общий случай и что, по существу, при этом применимы те же самые методы. Второй вопрос связан с временем первого прохождения и задачами о разорении. Стало модным связывать подобные темы с известной теорией Винера—Хопфа. Однако эти связи не так тесны, как их обычно изображают. Они будут обсуждены в § 3а и гл. XVIII, 4.

Открытие Спарре-Андерсеном в 1949 г. мощных комбинаторных методов в теории флуктуаций представило всю теорию случайных блужданий в новом свете. С тех пор отмечается очень быстрый прогресс, одним из стимулов которого является неожиданное открытие тесной связи между случайными блужданиями и теорией очередей¹⁾.

Литература по этой теме обширна и порой способна запутать читателя. Теория, излагаемая на следующих страницах, так элементарна и проста, что непосвященный читатель никогда не заподозрил бы, какими трудными были эти задачи до того, как была понята их истинная природа. Например, элементарные асимптотические оценки § 5 охватывают множество практически интересных результатов, которые ранее были получены глубокими методами и подчас с большой изобретательностью.

Параграфы 6—8 почти не зависят от первой части главы. Вряд ли нужно подчеркивать, что наше изложение является

¹⁾ Впервые такая связь была отмечена, по-видимому, Линдли (D. V. Lindley) в 1952 г. Он вывел интегральное уравнение, которое теперь отнесли бы к уравнениям типа Винера—Хопфа,

односторонним и оставляет в стороне ряд интересных аспектов теории случайных блужданий (таких, как связи с теорией потенциала или теорией групп¹⁾).

§ 1. ОСНОВНЫЕ ПОНЯТИЯ И ОБОЗНАЧЕНИЯ

Всюду в этой главе $X_1, X_2, \dots$ суть независимые случайные величины с одним и тем же распределением F , которое не сосредоточено ни на одной из полуосей (случай, когда $F(0)=0$ или $F(0)=1$, охватывается теорией восстановления). Случайное блуждание, порождаемое распределением F , — это последовательность случайных величин

$$S_0=0, \quad S_n=X_1+\dots+X_n. \quad (1.1)$$

Иногда мы будем рассматривать отрезок $(X_{j+1}, \dots, X_k)$ данной последовательности $\{X_j\}$. Ее частные суммы $0, S_{j+1}-S_j, \dots, S_k-S_j$ будут называться *отрезками случайного блуждания*. Индексы будут обычно интерпретироваться как значения временного параметра. Таким образом, момент n делит все случайное блуждание на *прошедший отрезок* и *остающийся отрезок*. Поскольку $S_0=0$, говорят, что случайное блуждание начинается в нуле. Добавляя постоянную a ко всем членам последовательности S_n , мы получим случайное блуждание, начинающееся в точке a . Таким образом, $S_n, S_{n+1}, \dots$ — случайное блуждание, порождаемое распределением F и начинающееся в S_n .

Предварительная информация. Если посмотреть на график случайного блуждания, можно заметить его выразительную особенность — точки, где S_n достигает *рекордного значения*, т. е. где S_n превосходит все ранее достигнутые значения $S_0, \dots, S_{n-1}$. Эти точки называются *лестничными точками*, согласно терминологии, введенной в VI,8 (см. рис. 1 в том параграфе). Теоретическая важность лестничных точек объясняется тем, что отрезки между ними являясь точными вероятностными копиями друг друга, и, следовательно, важные выводы относительно всего случайного блуждания можно получить при изучении первой лестничной точки.

В первом томе мы неоднократно рассматривали случайные блуждания, в которых X_k принимают значения $+1$ и -1 с вероятностями p и q соответственно. В таких блужданиях каждое последующее рекордное значение превосходит предыдущее на $+1$ и последовательные лестничные точки представляют собой просто первые прохождения через $1, 2, \dots$. В принятой нами термино-

¹⁾ Другие аспекты представлены в книге Спизера (1964), хотя там рассматриваются только арифметические распределения. О комбинаторных методах, пригодных в многомерном случае, см.: Hobby C., Pyke R., Combinatorial results in multidimensional fluctuation theory, Ann. Math. Statist., 34 (1963), 402—404.

логии мы бы сказали, что лестничные высоты известны заранее и только времена ожидания между последовательными лестничными точками требуют изучения. Они являются независимыми случайными величинами с тем же распределением, что и у времени первого прохождения через $+1$. Производящая функция этого распределения была найдена в 1, XI, (3.6), и имеет вид

$$[1 - \sqrt{1 - 4pq s^2}] / (2qs), \quad (1.2)$$

где $\sqrt{}$ обозначает положительный корень (см. также 1; XIV, 4; явная формула содержится в 1; XI, 3, г), и 1; XIV, 5). При $p < q$ времена первого прохождения являются несобственными случайными величинами, так как вероятность того, что когда-нибудь будет достигнуто положительное значение, равна p/q .

Одно и то же рекордное значение может быть повторено несколько раз, прежде чем будет достигнуто новое рекордное значение. Точки такого относительного максимума называются *слабыми* лестничными точками. [В простом случайном блуждании Бернулли первая слабая лестничная точка либо есть (1.1), либо имеет вид $(2r, 0)$.]

После этих предварительных замечаний введем формально лестничные величины, частично повторяя уже сказанное в гл. VI, 8. Определение основано на неравенствах, и потому существуют четыре типа лестничных величин, соответствующих четырем возможностям, именно $<$, $\leq$, $>$, $\geq$. Это приводит к двойной классификации, описываемой понятными терминами: *возрастающие* и *убывающие*¹⁾, строгие и слабые. Возрастающие и убывающие величины связаны столь же симметрично, как плюс и минус, максимум и минимум. Однако различие между строгими и слабыми величинами создает трудности в определениях и обозначениях. Простейший выход состоит в том, чтобы рассматривать только непрерывные распределения F , так как в этом случае строгие и слабые величины равны друг другу с вероятностью единица. Начинаящим можно рекомендовать именно этот путь и не делать различия между строгими и слабыми лестничными величинами, хотя такое различие неизбежно в общей теории, с одной стороны, и в таких примерах, как игра с бросанием монеты, — с другой.

Чтобы ввести необходимые понятия и соглашения, рассмотрим возрастающие строгие лестничные величины. Затем мы покажем, что теория слабых лестничных величин является простым следствием теории строгих лестничных величин. Убывающие лестничные величины не нуждаются в отдельной теории. Таким образом, мы будем, как правило, рассматривать возрастающие строгие лестничные величины, и в случаях, когда не будет опасности путаницы, мы будем опускать названия «возрастающие» и «строгие».

¹⁾ «Ascending» и «descending». В том случае, когда речь будет идти о точках, мы будем применять термины «верхние» и «нижние». — *Прим. перев.*

Возрастающие строгие лестничные величины. Рассмотрим последовательность точек (n, S_n) для $n=1, 2, \dots$ (нуль исключается). Первая строгая верхняя¹⁾ лестничная точка $(\mathcal{T}_1, \mathcal{H}_1)$ —это первый член указанной последовательности, для которого $S_n > 0$. Иными словами, $\mathcal{T}_1$ —это момент первого попадания на положительную (открытую) полуось, определяемый соотношениями

$$\{\mathcal{T}_1 = n\} = \{S_1 \leq 0, \dots, S_{n-1} \leq 0, S_n > 0\}, \quad (1.3)$$

а $\mathcal{H}_1 = S_{\mathcal{T}_1}$. Величина $\mathcal{T}_1$ называется первым *лестничным моментом*, а $\mathcal{H}_1$ —первой *лестничной высотой*. Эти случайные величины остаются *неопределенными*, если ни одно из событий (1.3) не наступает. Поэтому обе они могут быть несобственными²⁾.

Для совместного распределения $(\mathcal{T}_1, \mathcal{H}_1)$ мы примем обозначение

$$P\{\mathcal{T}_1 = n, \mathcal{H}_1 \leq x\} = H_n(x). \quad (1.4)$$

Тогда маргинальные распределения задаются равенствами

$$P\{\mathcal{T}_1 = n\} = H_n(\infty), \quad n = 1, 2, \dots, \quad (1.5)$$

$$P\{\mathcal{H}_1 \leq x\} = \sum_{n=1}^{\infty} H_n(x) = H(x). \quad (1.6)$$

Обе случайные величины имеют один и тот же дефект, а именно $1 - H(\infty) \geq 0$.

Отрезок случайного блуждания, следующий за первым лестничным моментом, является точной вероятностной копией всего случайного блуждания. Его первая лестничная точка является *второй* лестничной точкой всего случайного блуждания и обладает тем свойством, что

$$S_n > S_0, \dots, S_n > S_{n-1}. \quad (1.7)$$

Так мы ее и назовем—второй лестничной точкой случайного блуждания. Она имеет вид $(\mathcal{T}_1 + \mathcal{T}_2, \mathcal{H}_1 + \mathcal{H}_2)$, где пары $(\mathcal{T}_1, \mathcal{H}_1)$ и $(\mathcal{T}_2, \mathcal{H}_2)$ независимы и одинаково распределены (см. также гл. VI, 8). Поступая аналогично, мы можем определить третью, четвертую и т. д. лестничные точки нашего случайного блуждания. Таким образом, точка (n, S_n) является *верхней лестничной точкой*, если она удовлетворяет соотношению (1.7). Лестничная точка с номером r (если она существует) имеет вид $(\mathcal{T}_1 + \dots + \mathcal{T}_r, \mathcal{H}_1 + \dots + \mathcal{H}_r)$, где пары $(\mathcal{T}_k, \mathcal{H}_k)$ взаимно независимы и имеют одно и то же распределение (1.4) [см. рис. 1 в гл. VI, 8].

Для экономии обозначений мы не будем вводить никаких новых букв для сумм $\mathcal{T}_1 + \dots + \mathcal{T}_r$ и $\mathcal{H}_1 + \dots + \mathcal{H}_r$. Они образуют два (возможно, обрывающихся) процесса восстановления с «промежутками между восстановлениями» $\mathcal{T}_k$ и $\mathcal{H}_k$. В случай-

¹⁾ См. предыдущую сноску.

²⁾ Задачи 3—6 могут служить иллюстрацией, понятной без общей теории,

ном блуждании, конечно, только $\mathcal{F}_k$ играет роль времени. Отметим, что сами лестничные точки образуют двумерный процесс восстановления.

Обозначим через

$$\psi = \sum_{n=0}^{\infty} H^{n*} \quad (1.8)$$

меру, определяющую среднее число лестничных высот (здесь $H^{0*} = \psi_0$). Соответствующая несобственная функция распределения, задаваемая равенством $\psi(x) = \overline{\psi\{-\infty, x\}}$, равна нулю при $x < 0$, а для положительных x функция $\psi(x)$ равна единице плюс среднее число лестничных точек в полосе $\overline{\psi\{0, x\}}$ (время не ограничивается). Мы знаем из гл. VI,6, и гл. XI,1, что $\psi(x) < \infty$ при всех x и что в случае несобственных лестничных величин

$$\psi(\infty) = \sum_{n=0}^{\infty} H^n(\infty) = \frac{1}{1-H(\infty)}. \quad (1.9)$$

Наконец, введем обозначение ψ_0 для атомического распределения с единичной массой в нуле; таким образом, для любого интервала I

$$\psi_0\{I\} = 1, \text{ если } x \in I, \psi_0\{I\} = 0 \text{ в противном случае.} \quad (1.10)$$

Возрастающие слабые лестничные величины. Точка (n, S_n) есть слабая (верхняя) лестничная точка в том и только том случае, если $S_n \geq S_k$ при $k=0, 1, \dots, n$. Изучение строгих и слабых лестничных величин будет проходить параллельно, и мы будем систематически использовать те же самые буквы, пометая слабые величины чертой сверху. Так, $\overline{\mathcal{F}}_1$ — это наименьший индекс n , такой, что $S_1 < 0, \dots, S_{n-1} < 0$, но $S_n \geq 0$. Как отмечалось раньше, утомительная необходимость различать строгие и слабые величины отпадает, если распределение F непрерывно. В общем случае легко выразить распределение $\overline{H}$ слабых лестничных высот в терминах распределения H . Это дает нам возможность ограничиться единственным распределением, определенным в (1.6).

Первая слабая лестничная точка идентична первой строгой лестничной точке, за исключением того случая, когда случайное блуждание возвращается в нуль, пройдя только через отрицательные значения; в этом случае $\overline{\mathcal{H}}_1 = 0$, и мы положим $\zeta = \mathbf{P}\{\overline{\mathcal{H}}_1 = 0\}$. Таким образом

$$\zeta = \sum_{n=1}^{\infty} \mathbf{P}\{S_1 < 0, \dots, S_{n-1} < 0, S_n = 0\}. \quad (1.11)$$

(Это событие не может произойти, если $X_1 > 0$ и, следовательно, $0 \leq \zeta < 1$.) С вероятностью $1 - \zeta$ первая строгая лестничная точка

совпадает с первой слабой лестничной точкой. Следовательно,

$$\bar{H} = \zeta \psi_0 + (1 - \zeta) H. \quad (1.12)$$

Иначе говоря, распределение первой слабой лестничной высоты является смесью распределения H и атомического распределения, сосредоточенного в нуле.

Пример. В простом случайном блуждании Бернулли первая слабая лестничная высота равна 1 тогда и только тогда, когда первый шаг приводит в точку +1. Если первый шаг приводит в точку -1, то (условная) вероятность возвращения в 0 равна 1, если $p \geq q$, и равна p/q в противном случае. В первом случае $\zeta = q$, во втором $\zeta = p$. Возможные лестничные высоты равны 1 и 0, и они имеют вероятности p и q при $p \leq q$, тогда как при $p < q$ обе вероятности равны p . В последнем случае лестничная высота является несобственной случайной величиной. ►

Вероятность того, что случайное блуждание вернется в нуль k раз до первого попадания в $\bar{0}, \infty$, равна $\zeta^k (1 - \zeta)$. Математическое ожидание числа таких возвращений равно $1/(1 - \zeta)$. Эта величина дает нам ожидаемую кратность каждой лестничной высоты (т. е. число ее появлений до следующей лестничной высоты). Следовательно,

$$\bar{\psi} = \frac{1}{1 - \zeta} \psi \quad (1.13)$$

(см. задачу 7). Простота этих соотношений позволяет нам избежать явного использования распределения $\bar{H}$.

Убывающие лестничные величины. Строгие и слабые убывающие лестничные величины определяются по симметрии, т. е. путем замены знака $>$ на $<$. В тех редких случаях, когда потребуются специальное обозначение, мы будем помечать убывающие величины верхним индексом «минус». Так, первая строгая нижняя лестничная точка будет $(\mathcal{F}_1^-, \mathcal{H}_1^-)$ и т. д.

Вскоре мы увидим, что вероятности $P\{\bar{\mathcal{H}}_1 = 0\}$ и $P\{\bar{\mathcal{H}}_1^- = 0\}$ равны между собой, так как

$$\begin{aligned} P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = 0\} = \\ = P\{S_1 < 0, \dots, S_{n-1} < 0, S_n = 0\}. \end{aligned} \quad (1.14)$$

Отсюда следует, что аналоги (1.12) и (1.13) для убывающих лестничных величин зависят от той же величины ζ .

§ 2. ДВОЙСТВЕННОСТЬ. ТИПЫ СЛУЧАЙНЫХ БЛУЖДАНИЙ

В 1, гл. III, мы установили любопытные свойства флуктуаций, возникающих при бросании монеты. При этом были использованы простые комбинаторные рассуждения, основанные на рассмотрении величин $(X_1, \dots, X_n)$, взятых в обратном порядке. Эти

же приемы приведут нас теперь к важным и очень общим результатам.

При фиксированном n мы вводим n новых случайных величин равенствами $X_1^* = X_n, \dots, X_n^* = X_1$. Соответствующие частные суммы равны $S_k^* = S_n - S_{n-k}$, где $k = 0, \dots, n$. Совместные распределения величин $(S_0, \dots, S_n)$ и $(S_0^*, \dots, S_n^*)$ совпадают. Поэтому отображение $X_k \rightarrow X_k^*$ отображает любое событие A , связанное с $(S_0, \dots, S_n)$, в событие A^* равной вероятности. Это отображение легко представить себе наглядно, поскольку графики последовательностей $(0, S_1, \dots, S_n)$ и $(0, S_1^*, \dots, S_n^*)$ получаются один из другого поворотом на 180° .

Пример. а) Если $S_1 < 0, \dots, S_{n-1} < 0$, но $S_n = 0$, то $S_1^* > 0, \dots, S_{n-1}^* > 0$ и $S_n^* = 0$. Это доказывает справедливость соотношения (1.14), которое используется в § 1.

Теперь мы применим указанную выше процедуру «обращения» к событию $\{S_n > S_0, \dots, S_n > S_{n-1}\}$, определяющему верхнюю строгую лестничную точку. Двойственные события определяются неравенствами $S_n^* > S_{n-k}^*, k = 1, \dots, n$. Но событие $S_n^* > S_{n-k}^*$ совпадает с событием $S_k > 0$. Таким образом, для всякого конечного интервала $I \subset 0, \infty$

$$\begin{aligned} P\{S_n > S_j \text{ при } j = 0, \dots, n-1 \text{ и } S_n \in I\} = \\ = P\{S_j > 0 \text{ при } j = 1, \dots, n \text{ и } S_n \in I\}. \end{aligned} \quad (2.1)$$

Левая часть равна вероятности того, что найдется лестничная точка с абсциссой n и ординатой, принадлежащей I . Правая часть равна вероятности того, что в момент n произойдет попадание в I и при этом до момента n не будет заходов на замкнутую полупрямую $-\infty, 0$. Проведем суммирование по n в (2.1) и рассмотрим результат. Слева мы получим $\psi\{I\}$ в согласии с определением (1.8) меры ψ . Справа получим математическое ожидание числа попаданий в I до первого попадания в $-\infty, 0$. Эта величина конечна, так как $\psi\{I\} < \infty$. Нами установлена тем самым основная

Лемма о двойственности. Мера ψ допускает две интерпретации. Для каждого конечного интервала $I \subset 0, \infty$

а) $\psi\{I\}$ есть математическое ожидание числа лестничных точек в I .

б) $\psi\{I\}$ есть математическое ожидание числа попаданий $S_n \in I$, таких, что $S_k > 0, k = 1, 2, \dots, n$.

Эта простая лемма предоставляет нам возможность доказать элементарным способом теоремы, которые в ином случае потребовали бы глубоких аналитических методов. Из-за своей аналитической формулировки лемма не выглядит привлекательной, однако

ее непосредственные следствия оказываются удивительными и противоречащими наивной интуиции.

Пример. б) Простое случайное блуждание. В случайном блуждании, описанном в примере из § 1, лестничная точка с ординатой k существует тогда и только тогда, когда при некотором n наступает событие $\{S_n = k\}$, вероятность которого равна 1 при $p \geq q$ и $(p/q)^k$ при $p \leq q$. По лемме о двойственности это означает, что в симметричном случайном блуждании математическое ожидание числа попаданий в состояния $k \geq 1$ до первого возвращения в нуль равно единице при всех k . Фантастичность этого результата можно лучше уяснить в терминологии игры с бросанием монеты. Наше утверждение состоит в том, что до первого возвращения на нулевой уровень средний накопленный выигрыш Петра принимает когда-либо любое значение k . Это утверждение обычно порождает недоверие, однако его можно проверить непосредственным вычислением (см. задачу 2). (Наш прежний вывод о бесконечности математического ожидания времени до первого возвращения в 0 получается теперь суммированием по k .) ►

В симметричном случайном блуждании Бернулли (бросание монеты) каждое из значений ± 1 достигается с вероятностью единица, однако математическое ожидание времени ожидания для каждого из этих событий бесконечно. В следующей теореме показано, что это свойство не является особенностью игры в монету, поскольку аналогичное утверждение справедливо для всех случайных блужданий, в которых и положительные, и отрицательные значения допускаются с вероятностью единица.

Теорема 1. *Существуют только два типа случайных блужданий.*

(i) *Осциллирующий тип: возрастающий и убывающий процессы восстановления возвратны; S_n колеблется с вероятностью 1 между $-\infty$ и ∞ и*

$$E(\mathcal{F}_1) = \infty, \quad E(\mathcal{F}_1^-) = \infty. \quad (2.2)$$

(ii) *Уходящий тип (скажем, в $-\infty$). Возрастающий процесс восстановления обрывается; убывающий процесс является собственным; с вероятностью единица S_n уходит в $-\infty$ и достигает конечного максимума $M \geq 0$; выполняются соотношения (2.5) и (2.7).*

[Блуждания типа (ii) очевидным образом невозвратны, но среди блужданий типа (i) имеются как возвратные, так и невозвратные; см. конец гл. VI, 10.]

Доказательство. Тожество (2.1) выполняется и при замене строгих неравенств слабыми. Для $I = 0, \infty$ получаем

$$\begin{aligned} P\{S_n \geq S_k \text{ при } 0 \leq k \leq n\} &= P\{S_k \geq 0 \text{ при } 0 \leq k \leq n\} = \\ &= 1 - P\{\mathcal{F}_1^- \leq n\}. \end{aligned} \quad (2.3)$$

Левая часть есть просто вероятность того, что (n, S_n) будет сла-

бой верхней лестничной точкой. Эти вероятности дают в сумме $\bar{\psi}(\infty) \leq \infty$, и в силу равенства (1.13) мы имеем, таким образом,

$$\frac{1}{1-\xi} \psi(\infty) = \sum_{n=0}^{\infty} [1 - P\{\mathcal{T}_1^- \leq n\}]. \quad (2.4)$$

Когда убывающий лестничный процесс — несобственный, то члены ряда не стремятся к нулю и ряд расходится. В этом случае $\psi(\infty) = \infty$; это означает, что возрастающий лестничный процесс возвратен. Таким образом, мы получили аналитическое доказательство интуитивно очевидного факта: невозможно, чтобы оба лестничных процесса, и возрастающий, и убывающий, были обрывающимися.

Если $\mathcal{T}_1^-$ — собственная величина, то (2.4) сводится к

$$E(\mathcal{T}_1^-) = \frac{1}{1-\xi} \psi(\infty) = \frac{1}{(1-\xi)(1-H(\infty))}, \quad (2.5)$$

с очевидной интерпретацией в случае $H(\infty) = 1$.

Мы видим, что $E(\mathcal{T}_1^-) < \infty$ в том и только том случае, когда $H(\infty) < 1$, т. е. когда возрастающая лестничная величина $\mathcal{T}_1^-$ является несобственной. Таким образом, или одна из этих величин будет несобственной, или выполняется (2.2).

Если $E(\mathcal{T}_1^-) < \infty$, то возрастающий процесс восстановления будет обрывающимся. С вероятностью единица существует последняя лестничная точка, и максимум

$$M = \max \{S_0, S_1, \dots\} \quad (2.6)$$

конечен. При условии, что n -я лестничная точка существует, вероятность ей быть последней равна $1 - H(\infty)$. Поэтому [см. гл. XI, (6.3)]

$$P\{M \leq x\} = [1 - H(\infty)] \sum_{n=0}^{\infty} H^{n*}(x) = [1 - H(\infty)] \psi(x). \quad (2.7) \blacktriangleright$$

Условимся в следующей теореме писать $E(X) = +\infty$, если соответствующий интеграл расходится только в $+\infty$, или, что то же самое, если $P\{X < t\}$ интегрируемо на $-\infty, 0$.

Теорема 2. (i) Если $E(X_1) = 0$, то $\mathcal{H}_1$ и $\mathcal{T}_1^-$ — собственные величины¹⁾ и $E(\mathcal{T}_1^-) = \infty$.

(ii) Если $E(X_1)$ конечно и положительно, то $\mathcal{H}_1$ и $\mathcal{T}_1^-$ суть собственные величины с конечными математическими ожиданиями и

$$E(\mathcal{H}_1) = E(\mathcal{T}_1^-) E(X_1). \quad (2.8)$$

Случайное блуждание уходит в $+\infty$.

¹⁾ Теорема 4 из гл. VI, 10, содержит более сильный результат: при $E(X_1) = 0$ случайное блуждание возвратно,

(iii) Если $E(X_1) = +\infty$, то $E(\mathcal{H}_1) = \infty$ и случайное блуждание уходит в $+\infty$.

(iv) В других случаях либо случайное блуждание уходит в $-\infty$ (в этом случае $\mathcal{F}_1$ и $\mathcal{H}_1$ — несобственные величины), либо $E(\mathcal{H}_1) = \infty$.

Тождество (2.8) было впервые открыто А. Вальдом в значительно более общей обстановке, о чем будет идти речь в гл. XVIII, 2. Следующее ниже доказательство основано на усиленном законе больших чисел из гл. VII, 8. Чисто аналитические доказательства будут даны в свое время (см. теорему 3 из § 7 и задачи 9—11, а также гл. XVIII, 4).

Доказательство. Если n совпадает с k -м лестничным моментом, то имеем тождество

$$\frac{S_n}{n} = \frac{(\mathcal{H}_1 + \dots + \mathcal{H}_k)/k}{(\mathcal{F}_1 + \dots + \mathcal{F}_k)/k}. \quad (2.9)$$

Теперь заметим, что усиленный закон больших чисел применим и в том случае, когда $E(X_1) = +\infty$, что можно показать с помощью очевидного урезания.

(i) Пусть $E(X_1) = 0$. При $k \rightarrow \infty$ левая часть в (2.9) стремится к 0. Отсюда следует, что знаменатель стремится к бесконечности. Поэтому $\mathcal{F}_1$ — собственная величина и $E(\mathcal{F}_1) = \infty$.

(ii) Если $0 < E(X_1) < \infty$, то по усиленному закону больших чисел случайное блуждание уходит в ∞ . В силу (2.5) это означает, что $\mathcal{F}_1$ — собственная величина и $E(\mathcal{F}_1) < \infty$. Поэтому числитель и знаменатель в (2.9) стремятся к конечным пределам и (2.8) теперь следует из теоремы, обратной закону больших чисел (теорема 4 в гл. VII, 8).

(iii) Если $E(X_1) = +\infty$, то с помощью аналогичных доводов можно показать, что $E(\mathcal{H}_1) = \infty$.

(iv) В остальных случаях мы покажем, что если $\mathcal{H}_1$ — собственная величина и $E(\mathcal{H}_1) < \infty$, случайное блуждание уходит в $-\infty$. Рассматривая первый шаг случайного блуждания, видим, что при $x > 0$

$$P\{\mathcal{H}_1 > x\} \geq P\{X_1 > x\}. \quad (2.10)$$

Если $\mathcal{H}_1$ — несобственная величина, случайное блуждание уходит в $-\infty$. Если она собственная, то интеграл от левой части в (2.10) по области $\overline{0, \infty}$ равен $E(\mathcal{H}_1)$. Если $E(\mathcal{H}_1) < \infty$, то $E(X_1)$ либо конечно, либо равно $-\infty$. Случай $E(X_1) \geq 0$ уже был разобран, а если $E(X_1) < 0$ (или равно $-\infty$), то случайное блуждание уходит в $-\infty$.

Из (2.10) и аналогичного неравенства для $x < 0$ следует, что если $\mathcal{H}_1$ и $\mathcal{H}_1^-$ — собственные величины и имеют конечные математические ожидания, то $P\{|X_1| > x\}$ интегрируемо и, следовательно, $E(X_1)$ существует (лемма 2 из гл. V, 6). При $E(X_1) \neq 0$ одна из лестничных величин должна быть несобственной и, таким образом, мы имеем

Следствие. Если $\mathcal{H}_1$ и $\mathcal{H}_1^-$ являются собственными величинами и имеют конечные математические ожидания, то $E(X_1)=0$.

Обратное неверно. Однако если $E(X_1)=0$ и $E(X_1^2)<\infty$, то $\mathcal{H}_1$ и $\mathcal{H}_1^-$ имеют конечные математические ожидания. (См. задачу 10. Более точный результат содержится в теореме 1 из гл. XVIII, 5, где точки первого попадания S_n и S_n^- суть случайные величины с распределениями $\mathcal{H}_1$ и $\mathcal{H}_1^-$.)

§ 3. РАСПРЕДЕЛЕНИЕ ЛЕСТНИЧНЫХ ВЫСОТ. ФАКТОРИЗАЦИЯ ВИНЕРА — ХОПФА

Вычисление распределений лестничных высот, т. е. H и $\bar{H}$, на первый взгляд кажется устрашающей проблемой, и именно так ее первоначально воспринимали. Однако лемма о двойственности приводит к простому решению. Идея состоит в том, что первое вхождение, скажем, в $-\infty, 0$ следует рассматривать вместе с предшествующим ему отрезком случайного блуждания. Мы приходим к изучению видоизмененного случайного блуждания $\{S_n\}$, обрывающегося в момент первого вхождения в $-\infty, 0$. Обозначим через ψ_n несобственное распределение вероятностей для положения в момент n в этом ограниченном случайном блуждании. Иными словами, для любого интервала I и $n=1, 2, \dots$ положим

$$\psi_n\{I\} = P\{S_1 > 0, \dots, S_n > 0, S_n \in I\} \quad (3.1)$$

(заметим, что отсюда следует равенство $\psi_n\{-\infty, 0\} = 0$). Как и прежде, ψ_0 обозначает распределение вероятностей, сосредоточенное в нуле. Формула (2.1) показывает, что $\psi_n\{I\}$ равна вероятности того, что (n, S_n) есть лестничная точка с $S_n \in I$. Суммируя по всем n , получаем

$$\psi\{I\} = \sum_{n=0}^{\infty} \psi_n\{I\}, \quad (3.2)$$

где ψ — «функция восстановления», введенная в (1.8). Другими словами, для любого интервала I , расположенного на открытой положительной полуоси, $\psi\{I\}$ равно математическому ожиданию числа (строгих верхних) лестничных точек, ординаты которых принадлежат I . Для интервалов I отрицательной полуоси положим $\psi\{I\} = 0$. Заметим, что ряд (3.2) сходится для любого ограниченного интервала I (но может расходиться для $I = \overline{0, \infty}$). Этот неожиданный результат делает всю последующую теорию невероятно простой.

Изучение первого вхождения в $-\infty, 0$ — это изучение слабого убывающего лестничного процесса и (в обозначениях § 1) точки

первого входа $\overline{\mathcal{H}}_1^-$ и ее распределения $\overline{H}^-$. Для облегчения типографского набора заменим $\overline{H}^-$ на ρ . Обозначим через $\rho_n \{I\}$ вероятность того, что первое вхождение в $\overline{-\infty, 0}$ происходит в момент n и внутри интервала I , т. е. при $n=1, 2, \dots$

$$\rho_n \{I\} = P \{S_1 > 0, \dots, S_{n-1} > 0, S_n \leq 0, S_n \in I\} \quad (3.3)$$

(отсюда вытекает, что $\rho_n \{0, \infty\} = 0$; ρ_0 не определено). На этот раз очевидно, что ряд

$$\rho \{I\} = \sum_{n=1}^{\infty} \rho_n \{I\} \quad (3.4)$$

сходится и определяет (возможно, несобственное) распределение точки первого вхождения (т. е. $\rho \{I\} = \overline{\mathcal{H}}_1^- \{I\}$).

Легко вывести рекуррентные соотношения для ψ_n и ρ_n . В самом деле, условная вероятность события $S_{n+1} \in I$ при условии, что $S_n = y$, равна $F \{I - y\}$, где $I - y$ — интервал, получаемый сдвигом I на $-y$. Таким образом,

$$\rho_{n+1} \{I\} = \int_{0-}^{\infty} \psi_n \{dy\} F \{I - y\}, \text{ если } I \subset \overline{-\infty, 0}, \quad (3.5a)$$

$$\psi_{n+1} \{I\} = \int_{0-}^{\infty} \psi_n \{dy\} F \{I - y\}, \text{ если } I \subset \overline{0, \infty} \quad (3.5b)$$

(ноль дает вклад в интеграл только при $n=0$). Для ограниченных интервалов I сходимость $\sum \psi_n \{I\}$ обеспечивается леммой о двойственности, а $\sum \rho_n \{I\}$ всегда сходится к числу ≤ 1 . Суммы этих рядов равны ψ и ρ , и для них мы получаем

$$\rho \{I\} = \int_{0-}^{\infty} \psi \{dy\} F \{I - y\}, \text{ если } I \subset \overline{-\infty, 0}, \quad (3.6a)$$

$$\psi \{I\} = \int_{0-}^{\infty} \psi \{dy\} F \{I - y\}, \text{ если } I \subset \overline{0, \infty} \quad (3.6b)$$

с тем соглашением, что в (3.6b) рассматриваются только ограниченные интервалы I . Мы увидим, что соотношения (3.6) практически более полезны, чем представления ρ и ψ рядами. Иногда удобно заменить функции интервала ρ и ψ соответствующими функциями точки

$$\rho(x) = \rho \{ \overline{-\infty, x} \} \text{ и } \psi(x) = \psi \{ \overline{-\infty, x} \}.$$

Очевидно, что (3.6a) равносильно соотношению

$$\rho(x) = \int_{0-}^{\infty} \psi \{dy\} F(x-y), \quad x \leq 0. \quad (3.7a)$$

Из (3.6b) при $x > 0$ получаем

$$\psi(x) = 1 + \psi \{0, x\} = 1 + \int_{0-}^{\infty} \psi \{dy\} [F(x-y) - F(-y)].$$

Принимая во внимание (3.7a), мы видим, что (3.6b) эквивалентно

$$\psi(x) = 1 - \rho(0) + \int_{0-}^{\infty} \psi \{dy\} F(x-y), \quad x \geq 0. \quad (3.7b)$$

Для упрощения обозначений введем свертку

$$\psi \star F = \sum_{n=0}^{\infty} \psi_n \star F. \quad (3.8)$$

Так как ψ сосредоточено на $\overline{0, \infty}$, то значение $\psi \star F \{I\}$ равно сумме двух интегралов (3.6) и потому конечно. Вспоминая, что ψ имеет единичный атом в нуле, соединим два соотношения (3.6) в одно уравнение свертки

$$\rho + \psi = \psi_0 + \psi \star F. \quad (3.9)$$

Ввиду того что ρ и $\psi - \psi_0$ сосредоточены на $\overline{-\infty, 0}$ и $\overline{0, \infty}$ соответственно, соотношение (3.9) эквивалентно паре соотношений (3.6).

Мы будем рассматривать (3.9) как интегральное уравнение, определяющее неизвестные меры ρ и ψ . Большое число важных теоретических выводов можно получить непосредственно из (3.9). Мы отметим самую замечательную из подобных теорем в форме примера, для того чтобы подчеркнуть, что она не используется в дальнейшем и представляет собой отклонение от основного изложения.

Примеры. а) *Факторизация типа Винера—Хопфа.* Из определения (1.7) функции ψ следует, что она удовлетворяет уравнению восстановления

$$\psi = \psi_0 + \psi \star H. \quad (3.10)$$

Используя это равенство, покажем, что (3.9) можно переписать в эквивалентной форме:

$$F = H + \rho - H \star \rho. \quad (3.11)$$

Действительно, вычисляя свертку (3.9) с H , получаем

$$H \star \rho + \psi - \psi_0 = H - F + \psi \star F.$$

Вычитая это соотношение из (3.9), получим (3.11). Обратно, находя свертку (3.11) с ψ , получаем

$$\psi \star F = \psi - \psi_0 + \psi \star \rho - (\psi - \psi_0) \star \rho = \psi - \psi_0 + \rho,$$

что совпадает с (3.9).

Тождество (3.11) замечательно тем, что дает представление произвольного распределения F в терминах двух (возможно, несобственных) распределений вероятностей H и ρ , сосредоточенных на $\overline{0, \infty}$ и $\overline{-\infty, 0}$ соответственно. Первый интервал открыт, второй замкнут, но эту асимметрию можно поправить, выразив вероятности первого попадания ρ в $\overline{-\infty, 0}$ через вероятности H^- первого попадания в $\overline{-\infty, 0}$. Соотношение между этими вероятностями дается формулой, аналогичной (1.12) при $x < 0$, а именно

$$\rho = \xi \psi_0 + (1 - \xi) H^-,$$

где ξ определено соотношением (1.11) (последнее соотношение справедливо, даже если поменять знак неравенства на противоположный; см. пример 2, а)).

Подставляя это выражение в (3.11), после тривиальных преобразований получаем

$$\psi_0 - F = (1 - \xi) [\psi_0 - H] \star [\psi_0 - H^-]. \quad (3.12)$$

Конечно, если функция распределения F непрерывна, то соотношения (3.11) и (3.12) совпадают.

Различные варианты этой формулы были открыты независимо один от другого различными методами, и все они вызвали восхищение. Одно видоизменение см. в задаче 19, другое в терминах преобразований Фурье дано в гл. XVIII, 3. Связь с методами Винера—Хопфа обсуждается в § 3а. Тождество Вальда (2.8) является простым следствием (3.11). (См. задачу 11 и гл. XVIII, 2.)

б) Явные выражения для H и H^- обычно трудно получить. Интересное распределение, для которого вычисления особенно просты, было найдено в примере VI, 8, б). Если F —свертка двух показательных распределений, сосредоточенных на $\overline{0, \infty}$ и на $\overline{-\infty, 0}$ соответственно, то его плотность имеет вид

$$f(x) = \begin{cases} \frac{ab}{a+b} e^{ax}, & x < 0, \\ \frac{ab}{a+b} e^{-bx}, & x > 0. \end{cases}$$

Предположим, что $b \leq a$, так что $E(X) \geq 0$. Тогда H и H^- имеют плотности, равные be^{-bx} и be^{ax} . Здесь $H \star H^- = (b/a) F$, и (3.11) тривиально верно. 

(Другие примеры, содержащие явные решения, см. в задаче 9.)

Обратимся теперь к (3.9), как к интегральному уравнению относительно неизвестных мер ρ и ψ . Мы покажем, что его решение единственно, если принять во внимание свойства, которые ρ и ψ обязаны иметь в этом контексте. Условимся говорить для краткости, что пара (ρ, ψ) имеет *вероятностный смысл*, если ρ —(возможно, несобственное) распределение вероятностей, сосредоточенное на $-\infty, 0$, а ψ — ψ_0 —мера, сосредоточенная на $0, \infty$ и такая, что для каждого интервала I значения $\psi\{I+t\}$ ограничены (последнее условие вытекает из теорем восстановления, так как $\psi = \sum H^{n*}$).

Теорема 1. *Уравнение свертки (3.9) [или равносильная ему пара уравнений (3.6)] обладает только одним решением (ρ, ψ) , имеющим вероятностный смысл.*

Отсюда вытекает, что ρ есть распределение точки первого вхождения в $-\infty, 0$ и $\psi = \sum H^{n*}$, где H —распределение точки первого вхождения в $0, \infty$.

Доказательство. Пусть ρ^* и ψ^* —две неотрицательные меры, удовлетворяющие (3.6), и $\psi^* \geq \psi_0$. Из (3.6б) по индукции получаем $\psi^* \geq \psi_0 + \dots + \psi_n$ при любом n . Поэтому наше решение ψ минимально в том смысле, что для любого другого решения ψ^* с единичным атомом в нуле $\psi^*\{I\} \geq \psi\{I\}$ для всех интервалов. Другими словами, $\delta = \psi^* - \psi$ есть *мера*. Из (3.6а) видно, что это же верно и относительно $\gamma = \rho^* - \rho$. Так как и (ρ, ψ) , и (ρ^*, ψ^*) удовлетворяют (3.9), мы имеем

$$\delta + \gamma = \delta \star F. \quad (3.13)$$

Положим $z(t) = \delta(I+t)$, где I —фиксированный конечный интервал. Возможны два случая. Если ρ —собственное распределение, то из $\rho^* \geq \rho$ вытекает, что $\rho^* = \rho$ и, следовательно, $\gamma = 0$. Тогда функция z будет ограниченным решением уравнения свертки $z = F \star z$, и в этом случае по индукции получаем

$$z(t) = \int_{-\infty}^{+\infty} z(t-y) F^{n*} \{dy\} \quad (3.14)$$

при всех n . Теперь $z \geq 0$ и $z(t) = 0$ для любого t , такого, что $I+t$ лежит на отрицательной полуоси. Из (3.14) ясно, что при таких t $z(t-y) = 0$ для каждого значения y , которое является точкой роста для некоторой функции F^{n*} . По лемме 2 из гл. V, 4а, множество таких y всюду плотно, и мы приходим к заключению, что z тождественно равно нулю.

Для несобственного распределения ρ мы знаем только, что $\gamma \geq 0$, и тогда из (3.13) следует лишь то, что $z \leq F \star z$. В этом случае (3.14) выполнено с заменой знака равенства на $\leq$. Но

тогда случайное блуждание уходит в ∞ , и, таким образом, масса F_n^* стремится сосредоточиться «вблизи» ∞ . Кроме того, $z(t-y) = 0$ для всех достаточно больших y , и поэтому z должно тождественно обращаться в нуль. Таким образом, $\psi^* = \psi$, как и утверждалось.

3а. Интегральное уравнение Винера — Хопфа

Связь между интегральным уравнением (3.9) и обычным уравнением Винера — Хопфа можно лучше всего объяснить, начав с вероятностной задачи, в которой встречается последнее уравнение.

Пример. в) *Распределение максимума.* Предположим для простоты, что распределение F имеет плотность f и отрицательное математическое ожидание. Случайное блуждание $\{S_n\}$ уходит в $-\infty$, и с вероятностью 1 определена конечнозначная случайная величина

$$M = \max[0, S_1, S_2, \dots]. \quad (3.15)$$

Найдем ее распределение вероятностей $M(x) = P\{M \leq x\}$, которое по определению сосредоточено на $\overline{0, \infty}$. Событие $\{M \leq x\}$ происходит тогда и только тогда, когда $X_1 = y \leq x$ и $\max[0, X_2, X_2 + X_3, \dots] \leq x - y$. Суммируя по всем возможным значениям y , получаем

$$M(x) = \int_{-\infty}^x M(x-y) f(y) dy, \quad x > 0, \quad (3.16)$$

или, что то же самое,

$$M(x) = \int_0^{\infty} M(s) f(x-s) ds, \quad x > 0. \quad (3.17)$$

С другой стороны, мы знаем из (2.7), что $M(x) = [1 - H(\infty)] \psi(x)$. Мы видели, что ψ удовлетворяет интегральному уравнению (3.7б), в котором теперь следует принять $\rho(0) = 1$. Простое интегрирование по частям показывает теперь, что (3.7б) и (3.17) совпадают. ►

Стандартной формой уравнения Винера — Хопфа является (3.17). Наш пример указывает путь, на котором это уравнение может возникать в теории вероятностей. Однако ссылки на общую теорию Винера — Хопфа могут запутать положение, так как мы ограничиваемся положительными функциями и мерами, что меняет (и упрощает) характер проблемы.

Искусный метод¹⁾, использованный Винером и Хопфом для

¹⁾ Около 1931 г. Обширная литература появилась после первого монографического изложения: Hopf E., Mathematical problems of radiative equilibrium, Cambridge tracts, 31, 1934.

(3.17), привлек усиленное внимание и был приспособлен для решения различных вероятностных задач, например, Г. Крамером для вывода асимптотических оценок вероятностей разорения. Этот метод включает громоздкий аналитический аппарат, и поэтому легкость, с которой подобные оценки получаются при нашем подходе, вызывает некоторое беспокойство. Но следует понять более глубокие причины этой легкости. Уравнение (3.17) представляет собой в лучшем случае лишь одно из двух уравнений (3.7), а при $\rho(0) < 1$ и того меньше. Взятое само по себе, уравнение (3.17) значительно более трудно для изучения, чем пара уравнений (3.7). Например, теорема единственности *неверна* для (3.17) даже в классе распределений вероятностей. Основная идея метода Винера—Хопфа состоит во введении вспомогательной функции, которая в общей теории не имеет никакого особого смысла. Этот остроумный прием¹⁾ на самом деле заменяет отдельное уравнение (3.17) парой уравнений, равносильной (3.7), но единственность при этом теряется. Мы двигались в противоположном направлении, начиная с очевидных рекуррентных соотношений для вероятностей, связанных с двумя неотделимыми друг

от друга задачами: изучение первого вхождения в $-\infty, 0$ и случайного блуждания на полупрямой $x > 0$ до момента этого вхождения. Таким путем мы получили интегральное уравнение (3.9) из известных соотношений. Единственность *вероятностного* решения было легко доказать. Доказательства сходимости, свойства решений, так же как и связь между распределением M максимума и «мерой восстановления» Φ , основаны на лемме о двойственности.

Возможность подхода к уравнению Винера—Хопфа (3.17), опирающегося на принцип двойственности, была отмечена Ф. Спитцером²⁾. При обычном способе соединять теорию Винера—Хопфа с вероятностными задачами начальным шагом служат формулы, связанные с нижеследующей формулой (9.3) (в форме, содержащей преобразования Фурье; мы вернемся к этому в гл. XVIII). В настоящее время имеется обширная литература, посвященная применению метода Винера—Хопфа к вероятностным задачам и расширению возможностей комбинаторных методов. Большая часть работ использует анализ Фурье³⁾.

¹⁾ У автора «tour de force» — (фр.) — *Прим. перев.*

²⁾ Spitzer F., The Wiener—Hopf equation whose kernel is a probability density, Duke Math. J., 24 (1957), 327—343.

³⁾ Краткий, но осмысленный обзор литературы невозможен; причина в ее полной неупорядоченности и в том, что методология многих статей находится под влиянием случайностей исторического развития. Обобщения, лежащие за пределами теории вероятностей, проиллюстрированы Бакстером [Baxter G., An operator identity, Pacific J. Math., 4 (1958), 649—663].

§ 4. ПРИМЕРЫ

Явные формулы для распределений первого вхождения в общем случае получить трудно. К счастью, имеется одно примечательное исключение из этого правила, обсуждаемое в примере а). На первый взгляд распределение F из этого примера представляется искусственным, однако этот тип распределений часто встречается в связи с процессами Пуассона, теорией очередей, задачами о разорении и т. д. Крайняя простота наших общих результатов делает непостижимым то, как много изобретательности и аналитического мастерства было затрачено (часто повторно) на анализ частных случаев.

Пример в) показывает шаг за шагом процесс полного решения для арифметического распределения F с рациональной производящей функцией. Эти вычисления приведены потому, что точно такой же метод пригоден для рациональных преобразований Лапласа или Фурье. Другой пример указан в задачах 3—6. Пример б) посвящен некоторым общим соотношениям, имеющим независимый интерес.

Мы придерживаемся обозначений предыдущего параграфа. Таким образом, H и ρ суть распределения точки первого вхождения в 0 , ∞ и $-\infty$, 0 соответственно. (Другими словами, H и ρ являются распределениями соответственно первой строгой верхней и слабой нижней лестничных высот.) Наконец, $\psi = \sum H^{n*}$ — это функция восстановления, соответствующая H . Нашим основным орудием будет уравнение (3.7а), утверждающее, что при $x < 0$ распределение точки первого вхождения в $-\infty$, 0 определяется формулой

$$\rho(x) = \int_0^{\infty} \psi\{dy\} F(x-y). \quad (4.1)$$

Примеры. а) Распределения с одним показательно убывающим хвостом появляются чаще, чем можно было ожидать. Так, в случайном блуждании примера гл. VI, 8, б), и в соответствующем процессе ожидания гл. VI, 9, д), оба хвоста показательны. Допустим для начала, что левый хвост распределения F — показательный, т. е. $F(x) = qe^{\beta x}$ при $x < 0$. Каково бы ни было ψ , (4.1) показывает, что $\rho(x) = Ce^{\beta x}$ при $x < 0$, где C — некоторая постоянная. Произведя это открытие, поменяем полуоси ролями (частью для облегчения ссылок на наши формулы, частью имея в виду наиболее важные применения к теории очередей). Предположим, таким образом, что

$$F(x) = 1 - re^{-\alpha x} \text{ при } x \geq 0, \quad (4.2)$$

не делая никаких ограничений при $x < 0$. Во избежание ненуж-

ных усложнений допустим, что F имеет конечное математическое ожидание μ и что F непрерывна. Из предварительного замечания вытекает, что распределение лестничной высоты H имеет плотность, пропорциональную $e^{-\alpha x}$. Рассмотрим теперь отдельно два случая.

(i) Если $\mu \geq 0$, то H — собственное распределение и, стало быть, при $x > 0$

$$H(x) = 1 - e^{-\alpha x}, \quad \psi(x) = 1 + \alpha x. \quad (4.3)$$

[Последнее равенство тривиально следует как из того, что $\psi = \sum H^n$, так и из уравнения восстановления (3.10).] Из (4.1) получаем

$$\rho(x) = F(x) + \alpha \int_{-\infty}^x F(s) ds, \quad x < 0. \quad (4.4)$$

Таким образом, мы имеем явные выражения для всех нужных нам вероятностей. Легкий подсчет показывает, что

$$\rho(0) = 1 - \alpha\mu. \quad (4.5)$$

Это частный случай (2.8), так как в силу (2.5) $(1 - \rho(0))^{-1} = E(\mathcal{T}_1)$.

(ii) Если $\mu < 0$, то соотношения (4.3) и (4.4) по-прежнему дают решение интегрального уравнения (3.9), однако (4.5) показывает, что это решение не может иметь вероятностного смысла при $\mu < 0$. Для отыскания правильного решения заметим, что H имеет плотность $h(x) = (\alpha - \kappa)e^{-\alpha x}$, где $0 < \kappa < \alpha$, так как распределение H — несобственное. Легко подсчитать, что $\psi'(x) = (\alpha - \kappa)e^{-\kappa x}$ при $x > 0$. Неизвестная константа κ определяется из условия $\rho(0) = 1$. Стандартные вычисления показывают, что κ совпадает с единственным положительным корнем уравнения (4.6). Найдя корень этого трансцендентного уравнения, мы снова получим явные выражения для H , ρ и ψ .

Читатель может легко проверить, что эта же теория применима и тогда, когда величины $X_1, X_2, \dots$, порождающие случайное блуждание, целочисленны и распределение F имеет геометрически убывающий правый хвост, т. е. если F приписывает целому числу $k > 0$ массу $q\beta^k$.

б) *Сопряженные случайные блуждания.* Допустим, что F имеет математическое ожидание $\mu \neq 0$ и что существует число $\kappa \neq 0$, для которого

$$\int_{-\infty}^{+\infty} e^{\kappa y} F(dy) = 1. \quad (4.6)$$

Пусть γ — произвольная мера на прямой. Свяжем с ней новую меру ${}^a\gamma$, полагая

$${}^a\gamma\{dy\} = e^{\kappa y} \gamma\{dy\}. \quad (4.7)$$

Мера ${}^aF$, «сопряженная» с F , представляет собой собственное распределение вероятностей. Мы будем говорить, что случайные блуждания, порождаемые ${}^aF$ и F , сопряжены друг с другом¹⁾. Легко видеть, что n -кратная свертка ${}^aF$ с собой сопряжена с F^{n*} , так что обозначение ${}^aF^{n*}$ не вызывает сомнений. Кроме того, рекуррентные формулы (3.5) показывают, что преобразования ${}^a\rho_n$ и ${}^a\psi_n$ имеют в новом случайном блуждании тот же самый вероятностный смысл, как ρ_n и ψ_n в старом. Вообще преобразования ${}^a\rho$, ${}^aH$, ${}^a\psi$ и т. д. имеют очевидное значение для случайного блуждания, порождаемого ${}^aF$. [Это можно увидеть и непосредственно из интегрального уравнения (3.9).]

Интеграл

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{yt} F\{dy\}$$

существует для всех t между 0 и κ , и в этом интервале функцию φ можно дифференцировать бесконечное число раз. Поскольку вторая производная положительна, φ является выпуклой функцией. Если $\varphi'(\kappa)$ существует, то из равенства $\varphi(0) = \varphi(\kappa)$ следует, что $\varphi'(0)$ и $\varphi'(\kappa)$ имеют противоположные знаки. Следовательно, случайные блуждания, порожденные распределениями F и ${}^aF$, имеют снос в противоположных направлениях. (Очевидно, это остается верным даже в том исключительном случае, когда ${}^aF$ имеет бесконечное математическое ожидание.)

Мы наметили, таким образом, широко применимый метод перевода результатов, касающихся случайного блуждания с $\mu < 0$, в результаты для случайных блужданий с положительным математическим ожиданием и обратно.

Если $\mu < 0$, то лестничная высота имеет несобственное распределение, но ${}^aH$ имеет собственное распределение. Это означает, что

$$\int_0^{\infty} e^{xy} H\{dy\} = 1. \quad (4.8)$$

Вся сила метода сопряженных случайных блужданий проистекает главным образом из этого замечания. В самом деле, в гл. XI, 6, мы установили, что, зная корень уравнения (4.8), можно дать прекрасные асимптотические оценки для возрастающего лестничного процесса. Эти оценки были бы иллюзорны, если бы они требовали

¹⁾ Это понятие применяли А. Хинчин, А. Вальд и др., но оно никогда не использовалось в полной мере. Преобразование (4.7) встречалось в теории восстановления (гл. XI, 6) и в форме, замаскированной применением производящих функций, в теореме 1 (iii) в 1; гл. XIII, 10. Оно появится снова в теории преобразований Лапласа, см. гл. XIII, (1.8). Уравнение (4.6) применяют и в теории Винера — Хопфа.

знания H , но мы только что видели, что корни уравнений (4.6) и (4.8) совпадают.

в) *Ограниченные арифметические распределения.* Пусть a и b — положительные числа и F — арифметическое распределение с шагом 1 и со скачками f_k в точках $k = -b, \dots, a$. Меры ψ и ρ также сосредоточены в целых точках. Их скачки в точке k мы обозначим ψ_k и ρ_k соответственно. Первое вхождение в $-\infty, 0$ осуществляется в целой точке $\geq -b$, так что $\rho_k = 0$ при $k < -b$.

Введем производящие функции

$$\Phi(s) = \sum_{k=-b}^a f_k s^k, \quad \Psi(s) = \sum_{k=0}^{\infty} \psi_k s^k, \quad R(s) = \sum_{k=-b}^0 \rho_k s^k. \quad (4.9)$$

Они отличаются от производящих функций из 1; гл. XI, тем, что Φ и R содержат также и отрицательные степени s , однако ясно, что основные свойства и правила действий при этом сохраняются. В частности, $\mu = \Phi'(1)$ есть математическое ожидание F . Свертке распределений соответствует перемножение производящих функций, поэтому основное интегральное уравнение (3.9) равносильно уравнению $\Psi + R = 1 + \Psi\Phi$ или

$$\Psi(s) = \frac{s^b (R(s) - 1)}{s^b (\Phi(s) - 1)}. \quad (4.10)$$

Числитель и знаменатель здесь — полиномы степени b и $a+b$ соответственно. Степенной ряд слева сходится при $|s| < 1$, так что все корни знаменателя, лежащие внутри единичного круга, должны быть в то же время и корнями числителя. Мы покажем, что это условие позволяет однозначно восстановить R и Ψ .

Предположим для определенности, что $\mu = 0$ (при $\mu \neq 0$ см. задачи 12, 13). Тогда $s = 1$ является двойным корнем уравнения $\Phi(s) = 1$. При $|s| = 1$ имеем $|\Phi(s)| \leq \Phi(1) = 1$, причем неравенство имеет место только тогда, когда $s^k = 1$ для всякого такого k , что $f_k > 0$. Поскольку предполагалось, что распределение F имеет скачок 1, то это верно лишь при $s = 1$ и, следовательно, других корней уравнения $\Phi(s) = 1$ на единичной окружности нет. Чтобы найти, сколько корней расположено внутри единичного круга, рассмотрим многочлен степени $a+b$, определенный формулой

$$P(s) = s^b [\Phi(s) - q], \quad q > 1.$$

При $|s| = 1$ имеем $|P(s)| \geq q - 1 > 0$ и

$$|P(s) + qs^b| = |\Phi(s)| \leq 1 < q |s^b|.$$

По теореме Руше¹⁾ отсюда следует, что многочлены $P(s)$ и qs^b имеют одинаковое число нулей, расположенных внутри единичного

¹⁾ См., например: Hille E., Analytic function theory, vol. I. section 9.2 (Ginn and Co., 1959) или Маркушевич А. И. Теория аналитических функций. — М.: Гостехиздат, 1950.

круга. Таким образом, P имеет в точности b корней, таких, что $|s| < 1$, и a корней, таких, что $|s| > 1$. Далее, $P(0) = 0$ и $P(1) = 1 - q < 0$, тогда как $P(s) > 0$ для больших s . Таким образом, P имеет два действительных корня $s' < 1 < s''$. При $q \rightarrow 1$ корни P стремятся к корням $\Phi(s) = 1$. Таким образом, наконец, приходим к выводу, что знаменатель в (4.10) имеет двойной корень 1 и, кроме того, $b-1$ корней $s_1, \dots, s_{b-1}$ с $|s_j| < 1$ и $a-1$ корней $\sigma_1, \dots, \sigma_{a-1}$ с $|\sigma_j| > 1$. Тогда знаменатель имеет вид $s^b (\Phi(s) - 1) = C(s-1)^2(s-s_1)\dots(s-s_{b-1})(s-\sigma_1)\dots(s-\sigma_{a-1})$. (4.11)

Корни $s_1, \dots, s_{b-1}$ должны совпадать с корнями числителя, и так как коэффициенты ψ_n ограничены, то же самое должно быть верно и для *одного* корня $s = 1$. Этим Ψ определяется с точностью до постоянного множителя. Но по определению $\Psi(0) = 1$, и мы получаем *искомую явную формулу*

$$\Psi(s) = \frac{1}{(1-s) \left(1 - \frac{s}{\sigma_1}\right) \dots \left(1 - \frac{s}{\sigma_{a-1}}\right)}. \quad (4.12)$$

Стандартное разложение на простейшие дроби приводит к явному выражению для ψ_n . Большое преимущество этого метода в том, что знание доминирующего корня приводит к хорошим асимптотическим оценкам (см. 1; гл. XI, 4).

Для производящей функции R вероятностей первого вхождения p_k из (4.10) и (4.12) получаем

$$R(s) = 1 + C(-1)^{a-1} \sigma_1 \dots \sigma_{a-1} \left(1 - \frac{1}{s}\right) \left(1 - \frac{s_1}{s}\right) \dots \left(1 - \frac{s_{b-1}}{s}\right). \quad (4.13)$$

[Коэффициент C определяется по (4.11) и зависит только от данного распределения $\{f_k\}$.] Снова разложение на простейшие дроби приводит к асимптотическим оценкам (продолжение см. в задачах 12—15).

§ 5. ПРИМЕНЕНИЯ

В гл. VI, 9, было показано, что основная задача теории очередей сводится к отысканию распределения M для

$$M = \max [0, S_1, \dots] \quad (5.1)$$

в случайном блуждании, порожденном величинами X_k с $\mu = E(X_k) < 0$. Примеры 9, а) — 9, в) гл. VI показывают, что это же распределение возникает и в других обстоятельствах, например в связи с задачей о разорении для обобщенного процесса Пуассона. В последнем случае, как и в теории очередей, основное распределение имеет вид

$$F = A \star B, \quad (5.2)$$

где A сосредоточено на $\overline{0, \infty}$, а B — на $\overline{-\infty, 0}$. Мы предположим, что A и B имеют конечные математические ожидания a и $-b$, так что F имеет математическое ожидание $\mu = a - b$. Мы предположим также, что распределение F непрерывно. Это позволит избежать утомительной необходимости различать строгие и слабые лестничные величины.

Мы обозначим, так же как в двух предыдущих параграфах, распределения для верхних и нижних лестничных высот через H и ρ соответственно. Иными словами, H и ρ суть распределения для первых вхождений в $\overline{0, \infty}$ и $\overline{-\infty, 0}$ (а также в соответствующие замкнутые интервалы). В примере 3, в) и в (2.7) было показано, что при $\mu < 0$

$$M(x) = \frac{\Psi(x)}{\Psi(\infty)} = [1 - H(\infty)] \sum_0^{\infty} H^{n*}(x). \quad (5.3)$$

Пример 4, а) содержит явную формулу¹⁾ для этого распределения, справедливую, если один из хвостов F — показательный, т. е.

$$F(x) = 1 - pe^{-\alpha x} \quad \text{при } x > 0 \quad (5.4)$$

или $F(x) = qe^{\alpha x}$ при $x < 0$.

По весьма счастливому совпадению условие (5.4) выполняется, если F имеет вид (5.2) с

$$A(x) = 1 - e^{-\alpha x} \quad \text{при } x > 0. \quad (5.5)$$

Тогда

$$\rho = \int_{-\infty}^0 e^{\alpha y} B\{dy\}. \quad (5.6)$$

Наши простые результаты применимы поэтому к теории очередей, если поступающий поток является пуассоновским или время обслуживания распределено показательно. Кроме того, упомянутое условие выполнено и в задаче о разорении для обобщенного процесса Пуассона. Существует необъятная прикладная литература, разбирающая отдельные задачи при тех или иных предположениях относительно распределения B , иногда в замаскированной форме, как в задаче о разорении. Оказывается, большая общность и значительно большая простота достигаются при использовании только одного условия (5.4) вместо комбинации (5.5) и (5.2). Мы имеем здесь прекрасный пример экономии мысли, присущей

¹⁾ Другая явная формула указана в примере 4, в) для случая арифметического распределения F с конечным числом атомов. Эта явная формула слишком сложна для практического использования, но разложение на простейшие дроби приводит к хорошим асимптотическим оценкам, если известен доминирующий корень знаменателя. Такой же метод применим в случае рациональных характеристических функций. Это замечание охватывает множество частных случаев, разобранных в литературе.

общей теории, когда понимание не затемняется случайными чертами отдельных примеров.

Примеры. а) *Формула Хинчина — Полачека.* Допустим, что F имеет вид (5.2), где A удовлетворяет (5.5) и $\mu = 1/\alpha - b > 0$. Случайное блуждание уходит в ∞ , и мы должны заменить в (5.1) максимум на минимум. Это равносильно замене H в (5.3) распределением r из (4.4). Простое интегрирование по частям показывает, что при $x < 0$

$$\rho(x) = \alpha \int_{-\infty}^x B(y) dy. \quad (5.7)$$

Здесь $\rho(0) = \alpha b$, так что при $x < 0$

$$P\{\min(S_0, S_1, \dots) \leq x\} = (1 - \alpha b) \sum_0^{\infty} \rho^{n*}(x). \quad (5.8)$$

Это известная формула Хинчина — Полачека, которую неоднократно открывали заново в частных случаях, неизменно используя метод преобразования Лапласа [неприменимый к более общим распределениям типа (5.4)]. Мы вернемся к этой формуле в задачах 10, 11 гл. XVIII, 7.

б) *Двойственный случай.* Рассмотрим такие же распределения, как и в предыдущем примере, но предположим, что $\mu < 0$. Как было показано во второй части примера 4, а), в этом случае

$$P\{\max(S_0, S_1, \dots) \leq x\} = \frac{x}{\alpha} \psi(x) = 1 - \left(1 - \frac{x}{\alpha}\right) e^{-\kappa x}, \quad (5.9)$$

где κ — единственный положительный корень «характеристического уравнения» (4.6). [Этот результат может быть получен также и методом сопряженных случайных блужданий; достаточно вспомнить, что при $\mu \geq 0$ $\psi(x) = 1 + \alpha x$ для $x > 0$.] В применении к теории очередей (5.9) означает, что если *время обслуживания распределено показательным, то распределение времен ожидания стремится к показательному пределу.*

в) *Асимптотические оценки.* Метод, использующий сопряженное случайное блуждание, который описан в § 4б, легко приводит к полезным оценкам для хвоста распределения

$$M(x) = P\{\max(S_0, S_1, \dots) \leq x\}. \quad (5.10)$$

Следующий простой метод заменяет множество сложных вычислений, предназначенных для решения отдельных задач в литературе прикладного характера. Этот метод представляет собой частный случай общей теории из гл. XI, 6, но здесь для удобства его изложение будет совершенно самостоятельно.

Распределение M дано в (5.3), где через ψ обозначена «мера восстановления», соответствующая несобственному распределению H . Сопряженному собственному распределению ${}^aH$ соответ-

вует мера ${}^a\psi$, заданная равенством ${}^a\psi\{dx\} = e^{\kappa x} \psi\{dx\}$, где κ дано в (4.6) или (4.8). Таким образом, можно переписать (5.3) в виде

$$M\{dx\} = [1 - H(\infty)] e^{-\kappa x} {}^a\psi\{dx\}. \quad (5.11)$$

В силу основной теоремы восстановления из гл. XI, 1, мера ${}^a\psi$ распределена асимптотически равномерно с плотностью β^{-1} , где

$$\beta = \int_0^{\infty} x e^{\kappa x} H\{dx\}. \quad (5.12)$$

Интегрируя (5.11) в пределах от t до ∞ , мы видим, таким образом, что при $t \rightarrow \infty$

$$1 - M(t) \sim \frac{1 - H(\infty)}{\beta \kappa} e^{-\kappa t}, \quad (5.13)$$

если только $\beta < \infty$. [В противном случае $1 - M(t) = o(e^{-\kappa t})$.]

Постоянная β зависит от распределения H , явный вид которого обычно неизвестен, однако показатель степени κ зависит только от заданного распределения F . Следовательно, в худшем случае (5.13) дает оценку, в которую входит неизвестный множитель, но даже этот результат нелегко получить другими методами. Следующий пример посвящен важным приложениям.

г) *Оценка Крамера для вероятностей разорения.* Применим теперь предыдущий результат к примеру а). Здесь случайное блуждание уходит в $+\infty$, и, следовательно, положительную и отрицательную полуоси можно поменять местами. Это означает, что $\kappa < 0$ и распределение H надо заменить распределением (5.7) первого попадания в $-\infty, 0$. Таким образом, (5.12) принимает вид

$$\beta = \alpha \int_{-\infty}^0 e^{-|\kappa| y} |y| B(y) dy. \quad (5.14)$$

Мы видели, что $\rho(0) = \alpha b$, и, таким образом, (5.13) эквивалентно утверждению, что при $x \rightarrow \infty$

$$P\{\min(S_0, S_1, \dots) \leq x\} \sim \frac{1 - \alpha b}{|\kappa| \beta} e^{|\kappa| x}. \quad (5.15)$$

Эта формула имеет много приложений. В теории очередей левая часть (5.15) представляет собой предельное распределение времени ожидания n -го клиента (см. теорему в гл. VI, 9). В примере VI, 9, г), было показано, что к этой задаче теории очередей может быть сведена основная задача о разорении из гл. VI, 5. В других обозначениях эта же задача рассматривается в примере XI, 7, а) ¹⁾.

¹⁾ Нашему несобственному распределению ρ , сосредоточенному на $-\infty, 0$, в гл. VI, 7, соответствует сосредоточенное на $0, \infty$ несобственное распределение L с плотностью $(\alpha/c)(1 - F(x))$.

Поэтому неудивительно, что оценку (5.15) неоднократно выводили при специальных предположениях, но задача оказывается более простой в ее естественной общей постановке. В общей теории оценка (5.15) эквивалентна известной оценке, которая для вероятности разорения в математической теории страхования была предложена Крамером¹⁾.

§ 6. ОДНА КОМБИНАТОРНАЯ ЛЕММА

Распределение лестничных моментов связано с одной простой комбинаторной леммой. Вероятностная часть рассуждений будет более ясной, если мы выделим эту лемму особо.

Пусть $x_1, \dots, x_n$ — n чисел. Рассмотрим их частные суммы

$$s_0 = 0, \dots, s_n = x_1 + \dots + x_n.$$

Мы скажем, что $v > 0$ есть *лестничный индекс*, если $s_v > s_0, \dots, s_v > s_{v-1}$, т. е. если s_v превосходит все предыдущие частные суммы. Ясно, что имеется n лестничных индексов, если все x_v положительны, и нет ни одного, если все x_v отрицательны.

Рассмотрим n циклических перестановок $(x_1, \dots, x_n)$, $(x_2, \dots, x_n, x_1)$, $\dots$, $(x_n, x_1, \dots, x_{n-1})$ и занумеруем их числами $0, \dots, n-1$. Частные суммы $s_k^{(v)}$ в перестановке с номером v равны

$$s_k^{(v)} = \begin{cases} s_{v+k} - s_v & \text{при } k = 1, \dots, n-v, \\ s_n - s_v + s_{k-n+v} & \text{при } k = n-v+1, \dots, n. \end{cases} \quad (6.1)$$

Лемма 1. Пусть $s_n > 0$. Обозначим через r число циклических перестановок, в которых n является лестничным индексом. Тогда $r \geq 1$, и в каждой такой циклической перестановке будет ровно r лестничных индексов.

Примеры. Для $(-1, -1, -1, 0, 1, 10)$ мы имеем $r=1$. Данный порядок — единственный, в котором последняя частная сумма максимальна. Для $(-1, 4, 7, 1)$ мы имеем $r=3$; перестановки с номерами 0, 2 и 3 содержат по три лестничных индекса.

Доказательство. Выберем v так, что s_v максимально. Если таких индексов несколько, то возьмем наименьший из них. Другими словами,

$$s_v > s_1, \dots, s_v > s_{v-1}, \quad s_v \geq s_{v+1}, \dots, s_v \geq s_n. \quad (6.2)$$

Из (6.1) видно, что при этом в v -й перестановке последняя частная сумма строго максимальна, так что n есть лестничный индекс.

¹⁾ Более новый вывод методом Винера—Хопфа в комплексной плоскости см. в статье Крамера, цитированной в гл. VI, 5. Наша оценка (5.10) совпадает с формулой (57) у Крамера.

Таким образом, $r \geq 1$. Не ограничивая общности, допустим, что n есть лестничный индекс в исходной перестановке, т. е. $s_n > s_j$ при всех j . Тогда все величины в первой строке (6.1) будут $< s_n$. Вторая строка показывает, что n будет лестничным индексом для ν перестановок тогда и только тогда, когда $s_\nu > s_1, \dots, s_\nu > s_{\nu-1}$, т. е. когда в исходной перестановке ν будет лестничным индексом. Таким образом, число перестановок, в которых n есть лестничный индекс, равно числу лестничных индексов. Лемма доказана. ►

Слабые лестничные индексы определяются аналогично, с той разницей, что строгие неравенства $>$ заменяются на $\geq$. В применении к ним предыдущие рассуждения показывают, что верна

Лемма 2. Если $s_n \geq 0$, то лемма 1 применима и к слабым лестничным индексам.

§ 7. РАСПРЕДЕЛЕНИЕ ЛЕСТНИЧНЫХ МОМЕНТОВ

В предыдущих параграфах наше внимание было сосредоточено на лестничных высотах. Теперь мы перейдем к лестничным моментам. Рассмотрим вероятность того, что n является моментом первого вхождения в $0, \infty$, т. е.

$$\tau_n = P\{S_1 \leq 0, \dots, S_{n-1} \leq 0, S_n > 0\}. \quad (7.1)$$

Другими словами, $\{\tau_n\}$ есть (возможно, несобственное) распределение первого лестничного момента $\mathcal{T}_i$. Введем производящую функцию

$$\tau(s) = \sum_{n=1}^{\infty} \tau_n s^n, \quad 0 \leq s \leq 1. \quad (7.2)$$

Следующая ниже важная теорема показывает, что распределение $\{\tau_n\}$ полностью определяется по вероятностям $P\{S_n > 0\}$ и наоборот. Теорема была найдена Спарре-Андерсеном. Данное им остроумное, но крайне сложное доказательство было постепенно упрощено другими авторами. Мы получаем эту теорему как простое следствие нашей комбинаторной леммы. [Более строгие варианты содержатся в (9.3) и будут рассмотрены с помощью методов Фурье в гл. XVIII.]

Теорема 1_ε

$$\log \frac{1}{1-\tau(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n > 0\}. \quad (7.3)$$

Замечание. Теорема и ее доказательство остаются верными, если в (7.1) и (7.3) знаки неравенств $>$ и $\leq$ заменить на $\geq$ и $<$ соответственно. В этом случае $\{\tau_n\}$ обозначает распределение первого слабого лестничного момента,

Доказательство. Для каждой выборочной точки $(X_1, X_2, \dots, X_n)$ рассмотрим n циклических перестановок $(X_v, \dots, X_n, X_1, \dots, X_{v-1})$ и обозначим соответствующие частные суммы через $S_0^{(v)}, \dots, S_n^{(v)}$. Фиксируем целое r и определим n случайных величин $Y^{(v)}$ следующим образом: $Y^{(v)} = 1$, если n есть r -й лестничный индекс для $(S_1^{(v)}, \dots, S_n^{(v)})$, и $Y^{(v)} = 0$ в других случаях. Случаю $v=1$ соответствует последовательность $\{S_0, \dots, S_n\}$, поэтому

$$P\{Y^{(1)} = 1\} = \tau_n^{(r)}, \quad (7.4)$$

где $\{\tau_n^{(r)}\}$ — это распределение r -го лестничного момента. Этот момент есть сумма r независимых случайных величин, распределенных так же, как $\mathcal{S}_1$, и, следовательно, $\tau_n^{(r)}$ — это коэффициент при s^n в $\tau^r(s)$. По причине симметрии все случайные величины $Y^{(v)}$ имеют одно и то же распределение. Поскольку они принимают лишь два значения 0 и 1, то из (7.4) следует, что

$$\tau_n^{(r)} = E(Y^{(1)}) = \frac{1}{n} E(Y^{(1)} + \dots + Y^{(n)}). \quad (7.5)$$

По лемме § 6 сумма $Y^{(1)} + \dots + Y^{(n)}$ может принимать только значения 0 и r , откуда

$$\frac{1}{r} \tau_n^{(r)} = \frac{1}{n} P\{Y^{(1)} + \dots + Y^{(n)} = r\}. \quad (7.6)$$

Для фиксированного n и $r=1, \dots$ события в правой части (7.6) взаимно исключают друг друга и их объединение есть событие $\{S_n > 0\}$. Суммируя по всем r , получаем

$$\sum_{r=1}^{\infty} \frac{1}{r} \tau_n^{(r)} = \frac{1}{n} P\{S_n > 0\}. \quad (7.7)$$

Умножая (7.7) на s^n и суммируя по n , находим

$$\sum_{r=1}^{\infty} \frac{1}{r} \tau^r(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n > 0\}, \quad (7.8)$$

что совпадает с утверждением (7.3). ►

Следствие. Если F симметрично и непрерывно, то

$$\tau(s) = 1 - \sqrt{1-s}. \quad (7.9)$$

Доказательство. Все вероятности $P\{S_n > 0\}$ в (7.3) равны $1/2$, поэтому правая часть (7.3) равна $\log(1/\sqrt{1-s})$. ►

Интересно обобщить этот результат, предполагая только, что

$$P\{S_n > 0\} \rightarrow \frac{1}{2}. \quad (7.10)$$

Это так, когда распределение величин S_n/a_n стремится к нормальному распределению $\mathcal{N}$. Мы предположим несколько больше, чем (7.10), а именно, что

сходится (не обязательно абсолютно) ряд

$$\sum_{n=1}^{\infty} \frac{1}{n} \left[P \{S_n > 0\} - \frac{1}{2} \right] = c, \quad (7.11)$$

В гл. XVIII, 5, будет показано, что (7.11) верно всякий раз, когда F имеет нулевое математическое ожидание и конечную дисперсию.

Следующая теорема приведена не только потому, что она имеет самостоятельный интерес, но и как иллюстрация к использованию уточненных тауберовых теорем,

Теорема 1а. Если (7.11) справедливо, то

$$P \{ \mathcal{F}_1 > n \} \sim \frac{1}{\sqrt{\pi}} e^{-c} \frac{1}{\sqrt{n}}. \quad (7.12)$$

Таким образом, когда F имеет нулевое математическое ожидание и конечную дисперсию, распределение $\{\tau_n\}$ очень близко к распределению, которое встречается при случайном блуждании Бернулли,

Доказательство. В силу (7.3) мы видим, что при $s \rightarrow 1$

$$\log \frac{\sqrt{1-s}}{1-\tau(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \left[P \{S_n > 0\} - \frac{1}{2} \right] \rightarrow c, \quad (7.13)$$

Отсюда следует, что

$$\frac{1-\tau(s)}{1-s} \sim e^{-c} \frac{1}{\sqrt{1-s}}. \quad (7.14)$$

В левой части можно узнать производящую функцию вероятностей из (7.12). Они монотонно убывают, и, следовательно, (7.12) верно в силу последней части тауберовой теоремы 5 из гл. XIII, 5. ►

Теорема 2. Случайное блуждание уходит в $-\infty$ тогда и только тогда, когда

$$\sum_{n=1}^{\infty} \frac{1}{n} P \{S_n > 0\} < \infty. \quad (7.15)$$

Этот критерий остается справедливым¹⁾ при замене событий $\{S_n > 0\}$ на $\{S_n \geq 0\}$.

Доказательство. Снос в $-\infty$ происходит тогда и только тогда, когда возрастающие лестничные процессы обрываются, т. е. когда $\mathcal{F}_1$ имеет несобственное распределение. Это равносильно неравенству $\tau(1) < 1$. В этом случае обе части (7.3) остаются ограниченными при $s \rightarrow 1$. Отсюда видно, что условие (7.15) является необходимым и достаточным. Это же рассуждение применимо к слабым лестничным моментам, что обосновывает заключительное утверждение теоремы. ►

¹⁾ Мы увидим, что всегда $\sum n^{-1} P \{S_n = 0\} < \infty$ [см. 9, в)].

Мы знаем, что снос в $-\infty$ имеет место, если F имеет математическое ожидание $\mu < 0$. Однако не очевидно, что неравенство $\mu < 0$ влечет за собой (7.15). Проверка этого факта представляет собой превосходное аналитическое упражнение, имеющее и методологический интерес (см. задачу 16).

Эта теорема имеет удивительные следствия.

Примеры. а) Пусть F — строго устойчивое распределение с $F(0) = \delta < 1/2$. Интуитивно можно было бы ожидать сноса в ∞ , но на самом деле случайное блуждание здесь является блужданием осциллирующего типа. Действительно, ряд (7.15) сводится к $(1-\delta) \sum n^{-1}$ и расходится. Таким образом, случайная величина $\mathcal{F}_1$ является собственной. Однако те же самые рассуждения применимы к отрицательной полуоси, и мы видим, что убывающая лестничная величина $\mathcal{F}_1^-$ тоже является собственной.

б) Пусть F обозначает симметричное распределение Коши. Рассмотрим случайное блуждание, порожденное величинами $X'_n = X_n + 1$. Медиана распределения сумм $S'_n = S_n + n$ находится в точке n , и интуитивно можно было бы ожидать сильного сноса в ∞ . На самом деле вероятность $P\{S'_n > 0\}$ снова не зависит от n , и, как в предыдущем примере, мы приходим к выводу, что случайное блуждание принадлежит осциллирующему типу. ►

Теорема 3. *Лестничный момент $\mathcal{F}_1$ имеет конечное математическое ожидание (и собственное распределение) тогда и только тогда, когда случайное блуждание уходит в ∞ . При этом*

$$\log E(\mathcal{F}_1) = \log \sum k \tau_k = \sum_{n=1}^{\infty} \frac{1}{n} P\{S_n \leq 0\}. \quad (7.16)$$

(Во всех остальных случаях ряд расходится.)

Доказательство. Вычитая обе части равенства (7.3) из выражения $\log(1-s)^{-1}$, получаем при $0 < s < 1$

$$\log \frac{1-\tau(s)}{1-s} = \sum_{n=1}^{\infty} \frac{s^n}{n} [1 - P\{S_n > 0\}]. \quad (7.17)$$

При $s \rightarrow 1$ левая часть (7.17) сходится тогда и только тогда, когда $\mathcal{F}_1$ — собственная случайная величина и имеет конечное математическое ожидание. Правая часть (7.17) стремится к правой части в (7.16). По теореме 2 сходимость имеет место тогда и только тогда, когда случайное блуждание уходит в ∞ . ►

В заключение мы покажем, что производящая функция из теоремы 1 имеет альтернативную вероятностную интерпретацию, которая приводит к замечательному закону арксинуса.

Теорема 4. *Производящая функция вероятностей*

$$p_n = P\{S_1 > 0, S_2 > 0, \dots, S_n > 0\} \quad (7.18)$$

определяется формулой

$$p(s) = \frac{1}{1 - \tau(s)} \quad (7.19)$$

или

$$\log p(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbf{P}\{S_n > 0\}. \quad (7.20)$$

Из соображений симметрии вероятности

$$q_n = \mathbf{P}\{S_1 \leq 0, \dots, S_n \leq 0\} \quad (7.21)$$

имеют производящую функцию q , определяемую формулой

$$\log q(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbf{P}\{S_n \leq 0\}. \quad (7.22)$$

(Ср. с задачей 21.)

Доказательство. Используем лемму о двойственности из § 2. Из теории рекуррентных событий ясно, что (7.19) является производящей функцией для вероятностей p_n того, что n будет лестничным моментом:

$$p_n = \mathbf{P}\{S_n > S_0, \dots, S_n > S_{n-1}\}. \quad (7.23)$$

Меняя порядок случайных величин X_j , получаем двойственное соотношение (7.18). [Этот случай содержится на самом деле в (2.1) при $I = 0, \infty$.]

§ 8. ЗАКОН АРКСИНУСА

Одна из удивительных черт случайных колебаний при бросании монеты находит свое выражение в двух законах арксинуса (1, гл. III, 4 и 8). Первый из них имеет следствием, что число положительных членов в последовательности $S_1, S_2, \dots, S_n$ с большой вероятностью будет ближе к 0 или n , чем к $n/2$ ($n/2$ соответствовало бы наивным ожиданиям). Второй закон приводит к такому же выводу относительно положения максимального члена. Мы покажем теперь, что эти законы верны для всех симметричных и многих других распределений. Это открытие доказывает общезначимость и применимость рассуждений в I; III.

В последующих формулировках нам будет мешать то, что максимум может достигаться несколько раз и что частные суммы могут обращаться в нуль. Эти возможности можно не рассматривать, если распределение F непрерывно, так как в этом случае вероятность равенства двух любых частных сумм равна нулю. (Мы советуем читателю ограничиться только этим случаем.) В общей теории мы условимся рассматривать номер *первой* мак-

симальной суммы, т. е. индекс k , для которого

$$S_k > S_0, \dots, S_k > S_{k-1}, \quad S_k \geq S_{k+1}, \dots, S_k \geq S_n. \quad (8.1)$$

Здесь n фиксировано и k пробегает значения $0, 1, \dots, n$. При некотором $k \leq n$ событие (8.1) должно произойти, так как мы можем определить (собственную) случайную величину K_n как номер первого максимума, т. е. номер, при котором верно (8.1). (Здесь $S_0 = 0$.) Для наступления (8.1) необходимо и достаточно одновременное наступление событий

$$\{S_k > S_0, \dots, S_k > S_{k-1}\} \text{ и } \{S_{k+1} - S_k \leq 0, \dots, S_n - S_k \leq 0\}.$$

Первое зависит только от случайных величин $X_1, \dots, X_k$, а второе — от величин $X_{k+1}, \dots, X_n$, поэтому они независимы. Но эти события суть события из последней теоремы, следовательно, верна

Лемма 1. Для всех k и n

$$P\{K_n = k\} = p_k q_{n-k}. \quad (8.2)$$

Допустим теперь, что $P\{S_n > 0\} = P\{S_n \leq 0\} = 1/2$ при всех n . Тогда правые части в (7.15) и (7.17) превращаются в $\frac{1}{2} \log(1-s)^{-1}$, так что

$$p(s) = q(s) = 1/\sqrt{1-s}.$$

Таким образом,

$$p_k q_{n-k} = \binom{-1/2}{k} \binom{-1/2}{n-k} (-1)^n, \quad (8.3)$$

что можно переписать в более удобной форме:

$$p_k q_{n-k} = \binom{2k}{k} \binom{2n-2k}{n-k} \frac{1}{2^{2n}}. \quad (8.4)$$

Это выражение использовалось в I; III, (4.1), для вывода дискретной формы распределения арксинуса, которое было найдено в I; III, 4 и 8, и которому подчиняются различные случайные величины, связанные с бросанием монеты; его предельная форма была выведена в I; III, (4.4). В частности, используя закон арксинуса, из I; III, 8, е), мы можем установить следующий результат.

Теорема 1. Если распределение F симметрично и непрерывно, то распределение вероятностей K_n (номера первого максимума среди $S_0, S_1, \dots, S_n$) остается тем же самым, что и в игре с бросанием монеты, и задается формулами (8.3) или (8.4). При $n \rightarrow \infty$ и фиксированном α , $0 < \alpha < 1$,

$$P\{K_n < n\alpha\} \rightarrow 2 \frac{1}{\pi} \arcsin \sqrt{\alpha}. \quad (8.5)$$

Предельное распределение имеет плотность $1/(\pi \sqrt{\alpha(1-\alpha)})$, ко-

торая неограничена вблизи точек 0 и 1 имеет минимум в точке $1/2$. Это показывает, что значения нормированного максимума K_n/n располагаются вблизи точек 0 и 1 с вероятностью, заметно большей, чем вблизи точки $1/2$. Более подробное обсуждение см. 1; гл. III, 4 и 8. Другую форму теоремы см. в задаче 22.

Эту теорему можно обобщить точно так же, как теорему 1 из § 7.

Теорема 1a¹. Если ряд

$$\sum_{n=1}^{\infty} \frac{1}{n} [P\{S_n > 0\} - 1/2] = c \quad (8.6)$$

сходится, то при $n \rightarrow \infty$ и $n-k \rightarrow \infty$

$$P\{K_n = k\} \sim \binom{2k}{k} \binom{2n-2k}{n-k} \frac{1}{2^{2n}} \quad (8.7)$$

и, следовательно, справедлив закон арксинуса (8.5).

В гл. XVIII, 5, будет показано, что ряд (8.6) сходится, если F имеет нулевое математическое ожидание и конечную дисперсию. К этим распределениям применим, следовательно, закон арксинуса,

Доказательство. Из (7.20) и элементарной теоремы Абеля для степенных рядов мы заключаем, что при $s \rightarrow 1$

$$\log(p(s) \sqrt{1-s}) = \sum_{n=1}^{\infty} \frac{s^n}{n} \left[P\{S_n > 0\} - \frac{1}{2} \right] \rightarrow c \quad (8.8)$$

и, таким образом,

$$p(s) \sim e^c \cdot (1-s)^{-1/2}. \quad (8.9)$$

По определению (7.18) вероятности p_n монотонно убывают, и, следовательно, из последней части тауберовой теоремы 5 гл. XIII, 5, вытекает, что коэффициенты обоих степенных рядов в (8.9) ведут себя одинаковым асимптотическим образом. Следовательно,

$$p_n \sim e^c \binom{-1/2}{n} (-1)^n, \quad n \rightarrow \infty. \quad (8.10)$$

Для q_n мы получаем то же самое соотношение с заменой s на $-s$, и, следовательно, утверждение (8.7) следует из (8.2). Вывод закона арксинуса основан только на асимптотическом соотношении (8.7), а не на тождестве (8.4). ►

Теорема 1 и ее доказательство переносятся на произвольные строго устойчивые распределения. Если $P\{S_n > 0\} = \delta$ не зависит от n , мы получаем из (7.20) и (7.22)

$$p(s) = (1-s)^{-\delta}, \quad q(s) = (1-s)^{\delta-1}, \quad n \rightarrow \infty \quad (8.11)$$

и, таким образом,

$$P\{K_n = k\} = p_k q_{n-k} = (-1)^n \binom{-\delta}{k} \binom{\delta-1}{n-k}. \quad (8.12)$$

Предельная теорема (8.5) имеет место, если заменить распределение аркси-

¹) Эта теорема была доказана трудоемкими вычислениями Спарре-Андерсеном. Наблюдение, что тауберова теорема устраняет все сложности, принадлежит Спицеру. Обобщение см. в 9, г).

нуса в правой части распределением с плотностью

$$\frac{\sin n\delta}{\pi} \frac{1}{x^{1-\delta} (1-x)^\delta}, \quad 0 < x < 1. \quad (8.13)$$

Теорема 1а распространяется на распределения, принадлежащие к области притяжения устойчивого распределения.

В 1; гл. III, нам пришлось доказывать оба закона арксинуса по отдельности, однако следующая теорема показывает, что они эквивалентны. Теорема 2 (для непрерывных распределений) была исходной точкой исследований Спарре-Андерсена, нашедшего новый подход к теории флуктуаций. Первоначальное доказательство было крайне сложным. Сейчас существует несколько доказательств, но приводимое здесь кажется простейшим.

Теорема 2. Число Π_n строго положительных членов последовательности $S_1, \dots, S_n$ имеет то же самое распределение (8.2), что и номер K_n первого максимального числа среди $S_0=0, S_1, \dots, S_n$.
(См. задачу 23.)

Эта теорема будет сведена к чисто комбинаторной лемме. Пусть $x_1, \dots, x_n$ — произвольные (не обязательно различные) действительные числа. Положим

$$s_0=0, \quad s_k = x_1 + \dots + x_k. \quad (8.14)$$

Максимум среди $s_0, \dots, s_n$ может достигаться несколько раз, и, следовательно, мы должны различать номера первого и последнего максимального членов.

Теперь рассмотрим $n!$ перестановок чисел $x_{i_1}, \dots, x_{i_n}$ (некоторые из них могут быть одинаковы). Каждой из перестановок мы поставим в соответствие последовательность из $n+1$ частных сумм $0, x_{i_1}, \dots, x_{i_1} + \dots + x_{i_n}$.

Пример. а) Пусть $x_1 = x_2 = 1$ и $x_3 = x_4 = -1$. Только 6 перестановок $(x_{i_1}, \dots, x_{i_n})$ различимы между собой, но каждая из них представляет четыре перестановки индексов. В перестановке $(1, 1, -1, -1)$ три частных суммы строго положительны, а (единственный) максимум появляется на третьем месте. В перестановке $(-1, -1, 1, 1)$ нет ни одной положительной частной суммы, а последняя равна нулю. Первому максимуму соответствует номер 0, последнему — 4. ►

Будет доказано, что теорема 2 является простым следствием следующей леммы.

Лемма 2. Пусть r — такое целое число, что $0 \leq r \leq n$. Число A_r перестановок, которым соответствуют в точности r строго положительных частных сумм, равно числу B_r переста-

новок, в которых первый максимум среди этих частных сумм появляется на r -м месте.

(См. задачу 24.)

Доказательство¹⁾. Будем рассуждать по индукции. Утверждение леммы верно для $n=1$, так как если $x_1 > 0$, то $A_1 = B_1 = 1$ и $A_0 = B_0 = 0$, а если $x_1 \leq 0$, то $A_1 = B_1 = 0$ и $A_0 = B_0 = 1$. Предположим, что лемма верна, если заменить n на $n-1 \geq 1$. Обозначим через $A_r^{(k)}$ и $B_r^{(k)}$ числа, соответствующие A_r и B_r при замене строки $(x_1, \dots, x_n)$ на строку из n чисел x , полученную выбрасыванием x_k . Тогда по предположению индукции $A_r^{(k)} = B_r^{(k)}$ при $1 \leq k \leq n$ и $r = 0, \dots, n-1$. Это верно также при $r = n$, поскольку тривиально $A_n^{(k)} = B_n^{(k)} = 0$.

а) Предположим, что $x_1 + \dots + x_n \leq 0$. Все $n!$ перестановок $(x_1, \dots, x_n)$ можно получить выбором элемента x_k на последнем месте и перестановкой оставшихся $n-1$ элементов. Поскольку n -я частная сумма ≤ 0 , ясно, что число положительных частных сумм и номер первого максимального числа зависят только от первых $n-1$ элементов. Таким образом,

$$A_r = \sum_{k=1}^n A_r^{(k)}, \quad B_r = \sum_{k=1}^n B_r^{(k)}, \quad (8.15)$$

и, следовательно, $A_r = B_r$ по предположению индукции.

б) Предположим, что $x_1 + \dots + x_n > 0$. Тогда n -я частная сумма положительна, и из предыдущих рассуждений ясно, что в этом случае

$$A_r = \sum_{k=1}^r A_{r-k-1}^{(k)}. \quad (8.16)$$

Чтобы получить аналогичную рекуррентную формулу для B_r , рассмотрим перестановки $(x_k, x_1, \dots, x_{j_{n-1}})$, начинающиеся с элемента x_k . Поскольку n -я частная сумма положительна, максимальные члены последовательности частных сумм имеют положительные индексы. Ясно, что первый максимум появляется на месте r ($1 \leq r \leq n$) тогда и только тогда, когда первый мак-

¹⁾ Это доказательство принадлежит А. У. Джозефу (A. W. Joseph) из Бирмингема (Англия). Крайняя простота доказательства может шокировать, если припомнить, что в 1949 г. сенсационное открытие Спарре-Андерсеном теоремы 2 было встречено с недоверием, а его первоначальное доказательство было крайне запутанным и сложным. Автор свел теорему к чисто комбинаторной лемме 2 и дал [ее] элементарное доказательство (см. первое издание настоящей книги). Доказательство Джозефа не только проще, но является первым конструктивным доказательством, устанавливающим взаимно однозначное соответствие между двумя типами перестановок. Идея применения этого приема в лемме 3 принадлежит М. Т. Л. Бизли (M. T. L. Bizley) из Лондона (Англия). Автор благодарен гг. Джозефу и Бизли за разрешение использовать их неопубликованные результаты (они были сообщены мне, когда рукопись была уже в печати).

симум частных сумм для $(x_{j_1}, \dots, x_{j_{n-1}})$ появляется на месте $r-1$. Таким образом,

$$B_r = \sum_{k=1}^r B_{r-1}^{(k)}. \quad (8.17)$$

Сравнение (8.16) и (8.17) снова показывает, что $A_r = B_r$, и это завершает доказательство. ►

Вскоре мы увидим, что проведенные рассуждения приводят к дальнейшим результатам, но сначала мы вернемся к доказательству теоремы 1.

Доказательство теоремы 1. Мы поступим так же, как при доказательстве теоремы 1 из § 7. Рассмотрим $n!$ перестановок $(x_{i_1}, \dots, x_{i_n})$ и перенумеруем их так, чтобы перестановка с естественным порядком $(x_1, \dots, x_n)$ имела номер 1. Для фиксированного целого числа r , $0 \leq r \leq n$, положим $Y^v = 1$, если перестановка с номером v имеет ровно r положительных частных сумм, и положим $Y^v = 0$ в противном случае. По соображениям симметрии $n!$ случайных величин $Y^{(v)}$ имеют одинаковое распределение и, следовательно,

$$P\{\Pi_n = r\} = P\{Y^{(1)} = 1\} = E(Y^{(1)}) = \frac{1}{n!} \sum E(Y^{(v)}). \quad (8.18)$$

Аналогично

$$P\{K_n = r\} = \frac{1}{n!} \sum E(Z^{(v)}), \quad (8.19)$$

где $Z^{(v)} = 1$, если в перестановке с номером v первая максимальная частная сумма имеет индекс r , и $Z^{(v)} = 0$ в противном случае. В силу леммы 2 суммы $\sum Y^{(v)}$ и $\sum Z^{(v)}$ равны и, следовательно, вероятности в (8.18) и (8.19) одинаковы. ►

Замечание о преобразованиях Спарре-Андерсена. Из леммы 2 следует, что существует такое преобразование, что каждая строка $(x_{i_1}, \dots, x_{i_n})$ из n действительных чисел отображается в перестановку $(x_{i_1}, \dots, x_{i_n})$ таким образом, что: (i) если среди частных сумм s_k в (8.14) ровно r ($0 \leq r \leq n$) строго положительных, то максимальная из частных сумм для $(x_{i_1}, \dots, x_{i_n})$ появляется в первый раз с индексом r ; и (ii) преобразование обратимо (или взаимно однозначно). Такие преобразования будут названы именем Спарре-Андерсена, хотя он имел дело с независимыми случайными величинами, не зная о возможности свести теорему 2 к чисто комбинаторной лемме 2. При внимательном изучении доказательства леммы 2 обнаруживается, что оно неявно содержит способ построения преобразования Спарре-Андерсена. Процедура является рекуррентной, причем первый шаг задается правилом: если $s_n \leq 0$, оставляем $(x_{i_1}, \dots, x_{i_n})$ без изменений, но

если $s_n > 0$, заменяем $(x_1, \dots, x_n)$ на циклическую перестановку $(x_n, x_1, \dots, x_{n-1})$. Следующий шаг состоит в применении того же самого правила к $(x_1, \dots, x_{n-1})$. Желаемая перестановка $(x_{i_1}, \dots, x_{i_n})$ получается через $n-1$ шагов.

Примеры. б) Пусть $(x_1, \dots, x_6) = (-1, 2, -1, 1, 1, -2)$. На первом шаге не происходит никаких изменений, но на втором шаге приходим к $(1, -1, 2, -1, 1, -2)$. Поскольку $s_4 = 1$, третий шаг приводит к $(1, 1, -1, 2, -1, -2)$, а четвертый шаг не вносит изменений, поскольку $s_3 = 0$. Так как $s_2 = 1$, последний шаг приводит к перестановке $(1, 1, 2, -1, -1, -2)$. Единственная максимальная частная сумма появляется на третьем месте, и в первоначальной перестановке в точности три частные суммы положительны.

в) Предположим, что $x_j \leq 0$ при всех j . Начальная и последняя перестановки x_j одинаковы. Ни одна частная сумма не является положительной, и сумма $s_0 = 0$ представляет собой максимум (который повторяется, если $x_1 = 0$). ►

Предпочтительнее заменить рекуррентное построение непосредственным описанием конечного результата, что мы сделаем в следующей лемме. Мы даем здесь новое доказательство, не опирающееся на предыдущую лемму и представляющее самостоятельный интерес (см. также задачу 24).

Лемма 3. Пусть $(x_1, \dots, x_n)$ — перестановка действительных чисел, такая, что частные суммы $s_{v_1}, \dots, s_{v_r}$ положительны, а все другие отрицательны или равны нулю; здесь $v_1 > v_2 > \dots > v_r > 0$. Запишем $x_{v_1}, \dots, x_{v_r}$, а за ними оставшиеся x_j в их первоначальном порядке. (Если все частные суммы ≤ 0 , то $r = 0$ и порядок остается неизменным.) Первый из максимумов частных сумм в новой перестановке появляется на r -м месте, а определенное таким образом преобразование является взаимно однозначным.

Доказательство. Обозначим новую перестановку через $(\xi_1, \dots, \xi_n)$, а частные суммы элементов через $\sigma_0, \dots, \sigma_n$. Каждому индексу $j \leq n$ соответствует единственный индекс k , такой, что $\xi_j = x_k$. В частности, $\xi_1, \dots, \xi_r$ совпадают с $x_{v_1}, \dots, x_{v_r}$, взятыми в указанном порядке.

Сначала рассмотрим такое j , что $s_k \leq 0$. Из построения ясно, что $j \geq k$ и элементы $\xi_{j-k+1}, \dots, \xi_j$ представляют собой результат перестановки $x_1, \dots, x_k$. Таким образом, $\sigma_j = \sigma_{j-k} + s_k \leq \sigma_{j-k}$, и, значит, σ_j не может быть первой максимальной частной суммой.

Если $r = 0$, то отсюда следует, что первый максимум среди σ_j появляется при $j = 0$. Когда $r > 0$, первый максимум появляется на одном из мест $0, 1, \dots, r$, и мы покажем, что возможно только место r . Действительно, если $j \leq r$, то из построения сле-

дует, что v_j элементов $\xi_j, \dots, \xi_{j-1+v_j}$ совпадают с некоторой перестановкой элементов $(x_i, \dots, x_{v_j})$. Таким образом, $\sigma_{j-1+v_j} = \sigma_{j-1} + s_{v_j} > \sigma_{j-1}$, и, следовательно, максимум не может появиться на месте с номером $j-1$.

Для завершения доказательства леммы нам остается показать, что полученное преобразование взаимно однозначно. На самом деле обратное для $\xi_i, \dots, \xi_n$ преобразование может быть построено по следующему рекуррентному правилу. Если все $\sigma_j \leq 0$, оставьте перестановку без изменений. В противном случае пусть k — наибольший индекс, такой, что $\sigma_k > 0$. Замените $(\xi_i, \dots, \xi_n)$ на $(\xi_2, \dots, \xi_k, \xi_1, \xi_{k+1}, \dots, \xi_n)$ и примените ту же процедуру к $(\xi_2, \dots, \xi_k)$. ►

Замечание о симметрично зависимых величинах. Уместно заметить, что доказательство не предполагает независимости величин X_j . Нужно лишь, чтобы все $n!$ перестановок $(X_{i_1}, \dots, X_{i_n})$ были одинаково распределены. Другими словами, теорема 2 остается справедливой для каждой системы n симметрично зависимых случайных величин (см. гл. VII, 4). Конечно, общее распределение K_n и Π_n будет зависеть от совместного распределения X_j . Интересен следующий пример. Пусть $X_1, X_2, \dots$ независимы и имеют одно и то же распределение F . Положим $Y_k = X_k - S_n/n$, $k=1, \dots, n$. Величины $Y_1, \dots, Y_n$ симметрично зависимы, и их частные суммы равны

$$\Sigma_k = S_k - \frac{k}{n} S_n, \quad k=1, \dots, n-1. \quad (8.20)$$

На графике $(S_0, S_1, \dots, S_n)$ величина Σ_k равна вертикальному отклонению вершины S_k от хорды, соединяющей начало координат и конечную точку (n, S_n) .

Допустим теперь, что распределение F непрерывно (с тем, чтобы избежать необходимости различать первый и последний максимумы). С вероятностью единица последовательность $0, \Sigma_1, \dots, \Sigma_{n-1}$ имеет единственный максимум. Циклической перестановке $(Y_2, \dots, Y_n, Y_1)$ соответствуют частные суммы $0, \Sigma_2 - \Sigma_1, \dots, \Sigma_{n-1} - \Sigma_1, -\Sigma_1$, и ясно, что положение максимума смещается при этом на одно место вперед в циклическом порядке (если первоначальный максимум был на нулевом месте, то $\Sigma_k < 0$ при $k=1, \dots, n-1$, а новый максимум будет на $(n-1)$ -м месте). В n циклических перестановках максимум будет появляться ровно один раз на каждом месте, и его положение равномерно распределено на множестве $\{0, 1, \dots, n-1\}$. Мы получаем таким образом следующую теорему, принадлежащую Спарре-Андерсену и связанную с теоремой 3 из I; гл. III, 9, касающейся бросаний монеты.

Теорема 3. В любом случайном блуждании, порождаемом непрерывным распределением F , и при любом n число вершин среди $S_1, \dots, S_{n-1}$, которые лежат выше хорды, соединяющей $(0, 0)$ и (n, S_n) , равномерно распределено на множестве $0, 1, \dots, n-1$.

(Это же верно и для номера максимально удаленной от хорды вершины.)

§ 9. РАЗЛИЧНЫЕ ДОПОЛНЕНИЯ

а) Совместные распределения

Для того чтобы получить совместное распределение лестничных величин, достаточно только изменить обозначения в рассуждениях, приводящих к теореме 1 из § 7. Приспосабливая обо-

значения § 1, обозначим через I интервал в $\overline{0, \infty}$ и через $H_n^{(r)}(I)$ вероятность того, что n совпадает с r -м лестничным моментом и $S_n \in I$. Положим

$$H\{I, s\} = \sum_{n=1}^{\infty} s^n H_n\{I\}, \quad 0 \leq s \leq 1. \quad (9.1)$$

По индукции можно установить, что при фиксированном s

$$H^{r*}\{I, s\} = \sum_{n=1}^{\infty} s^n H_n^{(r)}\{I\}. \quad (9.2)$$

Из рассуждений, приведших к установлению (7.3), легко получаем следующий результат Г. Бакстера, который сводится к (7.3) при $I = \overline{0, \infty}$.

Теорема. При $I \subset \overline{0, \infty}$ и $0 \leq s \leq 1$

$$\sum_{r=1}^{\infty} \frac{1}{r} H^{r*}\{I, s\} = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n \in I\}. \quad (9.3)$$

Более простая и доступная для изучения форма будет выведена в гл. XVIII, 3.

б) Интерпретация производящих функций с помощью «исчезновения»¹⁾

Указываемая здесь интерпретация может помочь интуиции и упростить формальные вычисления. Рассмотрим при фиксированном s , $0 < s < 1$, *несобственное* случайное блуждание, которое может оборваться на каждом шаге с вероятностью $1-s$, а в противном случае определяется распределением sF . Тогда $s^n F^{n*}\{I\}$ равно вероятности попасть в I в момент n . Дефект $1-s^n$ равен вероятности «исчезновения» процесса до момента n . Здесь проходят все прежние рассуждения с той разницей, что все распределения становятся несобственными. В частности, (9.1) представляет собой распределение первой лестничной высоты в нашем случайном блуждании с исчезновением. Функция (9.2) аналогична H^{r*} (см. § 2, 3). Производящая функция $\tau(s)$ равна вероятности появления лестничного индекса.

в) Рекуррентное событие

$$\{S_1 \leq 0, \dots, S_{n-1} \leq 0, S_n = 0\} \quad (9.4)$$

означает возвращение в нуль без предварительного захода на правую полуось. Оно появлялось в § 1 в определении слабых

¹⁾ В оригинале «mortality» — смертность. — Прим. перев.

лестничных величин. Обозначим ω_n вероятность первого наступления события (9.4) в момент n :

$$\omega_n = P\{S_1 < 0, \dots, S_{n-1} < 0, S_n = 0\}. \quad (9.5)$$

Если $\omega(s) = \sum \omega_r s^r$, то ω' будет производящей функцией для r -го наступления нашего события, а $1/[1 - \omega(s)]$ будет производящей функцией для вероятностей (9.4). Упрощенный вариант доказательства (7.3) приводит к основному тождеству

$$\log \frac{1}{1 - \omega(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n = 0\}. \quad (9.6)$$

Сравнивая его с (7.3), (7.16), (7.22) и т. д., мы видим, как легко можно перейти от слабых лестничных величин к строгим и обратно. Формула (9.6) подтверждает замечание § 1, что вероятности (9.4) не меняются, если все неравенства заменить противоположными.

г) Обобщение на произвольные интервалы

Теория § 3 с тривиальными изменениями в обозначениях применима в ситуации, когда полупрямая $\overline{0, \infty}$ заменяется произвольным интервалом A , а полупрямая $-\infty, \overline{0}$ — дополнением A' к A . Сохраняется, в частности, без изменений интегральное уравнение Винера—Хопфа. Читателю предлагается проследить за деталями; полностью это сделано в гл. XVIII, 1 (см. также задачу 15).

§ 10. ЗАДАЧИ

1. Рассмотрим биномиальное случайное блуждание (пример 2, б). Пусть e_k — математическое ожидание числа номеров $n \geq 0$, таких, что $S_n = k$, $S_1 \geq 0, \dots, S_{n-1} \geq 0$ (это событие означает попадание в k без предварительного захода на отрицательную полуось). Обозначим через f вероятность достижения вначания -1 , т. е. $f = 1$, если $q \geq p$ и $f = q/p$ в противном случае. Взяв за новое начало координат точку $(1, 1)$, докажем, что $e_0 = 1 + p f e_0$ и $e_k = p(e_{k-1} + f e_k)$ при $k \geq 1$. Приходим к выводу, что при $k \geq 0$

$$e_k = p^{-1}, \text{ если } p \geq q, \quad e_k = (p/q)^k q^{-1}, \quad \text{если } p \leq q.$$

2. Продолжение. Пусть a_k при $k \geq 1$ — математическое ожидание числа индексов $n \geq 1$, таких, что $S_n = k$, $S_1 > 0, \dots, S_{n-1} > 0$ (это событие означает попадание в k , предшествующее первому возвращению в начало координат). Показать, что $a_k = p e_{k-1}$ и, следовательно,

$$a_k = 1, \text{ если } p \geq q, \quad a_k = (p/q)^k, \text{ если } p \leq q.$$

Это дает прямое доказательство парадоксального результата примера 2, б).

Примечание. Нижеследующие задачи 3—6 могут служить введением к настоящей главе и могут быть решены до ее изучения. Они содержат также примеры явных решений основных интегральных уравнений. Сверх того, они показывают силу и красоту метода производящих функций [попробуйте решить уравнение (1) непосредственно!].

3. Случайные величины X_k , порождающие случайное блуждание, имеют одно и то же арифметическое распределение, приписывающее вероятности $f_1, f_2, \dots$ целым числам $1, 2, \dots$ и вероятность q числу -1 ($q + f_1 + f_2 + \dots = 1$). Обозначим через λ_r ($r = 1, 2, \dots$) вероятность того, что первый положительный член последовательности $S_1, S_2, \dots$ имеет величину r (иными словами $\{\lambda_r\}$ есть распределение первой лестничной высоты). Покажите, что

(а) Числа λ_r удовлетворяют рекуррентным соотношениям

$$\lambda_r = f_r + q(\lambda_{r+1} + \lambda_1 \lambda_r). \quad (1)$$

(б) Производящие функции связаны равенством

$$\lambda(s) = 1 - \frac{f(s) + qs^{-1} - 1}{\lambda_1 q + qs^{-1} - 1}, \quad 0 < s < 1. \quad (2)$$

(в) Если $E(X_k) = \mu = f'(1) - q > 0$, то существует единственный корень $0 < \sigma < 1$ уравнения

$$f(s) + \frac{q}{s} = 1. \quad (3)$$

Из того что λ — монотонная функция и $\lambda < 1$ на $\overline{0, 1}$, выводится, что

$$\lambda(s) = s \frac{\sigma f(s) - f(\sigma)}{q s - \sigma}. \quad (4)$$

Это эквивалентно

$$\lambda_r = \frac{[f_r \sigma + f_{r+1} \sigma^2 + \dots]}{q}. \quad (5)$$

(г) Если $E(X_k) < 0$, то соответствующее решение получаем, полагая в (2) $\lambda_1 = (1-q)/q$. При этом (4) и (5) выполняются с $\sigma = 1$.

4. Решите предыдущую задачу (изменив ее там, где нужно) для слабых лестничных высот. Другими словами, рассмотрите вместо λ_r вероятность того, что первый из неотрицательных членов последовательности $S_1, S_2, \dots$ принимает значение r ($r = 0, 1, \dots$). Покажите, что (1) и (4) следует заметить на

$$\gamma_r = f_r + \frac{q}{1 - \gamma_0} \gamma_{r+1} \quad (1a)$$

и

$$\gamma(s) = 1 - \frac{q}{\sigma} + s \frac{f(s) - f(\sigma)}{s - \sigma}. \quad (4a)$$

5. Пусть в случайном блуждании задачи 3 x обозначает вероятность того, что $S_n < 0$ при некотором n . Покажите (независимо от задачи 3), что x удовлетворяет уравнению (3) и что, следовательно, $x = \sigma$.

6. Продолжение. Покажите, что вероятность выполнения при каком-нибудь $n > 0$ неравенства $S_n \leq 0$ равна $q + f(\sigma) = 1 - q(\sigma^{-1} - 1)$. Проверьте, что $\lambda'(1) = \mu \sigma [q(1 - \sigma)]^{-1}$. Это — частный случай (2.8) (или тождество Вальда).

7. Выведите (1.13) из (1.12) непосредственным вычислением.

8. Вероятности достижения. Пусть $t \geq 0$ и $\xi > 0$. Обозначим $G(t, \xi)$ вероятность того, что первая сумма S_n , превосходящая t , не будет превышать $t + \xi$. Докажите, что G удовлетворяет интегральному уравнению

$$G(t, \xi) = F(t + \xi) - F(t) + \int_{-\infty}^{t+\xi} G(t-y, \xi) F(dy).$$

В случае неединственности G является минимальным решением. Распределение H лестничной высоты однозначно определяется формулой $H(\xi) = G(0, \xi)$.

9. Пусть H — непрерывное распределение, сосредоточенное на $\overline{0, \infty}$, а H^- — (возможно, несобственное) непрерывное распределение, сосредоточенное на $-\infty, 0$. Предположим, что

$$H + H^- - H \star H^- = F \quad (6)$$

есть распределение вероятностей. Из теоремы единственности § 3 следует, что H и H^- являются распределениями точек первого попадания $\mathcal{H}$ и $\mathcal{H}^-$ в случайном блуждании, порожденном F . Таким способом возможно найти распределения F , допускающие явное представление в виде (6). Это имеет место в том случае, когда $0 < q \leq 1$ и H и H^- имеют плотности, определяемые либо формулой

$$(a) \quad be^{-bx} \text{ при } x > 0, \quad q \text{ при } -1 < x < 0,$$

либо формулой

$$(b) \quad b^{-1} \text{ при } 0 < x < b, \quad q \text{ при } -1 < x < 0,$$

В случае (а) распределение F имеет плотность, равную

$$(b-q)e^{-bx} + qe^{-b(x+1)} \text{ при } x > 0 \text{ и } qe^{-b(x+1)} \text{ при } -1 < x < 0,$$

В случае (б), если $b > 1$, плотность F равна $qb^{-1}(1+x)$ при $-1 < x < 0$, qb^{-1} при $0 < x < b-1$ и $qb^{-1}(b-x)$ при $b-1 < x < b$. В любом из этих двух случаев F имеет нулевое математическое ожидание тогда и только тогда, когда $q=1$.

10. Покажите, исходя из (3.11), что если распределения H и H^- собственные и имеют дисперсии, то $E(X_1)=0$ и $E(X_1^2)=-2E(\mathcal{H}_1)E(\mathcal{H}_1^-)$.

11. Аналитическое доказательство тождества Вальда (2.8). Выведите из (3.11), что

$$1-F(x) = [1-\rho(0)][1-H(x)] + \int_{-\infty}^{0+} \rho\{dy\} [H(x-y)-H(x)],$$

$$F(x) = \int_{-\infty}^x \rho\{dy\} [1-H(x-y)]$$

при $x > 0$ и $x < 0$ соответственно. Установите, что F имеет положительное математическое ожидание μ тогда и только тогда, когда H имеет конечное математическое ожидание ν и $\rho(0) < 1$. Выведите интегрированием на $-\infty, \infty$, что $\mu = [1-\rho(0)]\nu$ (что эквивалентно 2.8)).

12. К примеру 4, в). Если $\mu > 0$, то знаменатель имеет положительный корень $s_0 < 1$, ровно $b-1$ комплексных корней в круге $|s| < s_0$ и $a-1$ комплексных корней в области $|s| > 1$.

Случай $\mu < 0$ можно описать заменой s на $1/s$.

13. Производящая функция для верхней лестничной высоты в примере 4, в) равна

$$\chi(s) = 1 - (1-s) \left(1 - \frac{s}{\sigma_1}\right) \dots \left(1 - \frac{s}{\sigma_{a-1}}\right).$$

Для нижних лестничных высот результат получается заменой s/σ_k на σ_k/s .

14. К примеру 4, в). Допустим, что X_j принимают значения $-2, -1, 0, 1, 2$, с вероятностью $1/5$ каждое. Покажите, что верхняя лестничная высота имеет распределение, для которого

$$\lambda_1 = \frac{1 + \sqrt{5}}{3 + \sqrt{5}}, \quad \lambda_2 = \frac{2}{3 + \sqrt{5}}.$$

Для слабых высот $\bar{\lambda}_0 = \frac{1}{10}(7 - \sqrt{5})$, $\bar{\lambda}_1 = \frac{1}{10}(1 + \sqrt{5})$, $\bar{\lambda}_2 = 1/5$.

15. Обозначим в примере 4, в) через $\psi_k^{(n)}$ вероятность того, что первые n шагов не выводят из интервала $\overline{-B, A}$, а n -й шаг приводит в состояние k (таким образом, $\psi_k^{(n)} = 0$ при $k > A$ и $k < -B$; как обычно, $\psi_k^{(0)}$ равно 1 при $k=0$ и равно 0 в других случаях). Пусть $\psi_k = \sum \psi_k^{(n)}$ — математическое ожидание числа попаданий в k до первого выхода из $\overline{-B, A}$. Покажите, что

$$\psi_k = \sum_{v=-B}^A \psi_v f_{k-v} + \psi_k^{(0)}, \quad -B \leq k \leq A,$$

и что при $k > A$ и $k < -B$ величина

$$\rho_k = \sum_{v=-B}^A \psi_v f_{k-v}$$

равна вероятности того, что *первый выход* из интервала $\overline{-B, A}$ приводит в точку k . [Эта задача важна для последовательного анализа. Она иллюстрирует ситуацию, описанную в 9, г).]

16. Из теоремы 2 § 7 вытекает, что при $\mu < 0$ $\sum n^{-1} P\{S_n > 0\} < \infty$. Проведите следующее *прямое доказательство*. Достаточно показать (вспомните неравенство Чебышева), что

$$\sum_{|y| > n} \int F\{dy\} < \infty, \quad \sum_{-n}^n \frac{1}{n^2} \int y^2 F\{dy\} < \infty.$$

Первое утверждение очевидно. Для доказательства второго разобьем интеграл в сумму интегралов по областям $k-1 < |y| \leq k$, $k=1, \dots, n-1$. Меняя порядок суммирования, видим, что весь ряд $< 2E(|X|)$.

17. Покажите, что при бросании симметричной монеты

$$\sum \frac{s^n}{n} P\{S_n = 0\} = \log \frac{2}{1 + \sqrt{1-s^2}}.$$

Указание: представьте левую часть как интеграл от 0 до s от функции $[(1-x^2)^{-1/2} - 1]x^{-1}$.

18. Предположим, что случайное блуждание неозвратно, т. е. что $U\{I\} = \sum_0^\infty \tilde{F}^{n*}\{I\} < \infty$ для любого конечного интервала. Положим в обозначениях § 3 $\Phi = \sum_0^\infty \rho^{n*}$. Докажите справедливость уравнения восстановления

$$U = \Phi + U \star H.$$

Если ψ^- — аналог ψ для отрицательной полуоси, то $\psi^- = (1-\xi)\Phi$, как и в (1.13).

19. Выведите, что

$$U = \frac{1}{1-\xi} \psi \star \psi^-,$$

и покажите, что это эквивалентно разложению Винера — Хопфа (3.12).

20. Выведите тождество Вальда $E(\mathcal{H}_1) = E(\mathcal{F}_1)E(X_1)$ непосредственно из уравнения восстановления задачи 18.

21. К теореме 4 из § 7. Вероятности $\rho_n^* = P\{S_1 \geq 0, \dots, S_n \geq 0\}$ и $q_n^* = P\{S_1 < 0, \dots, S_n < 0\}$ имеют производящие функции, удовлетворяющие

соотношениям

$$\log p^*(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n \geq 0\} \text{ и } \log q_n^* = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n < 0\}.$$

22. О последнем максимуме. Вместо случайной величины K_n из § 8 рассмотрим номер K_n^* последнего максимума среди частных сумм $S_0, \dots, S_n$. Используя обозначения предыдущей задачи, докажите, что

$$P\{K_n^* = k\} = p_k^* q_{n-k}^*.$$

23. Альтернативная форма теоремы 2 из § 8. Число Π_n^* неотрицательных членов последовательности $S_0, \dots, S_n$ имеет такое же распределение, как и случайная величина K_n^* из предыдущей задачи. Доказать это, применив теорему 2 к $(-X_n - X_{n-1}, \dots, -X_1)$.

24. Комбинаторная лемма 2 из § 8 остается верной, если первый максимум среди $S_0, \dots, S_n$ заменить на последний максимум, а число положительных частных сумм — на число неотрицательных членов в последовательности $S_1, \dots, S_n$ (за исключением $S_0 = 0$). Доказательство остается тем же самым, за исключением очевидной замены знаков неравенств. Точно таким же образом на рассмотренный случай переносится лемма 3.

Преобразование Лапласа служит мощным практическим оружием. В то же самое время соответствующая теория интересна и сама по себе, и по связи ее с другими разделами, например с теорией полугрупп. Теорема о вполне монотонных функциях и основная тауберова теорема с полным правом рассматривались как жемчужины математического анализа. (Хотя приводимые доказательства просты и элементарны, первоначальные исследования в этом направлении требовали оригинальности и силы.) Резольвенты (§ 9, 10) являются основой теории полугрупп.

Материал этой главы предназначен для использования в различных целях. Поэтому были приложены значительные усилия, во-первых, чтобы сделать части настолько независимыми друг от друга, насколько это позволяет тема, и, во-вторых, чтобы дать читателю возможность опускать детали. Для дополнительного чтения может служить гл. XIV. Там же можно найти примеры. Остальная часть книги совершенно не зависит от настоящей главы.

Несмотря на то что нам часто будут встречаться правильно меняющиеся функции, мы будем пользоваться только совсем элементарной теоремой 1 из гл. VIII, 8.

§ 1. ОПРЕДЕЛЕНИЯ. ТЕОРЕМА НЕПРЕРЫВНОСТИ

Определение 1. Пусть F — собственное или несобственное распределение вероятностей, сосредоточенное на $\overline{0, \infty}$. Преобразованием Лапласа распределения F называют функцию φ , определенную для $\lambda \geq 0$ равенством

$$\varphi(\lambda) = \int_0^{\infty} e^{-\lambda x} F\{dx\}. \quad (1.1)$$

Здесь и далее подразумевается, что интервал интегрирования замкнут. Когда мы говорим о преобразовании Лапласа распределения F , мы неявно предполагаем, что F сосредоточено на

$0, \infty$. Как обычно, мы расширяем смысл терминов и говорим о «преобразовании Лапласа случайной величины X », понимая под этим преобразование Лапласа соответствующего распределения. Применяя привычное обозначение для математического ожидания, имеем

$$\varphi(\lambda) = E(e^{-\lambda X}). \quad (1.2)$$

Пример. а) Пусть X принимает значения $0, 1, \dots$ с вероятностями $p_0, p_1, \dots$. Тогда $\varphi(\lambda) = \sum p_n e^{-n\lambda}$. Так как производящая функция равна $P(s) = \sum p_n s^n$, то $\varphi(\lambda) = P(e^{-\lambda})$, и преобразование Лапласа отличается от производящей функции только заменой переменных $s = e^{-\lambda}$. Этим объясняется большое сходство свойств преобразований Лапласа и производящих функций.

б) *Гамма-распределение* с плотностью $f_\alpha(x) = (x^{\alpha-1}/\Gamma(\alpha))e^{-x}$ имеет преобразование Лапласа

$$\varphi_\alpha(\lambda) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-(\lambda+1)x} x^{\alpha-1} dx = \frac{1}{(\lambda+1)^\alpha}, \quad \alpha > 0. \quad (1.3)$$

Следующая теорема показывает, что распределение восстанавливается по соответствующему преобразованию: без этого польза преобразований Лапласа была бы ограниченной.

Теорема 1. (Единственность.) *Различным распределениям вероятностей соответствуют различные преобразования Лапласа.*

Первое доказательство. В (6.4) гл. VIII дана в явном виде формула обращения, которая позволяет вычислять F по ее известному преобразованию Лапласа. Эта формула будет выведена заново в § 4.

Второе доказательство. Положим $y = e^{-x}$. При изменении x от 0 до ∞ y меняется от 1 до 0. Полагая $G(y) = 1 - F(x)$ в точках непрерывности, получим распределение G , сосредоточенное на $[0, 1]$. Тогда

$$\varphi(\lambda) = \int_0^\infty e^{-\lambda x} F\{dx\} = \int_0^1 y^\lambda G\{dy\}; \quad (1.4)$$

это — очевидное следствие того, что римановы суммы $\sum e^{-\lambda x_k} [F(x_{k+1}) - F(x_k)]$ совпадают с римановыми суммами $\sum y_k^\lambda [G(y_k) - G(y_{k+1})]$ при $y_k = e^{-x_k}$. Мы знаем из гл. VII, 3, что распределение G однозначно определяется своими моментами. Эти моменты равны $\varphi(k)$. Поэтому значения $\varphi(1), \varphi(2), \dots$ опре-

деляют G , а вместе с тем и F . Этот результат сильнее, чем утверждение теоремы¹⁾.

Следующий ниже важный результат просто выводится из теоремы 1.

Теорема 2. (Теорема непрерывности.) Пусть F_n , $n = 1, 2, \dots$, — распределение вероятностей с преобразованием φ_n .

Если $F_n \rightarrow F$, где F — (возможно, несобственное) распределение с преобразованием φ , то $\varphi_n(\lambda) \rightarrow \varphi(\lambda)$ при $\lambda > 0$.

Обратно, если последовательность $\{\varphi_n(\lambda)\}$ сходится при каждом $\lambda > 0$ к пределу $\varphi(\lambda)$, то φ — преобразование (возможно, несобственного) распределения F и $F_n \rightarrow F$.

Предел F будет собственным в том и только том случае, когда $\varphi(\lambda) \rightarrow 1$ при $\lambda \rightarrow 0$.

Доказательство. Первая часть содержится в основной теореме о сходимости гл. VIII, 1. Для доказательства второй части привлечем теорему о выборе (теорема 1 из гл. VIII, 6). Пусть $\{F_{n_k}\}$ — последовательность, сходящаяся к (возможно, несобственному) распределению F . В силу первой части теоремы соответствующие преобразования сходятся к преобразованию Лапласа для F . Отсюда следует, что F является тем единственным распределением, которое имеет преобразование Лапласа φ . Поэтому все сходящиеся подпоследовательности сходятся к одному и тому же пределу F . Отсюда вытекает, что F_n сходится к F . Последнее утверждение ясно из (1.1).

Для ясности изложения мы сохраним всюду, где это возможно, букву F для обозначения распределения вероятностей. Вместо (1.1) мы могли бы взять более общий интеграл вида

$$\omega(\lambda) = \int_0^{\infty} e^{-\lambda x} U \{dx\}, \quad (1.5)$$

где мера U приписывает конечную массу $U \{I\}$ конечному интервалу I , но может приписывать бесконечную массу положительной полуоси. Как обычно, нам удобно описывать эту меру в терминах несобственной функции распределения, заданной равенством

$U(x) = U \{0, x\}$. В наиболее важном частном случае, когда U определяется как интеграл от функции $u \geq 0$, интеграл (1.5) сводится к интегралу

$$\omega(\lambda) = \int_0^{\infty} e^{-\lambda x} u(x) dx. \quad (1.6)$$

¹⁾ Более общим образом, вполне монотонная функция однозначно определяется своими значениями, вычисленными по последовательности точек $\{a_n\}$, таких, что ряд $\sum a_n^{-1}$ расходится. Однако если ряд сходится, то существуют две различные вполне монотонные функции, значения которых совпадают во всех точках a_n . Элементарное доказательство соответствующей теоремы см. у Феллера: On Müntz' theorem and completely monotone functions, Amer. Math. Monthly, vol. 75 (1968), 342—350.

Примеры. в) Если $u(x) = x^a$ с $a > -1$, то $\omega(\lambda) = \Gamma(a+1)/\lambda^{a+1}$ для всех $\lambda > 0$.

г) Если $u(x) = e^{ax}$, то $\omega(\lambda) = 1/(\lambda - a)$ для $\lambda > a > 0$. Интеграл (1.6) расходится для $\lambda \leq a$.

д) Если $u(x) = e^{x^2}$, то интеграл (1.6) расходится всюду.

е) Дифференцируя в (1.1), получаем

$$-\varphi'(\lambda) = \int_0^{\infty} e^{-\lambda x} x F\{dx\}, \quad (1.7)$$

что представляет собой интеграл вида (1.5) с $U\{dx\} = xF\{dx\}$. Этот пример демонстрирует, как интегралы вида (1.5) возникают естественным образом в связи с собственными распределениями вероятностей. ►

Мы будем интересоваться главным образом мерами U , полученными простыми преобразованиями из распределений вероятностей, и интеграл в (1.5) будет, как правило, сходящимся при всех $\lambda > 0$. Однако, исключая меры, для которых сходимость имеет место только при некотором λ , мы не получаем никакого выигрыша. Далее из $\omega(a) < \infty$ вытекает $\omega(\lambda) < \infty$ для всех $\lambda > a$, так что значения λ , для которых интеграл (1.5) сходится, заполняют промежутки a, ∞ .

Определение 2. Пусть U — мера, сосредоточенная на $[0, \infty)$. Если интеграл (1.5) сходится при $\lambda > a$, то определенная при $\lambda > a$ функция ω называется преобразованием Лапласа меры U .

Если U имеет плотность u , то преобразование Лапласа (1.6) меры U называется также обычным преобразованием Лапласа функции u .

Последнее соглашение принимается единственно ради удобства. Чтобы быть систематичным, следовало бы рассмотреть интеграл более общего типа

$$\int_0^{\infty} e^{-\lambda x} v(x) U\{dx\} \quad (1.8)$$

и называть его «преобразованием Лапласа функции v по отношению к мере U ». Тогда преобразование (1.6) получило бы название: «преобразование u по отношению к мере Лебега» (или обычной длине). Этот подход мог бы иметь теоретическое преимущество, если бы позволял рассматривать знакопеременные функции u и v . Однако для целей настоящей книги самый простой и самый ясный способ действий состоит в том, чтобы связывать понятие преобразования Лапласа только с мерами. Это мы и будем делать ¹⁾.

¹⁾ Терминология еще не вполне сложилась, и в литературе термин «преобразование Лапласа для F » может обозначать как (1.1), так и (2.6). Мы бы назвали (2.6) «обычным преобразованием Лапласа функции распределения F », но в учебниках, которые, как правило, рассматривают только такие преобразования, прилагательное «обычный» бывает опущено. В подобных случаях во избежание путаницы преобразование (1.1) называют преобразованием Лапласа — Стильтjesа.

Если U такая мера, что интеграл (1.5) сходится при $\lambda = a$, то функция

$$\omega(\lambda + a) = \int_0^{\infty} e^{-\lambda x} e^{-ax} U\{dx\} = \int_0^{\infty} e^{-\lambda x} U^*\{dx\} \quad (1.9)$$

определена при всех $\lambda > 0$ и представляет собой преобразование Лапласа ограниченной меры $U^*\{dx\} = e^{-ax}U\{dx\}$, а $\omega(\lambda + a)/\omega(a)$ является преобразованием Лапласа некоторого распределения вероятностей. Поэтому каждая теорема, касающаяся преобразований распределений вероятностей, автоматически распространяется на более широкий класс мер. Поскольку $\omega(\lambda + a)$ получается «сдвигом» графика ω , мы будем называть указанный полезный метод *принципом сдвига*. Заметим, что U однозначно определяется по U^* , а U^* — по $\omega(\lambda + a)$ при $\lambda > 0$. Поэтому мы можем обобщить теорему 1 следующим образом.

Теорема 1а. Мера U однозначно определяется значениями соответствующего преобразования Лапласа (1.5) на некотором интервале $a < \lambda < \infty$.

Следствие. Непрерывная функция и однозначно определяется значениями своего обычного преобразования Лапласа (1.6) в некотором интервале $a < \lambda < \infty$.

Доказательство. Преобразование однозначно определяет меру U как интеграл от u , а две различные непрерывные¹⁾ функции не могут приводить к одинаковым интегралам. ►

[Явная формула для u в терминах ω приводится в (6.6) гл. VII.]

Теорема непрерывности обобщается аналогичным образом на последовательности произвольных мер U_n , обладающих преобразованием Лапласа. Из того, что U_n имеет преобразование Лапласа, следует, что $U_n\{I\} < \infty$ для конечных интервалов I . Мы напомним из гл. VIII, 1, и VIII, 6, что последовательность таких мер называют сходящейся к мере U тогда и только тогда, когда $U_n\{I\} \rightarrow U\{I\} < \infty$ на каждом конечном интервале непрерывности U .

Теорема 2а. (Обобщенная теорема непрерывности.) Пусть U_n ($n=1, 2, \dots$) — меры с преобразованиями Лапласа ω_n . Если $\omega_n(\lambda) \rightarrow \omega(\lambda)$ при $\lambda > a$, то ω является преобразованием Лапласа некоторой меры U и $U_n \rightarrow U$.

Обратно, если $U_n \rightarrow U$ и последовательность $\{\omega_n(a)\}$ ограничена, то $\omega_n(\lambda) \rightarrow \omega(\lambda)$ при $\lambda > a$.

Доказательство. (а) Допустим, что $U_n \rightarrow U$ и что $\omega_n(a) < A$.

¹⁾ То же самое рассуждение показывает, что вообще u определяется с точностью до значений на произвольном множестве меры нуль.

Если $t > 0$ — точка непрерывности U , то

$$\int_0^t e^{-(\lambda+a)x} U_n \{dx\} \rightarrow \int_0^t e^{-(\lambda+a)x} U \{dx\}. \quad (1.10)$$

Левая часть отличается от $\omega_n(\lambda+a)$ самое большее на

$$\int_t^\infty e^{-(\lambda+a)x} U_n \{dx\} < Ae^{-\lambda t}; \quad (1.11)$$

ее можно сделать меньше ε , если выбрать t достаточно большим. Это означает, что верхний и нижний пределы $\omega_n(\lambda+a)$ отличаются менее чем на произвольное ε , и, следовательно, при каждом $\lambda > 0$ последовательность $\{\omega_n(\lambda+a)\}$ сходится к конечному пределу.

(б) Предположим теперь, что $\omega_n(\lambda) \rightarrow \omega(\lambda)$ при $\lambda > a$. При фиксированном $\lambda_0 > a$ функция $\omega_n(\lambda + \lambda_0)/\omega_n(\lambda_0)$ есть преобразование Лапласа распределения вероятностей $U_n^* \{dx\} = (1/\omega_n(\lambda_0)) \times \times e^{-\lambda_0 x} U_n \{dx\}$. Поэтому по теореме непрерывности U_n^* сходится к (возможно, несобственному) распределению U^* , а отсюда следует, что U_n сходится к мере U , такой, что $[U \{dx\} = \omega(\lambda_0) \times \times e^{\lambda_0 x} U^* \{dx\}$. ►

Следующий пример показывает необходимость условия ограниченности $\{\omega_n(a)\}$.

Пример, е) Пусть U_n приписывает массу e^n точке n и нуль — дополнению этой точки. Так как $U_n \{0, n\} = 0$, мы имеем $U_n \rightarrow 0$. Однако $\omega_n(\lambda) = e^{n(\lambda-n)} \rightarrow \infty$ при всех $\lambda > 0$. ►

Иногда говорят о *двустороннем преобразовании* Лапласа для распределения F , не сосредоточенного на положительной полуоси, а именно

$$\Phi(\lambda) = \int_{-\infty}^{+\infty} e^{-\lambda x} F \{dx\}. \quad (1.12)$$

Однако этот интеграл не обязан существовать при $\lambda \neq 0$. Если он существует для $\lambda \neq 0$, то $\Phi(-\lambda)$ называют обычно *производящей функцией моментов*, хотя на самом деле она служит производящей функцией для последовательности $\{\mu_n/n!\}$, где μ_n — n -й момент.

§ 2. ЭЛЕМЕНТАРНЫЕ СВОЙСТВА

В этом параграфе мы перечислим наиболее часто используемые свойства преобразований Лапласа. Аналогии с производящими функциями очевидны.

(i) *Свертки*. Пусть F и G — распределения вероятностей и U — их свертка, т. е.

$$U(x) = \int_0^x G(x-y) F \{dy\}. \quad (2.1)$$

Соответствующие преобразования Лапласа подчиняются правилу умножения

$$\omega = \varphi\gamma. \quad (2.2)$$

Последнее равносильно утверждению, что для независимых случайных величин $E(e^{-\lambda(x+y)}) = E(e^{-\lambda x}) E(e^{-\lambda y})$, что есть частный случай правила умножения математических ожиданий¹⁾.

Если F и G имеют плотности f и g , то U имеет плотность

$$u(x) = \int_0^x g(x-y)f(y)dy \quad (2.3)$$

и правило умножения (2.2) справедливо и для «обычных» преобразований Лапласа (1.6) для плотностей f , g и u .

Покажем, что правило умножения можно обобщить следующим образом. Пусть F и G — произвольные меры с преобразованиями Лапласа φ и γ , определенными при $\lambda > 0$. Тогда свертка U имеет преобразование Лапласа ω , заданное формулой (2.2). Отсюда, в частности, следует, что правило умножения применимо к «обычным» преобразованиям Лапласа любых двух интегрируемых функций f и g и их свертке (2.3).

Для доказательства сформулированного утверждения введем конечные меры F_n , полученные урезанием F : при $x \leq n$ положим $F_n(x) = F(x)$ и $F_n(x) = F(n)$ при $x > n$. Определим G_n аналогичным урезанием G . При $x < n$ свертка $U_n = F_n * G_n$ не будет отличаться от U , и, следовательно, не только $F_n \rightarrow F$ и $G_n \rightarrow G$, но и $U_n \rightarrow U$. Для соответствующих преобразований Лапласа имеет место равенство $\omega_n = \varphi_n \gamma_n$, откуда, полагая $n \rightarrow \infty$, мы получаем утверждение $\omega = \varphi\gamma$.

Примеры. а) *Гамма-распределения.* В примере 1, б) правило свертки $f_\alpha * f_\beta = f_{\alpha+\beta}$ отражается в очевидном соотношении $\varphi_\alpha \varphi_\beta = \varphi_{\alpha+\beta}$.

б) *Степени.* Обычное преобразование Лапласа для $u_\alpha(x) = x^{\alpha-1}/\Gamma(\alpha)$ равно $\omega_\alpha(\lambda) = \lambda^{-\alpha}$. Отсюда следует, что свертка (2.3) функций u_α и u_β равна $u_{\alpha+\beta}$.

Результат примера а) можно вывести отсюда с помощью принципа сдвига, так как $\varphi_\alpha(\lambda) = \omega_\alpha(\lambda+1)$.

в) Если $a > 0$, то $e^{-a\lambda}\omega(\lambda)$ есть преобразование Лапласа для меры с функцией распределения $U(x-a)$. Это очевидным образом следует из определения, но может рассматриваться и как частный случай теоремы о свертке, так как $e^{-a\lambda}$ является преобразованием распределения, сосредоточенного в точке a . ►

(ii) *Производные и моменты.* Если F — распределение вероят-

¹⁾ Обратное неверно: две случайные величины могут быть зависимыми, но распределение их суммы определяется по формуле свертки [см. пример 4, д) гл. II и задачу 1 в гл. III, 9].

ностей и φ —его преобразование Лапласа (1.1), то φ обладает производными всех порядков, причем

$$(-1)^n \varphi^{(n)}(\lambda) = \int_0^{\infty} e^{-\lambda x} x^n F\{dx\} \quad (2.4)$$

(как всегда, $\lambda > 0$). Дифференцирование под знаком интеграла законно, так как формальное дифференцирование приводит к ограниченному и непрерывному подынтегральному выражению.

Из (2.5) вытекает, в частности, что F имеет конечный n -й момент в том и только том случае, когда существует конечный предел $\varphi^n(0)$. Поэтому для случайной величины X мы можем написать

$$E(X) = -\varphi'(0), \quad E(X^2) = \varphi''(0) \quad (2.5)$$

с очевидным соглашением о смысле этих равенств в случае расходимости. Правило дифференцирования (2.4) верно и для произвольных мер F .

(iii) Интегрирование по частям приводит от (1.1) к

$$\int_0^{\infty} e^{-\lambda x} F(x) dx = \frac{\varphi(\lambda)}{\lambda}, \quad \lambda > 0. \quad (2.6)$$

Если F —распределение вероятностей, то (2.6) иногда предпочтительнее записывать в форме, содержащей «хвост» распределения,

$$\int_0^{\infty} e^{-\lambda x} [1 - F(x)] dx = \frac{1 - \varphi(\lambda)}{\lambda}. \quad (2.7)$$

Это соответствует формуле (1.6) из 1; гл. XI, для производящих функций.

(iv) Изменение масштаба. Из (1.2) выводим, что при любом фиксированном $a > 0$ $E(e^{-a\lambda X}) = \varphi(a\lambda)$, т. е. $\varphi(a\lambda)$ является преобразованием для распределения $F\{dx/a\}$ [с функцией распределения $F(x/a)$]. Это соотношение часто используется.

Пример. г) Закон больших чисел. Пусть $X_1, X_2, \dots$ —независимые случайные величины с одним и тем же преобразованием Лапласа φ . Положим $E(X_i) = \mu$. Преобразование Лапласа для суммы $X_1 + \dots + X_n$ равно φ^n . Поэтому преобразование для среднего арифметического $[X_1 + \dots + X_n]/n$ равно $\varphi^n(\lambda/n)$. Вблизи нуля $\varphi(\lambda) = 1 - \mu\lambda + o(\lambda)$ [см. (2.5)], так что при $n \rightarrow \infty$

$$\lim \varphi^n\left(\frac{\lambda}{n}\right) = \lim \left(1 - \frac{\mu\lambda}{n}\right)^n = e^{-\mu\lambda}. \quad (2.8)$$

Но $e^{-\mu\lambda}$ есть преобразование Лапласа для распределения, сосредоточенного в точке μ , следовательно, распределение $[X_1 + \dots + X_n]/n$ сходится к этому пределу. Мы получили (слабый) закон больших

чисел в форме Хинчина, не требующей существования дисперсии. Доказательство, правда, применимо только к неотрицательным случайным величинам, но оно иллюстрирует изящество метода, основанного на преобразованиях Лапласа. ►

§ 3. ПРИМЕРЫ

а) *Равномерное распределение.* Пусть F обозначает равномерное распределение на $0, 1$. Его преобразование Лапласа равно $\varphi(\lambda) = (1 - e^{-\lambda})/\lambda$. Используя формулу бинома, можно показать, что n -кратная свертка F^{n*} имеет преобразование

$$\varphi^n(\lambda) = \sum_{k=0}^n (-1)^k \binom{n}{k} e^{-\lambda k} \lambda^{-n}. \quad (3.1)$$

Функция λ^{-n} служит преобразованием для $U(x) = x^n/n!$. Пример 2, в) показывает, что функция $e^{-k\lambda} \lambda^{-n}$ соответствует $(x-k)_+^n/n!$, где x_+ обозначает функцию, равную 0 для $x \leq 0$ и x для $x \geq 0$. Таким образом,

$$F^{n*}(x) = \frac{1}{n!} \sum_{k=0}^n (-1)^k \binom{n}{k} (x-k)_+^n. \quad (3.2)$$

Эта формула была получена прямым вычислением в гл. I, (9.5), и предельным переходом в задаче 20 из 1; гл. XI.

б) *Устойчивое распределение с показателем $1/2$.* Функция распределения

$$G(x) = 2[1 - \mathfrak{N}(1/\sqrt{x})], \quad x > 0, \quad (3.3)$$

(где $\mathfrak{N}$ — стандартное нормальное распределение) имеет преобразование Лапласа

$$\gamma(\lambda) = e^{-\sqrt{2\lambda}}. \quad (3.4)$$

Этот результат можно получить, проделав элементарные вычисления, но это неинтересно, и мы предпочитаем вывести (3.4) из предельной теоремы 3 из 1; гл. III, 7, где с распределением G мы впервые встретились. Рассмотрим простое симметричное случайное блуждание (бросание монеты) и обозначим через T момент первого возвращения в нуль. Цитированная теорема утверждает, что G является предельным распределением для нормированной суммы $(T_1 + \dots + T_n)/n^2$, где $T_1, T_2, \dots$ — независимые случайные величины, распределенные, как T . В соответствии с 1; гл. XI, (3.14), производящая функция для T равна $f(s) = 1 - \sqrt{1 - s^2}$, и поэтому

$$\gamma(\lambda) = \lim [1 - \sqrt{1 - e^{-2\lambda/n^2}}]^n = \lim \left[1 - \frac{\sqrt{2\lambda}}{n} \right]^n = e^{-\sqrt{2\lambda}}. \quad (3.5)$$

Мы несколько раз отмечали, что *распределение G является устойчивым*, но опять-таки прямая проверка требует трудоемких вычислений. В то же время очевидно, что $\gamma^n(\lambda) = \gamma(n^2\lambda)$, т. е. $G^{n*}(x) = G(n^{-2}x)$, чем без всякого труда доказывается устойчивость G .

в) *Степенные ряды и смеси*. Пусть F — распределение вероятностей с преобразованием Лапласа $\varphi(\lambda)$. Мы постоянно сталкивались с распределениями вида

$$G = \sum_{k=0}^{\infty} p_k F^{k*}, \quad (3.6)$$

где $\{p_k\}$ — распределение вероятностей. Если $P(s) = \sum p_k s^k$ — производящая функция для $\{p_k\}$, то преобразование Лапласа для G равно, очевидно,

$$\gamma(\lambda) = \sum_{k=0}^{\infty} p_k \varphi^k(\lambda) = P(\varphi(\lambda)). \quad (3.7)$$

Это соотношение распространяется на все степенные ряды с положительными коэффициентами. Перейдем к примерам.

г) *Бесселевы плотности*. Мы видели в примере 7, с) гл. II, что при $r = 1, 2, \dots$ плотность

$$v_r(x) = e^{-x} \frac{x^r}{x} I_r(x) \quad (3.8)$$

соответствует распределению вида (3.6), где F имеет показательное распределение с $\varphi(\lambda) = 1/(\lambda + 1)$ и $\{p_k\}$ — распределение момента первого перехода через точку $r > 0$ в обычном симметричном случайном блуждании. Производящая функция последнего распределения равна

$$P(s) = \left(\frac{1 - \sqrt{1 - s^2}}{s} \right)^r \quad (3.9)$$

[см. 1; гл. XI, (3.6)]. Подставляя сюда $s = (1 + \lambda)^{-1}$, мы видим, что *обычное преобразование Лапласа для плотности вероятности (3.8) равно*

$$[\lambda + 1 - \sqrt{(\lambda + 1)^2 - 1}]^r. \quad (3.10)$$

Утверждение о том, что v_r является плотностью вероятности, преобразование которой равно (3.10), доказано нами только для $r = 1, 2, \dots$. Однако оно верно¹⁾ для всех $r > 0$ и в этом смысле представляет вероятностный интерес. В самом деле, отсюда вытекает формула свертки $v_r * v_s = v_{r+s}$ и, следовательно, *безграничная делимость v_r* (см. § 7).

¹⁾ Результат принадлежит Н. Weber. Очень трудное аналитическое доказательство заменено теперь элементарным доказательством; см. J. Soc. Industr. Appl. Math., vol. 14 (1966).

г) *Другие бесселевы плотности.* Выберем в (3.6) в качестве F показательное распределение с $\varphi(\lambda) = 1/(1+\lambda)$, а в качестве $\{p_k\}$ — распределение Пуассона с $P(s) = e^{-t+ts}$. Легко дать явную формулу для G , но, к счастью, эта задача уже была решена в примере 7, а) гл. II. Мы видели там, что плотность

$$w_p(x) = e^{-t-x} \sqrt{\left(\frac{x}{t}\right)^p} I_p(2\sqrt{tx}), \quad (3.11)$$

определенная в (7.2) гл. II, является сверткой нашего распределения G с гамма-плотностью $f_{1, p+1}$. Отсюда следует, что обычное преобразование Лапласа для w_p равно произведению γ на преобразование для $f_{1, p+1}$, т. е. на $(\lambda+1)^{-(p+1)}$. Соответственно *плотность вероятности (3.11) имеет преобразование Лапласа*

$$\frac{1}{(\lambda+1)^{p+1}} e^{-t+t/(\lambda+1)}. \quad (3.12)$$

При $t=1$, используя принцип сдвига (1.9), получаем, что $\sqrt{x^p} I_p(2\sqrt{x})$ имеет обычное преобразование $\lambda^{-p-1} e^{1/\lambda}$.

е) *Смеси показательных плотностей.* Пусть плотность f имеет вид

$$f(x) = \sum_{k=1}^n p_k a_k e^{-a_k x}, \quad p_k > 0, \quad \sum_{k=1}^n p_k = 1, \quad (3.13)$$

где для определенности мы будем считать, что $0 < a_1 < \dots < a_n$. Соответствующее преобразование Лапласа задается формулой

$$\varphi(\lambda) = \sum_{k=1}^n p_k \frac{a_k}{\lambda + a_k} = \frac{Q(\lambda)}{P(\lambda)}, \quad (3.14)$$

где P — полином степени n с корнями $-a_k$, а Q — полином степени $n-1$. Обратно, для любого полинома Q степени $n-1$ отношение $Q(\lambda)/P(\lambda)$ допускает разложение на элементарные дроби вида (3.14) с

$$a_r p_r = \frac{Q(-a_r)}{P'(-a_r)} \quad (3.15)$$

[см. (4.5) в 1; XI]. Для того чтобы (3.14) соответствовало смеси (3.13), необходимо и достаточно, чтобы $p_r > 0$ и $Q(0)/P(0) = 1$. Из вида графика P ясно, что $P'(-a_r)$ и $P'(-a_{r+1})$ имеют противоположные знаки и, следовательно, то же самое верно для $Q(-a_r)$ и $Q(-a_{r+1})$. Другими словами, необходимо, чтобы Q имело корень $-b_r$ между $-a_r$ и $-a_{r+1}$. Но поскольку Q не может иметь более чем $n-1$ корней $-b_r$, то делаем вывод, что эти корни должны удовлетворять неравенствам

$$0 < a_1 < b_1 < a_2 < b_2 < \dots < b_{n-1} < a_n. \quad (3.16)$$

Это соотношение гарантирует, что все p_r имеют один и тот же

знак. Таким образом, мы приходим к следующему заключению. Пусть P и Q — полиномы степени n и $n-1$ соответственно, и пусть $Q(0)/P(0)=1$. Для того чтобы отношение $Q(\lambda)/P(\lambda)$ было преобразованием Лапласа смеси (3.13) показательных плотностей, необходимо и достаточно, чтобы корни $-a$, полинома P и корни $-b$, полинома Q были различны и (в правильном порядке) удовлетворяли условию (3.16). ►

§ 4. ВПОЛНЕ МОНОТОННЫЕ ФУНКЦИИ. ФОРМУЛЫ ОБРАЩЕНИЯ

Как мы видели в гл. XII, 2, функция f , заданная на $\overline{0, 1}$, является производящей функцией для положительной последовательности $\{f_n\}$ тогда и только тогда, когда f абсолютно монотонна (т. е. когда f имеет положительные производные $f^{(n)}$ всех порядков). Аналогичная теорема имеет место и для преобразований Лапласа. Разница лишь в том, что теперь знаки производных чередуются.

Определение 1. Заданная на $\overline{0, \infty}$ функция φ называется вполне монотонной, если она имеет производные $\varphi^{(n)}$ всех порядков и

$$(-1)^n \varphi^{(n)}(\lambda) \geq 0, \quad \lambda > 0. \quad (4.1)$$

При $\lambda \rightarrow 0$ значения $\varphi^{(n)}(\lambda)$ приближаются к конечным или бесконечным пределам, которые мы обозначим $\varphi^{(n)}(0)$. Типичные примеры — функции $1/\lambda$ и $1/(1+\lambda)$.

С. Н. Бернштейну (1928) принадлежит следующая прекрасная теорема, которая послужила отправным пунктом многих исследований (ее доказательство упрощалось шаг за шагом).

Мы в состоянии дать крайне простое доказательство, поскольку подготовительная работа была совершена при характеристизации производящих функций, которая была получена в теореме 2 гл. VII, 2, как следствие закона больших чисел.

Теорема 1. Функция φ на $\overline{0, \infty}$ является преобразованием Лапласа распределения вероятностей F тогда и только тогда, когда она вполне монотонна и $\varphi(0)=1$.

Мы докажем вариант этой теоремы, который кажется более общим по форме, но фактически может быть выведен из первоначального варианта при использовании принципа сдвига, описанного в связи с (1.9).

Теорема 1а. Функция φ на $\overline{0, \infty}$ является вполне монотонной

тогда и только тогда, когда она имеет вид

$$\varphi(\lambda) = \int_0^{\infty} e^{-\lambda x} F\{dx\}, \quad \lambda > 0, \quad (4.2)$$

где F — (не обязательно конечная) мера на $\overline{0, \infty}$.

(По нашему первоначальному соглашению промежуток интегрирования всегда *замкнут*; наличие у F атома в нуле приводит к $\varphi(\infty) > 0$.)

Доказательство. Необходимость условия устанавливается формальным дифференцированием [ср. (2.4)]. Предположим теперь, что φ вполне монотонна. Рассмотрим $\varphi(a - as)$ при фиксированном $a > 0$ как функцию от s , $0 < s < 1$. Производные этой функции, очевидно, положительны, и по теореме 2 из гл. VII, 2, при $0 \leq s < 1$ справедливо разложение Тейлора

$$\varphi(a - as) = \sum_{n=0}^{\infty} \frac{(-a)^n \varphi^{(n)}(a)}{n!} s^n. \quad (4.3)$$

Таким образом,

$$\varphi_a(\lambda) = \varphi(a - ae^{-\lambda/a}) = \sum_{n=0}^{\infty} \frac{(-a)^n \varphi^{(n)}(a)}{n!} e^{-n\lambda/a} \quad (4.4)$$

есть преобразование Лапласа арифметической меры, приписывающей массу $(-a)^n \varphi^{(n)}(a)/n!$ точке n/a , где $n = 0, 1, 2, \dots$. При $a \rightarrow \infty$ $\varphi_a(\lambda) \rightarrow \varphi(\lambda)$. Поэтому по обобщенной теореме непрерывности существует такая мера F , что $F_a \rightarrow F$ и φ будет ее преобразованием Лапласа. ►

Мы доказали не только теорему 1а, но и соотношение $F_a \rightarrow F$, которое можно переписать в форме следующей важнейшей теоремы.

Теорема 2. (Формула обращения.) Если (4.2) выполняется при $\lambda > 0$, то во всех точках непрерывности ¹⁾

$$F(x) = \lim_{a \rightarrow \infty} \sum_{n \leq ax} \frac{(-a)^n}{n!} \varphi^{(n)}(a). \quad (4.5)$$

Последняя формула представляет значительный теоретический интерес и приводит к различным полезным заключениям. Примером может служить следующий критерий ограниченности, применяемый в теории полугрупп [см. задачу 13].

Следствие. Для того чтобы функция φ представлялась в форме

$$\varphi(\lambda) = \int_0^{\infty} e^{-\lambda x} f(x) dx, \quad \text{где } 0 \leq f \leq C, \quad (4.6)$$

¹⁾ Формула обращения (4.5) была получена в (6.4) гл. VII как прямое следствие закона больших чисел. В (6.6) гл. VII указана аналогичная формула для интегралов вида (4.6) с непрерывной (не обязательно положительной) f .

необходимо и достаточно, чтобы при всех $a > 0$

$$0 \leq \frac{(-a)^n \varphi^{(n)}(a)}{n!} \leq \frac{C}{a}. \quad (4.7)$$

Доказательство. Дифференцируя (4.6) под знаком интеграла, получаем (4.7) [ср. (2.4)]. Обратно, из (4.7) видно, что φ вполне монотонна и потому является преобразованием некоторой меры F . Из (4.7) и (4.5) выводим

$$F(x_2) - F(x_1) \leq C(x_2 - x_1)$$

для любых $x_1 < x_2$. Иными словами, F имеет ограниченные разностные отношения и, следовательно, представляет собой интеграл от функции $f \leq C$ (см. гл. V, 3). ►

Теорема 1 приводит к простым критериям того, будет ли данная функция преобразованием Лапласа некоторого распределения вероятностей. Для иллюстраций стандартных приемов докажем

Критерий 1. Если φ и ψ вполне монотонны, то таково же и их произведение $\varphi\psi$.

Доказательство. Мы покажем по индукции, что знаки производных $\varphi\psi$ чередуются. Допустим, что для каждой пары φ и ψ вполне монотонных функций знаки первых n производных $\varphi\psi$ чередуются. Так как $-\varphi'$ и $-\psi'$ вполне монотонны, то предположение индукции применимо к произведениям $-\varphi'\psi$ и $-\varphi\psi'$. Из равенства $-(\varphi\psi)' = -\varphi'\psi - \varphi\psi'$ заключаем, что знаки первых $n+1$ производных $\varphi\psi$ чередуются. Так как при $n=1$ утверждение тривиально верно, критерий доказан. ►

Точно так же доказывается полезный

Критерий 2. Если φ вполне монотонна и ψ — положительная функция с вполне монотонной производной, то $\varphi(\psi)$ вполне монотонна (в частности, функция $e^{-\psi}$ вполне монотонна).

Типичные применения даны в § 6, а также в следующем примере, который часто излагается в литературе с ненужными усложнениями.

Пример. а) Уравнение, встречающееся в теории ветвящихся процессов. Пусть φ — преобразование Лапласа для распределения вероятностей F с математическим ожиданием μ , $0 < \mu \leq \infty$. Пусть $c > 0$. Мы докажем, что уравнение

$$\beta(\lambda) = \varphi(\lambda + c - c\beta(\lambda)) \quad (4.8)$$

имеет единственный корень $\beta(\lambda) \leq 1$, а также, что β есть преобразование Лапласа для некоторого распределения B , которое будет собственным при $\mu c \leq 1$ и несобственным в других случаях.

(См. применения и ссылки в гл. XIV, 4.)

Доказательство. Рассмотрим уравнение

$$\varphi(\lambda + c - cs) - s = 0 \quad (4.9)$$

при фиксированном $\lambda > 0$ и $0 \leq s \leq 1$. Левая часть его — выпуклая функция, значение которой при $s = 1$ отрицательно, а при $s = 0$ — положительно. Отсюда следует существование единственного корня.

Докажем, что корень $\beta(\lambda)$ есть преобразование Лапласа. Положим $\beta_0 = 0$ и $\beta_{n+1} = \varphi(\lambda + c - c\beta_n)$. Тогда $\beta_0 \leq \beta_1 \leq 1$, и так как φ убывает, то $\beta_1 \leq \beta_2 \leq 1$. По индукции $\beta_n \leq \beta_{n+1} \leq 1$. Предел ограниченной монотонной последовательности $\{\beta_n\}$ удовлетворяет (4.8), и, следовательно, $\beta = \lim \beta_n$. Функция $\beta_1(\lambda) = \varphi(\lambda + c)$ вполне монотонна, и критерий 2 показывает, что $\beta_2, \beta_3, \dots$ также вполне монотонны. По теореме непрерывности то же самое верно для предела β , и, следовательно, β есть преобразование Лапласа меры B . Поскольку $\beta(\lambda) \leq 1$ для всех λ , то полная масса распределения B равна $\beta(0) \leq 1$. Остается только выяснить, при каких условиях $\beta(0) = 1$.

По построению $s = \beta(0)$ является *наименьшим* корнем уравнения

$$\varphi(c - cs) - s = 0. \quad (4.10)$$

Левая часть (4.10) как функция s является выпуклой; она положительна при $s = 0$ и равна нулю при $s = 1$. Отсюда следует, что второй корень $s < 1$ существует тогда и только тогда, когда производная в точке $s = 1$ положительна, т. е. тогда и только тогда, когда $-c\varphi'(0) > 1$. В ином случае $\beta(0) = 1$, и β есть преобразование Лапласа собственного распределения вероятностей B . Следовательно, распределение B является собственным тогда и только тогда, когда $-c\varphi'(0) = c\mu \leq 1$. ►

§ 5. ТАУБЕРОВЫ ТЕОРЕМЫ

Пусть U — мера, сосредоточенная на $\overline{0, \infty}$ и такая, что ее преобразование Лапласа

$$\omega(\lambda) = \int_0^{\infty} e^{-\lambda x} U\{dx\} \quad (5.1)$$

существует при $\lambda > 0$. Удобно описывать меру U в терминах соответствующей «функции распределения», определенной при $x \geq 0$, как $U\{0, x\}$. Мы увидим, что при весьма общих условиях поведение ω вблизи нуля однозначно определяет асимптотическое поведение $U(x)$ при $x \rightarrow \infty$, и наоборот. Исторически любое соотношение, описывающее асимптотическое поведение U в терминах ω , называется тауберовой теоремой, тогда как теоремы, описывающие поведение ω в терминах U , называются обычно абелевыми,

Мы не будем делать различия между этими двумя типами теорем, так как наши соотношения симметричны.

Чтобы избежать непривлекательных формул, содержащих обратные величины, введем положительные переменные t и τ , связанные соотношением

$$t\tau = 1. \quad (5.2)$$

Тогда $\tau \rightarrow 0$ при $t \rightarrow \infty$.

Поясним происхождение тауберовых теорем. При фиксированном t замена переменных $x = ty$ в (5.1) показывает, что $\omega(\tau\lambda)$ является преобразованием Лапласа для несобственной функции распределения $U(ty)$. Так как ω убывает, возможно найти последовательность $\tau_1, \tau_2, \dots \rightarrow 0$, такую, что, когда τ пробегает ее,

$$\frac{\omega(\tau\lambda)}{\omega(\tau)} \rightarrow \gamma(\lambda), \quad (5.3)$$

где $\gamma(\lambda)$ конечно по крайней мере при $\lambda > 1$. По обобщенной теореме [непрерывности предел γ есть преобразование Лапласа некоторой меры G , и когда t пробегает последовательность $t_k = 1/\tau_k$, то

$$\frac{U(tx)}{\omega(\tau)} \rightarrow G(x) \quad (5.4)$$

во всех точках непрерывности G . При $x = 1$ мы видим, что асимптотическое поведение $U(t)$ при $t \rightarrow \infty$ тесно связано с поведением $\omega(t^{-1})$.

В принципе мы могли бы сформулировать это утверждение как некую всеобъемлющую тауберову теорему, но она была бы неудобна для практического использования. Для того чтобы достичь разумной простоты, мы рассмотрим только случай, когда (5.3) верно при *любом* способе приближения τ к нулю (т. е. когда ω правильно изменяется в окрестностях нуля). Элементарная лемма¹⁾ из гл. VIII, 8, утверждает, что предел γ необходимо имеет вид $\gamma(\lambda) = \lambda^{-\rho}$, где $\rho \geq 0$. Соответствующая мера определяется равенством $G(x) = x^\rho/\Gamma(\rho+1)$. Из (5.4) вытекает, что U правильно меняется и соответствующие показатели у ω и U равны по абсолютной величине. Мы сформулируем этот важнейший результат вместе с обратным утверждением.

Теорема 1. Пусть U — мера, преобразование Лапласа ω которой определено при $\lambda > 0$. Тогда каждое из соотношений

$$\frac{\omega(\tau\lambda)}{\omega(\tau)} \rightarrow \frac{1}{\lambda^\rho}, \quad \tau \rightarrow 0, \quad (5.5)$$

¹⁾ Эта лемма нужна только для объяснения формы соотношений (5.5) и (5.6) (которая иначе показалась бы неестественной). Теория правильно меняющихся функций не используется нигде в этом параграфе [кроме замечания, что (5.18) влечет за собой (5.16)].

и

$$\frac{U(tx)}{U(t)} \rightarrow x^\rho, \quad t \rightarrow \infty, \quad (5.6)$$

влечет за собой другое, а также соотношение

$$\omega(\tau) \sim U(t) \Gamma(\rho + 1). \quad (5.7)$$

Доказательство. а) Предположим (5.5). Левая часть (5.5) представляет собой преобразование Лапласа, соответствующее $U(tx)/\omega(\tau)$. По обобщенной теореме непрерывности отсюда следует

$$\frac{U(tx)}{\omega(\tau)} \rightarrow \frac{x^\rho}{\Gamma(\rho + 1)}. \quad (5.8)$$

При $x=1$ мы получаем (5.7), а подставляя это назад в (5.8), мы получаем (5.6).

б) Предположим (5.6). Переходя к преобразованиям Лапласа, мы получаем

$$\frac{\omega(\tau\lambda)}{U(t)} \rightarrow \frac{\Gamma(\rho + 1)}{\lambda^\rho} \quad (5.9)$$

в предположении, что применима обобщенная теорема непрерывности, т. е. в предположении, что левая часть ограничена при некотором λ . В условиях а) мы видим, что из (5.9) выводятся (5.7) и (5.5), поэтому для доказательства теоремы достаточно проверить, что отношения $\omega(\tau)/U(t)$ ограничены.

Разбивая область интегрирования точками $t, 2t, 4t, \dots$, получаем

$$\omega(\tau) \leq \sum_0^\infty e^{-2^n \tau} U(2^n t). \quad (5.10)$$

В силу (5.7) существует число t_0 , такое, что $U(2t) < 2^{\rho+1} U(t)$ при $t > t_0$. Последовательное применение этого неравенства дает

$$\frac{\omega(\tau)}{U(t)} \leq \sum_0^\infty 2^{n(\rho+1)} e^{-2^n \tau}, \quad (5.11)$$

так что левая часть действительно ограничена при $t \rightarrow \infty$. ►

Примеры. а) $U(x) \sim \log^2 x$ при $x \rightarrow \infty$ тогда и только тогда, когда $\omega(\lambda) \sim \log^2 \lambda$ при $\lambda \rightarrow 0$. Аналогично $U(x) \sim \sqrt{x}$ тогда и только тогда, когда $\omega(\lambda) \sim \sqrt{\pi/\lambda}$.

б) Пусть F — распределение вероятностей с преобразованием Лапласа φ . Мера $U\{dx\} = xF\{dx\}$ имеет преобразование $-\varphi'$. Поэтому если $-\varphi'(\lambda) \sim \mu\lambda^{-\rho}$ при $\lambda \rightarrow \infty$, то

$$U(x) = \int_0^x yF\{dy\} \sim \frac{\mu}{\Gamma(\rho+1)} x^\rho, \quad x \rightarrow \infty,$$

и наоборот. Это обобщает правило дифференцирования (2.4), которое заключается в (5.7) при $\rho = 0$. ►

Иногда полезно знать, в какой степени справедлива теорема в пределе при $\rho \rightarrow \infty$. Мы сформулируем соответствующий результат как

Следствие. Если при некотором $a > 1$ и $t \rightarrow \infty$

$$\text{или } \frac{\omega(\tau a)}{\omega(\tau)} \rightarrow 0, \quad \text{или } \frac{U(ta)}{U(t)} \rightarrow \infty, \quad (5.12)$$

то

$$\frac{U(t)}{\omega(\tau)} \rightarrow 0. \quad (5.13)$$

Доказательство. Если выполняется первое из соотношений (5.12), то $\omega(\tau\lambda)/\omega(\tau) \rightarrow 0$ при $\lambda > a$, и по обобщенной теореме непрерывности $U(tx)/\omega(\tau) \rightarrow 0$ при всех $x > 0$. Второе из соотношений (5.12) влечет (5.13), так как

$$\omega(\tau) \geq \int_0^{at} e^{-x/t} U\{dx\} \geq e^{-a} U(ta). \quad \blacktriangleright$$

В применениях более удобно формулировать теорему 1 в терминах медленно меняющихся функций. Напомним, что определенная на $\overline{0, \infty}$ положительная функция L называется *медленно меняющейся на бесконечности*, если при каждом фиксированном x

$$\frac{L(tx)}{L(t)} \rightarrow 1, \quad t \rightarrow \infty. \quad (5.14)$$

Говорят, что L медленно меняется в нуле, если это соотношение выполняется при $t \rightarrow 0$ (т. е. если $L(1/x)$ медленно меняется на бесконечности). Очевидно, что U удовлетворяет (5.6) тогда и только тогда, когда $U(x)/x^\rho$ медленно меняется на бесконечности. Аналогично (5.5) выполняется тогда и только тогда, когда $\lambda^\rho \omega(\lambda)$ медленно меняется в нуле. Поэтому теорема 1 может быть сформулирована следующим образом.

Теорема 2. Если L медленно меняется на бесконечности и $0 \leq \rho < \infty$, то каждое из соотношений

$$\omega(\tau) \sim \tau^{-\rho} L\left(\frac{1}{\tau}\right), \quad \tau \rightarrow 0, \quad (5.15)$$

и

$$U(t) \sim \frac{1}{\Gamma(\rho+1)} t^\rho L(t), \quad t \rightarrow \infty, \quad (5.16)$$

влечет за собой другое.

Теорема 2 имеет интересную историю. Переход от (5.16) к (5.15) (от меры к преобразованию Лапласа) называют абелевой теоремой, обратный переход — от (5.15) к (5.16) (от преобразования Лапласа к мере) — называют тауберовой

теоремой. В обычном изложении эти две теоремы отделены одна от другой, причем доказательство второй из них заметно сложнее. В знаменитой работе Харди и Литлвуда с помощью трудных вычислений был исследован случай $\omega(\lambda) \sim \lambda^{-\rho}$. В 1930 г. Карамата произвел сенсацию, дав упрощенное доказательство для указанного специального случая (это доказательство до сих пор приводится в учебниках по теории функции комплексного переменного и преобразованиям Лапласа). Вскоре после этого он ввел класс правильно меняющихся функций и доказал теорему 2. Доказательство было, однако, слишком сложным для того, чтобы включать его в учебники. Понятие медленного изменения было введено Р. Шмидтом около 1925 г. в связи с этими же вопросами. Наше доказательство упрощает теорию и делает ее цельной. Кроме того, оно приводит к малоизвестному, но полезному следствию.

Большим преимуществом нашего доказательства является то, что оно без изменений применимо к случаю, когда нуль и бесконечность меняются ролями, т. е. когда $\tau \rightarrow \infty$, а $t \rightarrow 0$. Этим путем мы приходим к двойственной теореме, связывающей поведение ω на бесконечности и поведение U в нуле [эта теорема будет использована в настоящей книге только для вывода (6.2)].

Теорема 3. *Последние две теоремы и следствие остаются верными, если нуль и бесконечность меняются ролями, т. е. если $\tau \rightarrow \infty$ и $t \rightarrow 0$.*

Теорема 2 содержит основной результат этого параграфа, однако для полноты изложения мы сделаем два добавления к ней. Прежде всего если U обладает плотностью $U' = u$, то желательно иметь оценки для u . Эту проблему нельзя решить во всей общности, так как «хорошая» функция U может иметь очень «плохую» плотность u . В большинстве применений, однако, плотность u будет «монотонной, начиная с некоторого места», т. е. монотонной на некотором интервале x_0, ∞ . Для таких плотностей верна

Теорема 4¹⁾. *Пусть $0 < \rho < \infty$. Если U имеет монотонную, начиная с некоторого места, производную u , то при $\lambda \rightarrow 0$ и при $x \rightarrow \infty$ соответственно*

$$\begin{aligned} \omega(\lambda) &\sim \frac{1}{\lambda^\rho} L\left(\frac{1}{\lambda}\right) \text{ тогда и только тогда,} \\ &\text{когда } u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x). \end{aligned} \quad (5.17)$$

(Более строгий в формальном отношении вариант см. в задаче 16.)

Доказательство. Утверждение леммы есть прямое следствие теоремы 2 и следующей ниже леммы.

Лемма. *Допустим, что U имеет монотонную, начиная с неко-*

¹⁾ Эта теорема включает известную тауберову теорему Э. Ландау. Наше доказательство дает новый пример того, как теорема о выборе устраняет аналитические сложности,

того места, плотность u . Если выполняется (5.16) и $\rho > 0$, то

$$u(x) \sim \rho U(x)/x, \quad x \rightarrow \infty. \quad (5.18)$$

[Обратно, (5.18) влечет за собой (5.16), даже если u не монотонна. Это утверждение получается из (9.6) гл. VIII при $Z = u$ и $\rho = 0$.]

Доказательство. Для $0 < a < b$ имеем

$$\frac{U(bt) - U(at)}{U(t)} = \int_a^b \frac{u(y)t}{U(t)} dy. \quad (5.19)$$

При $t \rightarrow \infty$ левая часть сходится к $b^\rho - a^\rho$. Для достаточно больших t подынтегральное выражение монотонно, и поэтому из (5.16) следует, что подынтегральное выражение остается ограниченным при $t \rightarrow \infty$. По теореме о выборе [гл. VIII, 6] существует такая последовательность $t_1, t_2, \dots \rightarrow \infty$, что, когда t пробегает ее,

$$\frac{u(y)t}{U(t)} \rightarrow \psi(y) \quad (5.20)$$

во всех точках непрерывности. Как вытекает из предыдущего, интеграл от ψ по $\overline{a, b}$ равен $b^\rho - a^\rho$, так что $\psi(y) = \rho y^{\rho-1}$. Так как этот предел не зависит от выбора последовательности $\{t_k\}$, то (5.20) верно при любом стремлении t к бесконечности. Для $y=1$ это соотношение сводится к (5.18). ►

Пример. в) Для распределения вероятностей F с характеристической функцией φ имеем [см. (2.7)]

$$\int_0^\infty e^{-\lambda x} [1 - F(x)] dx = [1 - \varphi(\lambda)]/\lambda. \quad (5.21)$$

Поскольку функция $1 - F$ монотонна, каждое из соотношений

$$1 - \varphi(\lambda) \sim \lambda^{1-\rho} L(1/\lambda) \quad \text{и} \quad 1 - F(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x) \quad (5.22)$$

($\rho > 0$) влечет другое. Полезность этого наблюдения будет продемонстрирована в следующем параграфе. ►

Применение этой теоремы иллюстрируется в следующем параграфе. В заключение мы покажем, как теорема 2 позволяет получить *тауберову теорему для степенных рядов* [она используется в (8.10) гл. XII и гл. XVII, 5].

Теорема 5. Пусть $q_n \geq 0$, и пусть

$$Q(s) = \sum_{n=0}^{\infty} q_n s^n \quad (5.23)$$

сходится при $0 \leq s < 1$. Если L медленно меняется на бесконечности и $0 \leq \rho < \infty$, то каждое из двух соотношений

$$Q(s) \sim \frac{1}{(1-s)^\rho} L\left(\frac{1}{1-s}\right), \quad s \rightarrow 1- \quad (5.24)$$

и

$$q_0 + q_1 + \dots + q_{n-1} \sim \frac{1}{\Gamma(\rho+1)} n^\rho L(n), \quad n \rightarrow \infty, \quad (5.25)$$

влечет другое.

Далее, если последовательность $\{q_n\}$ монотонна и $0 < \rho < \infty$, то (5.24) равносильно соотношению

$$q_n \sim \frac{1}{\Gamma(\rho)} n^{\rho-1} L(n), \quad n \rightarrow \infty. \quad (5.26)$$

Доказательство. Пусть U — мера с плотностью u , такой, что

$$u(x) = q_n \quad \text{при} \quad n \leq x < n+1. \quad (5.27)$$

Левая часть в (5.25) равна $U(n)$. Преобразование Лапласа меры U определяется равенством

$$\omega(\lambda) = \frac{1-e^{-\lambda}}{\lambda} \sum_{n=0}^{\infty} q_n e^{-n\lambda} = \frac{1-e^{-\lambda}}{\lambda} Q(e^{-\lambda}). \quad (5.28)$$

Таким образом, видно, что соотношения (5.24) и (5.25) равносильны соответственно соотношениям (5.15) и (5.16), и поэтому в силу теоремы 2 каждое из них влечет за собой другое. В свою очередь (5.26) есть прямое следствие теоремы 4. ►

Пример. г) Пусть $q_n = n^{\rho-1} \log^\alpha n$, где $\rho > 0$ и α произвольно. Последовательность $\{q_n\}$ монотонна, начиная с некоторого места, и поэтому (5.24) выполняется с $L(t) = \Gamma(\rho) \log^\alpha t$. ►

§ 6*). УСТОЙЧИВЫЕ РАСПРЕДЕЛЕНИЯ

Чтобы показать пользу тауберовых теорем, мы выведем общую формулу для устойчивых распределений, сосредоточенных на $\overline{0, \infty}$, и дадим полную характеристику соответствующих областей притяжения. Доказательства проводятся прямым методом и обладают замечательной простотой по сравнению с методами, необходимыми для распределений, не сосредоточенных на $\overline{0, \infty}$.

Теорема 1. При фиксированном $0 < \alpha < 1$ функция $\gamma_\alpha(\lambda) = e^{-\lambda^\alpha}$ является преобразованием Лапласа распределения G_α , обладающего следующими свойствами:

*) За исключением (6.2), результаты этого параграфа получены независимо от гл. IX и XVII. Устойчивые распределения были введены в гл. VI, 1.

G_α устойчиво; более точно, если $X_1, \dots, X_n$ — независимые случайные величины с распределением G_α , то $(X_1 + \dots + X_n)/n^{1/\alpha}$ снова имеет распределение G_α ; далее

$$x^\alpha [1 - G_\alpha(x)] \rightarrow \frac{1}{\Gamma(1-\alpha)}, \quad x \rightarrow \infty, \quad (6.1)$$

$$e^{x^{-\alpha}} G_\alpha(x) \rightarrow 0, \quad x \rightarrow 0. \quad (6.2)$$

Доказательство. По второму критерию § 4 функция γ_α вполне монотонна, так как $e^{-\lambda}$ вполне монотонна и λ^α имеет вполне монотонную производную. Поскольку $\gamma_\alpha(0) = 1$, мера G_α с преобразованием Лапласа γ_α имеет полную массу, равную единице. Свойство устойчивости очевидным образом вытекает из равенства $\gamma_\alpha^2(\lambda) = \gamma_\alpha(n^{1/\alpha}\lambda)$. Соотношение (6.1) является частным случаем (5.22), а (6.2) немедленно получается из теоремы 3 и следствия теоремы 1 из § 5. ►

Теорема 2. Пусть F — распределение вероятностей, сосредоточенное на $0, \infty$ и такое, что

$$F^n \star (a_n x) \rightarrow G(x) \quad (6.3)$$

(в точках непрерывности), где G — собственное распределение, не сосредоточенное в одной точке. Тогда

а) существует функция L , медленно меняющаяся на бесконечности¹⁾, и постоянная α , $0 < \alpha < 1$, такие, что

$$1 - F(x) \sim \frac{x^{-\alpha} L(x)}{\Gamma(1-\alpha)}, \quad x \rightarrow \infty. \quad (6.4)$$

б) Обратно, если F имеет вид (6.4), то можно выбрать числа a_n так, что

$$\frac{nL(a_n)}{a_n^\alpha} \rightarrow 1. \quad (6.5)$$

В этом случае выполняется (6.3) с $G = G_\alpha$.

Из сказанного следует, что возможные предельные распределения G в (6.3) отличаются от какого-нибудь G_α лишь масштабным множителем. В частности, отсюда видно, что не существует других устойчивых распределений, сосредоточенных на $0, \infty$. (См. лемму 1 гл. VIII, 2.)

Доказательство. Если φ и γ — преобразования Лапласа F и G соответственно, то (6.3) равносильно соотношению

$$-n \log \varphi\left(\frac{\lambda}{a_n}\right) \rightarrow -\log \gamma(\lambda). \quad (6.6)$$

¹⁾ То есть L удовлетворяет (5.14). Множитель $\Gamma(1-\alpha)$ в (6.4) введен лишь для удобства и приводит только к изменению в обозначениях.

По теореме гл. VIII, 8, отсюда следует, что $-\log \varphi$ правильно меняется в нуле, т. е.

$$-\log \varphi(\lambda) \sim \lambda^\alpha L\left(\frac{1}{\lambda}\right), \quad \lambda \rightarrow 0, \quad (6.7)$$

где L медленно меняется на бесконечности и $\alpha \geq 0$. Тогда из (6.6) $-\log \gamma(\lambda) = C\lambda^\alpha$. Так как G не сосредоточено в одной точке, то $0 < \alpha < 1$.

Теперь из (6.7) следует

$$\frac{1-\varphi(\lambda)}{\lambda} \sim \lambda^{\alpha-1} L\left(\frac{1}{\lambda}\right), \quad \lambda \rightarrow 0. \quad (6.8)$$

Ввиду (5.22) каждое из соотношений (6.4) и (6.8) влечет другое. Таким образом, (6.4) необходимо для выполнения (6.1).

Для доказательства части б) мы будем исходить из (6.4), которое, как только что было показано, влечет (6.8). Определим a_n при фиксированном n как нижнюю границу всех таких x , что $n[1-F(x)] \sim 1/\Gamma(1-\alpha)$. Тогда (6.5) выполняется. Используя это и медленное изменение L , заключаем из (6.8), что

$$1-\varphi(\lambda/a_n) \sim \lambda^\alpha a_n^{-\alpha} L(a_n/\lambda) \sim \lambda^\alpha/n. \quad (6.9)$$

Отсюда следует, что левая часть в (6.6) стремится к λ^α , и это завершает доказательство. ►

(См. задачу 26 о влиянии максимального слагаемого.)

§ 7*). БЕЗГРАНИЧНО ДЕЛИМЫЕ РАСПРЕДЕЛЕНИЯ

В соответствии с определением гл. VI, 3, распределение вероятностей U с преобразованием Лапласа ω называется безгранично делимым, если для любого $n=1, 2, \dots$ положительный корень n -й степени $\omega_n = \omega^{1/n}$ является преобразованием Лапласа некоторого распределения вероятностей.

Теорема 1. *Функция ω представляет собой преобразование Лапласа безгранично делимого распределения тогда и только тогда, когда $\omega = e^{-\psi}$, где ψ имеет вполне монотонную производную и $\psi(0) = 0$.*

Доказательство. Используем критерий 2 § 4. Мы видим, что если $\psi(0) = 0$ и ψ' вполне монотонна, то функция $\omega_n = e^{-\psi/n}$ является преобразованием Лапласа некоторого распределения вероятностей. Следовательно, условие теоремы достаточно.

Для доказательства необходимости предположим, что $\omega_n = e^{-\psi/n}$ есть при каждом n преобразование Лапласа некоторого распределения вероятностей. Положим

$$\psi_n(\lambda) = n[1 - \omega_n(\lambda)]. \quad (7.1)$$

* Материал этого параграфа не используется в дальнейшем.

Тогда $\psi_n \rightarrow \psi$ и производная $\psi'_n = -n\omega'_n$ вполне монотонна. По теореме о среднем $\psi_n(\lambda) = \lambda \psi'_n(\theta\lambda) \geq \lambda \psi'_n(\lambda)$, и так как $\psi_n \rightarrow \psi$, то последовательность $\{\psi'_n(\lambda)\}$ ограничена при каждом фиксированном $\lambda > 0$. Следовательно, из нее можно выбрать сходящуюся подпоследовательность, предел которой автоматически будет вполне монотонной функцией (по обобщенной теореме непрерывности). Таким образом, ψ является интегралом от вполне монотонной функции. Доказательство закончено. ►

Другой формой этой теоремы служит

Теорема 2. *Функция ω представляет собой преобразование Лапласа безгранично делимого распределения тогда и только тогда, когда она имеет вид $\omega = e^{-\psi}$, где*

$$\psi(\lambda) = \int_0^{\infty} \frac{1 - e^{-\lambda x}}{x} P\{dx\} \quad (7.2)$$

и P такая мера, что

$$\int_1^{\infty} \frac{1}{x} P\{dx\} < \infty. \quad (7.3)$$

Доказательство. В силу теоремы о представлении вполне монотонных функций условия теоремы 1 могут быть заменены следующими: должно быть $\psi(0) = 0$ и

$$\psi'(\lambda) = \int_0^{\infty} e^{-\lambda x} P\{dx\}, \quad (7.4)$$

где P — некоторая мера. Заменяя в интеграле верхний предел числом a , легко получим, что при любом $a > 0$

$$\psi(\lambda) \geq \int_0^a \frac{1 - e^{-\lambda x}}{x} P\{dx\}. \quad (7.5)$$

Отсюда вытекает, что правая часть (7.2) имеет смысл, и условие (7.3) выполнено. Формальное дифференцирование показывает, что правая часть (7.2) представляет собой первообразную от (7.4), равную в нуле нулю. ►

(См. задачи 17—23 и пример 9, а.)

Примеры. а) Обобщенное распределение Пуассона

$$U = e^{-c} \sum_{n=0}^{\infty} \frac{c^n}{n!} F^{n*} \quad (7.6)$$

имеет преобразование Лапласа, равное $e^{-c+c\psi}$, и (7.2) верно с $P\{dx\} = cx F\{dx\}$,

б) *Гамма-плотность*. $x^{\alpha-1}e^{-x}/\Gamma(\alpha)$ имеет преобразование $\omega(\lambda) = 1/(\lambda+1)^\alpha$. Здесь

$$\psi(\lambda) = a \int_0^\infty \frac{1-e^{-\lambda x}}{x} e^{-x} dx, \quad (7.7)$$

так как $\psi'(\lambda) = a(\lambda+1)^{-1} = \omega'(\lambda)/\omega(\lambda)$.

в) *Устойчивые распределения*. Для найденного в § 6 преобразования $\omega(\lambda) = e^{-\lambda^\alpha}$ имеем $\psi(\lambda) = \lambda^\alpha$ и

$$\lambda^\alpha = \frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty \frac{1-e^{-\lambda x}}{x^{\alpha+1}} dx, \quad (7.8)$$

что опять можно проверить дифференцированием.

г) *Бесселевы функции*. Рассмотрим плотность v_r из примера 3, г) с преобразованием Лапласа (3.10). Из вида преобразования ясно, что v_r равна n -кратной свертке плотности $v_{r/n}$ с собой, т. е. v_r безгранично делима. Формальное дифференцирование показывает, что в этом случае $\psi'(\lambda) = r/\sqrt{(\lambda+1)^2-1}$ и, как нетрудно проверить (см. задачу 6), что ψ' представима в форме (7.4) с

$$P\{dx\} = re^{-x}I_0(x)dx.$$

д) *Подчиненность*. Если ψ_1 и ψ_2 — положительные функции с вполне монотонными производными, то, применяя критерии § 4, легко видеть, что функция $\psi(\lambda) = \psi_1(\psi_2(\lambda))$ имеет те же самые свойства. Соответствующее безгранично делимое распределение представляет особый интерес. Чтобы понять это, обозначим через $Q_i^{(i)}$ распределение вероятностей с преобразованием Лапласа $e^{-i\psi_i(\lambda)}$ (где $i=1, 2$) и положим

$$U_t(x) = \int_0^\infty Q_s^{(2)}(x) Q_t^{(1)}\{ds\}. \quad (7.9)$$

(Результирующее распределение U_t получается рандомизацией параметра s в $Q_s^{(2)}$.) Преобразование Лапласа для U_t равно

$$\omega_t(\lambda) = \int_0^\infty e^{-s\psi_2(\lambda)} Q_t^{(1)}\{ds\} = e^{-t\psi_1(\lambda)}. \quad (7.10)$$

Читатель, знакомый с гл. X, 7, легко свяжет (7.9) с *подчиненностью процессов*: U_t подчинено $Q_s^{(2)}$ с управляющим процессом $Q_t^{(1)}$. Мы очень легко получили формулу для преобразования Лапласа нового процесса, хотя и при специальном предположении, что $Q_t^{(2)}$ сосредоточено на положительной полуоси.

Заслуживает внимания частный случай, когда $\psi_1(\lambda) = \lambda^\alpha$ и $\psi_2(\lambda) = \lambda^\beta$. Тогда $\psi(\lambda) = \lambda^{\alpha\beta}$. Таким образом, *устойчивый α -процесс, управляемый устойчивым β -процессом, будет устойчивый $\alpha\beta$ -процесс*.

Читатель может проверить, что, по существу, это утверждение повторяет формулировку задачи 10. Более общее предложение см. в примере 2, э) гл. VI. ►

е) Любая смесь показательных распределений безгранично делима ¹⁾. В наиболее общем случае такое распределение имеет плотность вида

$$f(x) = \int_0^{\infty} se^{-sx} U\{ds\}, \quad (7.11)$$

где U — некоторое распределение вероятностей. Для частного случая, когда U сосредоточено в конечном числе точек $0 < a_1 < \dots < a_n$, в примере 3, е) было показано, что преобразование Лапласа имеет вид

$$\varphi(\lambda) = C \cdot \frac{\lambda + b_1}{\lambda + a_1} \dots \frac{\lambda + b_{n-1}}{\lambda + a_{n-1}} \cdot \frac{1}{\lambda + a_n} \quad (7.12)$$

с $a_k < b_k < a_{k+1}$. Далее, функция

$$-\frac{d}{d\lambda} \log \frac{\lambda + b_k}{\lambda + a_k} = \frac{1}{\lambda + a_k} - \frac{1}{\lambda + b_k} = \frac{b_k - a_k}{(\lambda + a_k)(\lambda + b_k)} \quad (7.13)$$

является произведением двух вполне монотонных функций и поэтому сама вполне монотонна. Отсюда следует, что каждый множитель в (7.12) безгранично делим и потому то же самое справедливо для φ . Утверждение для смесей общего вида получается с помощью обычного перехода к пределу (см. задачи 20—23). ►

§ 8*). МНОГОМЕРНЫЙ СЛУЧАЙ

Понятие преобразования Лапласа распространяется на многомерный случай очевидным образом. Даже формула (1.1) не требует изменения, если интерпретировать x как вектор-столбец $(x_1, \dots, x_n)$, λ — как вектор-строку $(\lambda_1, \dots, \lambda_n)$, а $\lambda x = \lambda_1 x_1 + \dots + \lambda_n x_n$ — как скалярное произведение λ и x . В теории вероятностей многомерные преобразования используются в ограниченной степени.

Примеры. а) *Резольвентное уравнение.* Пусть f — непрерывная функция одной переменной с обычным преобразованием Лапласа $\varphi(\lambda)$. Рассмотрим функцию $f(s+t)$ от двух переменных s и t . Ее двумерное преобразование Лапласа равно

$$\omega(\lambda, \nu) = \int_0^{\infty} \int_0^{\infty} e^{-\lambda s - \nu t} f(s+t) ds dt. \quad (8.1)$$

¹⁾ Это неожиданное наблюдение было сделано Стейтелем, см. F. W. Steutel, Ann. Math. Statist., vol. 40 (1969), 1130—1131, и vol. 38 (1967), 1303—1305.

^{*)} Материал не используется в дальнейшем.

После замены переменных $s + t = x$, $-s + t = y$ интеграл сводится к

$$\frac{1}{2} \int_0^{\infty} e^{-\frac{1}{2}(\lambda+v)x} f(x) dx \int_{-x}^x e^{\frac{1}{2}(\lambda-v)y} dy = \frac{1}{\lambda-v} \int_0^{\infty} (e^{-vx} - e^{-\lambda x}) f(x) dx.$$

Таким образом,

$$\omega(\lambda, v) = -\frac{\varphi(\lambda) - \varphi(v)}{\lambda - v}. \quad (8.2)$$

Мы встретимся с этим соотношением в более «возвышенной» обстановке, как с основным *резольвентным уравнением* для полугрупп [см. (10.5)].

б) *Функции Миттаг-Леффлера*. Этот пример иллюстрирует пользу многомерных преобразований как технического средства вычисления некоторых простых одномерных преобразований. Мы докажем сейчас следующее *утверждение*:

Если F — устойчивое распределение с преобразованием Лапласа $e^{-\lambda^\alpha}$, то (при фиксированном t) распределение

$$G_t(x) = 1 - F\left(\frac{t}{x^{1/\alpha}}\right), \quad x > 0, \quad (8.3)$$

имеет своим преобразованием Лапласа функцию Миттаг-Леффлера

$$\sum_{k=0}^{\infty} \frac{(-\lambda)^k}{\Gamma(1+k\alpha)} t^{k\alpha}. \quad (8.4)$$

Этот результат довольно интересен, так как в различных предельных теоремах распределение G встречается вместе с F [см., например, гл. IX, (5.6)]. Прямая проверка кажется трудной, однако можно поступить следующим образом. Сначала фиксируем x , а t примем в качестве переменной. Обычное преобразование Лапласа $\gamma_x(v)$ (v — переменная) для $G_t(x)$ равно, очевидно, $(1 - e^{-v^\alpha x})/v$. С точностью до множителя $1/v$ это есть функция распределения (по x), и ее преобразование Лапласа равно

$$\frac{v^{\alpha-1}}{\lambda + v^\alpha}. \quad (8.5)$$

Это есть не что иное, как двумерное преобразование для функции (8.3). Но оно могло бы быть вычислено и другим путем — интегрированием сначала по x , а затем по t . Поэтому (8.5) можно рассматривать, как преобразование по t того преобразования, которое мы ищем. Разлагая (8.5) в геометрический ряд, видим, что (8.5) действительно является преобразованием для (8.4). Утверждение доказано.

Функция Миттаг-Леффлера (8.4) обобщает показательную функцию, к которой она сводится при $\alpha = 1$. ►

§ 9. ПРЕОБРАЗОВАНИЯ ЛАПЛАСА ДЛЯ ПОЛУГРУПП

Понятие преобразования Лапласа может быть распространено на абстрактные функции и интегралы ¹⁾. Однако мы рассмотрим только преобразования Лапласа для полугрупп операторов, связанных с марковскими процессами ²⁾. Мы вернемся к основным обозначениям и соглашениям гл. X, 8.

Пусть Σ — некоторое пространство (например, прямая, интервал или совокупность всех целых чисел), и пусть $\mathcal{L}$ — банахово пространство ограниченных функций на Σ с нормой $\|u\| = \sup |u(x)|$. Мы предполагаем, что из $u \in \mathcal{L}$ вытекает $|u| \in \mathcal{L}$. Пусть $\{\Omega(t), t > 0\}$ — непрерывная полугруппа сжатий $\mathcal{L}$. Иными словами, мы предполагаем, что для каждой функции $u \in \mathcal{L}$ существует функция $\Omega(t)u \in \mathcal{L}$ и что $\Omega(t)$ обладает следующими свойствами: из $0 \leq u \leq 1$ вытекает $0 \leq \Omega(t)u \leq 1$; $\Omega(t+s) = \Omega(t)\Omega(s)$ и $\Omega(h) \rightarrow \Omega(0) = 1$ (единица обозначает тождественное преобразование ³⁾).

Мы начнем с определения интегрирования. Пусть F — произвольное распределение вероятностей на $\overline{0, \infty}$. Мы желаем определить сжимающий оператор E , отображающий $\mathcal{L}$ в $\mathcal{L}$, который будем обозначать

$$E = \int_0^{\infty} \Omega(s) F\{ds\} \quad (9.1)$$

и который удовлетворяет равенству

$$\Omega(t)E = E\Omega(t) = \int_0^{\infty} \Omega(t+s) F\{ds\}. \quad (9.2)$$

(Не нужно забывать о зависимости E от распределения F .)

Для полугруппы, отвечающей марковскому процессу с переходными вероятностями $Q_t(x, \Gamma)$, этот оператор E будет порождаться стохастическим или субстохастическим ядром

$$\int_0^{\infty} Q_s(x, \Gamma) F\{ds\}. \quad (9.3)$$

Естественное (и почти тривиальное) определение оператора E

¹⁾ Плодотворная теория, охватывающая преобразования вида (9.6), была развита С. Бокнером, Completely monotone functions in partially ordered spaces, Duke Math. J., 9 (1942), 519—526. Распространение на произвольные семейства операторов см. у Э. Хилле и Р. Филлипса (1957).

²⁾ Построение минимального решения в гл. XIV, 7, может служить типичным примером излагаемых далее методов.

³⁾ Напомним из гл. X, 8, что сильная сходимость $T_n \rightarrow T$ эндоморфизмов означает, что $\|T_n u - T u\| \rightarrow 0$ для всех $u \in \mathcal{L}$. Мы говорим кратко «непрерывность», подразумевая «сильную непрерывность при $t \geq 0$ ».

для случая атомической F приводит к желаемому общему определению с помощью простого предельного перехода.

Пусть $p_j \geq 0$ и $p_1 + \dots + p_r = 1$. Линейная комбинация

$$E = p_1 \Omega(t_1) + \dots + p_r \Omega(t_r) \quad (9.4)$$

снова является сжатием и может рассматриваться как математическое ожидание $\Omega(t)$ по отношению к распределению вероятностей, приписываемому вес p_j в точке t_j . Таким путем определяется оператор (9.1) для специального случая дискретных распределений с конечным числом атомов. При этом (9.2) оказывается выполненным. В общем случае оператор (9.1) определяется предельным переходом, подобно тому, как это делается в теории интеграла Римана: интервал $0, \infty$ разбивают на интервалы $I_1, \dots, I_n$; выбирают $t_j \in I_j$ и образуют риманову сумму $\sum \Omega(t_k) F\{I_k\}$, которая определяет некоторое сжатие. Ввиду равномерной непрерывности [см. (8.7) гл. X] известное доказательство сходимости римановых сумм применимо без изменений. Таким путем (9.1) определяется как частный случай интеграла Бохнера.

Если полугруппа образована операторами перехода, т. е. если $\Omega(t)1 = 1$ для всех t , то $E1 = 1$. Обозначение (9.1) будет использоваться для E , а для функции Ew мы будем использовать обычную запись:

$$Ew = \int_0^\infty \Omega(s) w \cdot F\{ds\} \quad (9.5)$$

(хотя более логично было бы писать w вне знака интеграла). Значение функции $Ew(x)$ в данной точке x — это обычное математическое ожидание числовой функции $\Omega(s)w(x)$ по отношению к F .

В частном случае $F\{ds\} = e^{-\lambda s} ds$ оператор E называют *интегралом Лапласа полугруппы* или *резольвентой*. Этот оператор мы будем обозначать

$$\Re(\lambda) = \int_0^\infty e^{-\lambda s} \Omega(s) ds, \quad \lambda > 0. \quad (9.6)$$

Ввиду (9.2) оператор (резольвента) $\Re(\lambda)$ перестановочен с любым оператором $\Omega(t)$ полугруппы. Для выполнения равенства $\lambda \Re(\lambda)1 = 1$ необходимо и достаточно, чтобы $\Omega(t)1 = 1$ при всех t . Таким образом, оператор сжатия $\lambda \Re(\lambda)$ является оператором перехода тогда и только тогда, когда все $\Omega(s)$ суть операторы перехода.

Лемма. Задание $\Re(\lambda)w$ при всех $\lambda > 0$ и $w \in \mathcal{E}$ однозначно определяет полугруппу.

Доказательство. Значение

$$\Re(\lambda) w(x) = \int_0^{\infty} e^{-\lambda t} \Omega(t) w(x) \cdot dt$$

в каждой фиксированной точке x представляет собой обычное преобразование Лапласа числовой функции t , заданной как $\Omega(t) w(x)$. Эта функция, будучи непрерывной, однозначно определяется своим преобразованием Лапласа (см. следствие в § 1). Таким образом, $\Omega(t) w$ однозначно определяется при всех t и всех $w \in \mathcal{L}$. ▶

Преобразование Лапласа (9.6) приводит к простой характеристизации *инфинитезимального оператора* $\Re$ рассматриваемой полугруппы. По определению этого оператора [см. гл. X, 10] имеем

$$\frac{\Omega(h) - 1}{h} u \rightarrow \Re u, \quad h \rightarrow 0+, \quad (9.7)$$

если $\Re u$ существует (т. е. если норма разности правой и левой части (9.7) стремится к нулю).

Теорема 1. При фиксированном $\lambda > 0$

$$u = \Re(\lambda) w \quad (9.8)$$

тогда и только тогда, когда u принадлежит области определения $\Re u$

$$\lambda u - \Re u = w. \quad (9.9)$$

Доказательство. (i) Определим u равенством (9.8). Опираясь на свойство (9.2) математических ожиданий, получаем

$$\frac{\Omega(h) - 1}{h} u = \frac{1}{h} \int_0^{\infty} e^{-\lambda s} \Omega(s+h) w \cdot ds - \frac{1}{h} \int_0^{\infty} e^{-\lambda s} \Omega(s) w \cdot ds. \quad (9.10)$$

Замена переменной $s+h=t$ в первом интеграле приводит к

$$\begin{aligned} \frac{\Omega(h) - 1}{h} u &= \frac{e^{\lambda h} - 1}{h} \int_0^{\infty} e^{-\lambda t} \Omega(t) w \cdot dt - \frac{1}{h} \int_0^h e^{\lambda(h-t)} \Omega(t) w \cdot dt = \\ &= \frac{e^{\lambda h} - 1}{h} (u - \lambda^{-1} w) - \frac{1}{h} \int_0^h e^{\lambda(h-t)} (\Omega(t) w - w) dt. \end{aligned} \quad (9.11)$$

Так как $\|\Omega(t) w - w\| \rightarrow 0$ при $t \rightarrow 0$, то второй член справа стремится по норме к нулю. Поэтому вся правая часть стремится к $\lambda(u - \lambda^{-1} w)$. Таким образом, (9.9) верно.

(ii) Обратно, предположим, что $\Re u$ существует, т. е. что выполняется (9.7). Так как оператор сжатия $\lambda \Re(\lambda)$ перестановочен

со всеми операторами полугруппы, то из (9.7) вытекает

$$\frac{\Omega(h)-1}{h} \mathfrak{N}(\lambda) u \rightarrow \mathfrak{N}(\lambda) \mathfrak{A}u. \quad (9.12)$$

Но, как мы только что видели, левая часть стремится к $\lambda \mathfrak{N}(\lambda) u = u$. В полученном в результате тождестве u представляется как преобразование Лапласа функции w из (9.9). ►

Следствие 1. Для данного $w \in \mathcal{L}$ существует ровно одно решение u уравнения (9.9).

Следствие 2. Две различные полугруппы не могут иметь один и тот же инфинитезимальный оператор $\mathfrak{A}$.

Доказательство. Знание оператора Ω дает возможность найти преобразование Лапласа $\mathfrak{N}(\lambda)w$ для всех $w \in \mathcal{L}$. По доказанной выше лемме это однозначно определяет все операторы полугруппы.

Соблазнительно было бы получить тауберовы теоремы, аналогичные приведенным в § 5. Однако мы ограничимся совсем простым результатом.

Теорема 2. При $\lambda \rightarrow \infty$

$$\lambda \mathfrak{N}(\lambda) \rightarrow 1. \quad (9.13)$$

Доказательство. Для произвольного $w \in \mathcal{L}$ имеем

$$\|\lambda \mathfrak{N}(\lambda) w - w\| \leq \int_0^{\infty} \|\Omega(t)w - w\| \cdot \lambda e^{-\lambda t} dt. \quad (9.14)$$

При $\lambda \rightarrow \infty$ распределение вероятностей с плотностью $\lambda e^{-\lambda t}$ сходится к распределению, сосредоточенному в нуле. Подынтегральное выражение ограничено и стремится к нулю при $t \rightarrow 0$. Поэтому интеграл (9.14) стремится к нулю и (9.13) верно. ►

Следствие 3. Инфинитезимальный оператор $\mathfrak{A}$ имеет область определения, плотную в $\mathcal{L}$.

Доказательство. Из (9.13) следует, что каждое $w \in \mathcal{L}$ может быть представлено как сильный предел последовательности элементов $\lambda \mathfrak{N}(\lambda)w$. По теореме 1 эти элементы принадлежат области определения $\mathfrak{A}$. ►

Примеры. а) *Безгранично делимые полугруппы.* Пусть U — безгранично делимое распределение с преобразованием Лапласа $\omega = e^{-\psi(\lambda)}$, где $\psi(\lambda)$ дано в (7.2). Распределение U_t с преобразованием Лапласа

$$\int_0^{\infty} e^{-\lambda x} U_t \{dx\} = e^{-t\psi(\lambda)} = \exp \left(-t \int_0^{\infty} \frac{1-e^{-\lambda x}}{x} P \{dx\} \right) \quad (9.15)$$

снова безгранично делимо, и соответствующие операторы свертки

$\mathfrak{A}(t)$ образуют полугруппу. Чтобы найти ее инфинитезимальный оператор¹⁾, возьмем какую-либо ограниченную непрерывно дифференцируемую функцию v . Тогда, очевидно,

$$\frac{\mathfrak{A}(t)-1}{t} v(x) = \int_0^{\infty} \frac{v(x-y)-v(x)}{y} \cdot \frac{1}{t} y U_t \{dy\}. \quad (9.16)$$

Дифференцирование (9.15) показывает, что мера $t^{-1}yU_t\{dy\}$ имеет преобразование $\psi'(\lambda)e^{-t\psi(\lambda)}$, которое стремится к $\psi'(\lambda)$ при $t \rightarrow 0$. Но ψ' есть преобразование меры P , так что наша мера сходится к P . Первая дробь под знаком интеграла (9.16) является (при фиксированном x) ограниченной и непрерывной функцией от y . Поэтому мы получаем соотношение

$$\mathfrak{A}v(x) = \int_0^{\infty} \frac{v(x-y)-v(x)}{y} P\{dy\}. \quad (9.17)$$

Таким образом, мы имеем новую интерпретацию меры P , входящей в каноническое представление безгранично делимого распределения.

б) *Подчиненные полугруппы.* Пусть $\{\Omega(t)\}$ обозначает произвольную марковскую полугруппу, и пусть U_t — безгранично делимое распределение предыдущего примера. Рандомизацией параметра t можно получить, как объяснено в гл. X, 7, новую марковскую полугруппу $\{\Omega^*(t)\}$. В принятых здесь обозначениях

$$\Omega^*(t) = \int_0^{\infty} \Omega(s) U_t \{ds\}. \quad (9.18)$$

Полагая для краткости

$$V(s, x) = \frac{\Omega(s)-1}{s} v(x), \quad (9.19)$$

имеем

$$\frac{\Omega^*(t)-1}{t} v(x) = \int_0^{\infty} V(s, x) \cdot \frac{1}{t} s U_t \{ds\}. \quad (9.20)$$

Пусть v принадлежит области определения $\mathfrak{A}$. При фиксированном x функция V непрерывна всюду, включая точку $s=0$, поскольку $V(s, x) \rightarrow \mathfrak{A}v(x)$ при $s \rightarrow 0$. Мы видели в предыдущем примере, что при $t \rightarrow 0$ $t^{-1}sU_t\{ds\} \rightarrow P\{ds\}$. Таким образом, правая часть в (9.20) сходится к некоторому пределу. Поэтому $\mathfrak{A}^*v$

¹⁾ Этот вывод дан только как иллюстрация. Вид оператора известен из гл. IX и может быть получен переходом к пределу от обобщенных пуассоновских распределений.

существует и задается формулой

$$\mathfrak{A}^*v(x) = \int_0^\infty V(s, x) P\{ds\}. \quad (9.21)$$

Мы приходим к заключению, что области определения $\mathfrak{A}$ и $\mathfrak{A}^*$ совпадают и

$$\mathfrak{A}^* = \int_0^\infty \frac{\mathfrak{D}(s) - 1}{s} P\{ds\} \quad (9.22)$$

в том смысле, что (9.21) выполняется для всех v в области определения $\mathfrak{A}$. ▶

§ 10. ТЕОРЕМА ХИЛЛЕ — ИОСИДЫ

Известная и весьма полезная теорема Хилле—Иосиды дает характеризацию инфинитезимальных операторов произвольных полугрупп. Мы изложим ее применительно к случаю сжимающих полугрупп. Эта теорема утверждает, что свойства инфинитезимальных операторов, описанные в предыдущем параграфе, присущи только им.

Теорема 1. (Хилле—Иосида). *Оператор $\mathfrak{A}$ с областью определения $\mathcal{D}' \subset \mathcal{D}$ представляет собой инфинитезимальный оператор некоторой непрерывной сжимаемой полугруппы $\mathfrak{D}(t)$ на $\mathcal{D}$ ($\mathfrak{D}(0) = 1$) тогда и только тогда, когда он обладает следующими свойствами*

(i) *уравнение*

$$\lambda u - \mathfrak{A}u = w, \quad \lambda > 0, \quad (10.1)$$

имеет для каждого $w \in \mathcal{D}$ ровно одно решение u ;

(ii) *если $0 \leq w \leq 1$, то $0 \leq \lambda u \leq 1$;*

(iii) *область определения $\mathcal{D}'$ оператора $\mathfrak{A}$ плотна в $\mathcal{D}$.*

Мы уже знаем, что каждый инфинитезимальный оператор обладает этими свойствами, т. е. условия теоремы необходимы. Далее обозначим решение (10.1) через $u = \mathfrak{R}(\lambda)w$. Тогда, как мы знаем, оператор $\mathfrak{R}(\lambda)$ совпадает с преобразованием Лапласа (9.6). Условия теоремы мы можем сформулировать теперь следующим образом.

(i') Оператор $\mathfrak{R}(\lambda)$ удовлетворяет соотношению

$$\lambda \mathfrak{R}(\lambda) - \mathfrak{A} \mathfrak{R}(\lambda) = 1. \quad (10.2)$$

Область определения $\mathfrak{R}(\lambda)$ совпадает с $\mathcal{D}$, а область значений — с $\mathcal{D}'$, т. е. с областью определения $\mathfrak{A}$.

(ii') Оператор $\lambda \mathfrak{R}(\lambda)$ является сжатием.

(iii') Область значений $\mathfrak{R}(\lambda)$ плотна в $\mathcal{D}$.

Как известно из теоремы 2 § 9, оператор $\mathfrak{R}(\lambda)$ должен удовлетворять соотношению

$$\lambda \mathfrak{R}(\lambda) \rightarrow 1, \quad \lambda \rightarrow \infty. \quad (10.3)$$

Из него вытекает, что каждая функция u является пределом своих собственных преобразований и, следовательно, область $\mathcal{L}'$ значений $\mathfrak{N}(\lambda)$ плотна в $\mathcal{L}$. Отсюда ясно, что (10.3) может заменить (iii'). Таким образом, три условия теоремы эквивалентны условиям (i'), (ii'), (10.3).

Пусть теперь задано семейство операторов $\mathfrak{N}(\lambda)$ с этими свойствами. Мы намерены построить искомую полугруппу как предел семейства псевдопуассоновских полугрупп. При построении используется

Лемма 1. Если w принадлежит $\mathcal{L}'$ — области определения оператора $\mathfrak{N}$, то

$$\mathfrak{N}\mathfrak{N}(\lambda)w = \mathfrak{N}(\lambda)\mathfrak{N}w. \quad (10.4)$$

Операторы $\mathfrak{N}(\lambda)$ и $\mathfrak{N}(\nu)$ перестановочны и удовлетворяют резольвентному уравнению

$$\mathfrak{N}(\lambda) - \mathfrak{N}(\nu) = (\nu - \lambda) \mathfrak{N}(\lambda) \mathfrak{N}(\nu). \quad (10.5)$$

Доказательство. Положим $v = \mathfrak{N}u$. Поскольку как u , так и w принадлежат области $\mathcal{L}'$, то из (10.1) следует, что то же самое справедливо для v , и

$$\lambda v - \mathfrak{N}v = \mathfrak{N}w.$$

Таким образом, получим $v = \mathfrak{N}(\lambda) \mathfrak{N}w$, т. е. то же самое, что и в правой части (10.4).

Далее, определим z как единственное решение уравнения $\nu z - \mathfrak{N}z = w$. Вычитая это равенство из (10.1), после простых преобразований получим

$$\lambda(u - z) - \mathfrak{N}(u - z) = (\nu - \lambda)z,$$

что совпадает с (10.5). Перестановочность операторов вытекает из симметрии предыдущего соотношения. ►

Для построения нашей полугруппы вспомним (см. теорему 1 в гл. X, 9), что произвольному оператору сжатия T и $a > 0$ соответствует полугруппа сжатий вида

$$e^{at(T-1)} = e^{-at} \sum_{n=0}^{\infty} \frac{(at)^n}{n!} T^n. \quad (10.6)$$

Порождает эту полугруппу оператор $a(T-1)$, который представляет собой эндоморфизм.

Применим этот результат для $T = \lambda \mathfrak{N}(\lambda)$. Положим для сокращения обозначений

$$\mathfrak{A}_\lambda = \lambda[\lambda \mathfrak{N}(\lambda) - 1] = \lambda \mathfrak{N}(\lambda), \quad \mathfrak{Q}_\lambda(t) = e^{t \mathfrak{A}_\lambda}. \quad (10.7)$$

Эти операторы, определенные при $\lambda > 0$, перестановочны друг

с другом и при фиксированном λ оператор $\mathfrak{A}_\lambda$ порождает псевдопуассоновскую полугруппу сжатий $\mathfrak{A}_\lambda(t)$.

Из (10.4) следует, что $\mathfrak{A}_\lambda u \rightarrow \mathfrak{A}u$ при всех u из $\mathscr{D}'$, области определения данного оператора $\mathfrak{A}$.

Мы можем забыть теперь о специальном способе построения $\mathfrak{A}_\lambda$ и получить теорему Хилле—Иосиды как частный случай более общей предельной теоремы, которая полезна и сама по себе. В этой теореме λ может пробегать и множество целых чисел.

Лемма 2 (лемма об аппроксимации). Пусть $\{\mathfrak{D}_\lambda(t)\}$ — семейство псевдопуассоновских полугрупп, перестановочных одна с другой и порожденных эндоморфизмами $\mathfrak{A}_\lambda$.

Если $\mathfrak{A}_\lambda u \rightarrow \mathfrak{A}u$ на плотном множестве $\mathscr{D}'$, то

$$\mathfrak{D}_\lambda(t) \rightarrow \mathfrak{D}(t), \quad \lambda \rightarrow \infty, \quad (10.8)$$

где $\{\mathfrak{D}(t)\}$ — полугруппа сжатий, с инфинитезимальным оператором, совпадающим на $\mathscr{D}'$ с $\mathfrak{A}$.

Далее для $u \in \mathscr{D}'$

$$\|\mathfrak{D}(t)u - \mathfrak{D}_\lambda(t)u\| \leq t \|\mathfrak{A}u - \mathfrak{A}_\lambda u\|. \quad (10.9)$$

Доказательство. Для любых двух перестановочных сжатий верно тождество

$$S^n - T^n = (S^{n-1} + \dots + T^{n-1})(S - T),$$

и потому

$$\|S^n u - T^n u\| \leq n \|Su - Tu\|. \quad (10.10)$$

В применении к операторам $\mathfrak{D}_\lambda(t/n)$ это неравенство дает (после очевидных преобразований)

$$\|\mathfrak{D}_\lambda(t)u - \mathfrak{D}_\nu(t)u\| \leq t \left\| \frac{\mathfrak{D}_\lambda(t/n) - 1}{t/n} u - \frac{\mathfrak{D}_\nu(t/n) - 1}{t/n} u \right\|. \quad (10.11)$$

Полагая $n \rightarrow \infty$, получаем

$$\|\mathfrak{D}_\lambda(t)u - \mathfrak{D}_\nu(t)u\| \leq t \|\mathfrak{A}_\lambda u - \mathfrak{A}_\nu u\|. \quad (10.12)$$

Это показывает, что для $u \in \mathscr{D}'$ последовательность $\{\mathfrak{D}_\lambda(t)u\}$ равномерно сходится при $\lambda \rightarrow \infty$. Так как $\mathscr{D}'$ плотно в $\mathscr{D}$, то эта равномерная сходимости имеет место для всех u . Обозначая предел через $\mathfrak{D}(t)u$, мы видим, что $\mathfrak{D}(t)$ есть сжатие, удовлетворяющее (10.8). Полугрупповое свойство очевидно. Полагая в (10.12) $\nu \rightarrow \infty$, приходим к (10.9). Переписывая левую часть, как в (10.11), имеем

$$\left\| \frac{\mathfrak{D}(t) - 1}{t} u - \frac{\mathfrak{D}_\lambda(t) - 1}{t} u \right\| \leq \|\mathfrak{A}u - \mathfrak{A}_\lambda u\|. \quad (10.13)$$

Выберем λ столь большим, чтобы правая часть стала меньше ε . Для достаточно малых t второе разностное отношение в левой части отличается (по норме) от $\mathfrak{A}_\lambda u$ меньше чем на ε , а от $\mathfrak{A}u$ —

меньше чем на 3ε . Таким образом, для $u \in \mathcal{L}'$

$$\frac{\Omega(t) - 1}{t} u \rightarrow \mathfrak{U}u, \quad (10.14)$$

чем заканчивается доказательство. ►

Примеры. Диффузия. Пусть $\mathcal{L}$ — семейство непрерывных функций на прямой, обращающихся в нуль на $\pm\infty$. Чтобы перейти к привычным обозначениям, заменим λ на h^{-2} и положим $h \rightarrow 0$. Определим разностный оператор ∇_h формулой

$$\nabla_h u(x) = \frac{1}{h^2} \left[\frac{u(x+h) + u(x-h)}{2} - u(x) \right]. \quad (10.15)$$

Этот оператор имеет вид $h^{-2}(T-1)$, где T есть оператор перехода, и потому ∇_h порождает полугруппу $e^{t\nabla_h}$ операторов перехода (марковскую полугруппу). Операторы ∇_h перестановочны один с другим, и для функций, имеющих три ограниченные производные, $\nabla_h u \rightarrow \frac{1}{2}u''$ равномерно. Лемма 2 утверждает существование предельной полугруппы $\{\Omega(t)\}$, порождаемой таким оператором $\mathfrak{U}$, что $\mathfrak{U}u = \frac{1}{2}u''$ (по крайней мере для всех достаточно гладких u).

В этом частном случае, как мы знаем, $\{\Omega(t)\}$ есть полугруппа свертки с нормальными распределениями с дисперсией t , так что мы не получаем новой информации. Пример тем не менее указывает, как легко можно (иногда) установить существование полугруппы с заданным инфинитезимальным оператором. Проведенное выше рассуждение применимо, например, к более общим дифференциальным операторам и граничным условиям (см. задачи 24, 25). ►

Замечание по поводу резольвенты и полной монотонности. Теорема Хилле — Иосиды посвящена свойствам инфинитезимального оператора $\mathfrak{U}$, однако теорему можно сформулировать так, чтобы получить характеристику семейства $\{\Re(\lambda)\}$.

Теорема 2. (Альтернативная форма теоремы Хилле — Иосиды.) Для того чтобы семейство эндоморфизмов $\{\Re(\lambda); \lambda > 0\}$ было резольвентой полугруппы $\{\Omega(t)\}$ сжатий, необходимо и достаточно, чтобы а) резольвентное уравнение

$$\Re(\lambda) - \Re(\nu) = (\nu - \lambda) \Re(\lambda) \Re(\nu) \quad (10.16)$$

было удовлетворено, б) $\lambda \Re(\lambda)$ был сжатием и в) $\lambda \Re \lambda \rightarrow 1$ при $\lambda \rightarrow \infty$.

Доказательство. Уравнение (10.16) полностью совпадает с (10.5), тогда как условия б) и в) появляются выше в виде (ii) и (10.3). Поэтому все три условия необходимы.

Предполагая, что условия теоремы выполнены, определим оператор $\mathfrak{U}$ следующим образом. Выберем некоторое $\nu > 0$ и определим $\mathcal{L}'$ как область значений $\Re(\nu)$, т. е.: $u \in \mathcal{L}'$ тогда и только тогда, когда $u = \Re(\nu)w$ для некоторого $w \in \mathcal{L}$. Для такого u положим $\mathfrak{U}u = \lambda u - w$. Этим определен оператор $\mathfrak{U}$ в области определения $\mathcal{L}'$, который удовлетворяет тождеству

$$\nu \Re(\nu) - \mathfrak{U} \Re(\nu) = 1. \quad (10.17)$$

Покажем, что это тождество имеет место для всех λ , т. е.

$$\lambda \mathfrak{R}(\lambda) - \mathfrak{U} \mathfrak{R}(\lambda) = 1. \quad (10.18)$$

Левую часть (10.18) можно переписать в виде

$$(\lambda - \nu) \mathfrak{R}(\lambda) + (\nu - \mathfrak{U}) \mathfrak{R}(\lambda). \quad (10.19)$$

Используя (10.16) и тот факт, что $(\nu - \mathfrak{U}) \mathfrak{R}(\nu) = 1$, получим

$$(\nu - \mathfrak{U}) \mathfrak{R}(\lambda) = 1 + (\nu - \lambda) \mathfrak{R}(\lambda). \quad (10.20)$$

Отсюда и из (10.19) следует тождество (10.18). Таким образом, показано, что все условия теоремы Хилле—Иосиды удовлетворены, и это исчерпывает доказательство. ►

Предыдущая теорема показывает, что вся теория полугрупп строится на резольвентном уравнении (10.16) и поэтому интересно выяснить его смысл в терминах обычных преобразований Лапласа. Из (10.16) ясно, что $\mathfrak{R}(\lambda)$ непрерывно зависит от λ в том смысле, что $\mathfrak{R}(\nu) \rightarrow \mathfrak{R}(\lambda)$ при $\nu \rightarrow \lambda$. Однако мы можем сделать следующий шаг и определить производную $\mathfrak{R}'(\lambda)$:

$$\mathfrak{R}'(\lambda) = \lim_{\nu \rightarrow \lambda} \frac{\mathfrak{R}(\nu) - \mathfrak{R}(\lambda)}{\nu - \lambda} = -\mathfrak{R}^2(\lambda). \quad (10.21)$$

Та же самая процедура, повторенная еще раз, показывает, что правая часть в (10.21) имеет производную, равную $-2\mathfrak{R}(\lambda) \mathfrak{R}'(\lambda)$. Продолжая по индукции, получаем, что $\mathfrak{R}(\lambda)$ имеет производные $\mathfrak{R}^{(n)}(\lambda)$ всех порядков и

$$(-1)^n \mathfrak{R}^{(n)}(\lambda) = n! \mathfrak{R}^{n+1}(\lambda). \quad (10.22)$$

Пусть теперь u — произвольная функция на $\mathcal{X}$, такая, что $0 \leq u \leq 1$. Возьмем произвольную точку x и положим $\omega(\lambda) = \mathfrak{R}(\lambda) u(x)$. Правая часть в (10.22) представляет собой положительный оператор, норма которого не превосходит $n!/\lambda^{n+1}$ и поэтому функция ω вполне монотонна и $|\omega^{(n)}(\lambda)| \leq n!/\lambda^{n+1}$. Из следствия в § 4 теперь следует, что ω есть обычное преобразование Лапласа со значениями между 0 и 1. Если мы обозначим эту функцию через $\mathfrak{Q}(t) u(x)$, то она определяет $\mathfrak{Q}(t)$ — оператор сжатия. Сравнивая резольвентное уравнение (10.16) с (8.2), убеждаемся в том, что оно влечет полугрупповое свойство

$$\mathfrak{Q}(t+s) u(x) = \mathfrak{Q}(t) \mathfrak{Q}(s) u(x). \quad (10.23)$$

Итак, мы видим, что основные свойства теории полугрупп могут быть выведены из (10.16) при использовании только классических преобразований Лапласа для обычных функций. В частности, резольвентное уравнение оказывается только абстрактным эквивалентом элементарного примера 8, а).

Абстрактная теория лишь повторяет теоремы, касающиеся обычных преобразований Лапласа. Чтобы подчеркнуть это еще раз, мы докажем формулу обращения.

Теорема 3. При фиксированном $t > 0$ и $\lambda \rightarrow \infty$

$$\frac{(-1)^{n-1}}{(n-1)!} \mathfrak{R}^{(n-1)} \left(\frac{n}{t} \right) \left(\frac{n}{t} \right)^n \rightarrow \mathfrak{Q}(t). \quad (10.24)$$

Доказательство. Из определения $R(\lambda)$ как преобразования Лапласа для $\mathfrak{Q}(t)$ (см. равенство 9.6) следует, что

$$(-1)^n \mathfrak{R}^{(n)}(\lambda) = \int_0^\infty e^{-\lambda s} s^n \mathfrak{Q}(s) ds. \quad (10.25)$$

Левая часть (10.24) является интегралом от $\mathfrak{Q}(s)$ по мере с плотностью $\frac{e^{-ns/t} (ns/t)^{n-1}}{t(n-1)!}$, которой соответствуют математическое ожидание t и дис-

персия t^2/n . При $n \rightarrow \infty$ эта мера сходится к распределению, сосредоточенному в t , и ввиду непрерывности $\Sigma(s)$ отсюда вытекает (10.24), так же как и в случае обычных функций [формула (10.24)—то же самое, что (1.6) гл. VII].

§ 11. ЗАДАЧИ

1. Пусть F_q —геометрическое распределение, приписывающее массу q^n точке nq , $n=0, 1, \dots$. Покажите, что при $q \rightarrow 0$ распределение F_q стремится к показательному распределению $1-e^{-x}$ и что соответствующее преобразование Лапласа стремится к $1/(\lambda+1)$.

2. Покажите, что обычные преобразования Лапласа функций $\cos x$ и $\sin x$ суть $\lambda/(\lambda^2+1)$ и $1/(\lambda^2+1)$. Докажите, что выражение $(1+a^{-2})e^{-x}(1-\cos ax)$ есть плотность вероятности с преобразованием Лапласа $(1+a^2)(\lambda+1)^{-1} \times [(\lambda+1)^2+a^2]^{-1}$. *Указание.* Используйте формулу $e^{ix} = \cos x + i \sin x$ или два раза последовательно проинтегрируйте по частям.

3. Пусть ω —преобразование (1.5) меры U ; ω интегрируема на $0, 1$ и $1, \infty$ тогда и только тогда, когда функция $1/x$ интегрируема по отношению к U на $1, \infty$ и $0, 1$ соответственно.

4. *Равенство Парсеваля.* Если X и Y —независимые случайные величины с распределениями F и G и преобразованиями φ и γ , то преобразование для XY равно

$$\int_0^\infty \varphi(\lambda y) G\{dy\} = \int_0^\infty \gamma(\lambda y) F\{dy\}.$$

5. Пусть F —распределение с преобразованием φ . Если $a > 0$, то $\varphi(\lambda+a)/\varphi(a)$ является преобразованием распределения $e^{-ax}F\{dx\}/\varphi(a)$. Установите, что при фиксированном $t > 0$ функция $1) \exp[-t\sqrt{2\lambda+a^2+at}]$ есть преобразование безгранично делимого распределения с плотностью

$$\frac{t}{\sqrt{2\pi x^3}} \exp\left[-\frac{1}{2}\left(\frac{t}{\sqrt{x}} - a\sqrt{x}\right)^2\right].$$

6. Покажите, исходя из определения (7.1) гл. II, что *обычное преобразование Лапласа функции* $I_0(x)$ равно $\omega_0(\lambda) = 1/\sqrt{\lambda^2-1}$ при $\lambda > 1$. [Вспомните

$$\text{тождество } \binom{2n}{n} = \left(-\frac{1}{2}\right) \binom{2n}{n} (-4)^n.]$$

7. *Продолжение.* Покажите, что $I'_0 = I_1$ и что, следовательно, I_1 имеет обычное преобразование Лапласа, равное $\omega_1(\lambda) = \omega_0(\lambda) R(\lambda)$, где $R(\lambda) = \lambda - \sqrt{\lambda^2-1}$.

8. *Продолжение.* Покажите, что $2I'_n = I_{n-1} + I_{n+1}$ при $n=1, 2, \dots$. Отсюда по индукции выведите, что I_n имеет обычное преобразование $\omega_n(\lambda) = \omega_0(\lambda) R^n(\lambda)$.

9. Выведите из примера 3, д) посредством интегрирования, что $e^{1/\lambda} - 1$ есть *обычное преобразование Лапласа* функции $I_1(2\sqrt{x})/\sqrt{x}$.

10. Пусть X и Y —независимые случайные величины с преобразованиями Лапласа φ и $e^{-\lambda^\alpha}$ соответственно. Тогда $YX^{1/\alpha}$ имеет преобразование Лапласа $\varphi(\lambda^\alpha)$.

¹⁾ Эта формула встречается в применениях и неоднократно выводилась с помощью длинных вычислений.

11. Плотность f вероятностного распределения вполне монотонна тогда и только тогда, когда она представляет собой смесь показательных плотностей [т. е. если эта плотность имеет вид (7.11)]. *Указание.* Используйте задачу 3.

12. Проверьте формулу обращения (4.5) путем прямого вычисления для частных случаев $\varphi(\lambda) = 1/(\lambda + 1)$ и $\varphi(\lambda) = e^{-\lambda}$.

13. Покажите, что следствие в § 4 остается верным, если заменить f и $\varphi^{(n)}$ их абсолютными значениями.

14. Предполагая, что функция $e^{-x}I_n(x)$ монотонна на бесконечности, выведите из задачи 8, что

$$e^{-x}I_n(x) \sim \frac{1}{\sqrt{2\pi x}}, \quad x \rightarrow \infty.$$

15. Предположим, что $1 - \varphi(\lambda) \sim \lambda^{1-\rho} L(\lambda)$ при $\lambda \rightarrow 0$ ($\rho > 0$). Используя пример 5, в), покажите, что $1 - F^{n*}(x) \sim \rho x^{\rho-1} L(1/x)/\Gamma(\rho)$ при $x \rightarrow \infty$. [Сравните это с примером 8, в) из гл. VIII.]

16. В теореме 4 из § 5 достаточно, чтобы $u(x) \sim v(x)$, где функция v монотонна, начиная с некоторого места.

17. Каждое безгранично делимое распределение является пределом обобщенных пуассоновских распределений.

18. Допустим, что в каноническом представлении (7.2) некоторого безгранично делимого распределения $P(x) \sim x^c L(x)$ при $x \rightarrow \infty$, где $0 < c < 1$.

Тогда $1 - F(x) \sim \frac{c}{1-c} x^{c-1} L(x)$. [Продолжение см. в примере 4, г) гл. XVII.]

19. Пусть P — производящая функция некоторой целочисленной безгранично делимой случайной величины и φ — преобразование Лапласа некоторого распределения вероятностей. Тогда $P(\varphi)$ безгранично делима.

20. Преобразования Лапласа φ_n , соответствующие безгранично делимым распределениям, сходятся к преобразованию Лапласа φ некоторого распределения вероятностей тогда и только тогда, когда соответствующая мера P_n в каноническом представлении (7.2) сходится к P . Следовательно, предел последовательности безгранично делимых распределений сам безгранично делим.

21. Пусть F_n — смесь (7.11) показательных распределений, соответствующая «смешивающему» распределению U_n . Последовательность $\{F_n\}$ сходится к распределению вероятностей F тогда и только тогда, когда U_n сходится к распределению U . В таком случае F есть смесь, соответствующая «смешивающему» распределению U .

22. Распределение вероятностей с вполне монотонной плотностью безгранично делимо. *Указание.* Используйте задачи 11 и 21, а также пример 7, е).

23. Любая смесь геометрических распределений безгранично делима. *Указание.* Следуйте схеме примера 7, е).

24. Диффузия с поглощающим экраном. В примере § 10 примем $x > 0$ и положим в определении $\nabla_h u(x-h) = 0$ при $x-h \leq 0$. Покажите, что доказательство сходимости остается в силе, если $\mathcal{L}$ обозначает пространство непрерывных функций с $u(\infty) = 0$, $u(0) = 0$, причем последнее условие нельзя отбросить. Предельная полугруппа описана в примере 5, б) гл. X.

25. Отражающие экраны. В примере § 10 примем $x > 0$ и положим в определении $\nabla_h u(x-h) = u(x+h)$ при $x-h < 0$. Тогда $\nabla_h u$ сходится к пределу для каждой u с тремя ограниченными производными и с $u'(0) = 0$. Область $\mathcal{L}'$ (область определения оператора $\mathfrak{A}$) определяется этим граничным условием. Предельная полугруппа описана в примере 5, д) гл. X.

26. Влияние максимального члена при сходимости к устойчивым распределениям. Пусть $X_1, X_2, \dots$ — независимые случайные величины с одним и тем же распределением F , удовлетворяющим (6.4), т. е. принадлежащим области притяжения устойчивого закона G_α . Положим $S_n = X_1 + \dots + X_n$ и $M_n = \max[X_1, \dots, X_n]$. Докажите, что отношение S_n/M_n имеет преобразование

Лапласа $\omega_n(\lambda)$, сходящееся к ^{1})

$$\omega(\lambda) = \frac{e^{-\lambda}}{1 + \alpha \int_0^1 (1 - e^{-\lambda t}) t^{-\alpha-1} dt}, \quad (*)$$

Следовательно, $E(S_n/M_n) \rightarrow 1/(1-\alpha)$. Указание. Оценивая интеграл по области $X_j \leq X_1$, получаем

$$\omega_n(\lambda) = n e^{-\lambda} \int_0^\infty F\{dx\} \left(\int_0^\infty e^{-\lambda y/x} F\{dy\} \right)^{n-1}.$$

Произведем замену $y = tx$ и затем получим $x = a_n s$, где a_n удовлетворяет (6.5). Внутренний интеграл, как легко видеть, равен

$$1 - \frac{1 - F(a_n s)}{n[1 - F(a_n)]} - \frac{1}{n} \int_0^1 (1 - e^{-\lambda t}) \frac{F\{a_n dt\}}{1 - F(a_n)} + o\left(\frac{1}{n}\right) = 1 - \frac{s^{-\alpha} \psi(\lambda)}{n} + o\left(\frac{1}{n}\right),$$

где $\psi(\lambda)$ обозначает знаменатель в (*). Таким образом,

$$\omega_n(\lambda) \rightarrow e^{-\lambda} \int_0^\infty e^{-s-\alpha \psi(\lambda)} \cdot \frac{\alpha ds}{s^{\alpha+1}} = \omega(\lambda),$$

¹) Этот результат и его аналог для устойчивых распределений с показателем $\alpha \geq 1$ был получен Д. Дарлинггом в терминах характеристических функций, См. Trans. Amer. Math. Soc., 73 (1952), 95—107.

Эта глава может служить для дополнительного чтения после гл. XIII. Она охватывает несколько независимых друг от друга тем—от практических задач (§ 1, 2, 4, 5) до общих теорем существования § 7. Предельные теоремы § 3 иллюстрируют силу методов, развитых в связи с понятием правильно меняющихся функций. В последнем параграфе описываются средства анализа асимптотических свойств марковских процессов и времен первого прохождения для них.

§ 1. УРАВНЕНИЕ ВОССТАНОВЛЕНИЯ; ТЕОРИЯ

Вероятностные основы читатель может найти в гл. VI, 6, 7. Хотя теории восстановления целиком посвящена гл. XI, здесь мы излагаем независимый и значительно менее сложный подход. Сравнение методов и результатов поучительно. При известных основах теории преобразований Лапласа настоящий подход проще и прямо ведет к цели. Однако результат основной теоремы восстановления пока не удается получить с помощью преобразований Лапласа. С другой стороны, преобразования Лапласа позволяют легче получить предельные теоремы § 3 и явные решения типа, рассматриваемого в § 2.

Объектом нашего изучения будет интегральное уравнение

$$V(t) = G(t) + \int_0^t V(t-x) F\{dx\}, \quad (1.1)$$

в котором F и G —заданные монотонные непрерывные справа функции, равные нулю при $t < 0$. Мы рассмотрим их как функции распределения некоторых мер и предположим, что F не сосредоточено в нуле и что их преобразования Лапласа

$$\varphi(\lambda) = \int_0^\infty e^{-\lambda t} F\{dt\}, \quad \gamma(\lambda) = \int_0^\infty e^{-\lambda t} G\{dt\} \quad (1.2)$$

определены при $\lambda > 0$. Так же как и в предыдущей главе, все интервалы интегрирования считаются *замкнутыми*. Мы докажем, что (1.1) имеет ровно одно решение V ; это будет функция рас-

пределения, преобразование Лапласа ψ которой существует при всех $\lambda > 0$. Если G имеет плотность g , то V имеет плотность v , которая удовлетворяет интегральному уравнению

$$v(t) = g(t) + \int_0^t v(t-x) F\{dx\}, \quad (1.3)$$

получаемому из (1.1) дифференцированием.

Вспоминая правило свертки, мы получаем для преобразования Лапласа ψ распределения V (или обычного преобразования его плотности) соотношение $\psi = \gamma + \psi\varphi$, откуда формально следует

$$\psi(\lambda) = \frac{\gamma(\lambda)}{1 - \varphi(\lambda)}. \quad (1.4)$$

Чтобы доказать, что это формальное решение является преобразованием Лапласа некоторой меры (или плотности), мы рассмотрим отдельно три случая (из которых только первые два имеют вероятностное значение).

Случай (а). F есть распределение вероятностей, не сосредоточенное в нуле. Тогда $\varphi(0) = 1$ и $\varphi(\lambda) < 1$ при $\lambda > 0$. Соответственно ряд

$$\sum_0^\infty \varphi^n = \frac{1}{1 - \varphi} = \omega \quad (1.5)$$

сходится при $\lambda > 0$. Очевидно, что функция ω вполне монотонна и поэтому является преобразованием Лапласа некоторой меры U (теорема 1 из гл. XIII, 4). Далее, $\psi = \omega\gamma$ есть преобразование Лапласа для свертки $V = U \star G$, т. е.

$$V(t) = \int_0^t G(t-x) U\{dx\}. \quad (1.6)$$

Наконец, если G имеет плотность g , то V имеет плотность $v = U \star g$. Таким образом, мы доказали существование и единственность решения нашего интегрального уравнения. ►

Асимптотическое поведение V на бесконечности описывается тауберовой теоремой 2 из гл. XIII, 4. Рассмотрим типичный случай, когда $G(\infty) < \infty$ и F имеет конечное математическое ожидание μ . В окрестности нуля $\psi(\lambda) \sim \mu^{-1} G(\infty) \lambda^{-1}$, откуда следует, что

$$V(t) \sim \mu^{-1} G(\infty) \cdot t, \quad t \rightarrow \infty. \quad (1.7)$$

Из теоремы восстановления гл. XI, 1, вытекает более точный результат

$$V(t+h) - V(t) \rightarrow \mu^{-1} G(\infty) h,$$

но его нельзя получить из тауберовых теорем. [Последние дают лучший результат, если F не имеет математического ожидания; см. § 3.]

Случай (б). F — несобственное распределение, $F(\infty) < 1$. Допустим для простоты, что $G(\infty) < \infty$. Можно применить предшествующие рассуждения с заметным упрощением: так как $\varphi(0) = F(\infty) < 1$, то $\omega(0) < \infty$ и мера V на этот раз ограничена.

Случай (в). В этом случае $F(\infty) > 1$. Для малых λ знаменатель в (1.4) отрицателен, и $\omega(\lambda)$ для таких λ не может быть преобразованием Лапласа. К счастью, это обстоятельство не создает неудобств. Во избежание тривиальностей допустим, что F не имеет атома в нуле, так что $\varphi(\lambda) \rightarrow 0$ при $\lambda \rightarrow \infty$. В этом случае уравнение $\varphi(\kappa) = 1$ имеет единственный корень $\kappa > 0$. Рассуждения п. (а) применимы без изменений при $\lambda > \kappa$. Иными словами, существует единственное решение V , но его преобразование Лапласа ω сходится только при $\lambda > \kappa$. Для этих значений ω по-прежнему определяется формулой (1.4)¹.

§ 2. УРАВНЕНИЕ ТИПА УРАВНЕНИЯ ВОССТАНОВЛЕНИЯ: ПРИМЕРЫ

а) *Время ожидания пробелов в процессе Пуассона.* Пусть V обозначает распределение времени ожидания до окончания первого пробела длины ξ в процессе Пуассона с параметром c (т. е. в процессе восстановления с показательными временами между поступлениями). Распределение V изучалось аналитически в примере 7, б) гл. XI. Соответствующие практические ситуации (задержка пешехода или автомобиля, пытающегося пересечь поток автомашин, время блокировки в счетчиках Гейгера второго типа и т. п.) были описаны в гл. VI, 7. Здесь мы выведем нужное нам уравнение восстановления заново.

Время ожидания, если ожидание началось в момент 0, необходимо превосходит ξ . Оно меньше t , $t > \xi$, если до момента ξ не будет ни одного поступления (вероятность чего $e^{-c\xi}$) или же если первое поступление происходит в момент $x < \xi$ и остающееся время ожидания не превосходит $t - x$. В силу свойства отсутствия последствия вероятность $V(t)$ того, что время ожидания не превосходит t , равна

$$V(t) = e^{-c\xi} + \int_0^{\xi} V(t-x) \cdot e^{-cx} c dx \quad (2.1)$$

при $t \geq \xi$ и $V(t) = 0$ при $t < \xi$. Несмотря на непривычный вид, (2.1) есть уравнение восстановления стандартного типа (1.1),

¹) Решение V имеет вид $V\{dx\} = e^{\kappa x} V^*\{dx\}$, где V^* есть решение [с преобразованием Лапласа $\psi^*(\lambda) = \psi(\lambda + \kappa)$] стандартного уравнения восстановления (1.1) с распределением F , замененным на собственное распределение вероятностей $F^*\{dx\} = e^{-\kappa x} F\{dx\}$, и с G , замененным на $G^*\{dx\} = e^{-\kappa x} G\{dx\}$.

в котором F имеет плотность $f(x) = ce^{-cx}$ на $0 < x < \xi$, в то время как G сосредоточено в точке ξ . Таким образом,

$$\varphi(\lambda) = \frac{c}{c+\lambda} (1 - e^{-(c+\lambda)\xi}), \quad \gamma(\lambda) = e^{-(c+\lambda)\xi} \quad (2.2)$$

и, следовательно, преобразование ψ для V равно

$$\psi(\lambda) = \frac{(c+\lambda)e^{-(c+\lambda)\xi}}{\lambda + ce^{-(c+\lambda)\xi}}. \quad (2.3)$$

Выражения (7.8) гл. XI для математического ожидания и дисперсии получаются отсюда простым дифференцированием¹⁾. То же самое верно и для моментов высших порядков.

Поучителен вывод из (2.3) *явной формулы* для решения. По причинам, которые станут понятны, мы рассмотрим *хвост* $1 - V(t)$ распределения. Его обычное преобразование Лапласа равно $[1 - \psi(\lambda)]/\lambda$ [см. XIII, (2.7)]. Последняя функция может быть разложена в геометрический ряд:

$$\frac{1 - \psi(\lambda)}{\lambda} = \xi \sum_{n=1}^{\infty} c^{n-1} \xi^{n-1} \left\{ \frac{1 - e^{-(c+\lambda)\xi}}{(c+\lambda)\xi} \right\}^n. \quad (2.4)$$

Выражение в фигурных скобках отличается от преобразования Лапласа $(1 - e^{-\lambda})/\lambda$ равномерного распределения только масштабным множителем ξ и заменой λ на $\lambda + c$. Как неоднократно отмечалось, эта замена соответствует умножению плотности на e^{-ct} . Таким образом,

$$1 - V(t) = e^{-ct} \sum_{n=1}^{\infty} c^{n-1} \xi^{n-1} f^{n*} \left(\frac{t}{\xi} \right), \quad (2.5)$$

где f^{n*} есть плотность n -кратной свертки равномерного распределения с собой. Используя (9.6) гл. I, окончательно получаем

$$1 - V(t) = e^{-ct} \sum_{n=1}^{\infty} \frac{(ct)^{n-1}}{(n-1)!} \sum_{k=0}^n (-1)^k \binom{n}{k} \left(1 - k \frac{\xi}{t} \right)_+^{n-1}. \quad (2.6)$$

Интересна связь этого результата с *теоремами о покрытии*. Как было показано в (9.9) гл. I, при фиксированных t и ξ внутренняя сумма представляет собой вероятность того, что, *выбирая наудачу $n-1$ точек в $0, t$, мы разобьем этот интервал на n частей, каждая из которых не превышает ξ* . Отметим, что время ожидания превосходит t тогда и только тогда, когда каждый подынтервал $0, t$ длины ξ содержит по крайней мере один момент поступления. Таким образом, (2.6) утверждает: если в процессе Пуассона на $0, t$ происходит ровно $n-1$ поступлений, то *условное распределение моментов поступления равномерно*. Если исходить из этого факта, то (2.6) будет следствием теоремы о покрытии. С другой стороны, (2.6) дает новое доказательство теоремы о покрытии с помощью *рандомизации*.

б) Задача о разорении в обобщенном процессе Пуассона. В качестве второго иллюстративного примера мы возьмем *интегро-*

¹⁾ Чтобы избежать громоздкого дифференцирования дробей, знаменатель лучше убрать, перенеся его в левую часть.

дифференциальное уравнение

$$R'(t) = \frac{\alpha}{c} R(t) - \frac{\alpha}{c} \int_0^t R(t-x) F\{dx\}, \quad (2.7)$$

в котором F — распределение вероятностей с конечным математическим ожиданием μ . Это уравнение было получено в гл. VI, 5, где мы отметили его связь с теорией страхования, задачами хранения запасов и т. п. Решение этого уравнения и его асимптотические свойства были найдены другими методами в примере 7, а) гл. XI.

Задача состоит в том, чтобы найти *распределение вероятностей* R , удовлетворяющее (2.7). Это уравнение связано с уравнением восстановления и может быть исследовано теми же приемами. Переходя к обычным преобразованиям Лапласа и отмечая, что

$$\rho(\lambda) = \int_0^\infty e^{-\lambda x} R(x) dx = \frac{1}{\lambda} \int_0^\infty e^{-\lambda x} R'(x) dx + \frac{1}{\lambda} R(0), \quad (2.8)$$

мы находим

$$\rho(\lambda) = \frac{R(0)}{1 - \frac{\alpha}{c} \frac{1 - \varphi(\lambda)}{\lambda}} \cdot \frac{1}{\lambda}, \quad (2.9)$$

где φ — преобразование Лапласа для F . Вспоминая, что $[1 - \varphi(\lambda)]/\lambda$ есть преобразование Лапласа $1 - F(x)$, мы видим, что первая дробь в правой части имеет форму (1.4). Следовательно, эта дробь есть преобразование Лапласа — Стильтеса некоторой меры R . Множитель $1/\lambda$ соответствует интегрированию. Поэтому $\rho(\lambda)$ является *обычным* преобразованием Лапласа некоторой функции распределения $R(x)$ [см. (2.8)]. Так как $R(x) \rightarrow 1$ при $x \rightarrow \infty$, то по теореме 4 из гл. XIII, 5, $\rho(\lambda) \rightarrow 1$ при $\lambda \rightarrow 0$. Из (2.9) мы получаем теперь значение неизвестной постоянной $R(0)$:

$$R(0) = 1 - \frac{\alpha}{c} \mu. \quad (2.10)$$

В соответствии с этим наша задача имеет *единственное решение* при $\alpha\mu < c$ и не имеет решений при $\alpha\mu \geq c$. Этот результат следовало бы ожидать, исходя из вероятностных соображений.

Формула (2.9) встречается и в теории очередей (под названием формулы Хинчина — Полачека, см. пример 5, а) гл. XII). Много статей посвящено выводу явных формул в частных случаях. В случае *чисто пуассоновского процесса* F сосредоточено в точке 1 и $\varphi(\lambda) = e^{-\lambda}$. Выражение для ρ теперь почти такое же, как (2.3), и тот же метод легко дает *явную формулу решения*

$$R(x) = \left(1 - \frac{\alpha}{c}\right) \sum_{k=0}^{\infty} \left(\frac{-\alpha}{c}\right)^k \frac{(x-k)_+^k}{k!} \exp\left(\frac{\alpha}{a}(x-k)_+\right). \quad (2.11)$$

Эта формула, хотя она практически бесполезна, интересна тем, что в нее входят экспоненциальные функции с *положительными* показателями, которые должны занятым образом «компенсировать» друг друга. Формула была известна в связи с задачами страхования¹⁾ с 1934 г., но неоднократно перетривалась.

§ 3. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ, ВКЛЮЧАЮЩИЕ РАСПРЕДЕЛЕНИЯ АРКСИНУСА

Стало привычным называть распределения, сосредоточенные на $0, 1$ и имеющие плотности

$$q_{\alpha}(x) = \frac{\sin \pi \alpha}{\pi} x^{-\alpha} (1-x)^{\alpha-1}, \quad 0 < \alpha < 1, \quad (3.1)$$

«обобщенными распределениями арксинуса», хотя на самом деле это специальные бета-распределения. Частный случай $\alpha = 1/2$ соответствует функции распределения $2\pi^{-1} \arcsin \sqrt{x}$, играющей важную роль в теории флуктуаций в случайных блужданиях. Все большее число исследований приводит к предельным распределениям, связанным с q_{α} . Сложность вычисления этих распределений делает появление q_{α} мистическим. Подлинная причина заключается в близкой связи q_{α} с функциями распределения, имеющими правильно меняющиеся хвосты, т. е.

$$1 - F(x) = x^{-\alpha} L(x), \quad 0 < \alpha < 1, \quad (3.2)$$

где $L(tx)/L(t) \rightarrow 1$ при $t \rightarrow \infty$. Для таких функций теорема восстановления может быть дополнена в том смысле, что для «функции восстановления» $U = \sum F^n$ выполняется соотношение

$$U(t) \sim \frac{1}{\Gamma(1-\alpha) \Gamma(1+\alpha)} \frac{t^{\alpha}}{L(t)}, \quad t \rightarrow \infty. \quad (3.3)$$

Другими словами, если F меняется правильно, то это же верно и для U . Известно (хотя и не очевидно), что постоянный множитель в (3.3) равен $(\sin \pi \alpha)/\pi \alpha$, так что соотношение (3.3) может быть переписано в виде

$$[1 - F(x)] U(x) \rightarrow \frac{\sin \pi \alpha}{\pi \alpha}, \quad x \rightarrow \infty. \quad (3.4)$$

Лемма. Если F удовлетворяет (3.2), то выполняется (3.4).

Доказательство. По тауберовой теореме 4 гл. XIII, 5,

$$1 - \varphi(\lambda) \sim \Gamma(1-\alpha) \lambda^{\alpha} L(1/\lambda), \quad \lambda \rightarrow 0.$$

Преобразование Лапласа для U равно $\sum \varphi^n = 1/(1-\varphi)$, и по теореме 2 из гл. XIII, 5, верно (3.3). ►

¹⁾ Явное решение в задаче о разорении до момента t дано Пайком (Pyke R., The supremum and infimum of the Poisson process, Ann. Math. Statist., 30 (1959), 568—576).

Рассмотрим теперь последовательность положительных независимых случайных величин X_k с одним и тем же распределением F и их частные суммы $S_n = X_1 + \dots + X_n$. При фиксированном $t > 0$ обозначим через N_t случайный индекс, такой, что

$$S_{N_t} \leq t \leq S_{N_t+1}. \quad (3.5)$$

Мы интересуемся длинами двух подынтервалов

$$Y_t = t - S_{N_t} \text{ и } Z_t = S_{N_t+1} - t.$$

Они были введены в гл. VI, 7 и названы там: «прошедшее время ожидания» и «остающееся время ожидания» в момент t . Причины интереса, который могут представлять эти величины, объяснялись ранее по-разному. В гл. XI, 4, было доказано, что при $t \rightarrow \infty$ случайные величины Y_t и Z_t имеют (одно и то же) собственное предельное распределение тогда и только тогда, когда F имеет конечное математическое ожидание. В противном случае $P\{Y_t \leq x\} \rightarrow 0$ при любом фиксированном $x > 0$. Аналогичное соотношение верно и для Z_t . Побочным результатом наших исследований является следующая интересная теорема (первоначальное доказательство которой представляло значительные аналитические трудности¹⁾.

Теорема. Если имеет место (3.2), то нормированная случайная величина Y_t/t имеет предельную плотность q_α , определяемую по (3.1), а Z_t/t имеет предельную плотность²⁾

$$p_\alpha(x) = \frac{\sin \pi \alpha}{\pi} \cdot \frac{1}{x^\alpha (1+x)}, \quad x > 0. \quad (3.6)$$

Доказательство. Неравенство $tx_1 < Y_t < tx_2$ осуществляется тогда и только тогда, когда при некоторых n и y , $1 - x_2 < y < 1 - x_1$, имеем $S_n = ty$ и $X_{n+1} > t(1 - y)$. Суммируя по всем возможным n и y , получаем

$$P\{tx_1 < Y_t < tx_2\} = \int_{1-x_2}^{1-x_1} [1 - F(t(1-y))] U\{t dy\}, \quad (3.7)$$

или, используя (3.4),

$$P\{tx_1 < Y_t < tx_2\} \sim \frac{\sin \pi \alpha}{\pi \alpha} \int_{1-x_2}^{1-x_1} \frac{1 - F(t(1-y))}{1 - F(t)} \cdot \frac{U\{t dy\}}{U(t)}, \quad (3.8)$$

¹⁾ Дынкин Е. Б., Некоторые предельные теоремы для сумм независимых случайных величин с бесконечными математическими ожиданиями, Изв. АН СССР, серия матем., 19 (1955), 247—266.

²⁾ Так как $S_{N_t+1} = Z_t + t$, то распределение Z_t/S_{N_t+1} получается из (3.6) заменой переменных $x = y/(1-y)$. Таким образом, мы видим, что величина Z_t/S_{N_t+1} имеет предельную плотность q_α .

Теперь $\frac{U(ty)}{U(t)} \rightarrow y^\alpha$, так что мера $U\{t dy\}/U(t)$ сходится к мере с плотностью $\alpha y^{\alpha-1}$. В то же время первый множитель под интегралом сходится к $(1-y)^{-\alpha}$. В силу монотонности наших функций сходимость равномерна. Поэтому

$$P\{tx_1 < Y_t < tx_2\} \rightarrow \frac{\sin \pi\alpha}{\pi} \int_{1-x_2}^{1-x_1} y^{\alpha-1} (1-y)^{-\alpha} dy, \quad (3.9)$$

чем доказано первое утверждение. Для верности $P\{Z_t > ts\}$ мы получаем такой же интеграл, но в пределах от 0 до $1/(1+s)$. Теперь дифференцированием получаем (3.6). ►

Примечательно, что плотность q_α становится бесконечной вблизи точек 0 и 1. Следовательно, наиболее вероятные значения величины Y_t/t лежат вблизи точек 0 и 1.

Легко так усовершенствовать наши рассуждения, чтобы получить утверждения, обратные к лемме и теореме. При этом станет видно, что условие (3.2) необходимо для существования предельного распределения величины Y_t/t . С другой стороны, (3.2) характеризует области притяжения устойчивых распределений. Этим и объясняется частое появление q_α в связи с подобными распределениями.

§ 4. ПЕРИОДЫ ЗАНЯТОСТИ И СООТВЕТСТВУЮЩИЕ ВЕТВЯЩИЕСЯ ПРОЦЕССЫ

В примере 4, а) гл. XIII было показано, что если φ есть преобразование Лапласа некоторого распределения вероятностей F с математическим ожиданием μ , то уравнение

$$\beta(\lambda) = \varphi(\lambda + c - c\beta(\lambda)), \quad \lambda > 0, \quad (4.1)$$

имеет единственное решение β . Сверх того, β является преобразованием Лапласа некоторого распределения B ; при этом B собственно, если $c\mu \leq 1$, и несобственно в противном случае. Этот простой и красивый результат находит все больше и больше применений. Стоит поэтому объяснить вероятностный смысл уравнения (4.1) и его применения.

Вывод уравнения (4.1) и аналогичных уравнений оказывается простым, если выражать вероятностные соотношения непосредственно в терминах преобразований Лапласа. Типичная ситуация такова. Рассмотрим случайную сумму $S_N = X_1 + \dots + X_N$, где X_i независимы и имеют преобразование Лапласа $\gamma(\lambda)$ и N — не зависящая от X случайная величина с производящей функцией $P(s)$. Преобразование Лапласа S_N равно, очевидно, $P(\gamma(\lambda))$ [см. пример 3, с) гл. XIII]. Для пуассоновской случайной величины N это преобразование Лапласа имеет вид $e^{-\alpha[1-\gamma(\lambda)]}$. Как мы неоднократно видели, в применениях параметр α часто считают

случайным с некоторым распределением U . Используя ту же терминологию, что и для распределений, мы можем сказать, что $e^{-\alpha[1-\gamma(\lambda)]}$ есть условное преобразование Лапласа для S_N при данном значении α нашего параметра. Безусловное преобразование Лапласа получается интегрированием по отношению к U . Благодаря специальному виду подынтегрального выражения результат равен, очевидно, $\omega(1-\gamma(\lambda))$, где ω обозначает преобразование Лапласа для U .

Примеры. а) *Периоды занятости*¹⁾. Клиенты (или вызовы) приходят к обслуживающему прибору (линии связи) в моменты, образующие процесс Пуассона с параметром c . Последовательные времена обслуживания предполагаются независимыми случайными величинами с одним и тем же распределением F . Допустим, что в момент 0 приходит вызов и линия свободна. Обслуживание начинается мгновенно. Вызовы, поступающие в моменты, когда линия занята, встают в очередь, и обслуживание продолжается без перерывов до тех пор, пока не исчезнет очередь. Под *периодом занятости* понимается интервал от нуля до первого момента, когда линия снова становится свободной. Его продолжительность является случайной величиной. Мы обозначим ее распределение и преобразование Лапласа соответственно через B и β .

В терминологии ветвящихся процессов клиент, начинающий период занятости, есть «предок». Клиенты, прибывающие в моменты, когда все еще обслуживается первый, суть его прямые «потомки» и т. д. При условии, что предок уходит в момент x , число N его прямых потомков распределено по закону Пуассона с математическим ожиданием cx . Обозначим X_j полное время обслуживания j -го прямого потомка и всех его потомков. Хотя соответствующие промежутки времени не обязаны следовать непосредственно один за другим, их общая сумма имеет, очевидно, такое же распределение, как и период занятости. Полное время обслуживания всех (прямых и непрямых) потомков равно поэтому $S_N = X_1 + \dots + X_N$, где X_j имеет преобразование Лапласа β и все случайные величины независимы. Чтобы получить период занятости, мы должны добавить время x обслуживания предка. Соответственно при данной продолжительности x времени обслужи-

¹⁾ То, что уравнение (4.1) возникает в связи с периодами занятости, было отмечено Кендаллом (Kendall D. G., Some problems in the theory of queues, J. Roy. Statist. Soc. (B), 13 (1951), 151—185). Красивая редукция к ветвящимся процессам была предложена Гудом (Good I. J.). Уравнение (4.1) равносильно уравнению

$$B(t) = \sum_0^t \int_0^x e^{-cx} \frac{(cx)^n}{n!} B^{n*}(t-x) F\{dx\},$$

которое часто называют *интегральным уравнением Такача*. Внутреннюю простоту теории не всегда понимают,

вания предка длительность периода занятости $x + S_N$ имеет (условное) преобразование Лапласа $e^{-x[\lambda + c - c\beta_1(\lambda)]}$. Параметр x имеет распределение F . Интегрирование по x приводит к (4.1).

Если распределение B несобственно, то его дефект $1 - B(\infty)$ равен вероятности того, что период занятости никогда не кончится (переполнение). Условие $c\mu \leq 1$ означает, что математическое ожидание полного времени обслуживания всех вызовов, поступающих за единицу времени, не должно превышать единицы. Нетрудно вычислить, исходя из (4.1), математическое ожидание и дисперсию B .

В частном случае показательных времен обслуживания имеем $F(t) = 1 - e^{-\alpha t}$ и $\varphi(\lambda) = \alpha/(\lambda + \alpha)$. В этом случае (4.1) превращается в квадратное уравнение, один из корней которого не ограничен на бесконечности. Поэтому решение β совпадает с другим корнем, а именно

$$\beta(\lambda) = \sqrt{\frac{\alpha}{c}} \left[\frac{\lambda + \alpha + c}{2\sqrt{\alpha c}} - \sqrt{\left(\frac{\lambda + \alpha + c}{2\sqrt{\alpha c}} \right)^2 - 1} \right]. \quad (4.2)$$

Это преобразование Лапласа встречалось в примере 3, в) гл. XIII. Принимая во внимание изменение масштабного параметра и принцип сдвига, найдем, что соответствующая плотность равна

$$\sqrt{\alpha/c} e^{-(\alpha+c)x} x^{-1} I_1(2\sqrt{\alpha c}x). \quad (4.3)$$

Тот же самый результат будет получен другим методом в примере 6, б); кроме того, он был использован в примере 9, д) гл. VI.

б) *Задержки в уличном движении*¹⁾. Допустим, что автомобили, проходящие данный участок дороги, образуют процесс Пуассона с параметром c . Пусть движение задерживается на время δ (красным светом или какой-либо другой причиной). Когда движение возобновляется, в очереди будут стоять K автомобилей; K имеет распределение Пуассона с параметром $c\delta$. Так как r -й автомобиль в очереди не может двинуться ранее $r-1$ автомобилей, находящихся впереди него, то каждый автомобиль создает задержку для всех следующих за ним автомобилей. Естественно предположить, что отдельные задержки суть независимые случайные величины с одним и тем же распределением F . При наличии ожидающих вновь прибывший автомобиль обязан стать в очередь, тем самым увеличивая общую задержку. Положение такое же, как в предыдущем примере, с той лишь разницей, что здесь имеется K «предков». Полная задержка, создаваемая автомобилем и всеми его прямыми и непрямыми потомками, имеет преобразование Лапласа β , удовлетворяющее (4.1), а полный «период занятости» — промежуток времени от возобновления движения до первого мо-

¹⁾ Этот пример возник под влиянием изучения Дж. Литлом числа автомобилей, задержанных при движении более медленным транспортом [Little J. D. C., Operations Res., 9 (1961), 39–52].

мента, когда нет ожидающих автомобилей, — имеет преобразование Лапласа

$$e^{-c\delta} \sum \frac{(c\delta)^k}{k!} \beta^k(\lambda) = e^{-c\delta [1 - \beta(\lambda)]}.$$

Легко вычислить математическое ожидание задержки. Этот результат можно использовать для выяснения эффекта, создаваемого последовательными светофорами, и т. п. (см. задачи 6 и 7).

§ 5. ДИФФУЗИОННЫЕ ПРОЦЕССЫ

Как мы знаем, в одномерном броуновском движении переходные вероятности нормальны, а время первого прохождения имеет устойчивое распределение с показателем $\frac{1}{2}$ (см. пример 2, е) гл. VI). Имея в своем распоряжении эти результаты, мы не вправе ожидать новой информации от использования преобразований Лапласа. Причиной возвращения заново к диффузионным уравнениям служит то, что метод сам по себе поучителен и применим к наиболее общим диффузионным уравнениям (с той разницей, что *явные* решения не удастся получить при произвольных коэффициентах). Для упрощения записи мы допустим с самого начала, что переходные вероятности Q_t имеют плотности q_t (хотя метод, который мы укажем, был бы применим без всяких специальных ограничений).

Мы начнем с частного случая броуновского движения. Пусть f — данная непрерывная ограниченная функция. Положим

$$u(t, x) = \int_{-\infty}^{+\infty} q_t(x, y) f(y) dy. \quad (5.1)$$

Нашим исходным пунктом будет установленный в примере 4, а) гл. X факт, что (по крайней мере для достаточно гладких f) u удовлетворяет диффузионному уравнению

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \frac{\partial^2 u(t, x)}{\partial x^2} \quad (5.2)$$

и начальному условию $u(t, x) \rightarrow f(x)$ при $t \rightarrow 0$. Вводя обычное преобразование Лапласа

$$\omega_\lambda(x) = \int_0^\infty e^{-\lambda t} u(t, x) dt, \quad (5.3)$$

мы выводим из (5.2), что¹⁾

$$\lambda \omega_\lambda - \frac{1}{2} \omega_\lambda'' = f, \quad (5.4)$$

¹⁾ Читатели, ознакомившиеся с параграфами, посвященными полугруппам, заметят, что мы имеем дело с марковской полугруппой, порождаемой дифференциальным оператором $\mathfrak{A} = \frac{1}{2} d^2/dx^2$. Дифференциальное уравнение (5.4) есть частный случай основного уравнения (10.1) гл. XII, встречающегося в теореме Хилле — Йосиды.

а из (5.1), что

$$\omega_\lambda(x) = \int_{-\infty}^{+\infty} K_\lambda(x, s) f(s) ds, \quad (5.5)$$

где $K_\lambda(x, y)$ — обычное преобразование Лапласа функции $q_t(x, y)$. В теории дифференциальных уравнений K_λ называют функцией Грина для уравнения (5.4). Мы покажем, что

$$K_\lambda(x, y) = \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{2\lambda}|x-y|}. \quad (5.6)$$

Справедливость этой формулы можно установить, если проверить, что (5.5) изображает требуемое решение дифференциального уравнения (5.4), однако это не объясняет, как была найдена эта формула.

Мы намерены вывести (5.6) из вероятностных соображений, применимых и к более общим уравнениям и приводящих к явным выражениям для времен первого прохождения (задача 9). Мы примем как данное, что траектория $X(t)$ непрерывно зависит от t . Пусть $X(0) = x$. Обозначим через $F(t, x, y)$ вероятность того, что точка y будет достигнута до момента t . Мы назовем F распределением *момента первого перехода* из x в y и обозначим его преобразование через $\varphi_\lambda(x, y)$.

Пусть $x < y < z$. Тогда событие $X(t) = z$ имеет место тогда и только тогда, когда первое прохождение через y происходит в некоторый момент $\tau < t$ и когда за ним следует переход из y в z за время $t - \tau$. Таким образом, $q_t(x, z)$ есть свертка $F(t, x, y)$ и $q_t(y, z)$, откуда

$$K_\lambda(x, z) = \varphi_\lambda(x, y) K_\lambda(y, z), \quad x < y < z. \quad (5.7)$$

Фиксируем точку y и выберем в качестве f функцию, сосредоточенную на y, ∞ . Умножаем (5.7) на $f(z)$ и интегрируем относительно z . Ввиду (5.5) в результате имеем

$$\omega_\lambda(x) = \varphi_\lambda(x, y) \omega_\lambda(y), \quad x < y. \quad (5.8)$$

В то же время в силу (5.4) при фиксированном y $\varphi_\lambda(x, y)$ удовлетворяет дифференциальному уравнению

$$\lambda \varphi_\lambda - \frac{1}{2} \frac{\partial^2 \varphi_\lambda}{\partial x^2} = 0, \quad x < y. \quad (5.9)$$

Ограниченное на $-\infty$ решение необходимо имеет вид $C_\lambda e^{-\sqrt{2\lambda}x}$. Так как по (5.8) $\varphi_\lambda(x, y) \rightarrow 1$ при $x \rightarrow y$, то $\varphi_\lambda(x, y) = e^{\sqrt{2\lambda}(x-y)}$ при $x < y$. По соображениям симметрии выводим отсюда, что *преобразование Лапласа времени первого перехода из x в y равно*

$$\varphi_\lambda(x, y) = e^{-\sqrt{2\lambda}|x-y|}. \quad (5.10)$$

Полагая $z = y$ в (5.7), видим, что

$$K_{\lambda}(x, y) = e^{-\sqrt{2\lambda}|x-y|} K_{\lambda}(y, y).$$

Поскольку K должно симметричным образом зависеть от x и y , отсюда следует, что $K(y, y)$ сводится к постоянной C_{λ} , зависящей только от λ . Таким образом, мы определили K_{λ} с точностью до постоянного множителя C_{λ} . Равенство $\sqrt{2\lambda}C_{\lambda} = 1$ легко получается из того, что значению $f = 1$ соответствует решение $\omega_{\lambda}(x) = 1/\lambda$. Это доказывает справедливость формулы (5.6).

Следующие далее примеры показывают, как вычислить вероятность того, что точка $y_1 > x$ *будет достигнута ранее, чем некоторая другая точка $y_2 < x$. В то же время эти примеры продемонстрируют, как учитываются *границные условия*.

Примеры. а) *Один поглощающий экран.* Броуновское движение на $0, \infty$ с поглощающим экраном в нуле получается остановкой обычного броуновского движения с $X(0) = x > 0$, когда оно достигает начала координат. Обозначим соответствующие вероятности перехода через $q_t^{\text{abs}}(x, y)$. Аналогично изменим другие обозначения.

В неограниченном броуновском движении плотность вероятности перехода от $x > 0$ в $y > 0$ с заходом в 0 равна свертке плотности первого перехода из x в 0 и $q_t(0, y)$. Соответствующее преобразование Лапласа равно $\varphi_{\lambda}(x, 0)K_{\lambda}(0, y)$. Следовательно, должно быть

$$K_{\lambda}^{\text{abs}}(x, y) = K_{\lambda}(x, y) - \varphi_{\lambda}(x, 0)K_{\lambda}(0, y), \quad (5.11)$$

где $x > 0, y > 0$. Это эквивалентно

$$K_{\lambda}^{\text{abs}}(x, y) = \frac{1}{\sqrt{2\lambda}} [e^{-\sqrt{2\lambda}|x-y|} - e^{-\sqrt{2\lambda}(x+y)}] \quad (5.12)$$

или

$$q_t^{\text{abs}}(x, y) = q_t(x, y) - q_t(x, -y) \quad (5.13)$$

в соответствии с решением (5.5) гл. X, найденным *методом отражения*.

Рассуждения, приводящие к (5.7), применимы без изменений и к процессу с поглощающим экраном. Поэтому из (5.12) мы выводим, что при $0 < x < y$

$$\varphi_{\lambda}^{\text{abs}}(x, y) = \frac{e^{\sqrt{2\lambda}x} - e^{-\sqrt{2\lambda}x}}{e^{\sqrt{2\lambda}y} - e^{-\sqrt{2\lambda}y}}. \quad (5.14)$$

Это¹⁾ есть преобразование Лапласа для вероятности того,

¹⁾ При фиксированном y $\varphi_{\lambda}^{\text{abs}}$ является решением дифференциального уравнения (5.9), которое обращается в нуль при $x=0$ и в 1 при $x=y$. В такой форме результат применим к любой тройке точек $a < x < b$ и $a > x > b$ и к более общим дифференциальным уравнениям.

что в неограниченном броуновском движении с $X(0) = x$ точка $y > x$ будет достигнута до момента t и до прохождения через нуль. При $\lambda \rightarrow 0$ приходим к выводу: вероятность того, что точка y будет достигнута ранее точки 0, равна x/y (точно так же, как в симметричном случайном блуждании, соответствующем схеме Бернулли, см. задачу о разорении в 1; гл. XIV, 2). (Продолжение см. в задаче 8.)

б) Два поглощающих экрана. Рассмотрим теперь броуновское движение, начинающееся в точке x интервала $\overline{0, 1}$ и оканчивающееся при достижении точки 0 или точки 1. Легче всего этот процесс получить из предыдущего процесса с одним поглощающим экраном, ставя дополнительный экран в точке 1. При этом рассуждения, приводящие к (5.11), применимы без изменений. Переходные плотности $q_t^\#(x, y)$ нового процесса имеют преобразование Лапласа $K_\lambda^\#$, определяемое соотношением

$$K_\lambda^\#(x, y) = K_\lambda^{\text{abs}}(x, y) - \varphi_\lambda^{\text{abs}}(x, 1) K_\lambda^{\text{abs}}(1, y), \quad (5.15)$$

где x и y принадлежат интервалу $\overline{0, 1}$. [Отметим, что выполняются граничные условия $K_\lambda^\#(0, y) = K_\lambda^\#(1, y)$.] Простые вычисления показывают, что

$$K_\lambda^\#(x, y) = \frac{e^{-\sqrt{2\lambda}|x-y|} - e^{-\sqrt{2\lambda}(2-|x-y|)} - e^{-\sqrt{2\lambda}(x+y)} - e^{-\sqrt{2\lambda}(2-x-y)}}{\sqrt{2\lambda}(1 - e^{-2\sqrt{2\lambda}})}. \quad (5.16)$$

Разлагая дробь $1/[1 - e^{-2\sqrt{2\lambda}}]$ в геометрический ряд, мы приходим к другому представлению:

$$K_\lambda^\#(x, y) = \frac{1}{\sqrt{2\lambda}} \sum_{n=-\infty}^{+\infty} [e^{-\sqrt{2\lambda}|x-y+2n|} - e^{-\sqrt{2\lambda}|x+y+2n|}]. \quad (5.17)$$

Последнее эквивалентно решению (5.7) гл. X, полученному с помощью принципа отражения. ►

Сходные рассуждения применимы и к более общему диффузионному уравнению

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} a(x) \frac{\partial^2 u(t, x)}{\partial x^2} + b(x) \frac{\partial u(t, x)}{\partial x}, \quad a > 0, \quad (5.18)$$

на конечном или бесконечном интервале. Вместо (5.4) мы получаем

$$\lambda \omega_\lambda - \frac{1}{2} a \omega_\lambda'' - b \omega_\lambda' = f, \quad (5.19)$$

и решение снова имеет форму (5.5) с функцией Грина K_λ вида (5.7). При этом $\varphi_\lambda(x, y)$ есть преобразование плотности первого перехода от x к $y > x$. При фиксированном y эта функция должна

удовлетворять дифференциальному уравнению, соответствующему (5.9), а именно

$$\lambda \varphi_\lambda - \frac{1}{2} a \varphi_\lambda'' - b \varphi_\lambda' = 0. \quad (5.20)$$

Эта функция должна быть ограничена на левом конце. Кроме того, $\varphi_\lambda(y, y) = 1$. Эти условия определяют φ_λ однозначно, за исключением случая, когда (5.20) имеет ограниченное решение. В последнем случае (как и в приведенных выше примерах) следует добавить надлежащие граничные условия (см. задачи 9, 10).

§ 6. ПРОЦЕССЫ РАЗМНОЖЕНИЯ И ГИБЕЛИ. СЛУЧАЙНЫЕ БЛУЖДЕНИЯ

В этом параграфе мы исследуем связь между процессами размножения и гибели из I; гл. XVII, 5, и рандомизированным случайным блужданием из гл. II, 7. Основная цель состоит в том, чтобы проиллюстрировать технику применения преобразований Лапласа и правильное использование граничных условий.

Рассмотрим простое случайное блуждание, начинающееся в нуле, в котором размер отдельного скачка равен 1 или -1 с вероятностями p и q соответственно. Промежутки между последовательными скачками предполагаются независимыми показательно распределенными случайными величинами с математическим ожиданием $1/c$. Вероятность $P_n(t)$ (находиться в состоянии n в момент t) была найдена в (7.7) гл. II, но мы проведем вычисления заново, с новой точки зрения. Составляя уравнение для $P_n(t)$, мы будем рассуждать следующим образом. Состояние $n \neq 0$ может наблюдаться в момент t только при условии, что до момента t произошел скачок. Пусть первый скачок происходит в момент $t-x$ и приводит в точку 1. Тогда (условная) вероятность состояния n в момент t равна $P_{n-1}(x)$. Таким образом, при $n = \pm 1, \pm 2, \dots$

$$P_n(t) = \int_0^t c e^{-c(t-x)} [p P_{n-1}(x) + q P_{n+1}(x)] dx. \quad (6.1a)$$

При $n=0$ к правой части следует добавить член e^{-ct} , соответствующий отсутствию скачков до момента t . Таким образом,

$$P_0(t) = e^{-ct} + \int_0^t c e^{-c(t-x)} [p P_{-1}(x) + q P_1(x)] dx. \quad (6.1b)$$

Точно так же P_n должны удовлетворять бесконечной системе уравнений свертки (6.1). Простое дифференцирование приводит к бесконечной системе дифференциальных уравнений¹⁾

$$P_n'(t) = -c P_n(t) + c p P_{n-1}(t) + c q P_{n+1}(t) \quad (6.2)$$

¹⁾ Эта система представляет собой частный случай системы уравнений I, гл. XVII, (5.2), для процессов размножения и гибели и может быть получена таким же путем.

с начальными условиями $P_0(0)=1$, $P_n(0)=0$ при $n \neq 0$.

Системы (6.1) и (6.2) эквивалентны, но последняя имеет то формальное преимущество, что специальная роль состояния $n=0$ проявляется только в начальных условиях. Для применения преобразований Лапласа безразлично, отправляемся ли мы от (6.1) или от (6.2).

Переходя к преобразованиям Лапласа, обозначим

$$\pi_n(\lambda) = \int_0^{\infty} e^{-\lambda t} P_n(t) dt. \quad (6.3)$$

Поскольку операция свертки соответствует перемножению преобразований Лапласа, а e^{-cx} имеет преобразование Лапласа $1/(c+\lambda)$, то система (6.1) эквивалентна

$$\pi_n(\lambda) = \frac{c}{c+\lambda} [p\pi_{n-1}(\lambda) + q\pi_{n+1}(\lambda)], \quad n \neq 0, \quad (6.4a)$$

$$\pi_0(\lambda) = \frac{1}{c+\lambda} + \frac{c}{c+\lambda} [p\pi_{-1}(\lambda) + q\pi_1(\lambda)]. \quad (6.4b)$$

(Аналогичный результат можно было бы получить из (6.2), так как справедливо равенство

$$\int_0^{\infty} e^{-\lambda t} P'_n(t) dt = -P_n(0) + \lambda \pi_n(\lambda),$$

которое получается интегрированием по частям.)

Система линейных уравнений (6.4) аналогична системе, уже появлявшейся в связи со случайными блужданиями в I; гл. XIV. Мы можем решить ее таким же методом. Квадратное уравнение

$$cqs^2 - (c+\lambda)s + cp = 0 \quad (6.5)$$

имеет корни

$$s_\lambda = \frac{c+\lambda - \sqrt{(c+\lambda)^2 - 4c^2pq}}{2cq} \quad \text{и} \quad \sigma_\lambda = (p/q) s_\lambda^{-1}. \quad (6.6)$$

Легко проверить, что при произвольных постоянных A_λ и B_λ линейные комбинации $\pi_n(\lambda) = A_\lambda s_\lambda^n + B_\lambda \sigma_\lambda^n$ удовлетворяют (6.4a) при $n=1, 2, \dots$. Коэффициенты можно выбрать так, чтобы получить для $\pi_0(\lambda)$ и $\pi_1(\lambda)$ правильные значения. Если π_0 и π_1 даны, то, исходя из (6.4a), можно вычислить последовательно $\pi_2, \pi_3, \dots$. Поэтому при $n \geq 0$ каждое решение имеет вид $\pi_n(\lambda) = A_\lambda s_\lambda^n + B_\lambda \sigma_\lambda^n$. При $\lambda \rightarrow \infty$ имеем $s_\lambda \rightarrow 0$, но $\sigma_\lambda \rightarrow \infty$. Так как наше решение $\pi_n(\lambda)$ остается ограниченным на бесконечности, то необходимо $B_\lambda = 0$ и

$$\pi_n(\lambda) = \pi_0(\lambda) s_\lambda^n, \quad n=0, 1, 2, \dots \quad (6.7a)$$

При $n \leq 0$ мы имеем по аналогии

$$\pi_n(\lambda) = \pi_0(\lambda) \sigma_\lambda^n = (p/q)^n \pi_0(\lambda) s_\lambda^{-n}, \quad n=0, -1, -2, \dots \quad (6.7b)$$

Подставляя в (6.46), наконец, получаем

$$\pi_0(\lambda) = \frac{1}{\sqrt{(c+\lambda)^2 - 4c^2pq}}, \quad (6.8)$$

так что все $\pi_n(\lambda)$ определяются единственным образом.

Из этих преобразований Лапласа можно извлечь максимум информации без знания явных формул самого решения. Например, поскольку перемножение преобразований Лапласа соответствует свертке, то из (6.7) следует, что при $n \geq 0$ вероятности P_n имеют вид $P_n = F_n \star P_0$, где F (возможно, несобственное) распределение вероятностей с преобразованием Лапласа, равным s_λ . С вероятностной точки зрения это означает следующее. Если в момент t случайное блуждание находится в точке n , то первое прохождение через n могло произойти в некоторый момент $\tau \leq t$. Тогда (условная) вероятность попасть в момент t снова в n равна $P_0(t-\tau)$. Таким образом, P_n является сверткой P_0 с распределением F_n времени первого прохождения через n . В свою очередь это время первого прохождения есть сумма n одинаково распределенных независимых случайных величин, а именно сумма времен (ожидания) между последовательными прохождениями через 1, 2, ... Это объясняет вид (6.7) и в то же время показывает, что s_λ^n (для $n > 0$) есть преобразование Лапласа распределения F_n времени первого прохождения через n . Это распределение является несобственным, кроме случая $p=q=\frac{1}{2}$, так как только в этом последнем случае $s_0=1$.

В рассмотренном случае мы, к счастью, имеем возможность обратить преобразование s_λ . Как было показано в примере 3, г) гл. XIII, функция $(\lambda - \sqrt{\lambda^2 - 1})^r$ представляет собой (при $\lambda > 1$) обычное преобразование Лапласа для $(r/x) I_r(x)$. Замена λ на $\lambda/2c\sqrt{pq}$ меняет лишь масштабный параметр, а замена λ на $\lambda + c$ соответствует умножению плотности на e^{-cx} . Отсюда следует, что при $n > 0$ s_λ^n есть обычное преобразование Лапласа для распределения F с плотностью

$$f_n(t) = \sqrt{\left(\frac{p}{q}\right)^n \frac{n}{t}} I_n(2c\sqrt{pqt}) e^{-ct}. \quad (6.9)$$

Это есть плотность времени первого прохождения через точку $n > 0$. Этот факт был установлен прямыми методами в (7.13) гл. II (и поэтому последнее рассуждение может рассматриваться как новый вывод преобразования Лапласа для $x^{-1}I_n(x)$).

Явное выражение для вероятностей $P_n(t)$ можно получить проще. В задаче 8 гл. XIII, II, мы нашли преобразование Лапласа для I_n , а подбор параметров, как указывалось, непосредственно ведет к точным формулам

$$P_n(t) = \sqrt{(p/q)^n} e^{-ct} I_n(2c\sqrt{pqt}), \quad n=0, \pm 1, \pm 2, \dots \quad (6.10)$$

С другой стороны, этот результат был получен прямыми методами (7.7) гл. II.

Как мы уже видели в 1; гл. XVII, 7, различные задачи связи и обслуживания приводят к той же самой системе дифференциальных уравнений (6.2) с той лишь разницей, что рассматриваются только неотрицательные значения n и что граничному значению $n=0$ соответствует иное уравнение. Как действует наш метод в подобных случаях, покажут два следующих примера.

Примеры. а) *Очередь к одному обслуживающему прибору.* Рассмотрим систему обслуживания с одним прибором, в которой вновь прибывшие вызовы становятся в очередь, если обслуживающий прибор занят. Стоящие системы описывается числом $n \geq 0$ вызовов в очереди, включая обслуживаемый вызов. Промежутки между вызовами и время обслуживания являются взаимно независимыми случайными величинами и имеют показательные распределения с плотностями $\lambda e^{-\lambda t}$ и $\mu e^{-\mu t}$ соответственно. Этот пример не что иное, как частный случай примера б) со многими обслуживающими приборами из 1; гл. XVII, 7, но мы заново выведем дифференциальные уравнения, чтобы разъяснить тесную связь с рассмотренной выше моделью случайного блуждания.

Предположим, что в данный момент времени имеется $n \geq 1$ вызовов в очереди. Последующие изменения состояния системы можно охарактеризовать $+1$ в случае, если прибывает новый вызов, и -1 , если закончится текущее время обслуживания. Время ожидания T изменения системы есть наименьшее из времен ожидания каждого из этих двух событий, а поэтому $P\{T > t\} = e^{-ct}$, где $c = \lambda + \mu$. Если изменения системы происходят с вероятностью $p = \lambda/c$, это $+1$, и с вероятностью $q = \mu/c$, это -1 . Другими словами, пока существует очередь, рассматриваемый процесс подобен случайному блужданию и, следовательно, дифференциальные уравнения (6.2) выполнены для $n \geq 1$. Однако когда обслуживание не происходит, изменение состояния может быть вызвано только поступлением нового вызова, а поэтому для $n=0$ дифференциальное уравнение принимает вид

$$P'_0(t) = -cpP_0(t) + cqP_1(t). \quad (6.11)$$

Решим эти дифференциальные уравнения в предположении, что обслуживающий прибор первоначально был свободен, т. е. $P_0(0) = 1$. При $n \geq 1$ преобразования Лапласа $\pi_n(\lambda)$ снова удовлетворяют уравнениям (6.4а), но при $n=0$ мы получаем из (6.11)

$$(cp + \lambda)\pi_0(\lambda) = 1 + cq\pi_1(\lambda). \quad (6.12)$$

Как и в общем случайном блуждании, имеем $\pi_n(\lambda) = \pi_0(\lambda) s_\lambda^n$ при $n \geq 1$, но ввиду (6.12)

$$\pi_0(\lambda) = \frac{1}{cp + \lambda - cqs_\lambda} = \frac{1 - s_\lambda}{\lambda}. \quad (6.13)$$

Таким образом,

$$\pi_n(\lambda) + \pi_{n+1}(\lambda) + \dots = \frac{1-s_\lambda}{\lambda} (s_\lambda^n + s_\lambda^{n+1} + \dots) = s_\lambda^n/n.$$

Мы нашли, что s^n является преобразованием Лапласа для распределения F_n с плотностью (6.9). Множитель $1/\lambda$ соответствует интегрированию, а поэтому при $n > 0$

$$P_n(t) + P_{n+1}(t) + \dots = F_n(t), \quad (6.14)$$

где F_n есть распределение с плотностью (6.9). При $n=0$ левая часть, очевидно, равна единице.

б) *Флуктуации в период занятости.* Рассмотрим по-прежнему один обслуживающий прибор, но только в период занятости. Другими словами, предполагается, что в момент 0 поступает вызов, причем на свободную линию, и что мы обрываем процесс, когда линия впервые снова становится свободной. Аналитически это означает, что теперь рассматриваются значения $n \geq 1$ и начальное условие имеет вид $P_1(0) = 1$. При $n \geq 2$ в уравнениях (6.2) ничто не меняется, однако в связи с отсутствием нулевого состояния в первом уравнении отпадает член $cpP_0(t)$. Таким образом, преобразования Лапласа $\pi_n(\lambda)$ удовлетворяют при $n \geq 2$ уравнениям (6.4a). Кроме того,

$$(\lambda + c)\pi_1(\lambda) = 1 + cq\pi_2(\lambda). \quad (6.15)$$

Как и прежде, при $n \geq 2$ получаем $\pi_n(\lambda) = \pi_1(\lambda)s_\lambda^{n-1}$. Однако $\pi_1(\lambda)$ следует определить, исходя из (6.15). Стандартные вычисления показывают, что $\pi_1(\lambda) = s_\lambda/(cp)$, и, следовательно, $\pi_n(\lambda) = s_\lambda^n/(cp)$. Таким образом, используя предыдущий пример, имеем окончательный результат: $P_n(t) = f_n(t)/(cp)$, где f_n определяется по (6.9).

Для того чтобы обеспечить периоду занятости конечную продолжительность, предположим, что $p < q$. Обозначим *продолжительность периода занятости* через T . Тогда $P\{T > t\} = P(t) = \sum P_n(t)$. Далее, имеет место равенство $P'(t) = -cqP_1(t)$. В его справедливости можно убедиться, суммируя дифференциальные уравнения или опираясь на следующие вероятностные соображения. Пренебрегая событиями с очень малыми вероятностями, получим, что период занятости закончится в интервале между t и $t+h$ тогда и только тогда, когда в момент t в очереди находится только один вызов и его обслуживание закончится в следующий интервал времени продолжительности h . Эти два условия имеют вероятности $P_1(t)$ и $cqh + o(h)$ соответственно, а поэтому плотность распределения T удовлетворяет условию $-P'(t) = cqP_1(t)$. Таким образом, *продолжительность периода занятости имеет плотность*

$$-P'(t) = \sqrt{\frac{q}{p}} \frac{1}{t} I_1(2c\sqrt{pq}t) e^{-ct}, \quad (6.16)$$

Этот результат был получен другим методом в конце примера 4, а) и был использован при изучении длины очереди в примере 9, д) гл. VI (см. задачу 13). ►

§ 7. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ КОЛМОГОРОВА ¹⁾

Мы вернемся теперь к марковским процессам, состояния которых описываются целыми числами 1, 2, ... Дифференциальные уравнения Колмогорова были выведены в 1; гл. XVII, 9, и заново в гл. X, 3. В этом параграфе дано независимое изложение, основанное на преобразованиях Лапласа. Чтобы сделать это изложение замкнутым, мы приведем новый вывод основных уравнений, на этот раз в форме *уравнений свертки*.

Основное предположение состоит в том, что если $X(\tau) = i$ в некоторый момент τ , то в течение некоторого промежутка времени $\tau \leq t < \tau + T$, длина которого имеет показательную плотность $c_i e^{-c_i x}$, значение $X(t)$ остается постоянным. Вероятность того, что $X(\tau + T) = j$, равна p_{ij} . Пусть дано, что $X(0) = i$. Тогда вероятность $P_{ik}(t)$ того, что $X(t) = k \neq i$, может быть вычислена суммированием по всем возможным значениям момента и размера *первого скачка*:

$$P_{ik}(t) = \sum_{j=1}^{\infty} \int_0^t c_i e^{-c_i x} p_{ij} P_{jk}(t-x) dx \quad (k \neq i). \quad (7.1a)$$

При $k = i$ мы должны добавить член, соответствующий отсутствию скачков

$$P_{ii}(t) = e^{-c_i t} + \sum_{j=1}^{\infty} \int_0^t c_i e^{-c_i x} p_{ij} P_{ji}(t-x) dx. \quad (7.1b)$$

Эти уравнения могут быть объединены, если использовать символ Кронекера δ_{ik} , равный 1 при $k = i$ и 0 при $k \neq i$.

¹⁾ Развита в этом параграфе теория без существенного изменения применима к общим скачкообразным процессам, описанным в гл. X, 3. В качестве хорошего упражнения можно предложить провести доказательство в терминах самих распределений вероятностей без использования преобразований Лапласа. Элегантность утрачивается, но теория легко обобщается при этом на нестационарный случай, где коэффициенты c_j и p_{jk} зависят от t . В этом виде теория была развита (для общих скачкообразных процессов) В. Феллером, см. Trans. Amer. Math. Soc., vol. 48 (1940), 488—515 [поправка в т. 58, 474].
Вероятностное изложение, опирающееся на изучение траекторий, см. в книге Чжуна, 1960.

Обобщение на *полумарковские процессы* см. в задаче 14,

Нашим отправным пунктом будет обратное уравнение (7.1)¹⁾; при любых данных $c_i > 0$ и стохастической матрице $\mathbf{p} = (p_{ik})$ мы ищем стохастические матрицы $P(t) = (P_{ik}(t))$, удовлетворяющие (7.1).

С другой стороны, если мы предположим, что в любой конечный интервал времени происходит лишь конечное число скачков, мы можем видоизменить наши рассуждения, рассматривая момент x последнего скачка до момента t . Вероятность скачка из j в k имеет плотность $\sum P_{ij}(x) c_j p_{jk}$, а вероятность отсутствия скачка между x и t равна $e^{-c_j(t-x)}$. Поэтому вместо (7.1) мы получаем *прямое уравнение*

$$P_{ik}(t) = \delta_{ik} e^{-c_i t} + \int_0^t \sum_{j=1}^{\infty} P_{ij}(x) c_j p_{jk} e^{-c_j(t-x)} dx. \quad (7.2)$$

Однако, как мы увидим, существуют процессы с *бесконечным числом скачков*, удовлетворяющие обратному уравнению. Следовательно, прямое уравнение не вытекает из основных предположений о рассматриваемом процессе. Это обсуждалось в гл. X, 3, а также в 1; XVII, 9.

В терминах преобразований Лапласа

$$P_{ik}(\lambda) = \int_0^{\infty} e^{-\lambda t} P_{ik}(t) dt \quad (7.3)$$

обратное уравнение (7.1) принимает вид

$$P_{ik}(\lambda) = \frac{\delta_{ik}}{\lambda + c_i} + \frac{c_i}{\lambda + c_i} \sum_{j=1}^{\infty} p_{ij} P_{jk}(\lambda). \quad (7.4)$$

Мы перейдем теперь к более удобным матричным обозначениям (правила действий над матрицами вполне применимы к бесконечным матрицам с неотрицательными элементами). Мы вводим матрицы $\Pi(\lambda) = (\Pi_{ik}(\lambda))$, $P(t) = (P_{ik}(t))$, $\mathbf{p} = (p_{ik})$ и диагональную матрицу $\mathbf{c}$ с элементами c_i . Через $\mathbf{1}$ мы обозначим вектор-столбец, все компоненты которого равны 1. Суммы по строкам матрицы A образуют вектор-столбец $A\mathbf{1}$. Наконец, I обозначает единичную матрицу.

Из (7.4) видно, что *обратные уравнения* (7.1) превращаются в

$$(\lambda + \mathbf{c}) \Pi(\lambda) = I + \mathbf{c} \mathbf{p} \Pi(\lambda), \quad (7.5)$$

а *прямые уравнения* — в

$$\Pi(\lambda) (\lambda + \mathbf{c}) = I + \Pi(\lambda) \mathbf{c} \mathbf{p}. \quad (7.6)$$

¹⁾ Замена переменных $y = t - x$ облегчает дифференцирование, производя которое найдем, что уравнения свертки (7.1) эквивалентны системе дифференциальных уравнений

$$P'_{ik}(t) = -c_i P_{ik}(t) + c_i \sum_j p_{ij} P_{jk}(t)$$

при начальных условиях $P_{ii}(0) = 1$ и $P_{ik}(0) = 0$ при $k \neq i$. Эта система согласуется с 1; гл. XVII, (9.14), с той разницей, что там коэффициенты c_i и p_{ij} зависели от времени и P_{ik} было функцией двух переменных τ и t , а не только разности $t - \tau$,

Чтобы построить минимальное решение, полагаем последовательно

$$(\lambda + c) \Pi^{(0)}(\lambda) = I, \quad (\lambda + c) \Pi^{(n+1)}(\lambda) = I + c \Pi^{(n)}(\lambda). \quad (7.7)$$

Для сумм по строкам матрицы $\lambda \Pi^{(n)}(\lambda)$ мы вводим обозначение

$$\lambda \Pi^{(n)}(\lambda) \mathbf{1} = \mathbf{1} - \xi^{(n)}(\lambda). \quad (7.8)$$

Подставляя в (7.7) и вспоминая, что $\mathbf{p}\mathbf{1} = \mathbf{1}$, получаем

$$(\lambda + c) \xi^{(n+1)}(\lambda) = c \Pi^{(n)}(\lambda). \quad (7.9)$$

Так как $\xi^{(0)} \geq 0$, то $\xi^{(n)}(\lambda) \geq 0$ при всех n . Поэтому матрицы $\lambda \Pi^{(n)}(\lambda)$ субстохастические. Их элементы суть неубывающие функции n . Поэтому существует конечный предел

$$\Pi^{(\infty)}(\lambda) = \lim_{n \rightarrow \infty} \Pi^{(n)}(\lambda), \quad (7.10)$$

и матрица $\lambda \Pi^{(\infty)}(\lambda)$ — стохастическая или субстохастическая.

Очевидно, что $\Pi^{(\infty)}(\lambda)$ удовлетворяет обратному уравнению (7.5) и что для любого другого неотрицательного решения $\Pi(\lambda)$ тривиально верно неравенство $\Pi(\lambda) \geq \Pi^{(0)}(\lambda)$. По индукции $\Pi(\lambda) \geq \Pi^{(n)}(\lambda)$ при всех n . Таким образом,

$$\Pi(\lambda) \geq \Pi^{(\infty)}(\lambda). \quad (7.11)$$

Менее очевидно, что $\Pi^{(\infty)}(\lambda)$ удовлетворяет и прямому уравнению (7.6). Для доказательства этого мы установим по индукции, что

$$\Pi^{(n)}(\lambda)(\lambda + c) = I + \Pi^{(n-1)}(\lambda) c. \quad (7.12)$$

Это верно при $n=1$. Предполагая справедливость (7.12) и производя подстановку в (7.7), получаем

$$(\lambda + c) \Pi^{(n+1)}(\lambda)(\lambda + c) = I + c + [I + c \Pi^{(n-1)}(\lambda)] c. \quad (7.13)$$

Выражение в скобках равно $(\lambda + c) \Pi^{(n)}(\lambda)$. Умножая (7.13) слева на $(\lambda + c)^{-1}$, приходим к (7.12) с заменой n на $n+1$. Следовательно, это соотношение верно при всех n , и поэтому $\Pi^{(\infty)}(\lambda)$ удовлетворяет прямому уравнению.

Повторяя рассуждение при выводе (7.11), мы видим, что любое неотрицательное решение прямого уравнения (7.6) удовлетворяет неравенству $\Pi(\lambda) \geq \Pi^{(\infty)}(\lambda)$. По этой причине $\Pi^{(\infty)}(\lambda)$ называется минимальным решением. Нами доказана, таким образом,

Теорема 1. Существует матрица $\Pi^{(\infty)}(\lambda) \geq 0$, у которой суммы по строкам не превосходят λ^{-1} и которая удовлетворяет уравнениям (7.5) и (7.6). Сверх того, для любого неотрицательного решения уравнения (7.5) или (7.6) выполняется неравенство (7.11).

Теорема 2. Минимальное решение является преобразованием Лапласа семейства субстохастических или стохастических мат-

риц $P(t)$, удовлетворяющих уравнению Чепмена — Колмогорова

$$P(s+t) = P(s)P(t), \quad (7.14)$$

а также прямому и обратному уравнениям (7.1), (7.2). При этом или все матрицы $P(t)$ и $\lambda P^{(n)}(\lambda)$ ($t > 0$, $\lambda > 0$) строго стохастические, или ни одна из них не обладает этим свойством.

Доказательство. Отбросим верхний индекс ∞ и будем писать $\Pi(\lambda)$ вместо $\Pi^{(\infty)}(\lambda)$. Из определения (7.7) ясно, что $\Pi_{ik}^{(n)}(\lambda)$ есть преобразование положительной функции $P_{ik}^{(n)}$, представляющее собой свертку конечного числа показательных распределений. В силу (7.8) суммы по строкам элементов $P^{(n)}(t)$ образуют монотонно возрастающую последовательность, ограниченную 1. Поэтому $\Pi(\lambda)$ есть преобразование матрицы $P(t)$, стохастической или субстохастической. Из (7.5), (7.6) ясно, что $P(t)$ удовлетворяет первоначальному прямому и обратному уравнениям; отсюда вытекает, что $P(t)$ непрерывно зависит от t . Ясно, что если для $P(t)$ i -я сумма по строкам < 1 при некотором t , то i -я сумма по строкам матрицы $\Pi(\lambda)$ будет $< \lambda^{-1}$ при всех λ и обратно.

Выразим уравнение (7.14) в терминах преобразований Лапласа. Для этого умножим его на $e^{-\lambda t - vs}$ и произведем интегрирование по t и s . Справа получим произведение матриц $\Pi(\lambda)\Pi(v)$. Левая часть легко оценивается с помощью подстановки $x = t + s$, $y = -t + s$. Окончательный результат таков:

$$-\frac{\Pi(v) - \Pi(\lambda)}{v - \lambda} = \Pi(\lambda)\Pi(v). \quad (7.15)$$

Обратно, (7.15) влечет за собой (7.14) [это рассуждение повторяется в примере 8, а) гл. XIII].

Для доказательства (7.15) рассмотрим матричное уравнение

$$(\lambda + c)Q = A + cPQ. \quad (7.16)$$

Если A и Q неотрицательны, то, очевидно, $Q \geq (\lambda + c)^{-1}A = \Pi^{(0)}(\lambda)A$ и по индукции $Q \geq \Pi^{(n)}(\lambda)A$ при всех n . Таким образом, $Q \geq \Pi(\lambda)A$. Теперь $\Pi(v)$ удовлетворяет (7.16) с $A = I + (\lambda - v)\Pi(v)$, и потому при $\lambda > v$

$$\Pi(v) \geq \Pi(\lambda) + (\lambda - v)\Pi(\lambda)\Pi(v). \quad (7.17)$$

С другой стороны, правая часть (7.17) удовлетворяет прямому уравнению (7.6) с λ , замененным на v . Поэтому она не меньше $\Pi(v)$, и в (7.17) имеет место знак равенства. Доказательство закончено ¹⁾. ▶

¹⁾ Уравнение (7.15) есть *резольвентное уравнение* для семейства сжатий $\lambda\Pi(\lambda)$ на банаховом пространстве ограниченных векторов-столбцов. Мы видели в гл. XIII, 10, что оно выполняется тогда и только тогда, когда область значений этих преобразований не зависит от λ . Это гарантируется их минимальным характером,

Чтобы понять, является ли матрица $\lambda \Pi^{(\infty)}(\lambda)$ строго стохастической¹⁾, мы вернемся к соотношениям (7.8) и (7.9). Так как элементы $\xi_i^{(n)}(\lambda)$ суть невозрастающие функции n , то существует предел $\xi(\lambda) = \lim \xi^{(n)}(\lambda)$, такой, что

$$\lambda \Pi^{(\infty)}(\lambda) \mathbf{1} = \mathbf{1} - \xi(\lambda) \quad (7.18)$$

и

$$(\lambda + c) \xi(\lambda) = c \rho \xi(\lambda), \quad 0 \leq \xi(\lambda) \leq \mathbf{1}. \quad (7.19)$$

С другой стороны, мы имеем

$$(\lambda + c) \xi^{(0)}(\lambda) = c \mathbf{1} = c \rho \mathbf{1}, \quad (7.20)$$

и потому $\xi^{(0)}(\lambda) \geq \xi(\lambda)$ для любого вектора, удовлетворяющего (7.19). Из (7.9) по индукции выводим $\xi^{(n)}(\lambda) \geq \xi(\lambda)$ для всех n , так, что вектор $\xi(\lambda)$ в (7.18) является максимальным вектором, удовлетворяющим (7.19).

Таким образом, верна

Теорема 3. Дефекты по строкам минимального решения описываются максимальным вектором $\xi(\lambda)$, удовлетворяющим (7.19).

Следовательно, матрица $\lambda \Pi^{(\infty)}(\lambda)$ будет строго стохастической тогда и только тогда, когда (7.19) влечет за собой $\xi(\lambda) = 0$.

Следствие 1. Если $c_i \leq M < \infty$ при всех i , то минимальное решение будет строго стохастическим (так что ни прямое, ни обратное уравнения не имеют других допустимых решений).

Доказательство. Дробь $c/(\lambda + c)$ является возрастающей функцией c . Из (7.19) по индукции получаем

$$\xi(\lambda) \leq \left(\frac{M}{\lambda + M} \right)^n \cdot \mathbf{1} \quad (7.21)$$

при всех n . Следовательно, $\xi(\lambda) = 0$. ▶

Если $A(\lambda)$ — матрица с элементами вида $\xi_i(\lambda) \eta_k(\lambda)$ с произвольными $\eta_k(\lambda)$, то $\Pi(\lambda) + A(\lambda)$ снова будет решением обратного уравнения (7.5). Всегда можно выбрать $A(\lambda)$ так, чтобы получить допустимые матрицы, удовлетворяющие уравнению Чепмена — Колмогорова. Эта процедура проиллюстрирована в следующем параграфе. Соответствующие процессы характеризуются наличием бесконечного числа скачков в конечном промежутке времени. Любопытно, что прямые уравнения могут выполняться, даже когда их интерпретация в терминах последнего скачка недопустима.

Таковы основные результаты. Мы закончим параграф критерием, полезным в приложениях и интересным с той точки зрения, что при его доказательстве появляются понятия теории потенциала: ядро Γ в (7.25) служит типичным примером потен-

¹⁾ Предостережение. Казалось бы, что формальное умножение прямого уравнения на вектор-столбец $\mathbf{1}$ должно привести к тождеству $\lambda \Pi(\lambda) \mathbf{1} = \mathbf{1}$. Однако получающийся ряд может расходиться. Эта процедура законна, если все c_i ограничены (следствие 1).

циала.

Допустим, что $c_i > 0$. Перепишем (7.19) в форме

$$\xi(\lambda) + \lambda c^{-1} \xi(\lambda) = p \xi(\lambda). \quad (7.22)$$

Умножая на p^k и суммируя по $k=0, \dots, n-1$, получаем

$$\xi(\lambda) + \lambda \sum_{k=0}^{n-1} p^k c^{-1} \xi(\lambda) = p^n \xi(\lambda). \quad (7.23)$$

Отсюда следует, что $p^n \xi(\lambda)$ монотонно зависит от n , так что $p^n \xi(\lambda) \rightarrow x$, где x — минимальный вектор-столбец, удовлетворяющий¹⁾

$$px = x, \quad \xi(\lambda) \leq x \leq 1. \quad (7.24)$$

Определим теперь матрицу (возможно, с бесконечными элементами) равенством

$$\Gamma = \sum_{k=1}^{\infty} p^k c^{-k}. \quad (7.25)$$

Полагая в (7.23) $n \rightarrow \infty$, получаем

$$\xi(\lambda) + \lambda \Gamma \xi(\lambda) = x. \quad (7.26)$$

Отсюда вытекает, что $\xi_k(\lambda) = 0$ при каждом k , для которого $\Gamma_{kk} = \infty$. Так будет, если k является возвратным состоянием цепи Маркова с матрицей p . Мы имеем

Следствие 2. Минимальное решение будет строго стохастическим (и потому единственным), если в дискретной цепи Маркова с матрицей p все состояния возвратны.

§ 8. ПРИМЕР: ЧИСТЫЙ ПРОЦЕСС РАЗМНОЖЕНИЯ

Вместо того чтобы продолжать изложение общей теории, мы подробно рассмотрим процессы, в которых возможны лишь переходы $i \rightarrow i+1$. Они дают хорошую иллюстрацию того, какими могут быть процессы в обстановке неединственности. Во избежание тривиальностей допустим, что $c_i > 0$ при всех i . По определению $p_{i, i+1} = 1$ и $p_{i, k} = 0$ для всех других пар индексов. Обратное и прямое уравнения сводятся теперь к

$$(\lambda + c_i) \Pi_{ik}(\lambda) - c_i \Pi_{i+1, k}(\lambda) = \delta_{ik} \quad (8.1)$$

и

$$(\lambda + c_k) \Pi_{ik}(\lambda) - c_{k-1} \Pi_{i, k-1}(\lambda) = \delta_{ik}, \quad (8.2)$$

где $\delta_{ik} = 1$ при $i=k$ и 0 в других случаях. Положим для крат-

¹⁾Нетрудно видеть, что x не зависит от λ и $\lambda \Pi^{(\infty)}(\lambda) x = x - \xi(\lambda)$.

кости

$$\rho_i = \frac{c_i}{c_i + \lambda}, \quad r_i = \frac{1}{c_i + \lambda}. \quad (8.3)$$

Функция ρ_i есть преобразование Лапласа (показательного) времени пребывания в состоянии i , а r_i есть обычное преобразование Лапласа для вероятности того, что это время превосходит t . В дальнейшем следует иметь в виду зависимость r_j и ρ_j от λ .

а) *Минимальное решение.* Легко проверить, что

$$\Pi_{ik}(\lambda) = \begin{cases} \rho_i \rho_{i+1} \dots \rho_{k-1} r_k & \text{для } k \geq i, \\ 0 & \text{для } k < i \end{cases} \quad (8.4)$$

является минимальным решением как для (8.1), так и для (8.2). В (8.4) отражается тот факт, что переходы из i в $k > i$ невозможны, а момент *прихода* в $k > i$ равен сумме $k-i$ независимых времен пребывания в $i, i+1, \dots, k-1$.

Пусть $P_{ik}(t)$ обозначают переходные вероятности процесса, определенного в (8.4). Докажем следующий важный результат, который был получен другими методами в 1; гл. XVII, 4.

Лемма. Если

$$\sum 1/c_n = \infty, \quad (8.5)$$

то

$$\sum_{k=i}^{\infty} P_{ik}(t) = 1 \quad (8.6)$$

для всех i и $t > 0$; в противном случае (8.6) не выполняется для всех i .

Доказательство. Заметим, что $\lambda r_k = 1 - \rho_k$, откуда

$$\lambda [\Pi_{i,i}(\lambda) + \dots + \Pi_{i,i+n}(\lambda)] = 1 - \rho_i \dots \rho_{i+n}. \quad (8.7)$$

Следовательно, (8.6) выполняется тогда и только тогда, когда для всех $\lambda > 0$

$$\rho_i \rho_{i+1} \dots \rho_n \rightarrow 0, \quad n \rightarrow \infty. \quad (8.8)$$

Далее, если $c_n \rightarrow \infty$, то $\rho_n \sim e^{-\lambda/c_n}$ и, следовательно, в этом случае (8.5) необходимо и достаточно для (8.8). С другой стороны, если c_n не стремится к бесконечности, то существует такое число $q < 1$, что $\rho_n < q$ для бесконечно многих n . Следовательно, и (8.8), и (8.5) выполняются. ►

В том случае, когда ряд в (8.5) расходится, ничего особенного не происходит: c_n однозначно определяют процесс размножения, удовлетворяющий основным исходным постулатам. Поэтому в дальнейшем предположим, что $\sum c_n^{-1} < \infty$.

Дефект $1 - \sum_k P_{ik}(t)$ равен вероятности того, что к моменту t

система прошла через все состояния или «достигла границы ∞ ». Момент достижения есть сумма времен пребывания в $i, i+1, \dots$. Этот ряд сходится с вероятностью единица, так как сходится в силу (8.6) ряд из соответствующих математических ожиданий. В процессе, начинающемся в состоянии i , момент достижения ∞ имеет преобразование Лапласа

$$\xi_i = \lim_{n \rightarrow \infty} \rho_i \rho_{i+1} \dots \rho_{i+n} \quad (8.9)$$

и ξ_i удовлетворяют уравнению (7.19), а именно

$$(\lambda + c_i) \xi_i = c_i \xi_{i+1}. \quad (8.10)$$

Для сумм по строкам получим из (8.7)

$$\lambda \sum_{k=i}^{\infty} \Pi_{ik}(\lambda) = 1 - \xi_i. \quad (8.11)$$

б) *Процессы с возвращением.* Отправляясь от процесса (8.4), можно определить ряд новых процессов следующим образом. Выберем числа q_i , такие, что $q_i \geq 0$, $\sum q_i = 1$. Мы требуем, чтобы после достижения ∞ система мгновенно приходила в состояние i с вероятностью¹⁾ q_i . Первоначальный процесс повторяется заново до второго достижения ∞ . Время между двумя достижениями ∞ есть случайная величина с преобразованием Лапласа

$$\tau(\lambda) = \sum q_i \xi_i. \quad (8.12)$$

Марковский характер процесса требует, чтобы после второго достижения ∞ процесс вел себя прежним образом. Теперь мы опишем переходные вероятности $P_{ik}^{\text{ret}}(t)$ нового процесса в терминах соответствующих преобразований Лапласа $\Pi_{ik}^{\text{ret}}(\lambda)$. Вероятность перехода из состояния i в момент 0 в состояние k в момент t без промежуточного захода в ∞ имеет преобразование (8.4). Вероятность достичь k после ровно одного захода в ∞ имеет преобразование Лапласа $\xi_i \sum_j q_j \Pi_{jk}(\lambda)$, и момент второго захода в ∞ имеет преобразование $\xi_i \tau(\lambda)$. Рассматривая последующие возвращения, мы приходим к формуле

$$\Pi_{ik}^{\text{ret}}(\lambda) = \Pi_{ik}(\lambda) + \xi_i \frac{1}{1 - \tau(\lambda)} \sum_j q_j \Pi_{jk}(\lambda), \quad (8.13)$$

где $[1 - \tau(\lambda)]^{-1} = \sum \tau^n(\lambda)$ учитывает многократные прохождения через ∞ . Тривиальные вычисления, использующие (8.11), показывают, что суммы по строкам в (8.13) равны $1/\lambda$, поэтому $\Pi_{ik}^{\text{ret}}(\lambda)$ суть преобразования строго стохастических матриц переходных вероятностей $P^{\text{ret}}(t)$.

¹⁾ Варианты процессов с возвращением получаются при $\sum q_i < 1$; при достижении ∞ процесс с вероятностью $1 - \sum q_i$ обрывается.

Легко проверить, что *новый процесс удовлетворяет обратным уравнениям* (8.1), но не удовлетворяет *прямым уравнениям* (8.2). Все происходит так, как и должно быть: постулаты, приводящие к прямому уравнению, нарушены, так как последний скачок не обязан существовать.

в) *Двусторонний процесс размножения*. Чтобы получить процесс, удовлетворяющий и прямому и обратному уравнениям, мы видоизменим процесс размножения, приняв в качестве состояний системы числа $0, \pm 1, \pm 2, \dots$. Все другие условия остаются прежними: постоянные $c_i > 0$ определены для всех целых i и переходы возможны только из i в $i \pm 1$. Предположим опять, что $\sum 1/c_n < \infty$, где суммирование идет от $-\infty$ до ∞ .

По-прежнему *минимальное* решение определяется по (8.4). Предел

$$\eta_k = \lim_{l \rightarrow -\infty} \Pi_{ik}(\lambda) = r_k \rho_{k-1} \rho_{k-2} \rho_{k-3} \dots \quad (8.14)$$

существует и может быть интерпретирован как преобразование «вероятности $P_{-\infty, k}(t)$ перехода из $-\infty$ в момент 0 в k в момент t ». При такой начальной точке процесс пробегает все состояния от $-\infty$ до ∞ и «достигает ∞ » в момент с преобразованием Лапласа $\xi_{-\infty} = \lim_{n \rightarrow -\infty} \xi_n$. Определим теперь новый процесс

следующим образом. Он начинается как процесс, соответствующий минимальному решению (8.4), но по достижении ∞ он перескакивает в $-\infty$ и таким образом продолжается неопределенно долго. Построение примера б) приводит к следующим преобразованиям переходных вероятностей:

$$\Pi_{ik}^{\#}(\lambda) = \Pi_{ik}(\lambda) + \frac{\xi_i \eta_k}{1 - \xi_{-\infty}}. \quad (8.15)$$

Легко проверить, что $\Pi_{ik}^{\#}$ удовлетворяет и обратному и прямому уравнениям (8.1) и (8.2). В то же время процесс удовлетворяет предположениям, приводящим к обратным уравнениям, но не предположениям, приводящим к прямым уравнениям.

§ 9. ВЫЧИСЛЕНИЕ ЭРГОДИЧЕСКИХ ПРЕДЕЛОВ И ВРЕМЕН ПЕРВОГО ПРОХОЖДЕНИЯ

Можно ожидать, что поведение при $t \rightarrow \infty$ переходных вероятностей $P_{ij}(t)$ в марковском процессе с целочисленными состояниями аналогично поведению переходных вероятностей за большое число шагов в дискретных цепях Маркова (с тем, однако, приятным упрощением, что помехи, создаваемые периодичностью, исчезают). В теореме 1 этот факт устанавливается как простое следствие эргодических теорем 1; гл. XV. Затем нашей основной задачей будет вычисление пределов для переходных вероятностей процессов § 7. Мы покажем также, как могут быть исследованы

времени первого прохождения. Используемые методы применимы в широком круге задач.

Теорема 1. Пусть $P(t)$ — семейство стохастических матриц, для которого

$$P(s+t) = P(s)P(t) \quad (9.1)$$

и $P(t) \rightarrow I$ при $t \rightarrow 0$. Если никакой из элементов P_{ik} не равен нулю тождественно¹⁾, то при $t \rightarrow \infty$

$$P_{ik}(t) \rightarrow u_k, \quad (9.2)$$

где или $u_k = 0$ при всех k , или

$$u_k > 0, \quad \sum_k u_k = 1 \quad (9.3)$$

и

$$\sum_j u_j P_{jk}(t) = u_k. \quad (9.4)$$

Вторая возможность осуществляется, если найдется вероятностный вектор $(u_1, u_2, \dots)$, удовлетворяющий при некотором $t > 0$ соотношению (9.4). В этом случае (9.4) верно при всех $t > 0$, и вероятностный вектор u определяется единственным способом.

(Как уже объяснялось в 1; гл. XVII, 6, важная особенность состоит в том, что пределы не зависят от i . Иначе говоря, влияние начальных условий асимптотически пренебрежимо.)

Доказательство. Фиксируем $\delta > 0$ и рассмотрим дискретную цепь Маркова с матрицей вероятностей перехода $P(\delta)$. Матрица перехода за n шагов будет $P^n(\delta) = P(n\delta)$. Если при некотором n все элементы $P_{ik}(n\delta)$ положительны, то цепь неприводима и непериодична, и по эргодической теореме 1; гл. XV, 7, наши утверждения верны для значений t , пробегающих подпоследовательность $\delta, 2\delta, 3\delta, \dots$. Так как любые два рациональных числа имеют бесчисленное множество общих кратных, то предел при $n \rightarrow \infty$ $P_{ik}(n\delta)$ будет одним и тем же для всех рациональных δ . Для завершения доказательства достаточно проверить, что $P_{ik}(t)$ есть равномерно непрерывная функция t , положительная для больших t . По (9.1)

$$P_{ii}(s)P_{ik}(t) \leq P_{ik}(s+t) \leq P_{ik}(t) + [1 - P_{ii}(s)] \quad (9.5)$$

[первое неравенство тривиально, а второе вытекает из того факта, что члены $P_{ij}(s)$ с $j \neq i$ в сумме дают $1 - P_{ii}(s)$]. Для достаточно малых s имеем $1 - \varepsilon \leq P_{ii}(s) \leq 1$, так что из (9.5) сле-

¹⁾ Это условие введено только во избежание тривиальных осложнений; этого же можно достигнуть, ограничиваясь надлежащим образом выбранными множествами состояний. Нетрудно видеть, что наши условия влекут строгую положительность $P_{ik}(t)$ при всех t .

дует равномерная непрерывность P_{ik} . Из (9.5) видно также, что если $P_{ik}(t) > 0$, то $P_{ik}(t+s) > 0$ в некотором интервале значений s фиксированной длины. Поэтому или P_{ik} тождественно равно нулю, или, начиная с некоторого момента, становится положительным. ►

Мы применим теперь этот результат к минимальному решению § 7, предполагая, что оно является строго стохастическим (и стало быть, *единственным*). В матричных обозначениях (9.4) означает $uP(t) = u$, или, в терминах преобразования Лапласа,

$$u\lambda\P(\lambda) = u. \quad (9.6)$$

Пусть вектор u удовлетворяет (9.6) при некотором частном значении $\lambda > 0$. Тогда резольвентное уравнение (7.15) показывает справедливость (9.6) при *всех* $\lambda > 0$, а вместе с тем и справедливость (9.4) при *всех* $t > 0$. Подставляя (9.6) в прямое уравнение (7.6), получаем

$$u\sigma p = u\sigma, \quad (9.7)$$

причем компоненты $u_k c_k$ конечны (хотя, возможно, не ограничены.) С другой стороны, если u — вероятностный вектор, удовлетворяющий (9.7), то из (7.12) по индукции следует, что $u\lambda\P^{(n)}(\lambda) \leq u$ при всех n и потому $u\lambda\P(\lambda) \leq u$. Так как матрица $\lambda\P(\lambda)$ строго стохастическая, то суммы компонент в правой и левой частях должны быть равны. Поэтому верно (9.6). Нами доказана

Теорема 2. Если минимальное решение является строго стохастическим (и потому *единственным*), то соотношения (9.2) выполняются с $u_k > 0$ тогда и только тогда, когда найдется вероятностный вектор u , удовлетворяющий (9.7).

Отсюда вытекает, в частности, что решение u уравнения (9.7) *единственно*.

Вероятностная интерпретация. Рассмотрим для определенности простейший случай, когда дискретная цепь с вероятностями перехода p_{ij} эргодична. Иными словами, мы предполагаем, что существует строго положительный вероятностный вектор $\alpha = (\alpha_1, \alpha_2, \dots)$, такой, что $\alpha p = \alpha$ и $p_{ik}^{(n)} \rightarrow \alpha_k$ при $n \rightarrow \infty$. Тогда ясно, что если $\sigma = \sum \alpha_k c_k^{-1} < \infty$, то вероятностный вектор с компонентами $u_k = \alpha_k c_k^{-1} / \sigma$ удовлетворяет (9.7). В то же время при $\sigma = \infty$ решение не существует.

Интуитивно ясно, что *переходы* в нашем процессе такие же, как и в дискретной цепи Маркова с матрицей p , однако моменты их осуществления иные. Рассмотрим, к примеру, какое-либо состояние и пометим его индексом 0. Последовательные *времена пребывания* в 0 чередуются с *временами отсутствия*, в продолжение которых система находится в состояниях $j > 0$. Число

переходов в состояние j регулируется матрицей p , а продолжительность пребывания в нем зависит от c_j . В дискретной цепи Маркова при большом числе переходов частоты попадания в j и 0 относятся как $\alpha_j : \alpha_0$. Следовательно, α_j / α_0 равно среднему числу попаданий в j за время отсутствия в 0. Средняя продолжительность каждого пребывания в j равна $1/c_j$. Поэтому за большой период наблюдения времени пребывания в состояниях j и 0 (а следовательно, и вероятности этих состояний) относятся как $\alpha_j c_j^{-1} : \alpha_0 c_0^{-1}$ или $u_j : u_0$.

Эти рассуждения могут быть сделаны строгими, даже в случае, когда $P_{ij}(t) \rightarrow 0$. В соответствии с теоремой Дермана, упомянутой в 1, гл. XV, 11, если p порождает неприводимую и возвратную цепь, то существует вектор α , такой, что $\alpha p = \alpha$. При этом α определяется однозначно с точностью до постоянного множителя $\alpha_k \geq 0$, но ряд $\sum \alpha_k$ может расходиться. Даже в этом случае отношения $\alpha_j : \alpha_0$ допускают данную выше частотную интерпретацию, так что рассуждения имеют общий характер. Если $\sum \alpha_k c_k^{-1} < \infty$, то (9.2) — (9.4) выполняются с u_k , пропорциональными $\alpha_k c_k^{-1}$. В противном случае $P(t) \rightarrow 0$ при $t \rightarrow \infty$. Интересно отметить, что пределы u_k могут быть положительны, даже если все состояния дискретной цепи нулевые.

Существование пределов $P_{ik}(\infty)$ можно установить также, привлекая аргументы теории восстановления, тесно связанные с поведением времен возвращения. Покажем, как можно вычислить распределение времен возвращения и первого прохождения. Припишем состояниям номера 0, 1, 2, ... и выделим состояние 0. Рассмотрим новый процесс, который совпадает с первоначальным до момента первого достижения 0, а затем остается в нуле навсегда. Другими словами, новый процесс получается из старого установкой в нуле поглощающего экрана. Обозначим переходные вероятности видоизмененного процесса через ${}^0P_{ik}(t)$. Тогда ${}^0P_{00}(t) = 1$. В терминах первоначального процесса ${}^0P_{i0}(t)$ равно вероятности первого перехода из $i \neq 0$ в 0 до момента t , а ${}^0P_{ik}(t)$ равно вероятности перехода из $i \neq 0$ в $k \neq 0$ без промежуточного захода в 0. Из вероятностных соображений ясно, что матрица ${}^0P(t)$ должна удовлетворять тем же самым обратному и прямому уравнениям, что и $P(t)$, с той разницей, что c_0 заменяется на 0. Теперь мы пойдем в обратном направлении: мы изменим обратное и прямое уравнения, заменяя c_0 на 0, и покажем, что единственное решение этих уравнений имеет требуемые свойства.

Пусть ξ — вектор, совпадающий с нулевым столбцом $\Pi(\lambda)$. Обратное уравнение показывает, что вектор

$$(\lambda + c - c\rho)\xi = \eta \quad (9.8)$$

имеет компоненты 1, 0, 0, ... Теперь обратное уравнение для ${}^0\Pi(\lambda)$ получается заменой c_0 на 0, так что если ξ обозначает нулевой столбец ${}^0\Pi(\lambda)$, то вектор (9.8) имеет компоненты $\eta_1 = \eta_2 = \dots = 0$, но $\eta_0 = \rho \neq 0$. Отсюда следует, что вектор с компонентами $\xi_k = \Pi_{k0}(\lambda) - \rho {}^0\Pi_{k0}(\lambda)$ удовлетворяет (9.8) с $\eta = 0$. Так

как матрица $\lambda \Pi(\lambda)$ строго стохастическая, то $\xi_k = 0$ при всех k (теорема 3 из § 7). Так как ${}^0\Pi_{00}(\lambda) = 1/\lambda$, мы имеем при $k \geq 0$

$$\Pi_{k0}(\lambda) = \lambda^0 \Pi_{k0}(\lambda) \Pi_{00}(\lambda). \quad (9.9)$$

Учитывая первое уравнение (9.8), мы видим, что

$$\Pi_{00}(\lambda) = \frac{1}{\lambda + c_0} + \frac{c_0}{\lambda + c_0} \sum_j p_{0j} \lambda^0 \Pi_{j0}(\lambda) \Pi_{00}(\lambda). \quad (9.10)$$

Уравнения (9.9) и (9.10) суть *уравнения восстановления* с очевидным вероятностным содержанием. В самом деле, пусть процесс начинается в точке $k > 0$. Тогда ${}^0\Pi_{k0}$ есть обычное преобразование Лапласа для вероятности ${}^0P_{k0}(t)$ того, что первое попадание в 0 произойдет до момента t . Следовательно, $\lambda^0 \Pi_{k0}(\lambda)$ представляет собой преобразование Лапласа распределения F_k *момента первого достижения нуля*. Таким образом, (9.9) утверждает, что $P_{k0}(t)$ есть свертка F_k и P_{00} . Событие $X(t) = 0$ осуществляется тогда и только тогда, когда первое попадание в 0 происходит в некоторый момент $x < t$ и когда $t - x$ единиц времени спустя система снова находится в 0.

Аналогично $\sum p_{0j} \lambda^0 \Pi_{j0}$ задает распределение F_0 времени, проведенного вне 0, т. е. времени между двумя последовательными пребываниями в состоянии 0. Следовательно, множитель при $\Pi_{00}(\lambda)$ в правой части (9.10) соответствует *времени ожидания первого возвращения в нуль*, если система первоначально в нем находится. (Это есть также распределение полного периода, равного сумме времени пребывания и времени отсутствия.) Уравнение восстановления (9.10) выражает $P_{00}(t)$ как сумму вероятности того, что время пребывания в нуле превосходит t , и вероятности события $X(t) = 0$ после первого возвращения в момент $x < t$. Если состояние 0 возвратно, то из (9.10) и теоремы восстановления следует, что

$$P_{00}(\infty) = \frac{1}{1 + c_0 \mu}, \quad (9.11)$$

где μ — математическое ожидание времени отсутствия в 0 и $c_0^{-1} + \mu$ — математическое ожидание длины полного цикла.

§ 10. ЗАДАЧИ

1. Положим в уравнении восстановления (1.3) $F'(t) = g(t) = e^{-t} t^{p-1} / \Gamma(p)$. Тогда

$$\Psi(\lambda) = \frac{1}{(\lambda + 1)^p - 1}. \quad (10.1)$$

Разложением на простейшие дроби покажите, что при целых $^1) p$

$$v(t) = \frac{1}{p} \sum_{k=0}^{p-1} a_k e^{-(1-a_k)t}, \quad (10.2)$$

где $a_k = e^{-i2\pi k/p}$ и $i^2 = -1$.

2. На линию обслуживания поступает пуассоновский поток вызовов с параметром α . Время обслуживания имеет функцию распределения G с преобразованием Лапласа γ . Пусть $H(t)$ обозначает вероятность того, что время обслуживания не превосходит t и в течение этого времени обслуживания не поступают новые вызовы. Покажите, что H является несобственным распределением с преобразованием Лапласа — Стильтеса, равным $\gamma(\lambda + \alpha)$.

3. Потерянные вызовы. Предположим, что линия обслуживания из предыдущей задачи свободна в момент 0. Обозначим через $U(t)$ вероятность того, что все прибывающие до момента t вызовы застают линию свободной. Выведите уравнение восстановления для U и покажите, что обычное преобразование Лапласа ω для U удовлетворяет линейному уравнению

$$\omega(\lambda) = \frac{1}{\lambda + \alpha} + \alpha \frac{1 - \gamma(\lambda + \alpha)}{(\lambda + \alpha)^2} + \frac{\alpha}{\lambda + \alpha} \gamma(\lambda + \alpha) \omega(\lambda).$$

Среднее время ожидания первого вызова, пришедшего в течение периода занятости, равно

$$\frac{1}{\alpha} + \frac{1}{\alpha[1 - \gamma(\alpha)]}.$$

4. Продолжение. Решите предыдущую задачу с помощью метода, описанного в задаче 17 гл. VI, 13, рассматривая U как распределение полного времени жизни для обрывающегося процесса восстановления с запаздыванием.

5. Если F имеет математическое ожидание μ и дисперсию σ^2 и если $\sigma\mu < 1$, то решение уравнения (4.1) для периода занятости имеет дисперсию $(\sigma^2 + \sigma\mu^3)/(1 - \sigma\mu)$.

6. В примере 4, б) производящая функция общего числа задержанных автомобилей равна $e^{c\phi(\psi(s)-1)}$, где

$$\psi(s) = s\phi(c - c\phi(s)). \quad (10.3)$$

7. Покажите, что если ϕ есть преобразование Лапласа собственного распределения вероятностей, то решение ψ уравнения (10.3) есть производящая функция некоторого (возможно, несобственного) распределения. Последнее будет собственным тогда и только тогда, когда F имеет математическое ожидание $\mu \leq 1/c$.

8. В процессе с поглощающим экраном из примера 5, а) обозначим через $F(t, x)$ вероятность того, что (при условии, что процесс начинается в точке x) поглощение произойдет в течении времени t . (Таким образом, F обозначает распределение полного времени жизни процесса.) Покажите, что обычное преобразование Лапласа для $1 - F$ выражается интегралом от функции $K^{abs}(x, y)$ по области $0 < y < \infty$. Докажите, что преобразование Лапласа — Стильтеса для F равно $e^{-V\sqrt{\lambda}x}$ в согласии с тем, что это преобразование удовлетворяет дифференциальному уравнению (5.9).

9. Исходя из (5.7), покажите, что функция Грина общего диффузионного

¹⁾ Корни знаменателя будут одни и те же при $p=n$ и $p=n/2$, но решения совсем различны. Это показывает, что популярное «разложение по корням знаменателя» требует осторожности при иррациональной функции ψ .

уравнения (5.19) в любом интервале необходимо имеет вид

$$K_{\lambda}(x, y) = \begin{cases} \frac{\xi_{\lambda}(x) \eta_{\lambda}(y)}{W(y)} & \text{при } x \leq y, \\ \frac{\eta_{\lambda}(x) \xi_{\lambda}(y)}{W(y)} & \text{при } x \geq y, \end{cases} \quad (10.4)$$

где ξ_{λ} и η_{λ} суть решения уравнения

$$\lambda \varphi - \frac{1}{2} a \varphi'' - b \varphi' = 0, \quad (*)$$

ограниченные соответственно на левом и правом концах. Если (*) не имеет ограниченных решений, то ξ_{λ} и η_{λ} определяются с точностью до мультипликативной постоянной, которую можно включить в W (в противном случае нужно наложить подходящие граничные условия).

Покажите, что функция ω_{λ} , определенная по (10.4) и (5.5), удовлетворяет дифференциальному уравнению (5.19) в том и только в том случае, когда W есть вронскиан:

$$W(y) = [\xi'_{\lambda}(y) \eta_{\lambda}(y) - \xi_{\lambda}(y) \eta'_{\lambda}(y)] a/2. \quad (10.5)$$

Решения ξ_{λ} и η_{λ} обязательно монотонны, и потому $W(y) \neq 0$.

10. *Продолжение.* При $x < y$ момент первого перехода из x в y имеет преобразование Лапласа $\xi_{\lambda}(x)/\xi_{\lambda}(y)$. При $x > y$ оно равно $\eta_{\lambda}(x)/\eta_{\lambda}(y)$.

11. Покажите, что метод, использованный в § 5 для диффузионных процессов, применим и к общему процессу размножения и гибели¹⁾.

12. Разберите пример 6, а) для случая $a > 1$ линий (вычисление a постоянных в явном виде затруднительно и не рекомендуется).

13. В условиях примера 6, б) выведите непосредственно из дифференциальных уравнений, что период занятости имеет математическое ожидание

$$\frac{1}{c(q-p)} \text{ и дисперсию } \frac{1}{c^2(q-p)^2}.$$

14. *Полумарковские процессы.* Полумарковский процесс с состояниями 1, 2, ... отличается от марковского тем, что времена пребывания могут зависеть и от конечного состояния: пусть система *входит* в состояние i в момент τ ; тогда условная вероятность того, что пребывание в i оборвется до момента $\tau + t$ скачком в k , равна $F_{ik}(t)$. Тогда $\sum_k F_{ik}(t)$ дает распределение времени пребывания, а $p_{ik} = F_{ik}(\infty)$ равно вероятности скачка в состояние k . Обозначим через $P_{ik}(t)$ вероятность состояния k в момент $t + \tau$ при условии, что вхождение в i произошло в момент τ . Выведите аналог *обратного* уравнения Колмогорова. В понятных обозначениях для преобразований Лапласа имеем

$$\Pi(\lambda) = \gamma(\lambda) + \Phi(\lambda) \Pi(\lambda),$$

где $\gamma(\lambda)$ — диагональная матрица с элементами $[1 - \sum_k \Phi_{ik}(\lambda)]/\lambda$. При $F_{ik}(t) = p_{ik}(1 - e^{-ct})$ это сводится к обратному уравнению (7.5). Построение минимального решения § 7 осуществимо²⁾.

¹⁾ Детали и граничные условия см.: W. Feller, The birth and death process as diffusion process, Jour. Math. Pures Appliques, vol. 38 (1959), 301 — 345.

²⁾ Детали см.: W. Feller, On semi-Markow processes, Proc. Nat. Acad. Sci., 51 (1964); 653—659. Полумарковские процессы были введены П. Леви и В. Смитом и были исследованы, в частности, Р. Пайком:

Эта глава содержит основы теории характеристических функций и совсем не зависит от гл. VI, VII, IX—XIV. Более глубокое изложение анализа Фурье отложено до гл. XIX.

§ 1. ОПРЕДЕЛЕНИЕ. ОСНОВНЫЕ СВОЙСТВА

Производящей функцией неотрицательной целочисленной случайной величины X называют функцию $E(s^X)$, определенную для $0 \leq s \leq 1$, т. е. математическое ожидание величины s^X . Как было показано в гл. XIII, замена переменной $s = e^{-\lambda}$ дает возможность изучать любые неотрицательные случайные величины. Пользу от таких характеристик мы получаем в первую очередь из-за мультипликативного свойства $s^{x+y} = s^x \cdot s^y$ и $e^{-\lambda(x+y)} = e^{-\lambda x} e^{-\lambda y}$. Этим свойством обладает и показательная функция от чисто мнимого аргумента, т. е. функция, определенная для действительных x равенством

$$e^{i\zeta x} = \cos \zeta x + i \sin \zeta x, \quad (1.1)$$

где ζ — действительное число и $i^2 = -1$. Так как эта функция ограничена, то ее математическое ожидание всегда существует. Использование $E(e^{i\zeta X})$ в качестве замены для производящей функции дает нам мощный универсальный метод, но это достигается ценой введения комплекснозначных функций и случайных величин. Заметим, однако, что независимая переменная остается точкой действительной прямой (или будет позже точкой пространства $\mathcal{R}^r$).

Под комплекснозначной функцией $w = u + iv$ понимается пара действительных функций u и v , определенных для действительных x . Математическое ожидание $E(w)$ — это не более чем сокращенное обозначение для $E(u) + iE(v)$. Как обычно, через $\bar{w} = u - iv$ мы обозначаем функцию, комплексно сопряженную к w , а через $|w|$ — абсолютную величину функции w (т. е. $|w|^2 = w\bar{w} = u^2 + v^2$). Основные свойства математических ожиданий сохраняются. Объяснения требует только теорема о среднем значении: если $|w| \leq a$, то $|E(w)| \leq a$. В самом деле, по неравен-

ству Шварца

$$|E(w)|^2 = (E(\dot{u}))^2 + (E(v))^2 \leq E(u^2) + E(v^2) = E(|w|^2) \leq a^2. \quad (1.2)$$

Две комплекснозначные случайные величины $W_j = U_j + iV_j$, $j=1, 2$, называются *независимыми*, если независимы пары (U_1, V_1) и (U_2, V_2) . Обычным разложением на действительную и мнимую части можно показать, что мультипликативное свойство $E(W_1 W_2) = E(W_1) \cdot E(W_2)$ выполняется (эта формула иллюстрирует преимущества записи в комплексной форме). После этих подготовительных замечаний дадим определение аналога производящей функции.

Определение. Пусть X — случайная величина с распределением вероятностей F . Характеристической функцией распределения F (или случайной величины X) называется функция φ , определенная для действительных ξ формулой

$$\varphi(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} F\{dx\} = u(\xi) + iv(\xi), \quad (1.3)$$

где

$$u(\xi) = \int_{-\infty}^{+\infty} \cos \xi x \cdot F\{dx\}, \quad v(\xi) = \int_{-\infty}^{+\infty} \sin \xi x \cdot F\{dx\}. \quad (1.4)$$

Для распределения F с плотностью f имеем, конечно,

$$\varphi(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} f(x) dx. \quad (1.5)$$

Замечание по терминологии. В анализе Фурье функцию φ принято называть *преобразованием Фурье—Стилтьеса* распределения F . Такие преобразования определены для любой конечной меры, и термин «характеристическая функция» подчеркивает, что значение меры на всем пространстве равно единице (для других мер характеристические функции не определены). Но интегралы типа (1.5) встречаются во многих ситуациях, и мы будем говорить, что (1.5) определяет *обычное преобразование Фурье* функции f . Характеристическая функция распределения F является обычным преобразованием Фурье плотности f (когда последняя существует), но термин «преобразование Фурье» применяется также и к другим функциям.

Для удобства ссылок мы перечислим некоторые основные свойства характеристических функций.

Лемма 1. Пусть $\varphi = u + iv$ — характеристическая функция случайной величины X с распределением F . Тогда

- φ непрерывна;
- $\varphi(0) = 1$ и $|\varphi(\xi)| \leq 1$ при всех ξ ;

в) $aX + b$ имеет характеристическую функцию

$$E(e^{i\zeta(aX+b)}) = e^{ib\zeta} \varphi(a\zeta). \quad (1.6)$$

(в частности, $\overline{\varphi} = \varphi(-i\zeta)$ служит характеристической функцией величины $-X$);

г) u — четная, а v — нечетная функция; характеристическая функция действительна тогда и только тогда, когда распределение F симметрично;

д) для всех ζ

$$0 \leq 1 - u(2\zeta) \leq 4(1 - u(\zeta)). \quad (1.7)$$

(Другие варианты леммы см. в задачах 1—3.)

Доказательство. а) Заметим, что $|e^{i\zeta x}| = 1$ и поэтому

$$|e^{i\zeta(x+h)} - e^{i\zeta x}| = |e^{i\zeta h} - 1|. \quad (1.8)$$

Правая часть не зависит от x и может быть сделана сколь угодно малой при достаточно малом h . Таким образом, φ на самом деле даже равномерно непрерывна. Свойство б) с очевидностью вытекает из теоремы о среднем значении, а свойство в) не требует пояснений. Для доказательства п. г) мы привлечем тот факт, что различным распределениям соответствуют различные характеристические функции. В самом деле, φ действительна тогда и только тогда, когда $\varphi = \overline{\varphi}$, т. е. тогда и только тогда, когда X и $-X$ имеют одинаковые характеристические функции. Но это равносильно тому, что X и $-X$ имеют одно и то же распределение, т. е. F симметрично. Наконец, чтобы доказать д), рассмотрим элементарное тригонометрическое соотношение

$$1 - \cos 2\zeta x = 2(1 - \cos^2 \zeta x) \leq 4(1 - \cos \zeta x), \quad (1.9)$$

справедливое, поскольку $0 \leq 1 + \cos \zeta x \leq 2$. Переходя к математическим ожиданиям, получаем (1.7). ►

Рассмотрим теперь две случайные величины X_1 и X_2 с распределениями F_1 и F_2 и характеристическими функциями φ_1 и φ_2 соответственно. Если X_1 и X_2 независимы, то мультипликативное свойство показательной функции приводит к равенству

$$E(e^{i\zeta(X_1+X_2)}) = E(e^{i\zeta X_1}) E(e^{i\zeta X_2}). \quad (1.10)$$

Этот простой результат часто употребляется, поэтому мы отметим его в форме леммы.

Лемма 2. *Свертка $F_1 \star F_2$ имеет характеристическую функцию $\varphi_1 \varphi_2$.*

Другими словами, сумме $X_1 + X_2$ двух независимых случайных величин соответствует произведение $\varphi_1 \varphi_2$ их характеристических функций¹⁾.

¹⁾ Обратное неверно: в гл. II, 4, д), и в задаче 1 гл. III, 9, показано, что в некоторых исключительных случаях сумма двух *зависимых* величин может иметь распределение $F_1 \star F_2$ и, следовательно, характеристическую функцию $\varphi_1 \varphi_2$.

Если X_2 имеет такое же распределение, как X_1 , то сумма $X_1 - X_2$ оказывается симметризованной величиной (см. гл. V, 5). Поэтому справедливо

Следствие. Функция $|\varphi|^2$ является характеристической функцией симметризованного распределения ${}^0F$.

Следующая лемма дает описание арифметических распределений.

Лемма 3. Если $\lambda \neq 0$, то три следующих утверждения эквивалентны:

- а) $\varphi(\lambda) = 1$;
- б) φ имеет период λ , т. е. $\varphi(\xi + n\lambda) = \varphi(\xi)$ для всех ξ и n ;
- в) все точки роста распределения F находятся среди точек $0, \pm h, \pm 2h, \dots$, где $h = 2\pi/\lambda$.

Доказательство. Если в) справедливо и F приписывает массу p_n точке nh , то

$$\varphi_n(\xi) = \sum p_n e^{inh\xi}.$$

Эта функция имеет период $2\pi/h$. Поэтому из в) следует б), которое в свою очередь влечет за собой а).

Обратно, если а) выполнено, то математическое ожидание отрицательной функции $1 - \cos \lambda x$ обращается в нуль, а это возможно только при условии, что $1 - \cos \lambda x = 0$ в каждой точке x , которая является точкой роста F . Таким образом, распределение F сосредоточено в точках, кратных $2\pi/\lambda$, и, следовательно, в) выполняется.

Доказанная лемма покрывает и крайний случай, когда распределение F полностью сосредоточено в нуле. Тогда $\varphi(\xi) = 1$ для всех ξ , так что каждое число будет периодом для φ . В общем случае из того, что λ — период для φ , вытекает, что периодами будут все кратные λ : $\pm\lambda, \pm 2\lambda, \dots$. Но для отличной от константы периодической функции φ существует наименьший положительный период. Его называют *истинным периодом*. Аналогично для арифметического распределения F существует *наибольшее* положительное h , для которого выполняется свойство в). Такое h называют *шагом* распределения F . Из леммы 3 следует, что шаг h и (истинный) период λ связаны равенством $\lambda h = 2\pi$. Итак, за исключением случаев, когда $\varphi(\xi) \neq 1$ для всех $\xi \neq 0$ или когда $\varphi(\xi) = 1$ тождественно, существует наименьшее $\lambda > 0$, такое, что $\varphi(\lambda) = 1$, но $\varphi(\xi) \neq 1$ для $0 < \xi < \lambda$.

Все сказанное можно сформулировать в более общем виде. Вместо $\varphi(\lambda) = 1$ предположим, что только $|\varphi(\lambda)| = 1$. Тогда существует такое действительное число b , что $\varphi(\lambda) = e^{ib\lambda}$, и мы можем применить предыдущий результат к случайной величине $X - b$ с характеристической функцией $\varphi(\xi)e^{-ib\xi}$, равной 1 при $\xi = \lambda$. Каждый период последней функции автоматически будет периодом для $|\varphi|$. Таким образом, доказана

Лемма 4. Существуют только следующие три возможности:

а) $|\varphi(\xi)| < 1$ для всех $\xi \neq 0$;
 б) $|\varphi(\lambda)| = 1$ и $|\varphi(\xi)| < 1$ для $1 < \xi < \lambda$. В этом случае $|\varphi|$ имеет период λ и существует такое действительное число b , что распределение $F(x+b)$ является арифметическим с шагом $h = 2\pi/\lambda$;

в) $|\varphi(\xi)| = 1$ для всех ξ . В этом случае $\varphi(\xi) = e^{ib\xi}$ и F сосредоточено в точке b .

Пример. Пусть распределение F сосредоточено в 0 и 1 и каждой из этих точек приписывает вероятность $1/2$. Тогда F является арифметическим распределением с шагом 1, а его характеристическая функция $\varphi(\xi) = (1 + e^{i\xi})/2$ имеет период 2π . Распределение $F(x + 1/2)$ сосредоточено в точках $\pm 1/2$. Оно имеет шаг $1/2$, и его характеристическая функция $\cos \xi/2$ имеет период 4π . ▶

§ 2. СПЕЦИАЛЬНЫЕ ПЛОТНОСТИ. СМЕСИ

Для облегчения ссылок мы даем таблицу характеристических функций десяти употребительных плотностей и указываем метод, каким получена каждая функция.

Замечания к таблице 1. 1) *Нормальная плотность.* Для тех, кто не боится интегрирования в комплексной плоскости, результат есть очевидное следствие подстановки $y = x - i\xi$. Чтобы доказать приведенную формулу, не выходя из области действительных чисел, можно использовать дифференцирование и интегрирование по частям, что даст $\varphi'(\xi) = -\xi\varphi(\xi)$. Так как $\varphi(0) = 1$, то $\ln \varphi(\xi) = -\frac{1}{2}\xi^2$, что и утверждается.

2)—3) *Равномерные плотности.* Вычисления здесь очевидны. Распределения (2) и (3) отличаются одно от другого только масштабными параметрами. Связь между их характеристическими функциями можно получить из формулы (1.6).

4) *Треугольная плотность.* Легко провести прямое вычисление, используя интегрирование по частям. Другой способ: заметим, что треугольная плотность является сверткой равномерной плотности на $-\frac{1}{2}a < x < \frac{1}{2}a$ с собой. В силу 3) ее характеристическая функция равна $\left(\frac{2}{a\xi} \sin \frac{a\xi}{2}\right)^2$.

5) Эта плотность получается применением формулы обращения (3.5) к треугольной плотности 4). (См. также задачу 4.) Характеристическая функция этой плотности весьма важна, так как многие доказательства в анализе Фурье опираются на использование характеристических функций, равных нулю вне некоторого конечного интервала.

6) *Гамма-плотности.* Можно воспользоваться подстановкой $y = x(1 - c\xi)$. Если же не выходить из области действительных

чисел, то можно разложить e^{ix} в степенной ряд. Для характеристической функции получаем на этом пути равенство

$$\Gamma(t) \sum_{n=0}^{\infty} \frac{(i\xi)^n}{n!} \int_0^{\infty} e^{-x} x^{n+t-1} dx = \sum_{n=0}^{\infty} \frac{\Gamma(n+t)}{n! \Gamma(t)} (i\xi)^n = \sum_{n=0}^{\infty} \binom{-t}{n} (-i\xi)^n,$$

в правой части которого стоит ряд для $(1-i\xi)^{-t}$ (бином с отрицательным показателем).

Таблица 1

№	Название	Плотности	Область определения	Характеристическая функция
1	Нормальная	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2}$	$-\infty < x < \infty$	$e^{-\frac{1}{2} \xi^2}$
2	Равномерная	$\frac{1}{a}$	$0 < x < a$	$\frac{e^{ia\xi} - 1}{ia\xi}$
3	Равномерная	$\frac{1}{2a}$	$ x < a$	$\frac{\sin a\xi}{a\xi}$
4	Треугольная	$\frac{1}{a} \left(1 - \frac{ x }{a}\right)$	$ x < a$	$2 \frac{1 - \cos a\xi}{a^2 \xi^2}$
5	—	$\frac{1}{\pi} \frac{1 - \cos ax}{ax^2}$	$-\infty < x < \infty$	$\begin{cases} 1 - \frac{ \xi }{a} & \text{при } \xi \leq a \\ 0 & \text{при } \xi > a \end{cases}$
6	Гамма	$\frac{1}{\Gamma(t)} x^{t-1} e^{-x}$	$x > 0, t > 0$	$\frac{1}{(1 - i\xi)^t}$
7	Двусторонняя показательная	$\frac{1}{2} e^{- x }$	$-\infty < x < \infty$	$\frac{1}{1 + \xi^2}$
8	Коши	$\frac{1}{\pi} \frac{t}{t^2 + x^2}$	$-\infty < x < \infty, t > 0$	$e^{-t \xi }$
9	Бесселева	$e^{-x} \frac{t}{x} I_t(x)$	$x > 0, t > 0$	$[1 - i\xi - \sqrt{(1 - i\xi)^2 - 1}]^t$
10	Гиперболический косинус	$\frac{1}{\pi \operatorname{ch} x}$	$-\infty < x < \infty$	$\frac{1}{\operatorname{ch}(\pi\xi/2)}$

В частном случае $t=1$ (показательное распределение) вычисления можно провести без выхода из действительной области, повторно применяя интегрирование по частям. По индукции то же самое можно проделать при всех целых t .

7) Двустороннее показательное распределение получается симметризацией показательного распределения, так что формула для характеристической функции получается из 6) при $t=1$. Это легко проверить и непосредственно, с помощью повторного интегрирования по частям.

8) *Распределение Коши*. Снова формула вытекает из предшествующей и формулы обращения (3.5). Прямая проверка представляет собой стандартное упражнение на отыскание вычетов.

9) *Бесселева плотность*. Соответствующая формула — это преобразование Фурье, аналогичное преобразованию Лапласа, полученному в гл. XIII, 3, г), и может быть доказана тем же путем.

10) *Гиперболический косинус*. Соответствующая функция распределения равна $F(x) = 1 - (2/\pi) \operatorname{arctg} e^{-x}$. Формула 10) не очень важна, но она любопытна тем, что в ней появляются «*взаимные пары*»: плотность и характеристическая функция получаются одна из другой линейным преобразованием аргумента и линейным преобразованием самой функции (нормальная плотность служит первым примером такой связи). Для вычисления характеристической функции можно разложить плотность в ряд

$$\frac{1}{2\pi} \sum (-1)^k e^{-(2k+1)|x|}.$$

Применяя результат 7) к отдельным членам ряда, получаем каноническое разложение характеристической функции на простейшие дроби. ►

(Дальнейшие примеры см. в задачах 5—8.)

Возвращаясь к общей теории, укажем метод построения новых характеристических функций, исходя из перечисленных. Принцип весьма прост и, как показывает пример б), с его помощью можно избежать длинных вычислений.

Лемма. Пусть $F_0, F_1, \dots$ — распределения вероятностей с характеристическими функциями $\varphi_0, \varphi_1, \dots$. Если $p_k \geq 0$ и $\sum p_k = 1$, то смесь

$$U = \sum p_k F_k \quad (2.1)$$

есть распределение вероятностей с характеристической функцией

$$\omega = \sum p_k \varphi_k. \quad (2.2)$$

Примеры. а) *Сумма случайного числа слагаемых*. Пусть $X_1, X_2, \dots$ — независимые случайные величины с одним и тем же распределением F и характеристической функцией φ . Пусть N — целочисленная случайная величина с производящей функцией $P(s) = \sum p_k s^k$ и не зависящая от всех X_j . Тогда случайная сумма $X_1 + \dots + X_N$ имеет распределение (2.1) с $F_k = F^{k*}$, и соответствующая характеристическая функция равна

$$\omega(\xi) = P(\varphi(\xi)). \quad (2.3)$$

Наибольшего внимания заслуживает частный случай — *обобщенное распределение Пуассона*. Здесь $p_k = e^{-t} t^k / k!$

$$\omega(\xi) = e^{-t + t\varphi(\xi)}. \quad (2.4)$$

Обычное распределение Пуассона получается в случае, когда F сосредоточено в точке 1, т. е. когда $\varphi(\xi) = e^{\xi}$.

б) *Выпуклые полигоны.* Из формулы 5 табл. 1 мы видим, что

$$\varphi(\xi) = \begin{cases} 1 - |\xi| & \text{для } |\xi| \leq 1, \\ 0 & \text{для } |\xi| \geq 1 \end{cases} \quad (2.5)$$

характеристическая. При любых положительных $a_1, \dots, a_n$ смесь

$$\omega(\xi) = p_1 \varphi\left(\frac{\xi}{a_1}\right) + \dots + p_n \varphi\left(\frac{\xi}{a_n}\right) \quad (2.6)$$

является четной характеристической функцией, график которой на интервале $0, \infty$ есть выпуклый полигон (см. рис. 1). Действительно, не ограничивая общности, можно предположить, что $a_1 < a_2 < \dots < a_n$. На интервале $0 < \xi < a_1$ график функции ω

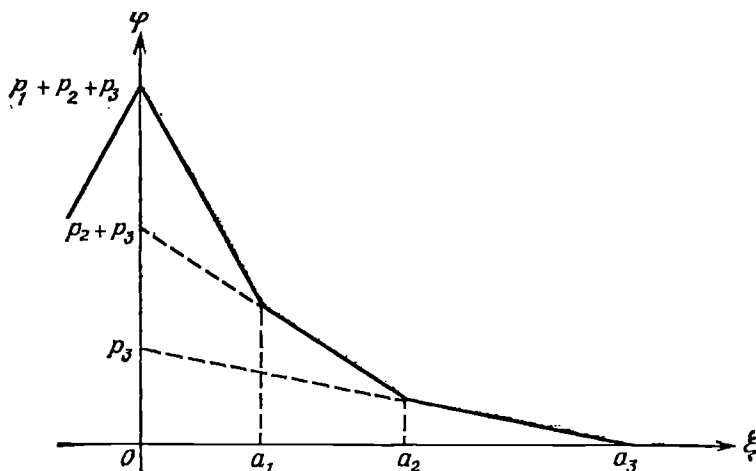


Рис. 1. Иллюстрация к примеру б).

имеет вид отрезка прямой с угловым коэффициентом $-(p_1/a_1 + \dots + p_n/a_n)$. На интервале между a_1 и a_n в этом выражении нужно отбросить член p_1/a_1 и т. д. Наконец, между a_{n-1} и a_n график совпадает с отрезком прямой с угловым коэффициентом $-p_n/a_n$. Следовательно, на $0, \infty$ график составлен из n конечных отрезков с убывающим наклоном и интервала a_n, ∞ оси ξ . Как легко видеть, каждый полигон с этими свойствами может быть графиком функции типа (2.6) (n «сторон» пересекают ось ω в точках $p_n, p_n + p_{n-1}, \dots, p_n + \dots + p_1 = 1$). Отсюда можно заключить, что любая четная функция $\omega \geq 0$ с $\omega(0) = 1$, график которой на $0, \infty$ есть выпуклый полигон, является характеристической функцией.

Простой переход к пределу приводит к известному критерию Поля [пример 3, б)] и указывает на естественный источник. Но даже и наш частный критерий приводит к удивительным и интересным результатам.

2а. Некоторые неожиданные явления

Мы несколько отступим от основной темы, чтобы ввести некоторые специальные типы характеристических функций с неожиданными и интересными свойствами. Начнем с предварительного замечания относительно арифметических распределений.

Предположим, что распределение G сосредоточено в точках вида π/L , $n=0, \pm 1, \dots$, кратных некоторому положительному фиксированному числу π/L . Точке π/L распределение G приписывает вероятность p_n , $n=0, \pm 1, \dots$. Характеристическая функция равна

$$\gamma(\xi) = \sum_{n=-\infty}^{+\infty} p_n e^{in\pi\xi/L} \quad (2.7)$$

и имеет период $2L$. Обычно нелегко отыскать явное выражение для γ , тогда как выразить вероятности p_r в терминах характеристической функции γ не сложно. Действительно, умножим выражение (2.7) на $e^{-ir\pi\xi/L}$. Вероятность p_n выступает как коэффициент периодической функции $e^{i(n-r)\pi\xi/L}$, интеграл от которой в интервале $[-L, L]$ обращается в нуль при $n \neq r$. Отсюда следует, что

$$p_r = \frac{1}{2L} \int_{-L}^L \gamma(\xi) e^{-ir\pi\xi/L} d\xi, \quad r=0, \pm 1, \dots \quad (2.8)$$

Теперь мы воспользуемся раньше времени следующим критерием из теоремы 1 гл. XIX, 4. Пусть γ — непрерывная периодическая функция с периодом $L > 0$ и такая, что $\gamma(0) = 1$. γ является характеристической функцией тогда и только тогда, когда все числа p_r из (2.8) неотрицательны. В этом случае $\{p_r\}$ автоматически является вероятностным распределением и (2.7) имеет место.

Пример. в) Выберем произвольное $L \geq 1$, и пусть γ — функция с периодом $2L$, которая при $|\xi| \leq L$ совпадает с характеристической функцией ϕ из (2.5). Тогда γ является характеристической функцией арифметического распределения, сосредоточенного в точках, кратных π/L . Фактически в силу симметрии (2.8) принимает вид

$$p_r = \frac{1}{L} \int_0^L \gamma(\xi) \cos(r\pi\xi/L) d\xi = \frac{1}{L} \int_0^L (1 - \xi) \cos(r\pi\xi/L) d\xi, \quad (2.9)$$

и простым интегрированием по частям находим, что

$$p_0 = 1/(2L), \quad p_r = L\pi^{-2} (1 - \cos r\pi/L) \geq 0, \quad r \neq 0, \quad (2.10)$$

Таким образом, нами получено целое семейство периодических характеристических функций, графики которых состоят из периодически повторяющихся правильных треугольников с основаниями $2nL - 1 < x < 2nL + 1$ и промежуточных отрезков ξ -оси (см. рис. 2). (Мы вернемся к этому примеру в более общей постановке при рассмотрении формулы суммирования Пуассона в гл. XIX, 5.)

Неожиданные примеры. (i) Две различные характеристические функции могут совпадать внутри конечного интервала $-a, a$. Это очевидное следст-

ние примеров б) и в) показывает существенную разницу между характеристическими и производящими функциями (или преобразованиями Лапласа).

(ii) Соотношение $F_1 \star F_1 = F \star F_2$ между тремя распределениями вероятностей не влечет¹⁾ в себя равенства $F_1 = F_2$.

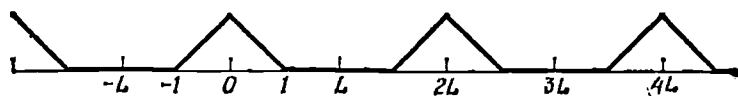


Рис. 2. Иллюстрация к примеру в).

Действительно, для функции φ , определенной в (2.5), имеем $\varphi\varphi_1 = \varphi\varphi_2$ для двух произвольных характеристических функций, совпадающих внутри интервала $[-1, 1]$. В частности, имеем $\varphi^2 = \varphi\gamma$ для любой периодической характеристической функции γ из примера в).

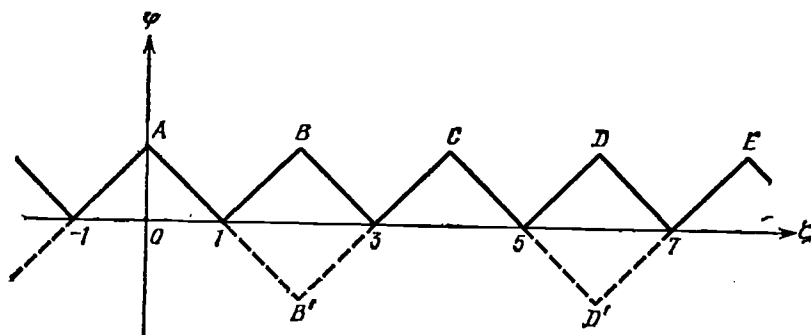


Рис. 3. Иллюстрация к неожиданному примеру ii).

(iii) Еще более удивительно то, что существуют две действительные характеристические функции φ_1 и φ_2 , такие, что всюду $|\varphi_2| = \varphi_1 \geq 0$. В самом деле, рассмотрим характеристическую функцию γ примера в) с $L=1$. Ее график на рис. 3 изображен сплошной ломаной линией. Мы видели, что соответствующее распределение приписывает нулю массу, равную $1/2$. Исключая этот атом и удваивая все другие массы, получаем распределение с характеристической функцией $2\gamma - 1$. Ее график изображен ломаной линией, колеблющейся в пределах ± 1 ; ее отрезки имеют наклон ± 2 . Отсюда следует, что $2\gamma(\frac{1}{2}\zeta) - 1$ есть характеристическая функция, график которой получается из графика функции γ отражением каждого второго треугольника относительно оси ζ (рис. 3). Таким образом, $\gamma(\zeta)$ и $2\gamma(\frac{1}{2}\zeta) - 1$ — две действительные характеристические функции, которые отличаются только знаками. (Подобное построение, относящееся к рис. 2, см. в задаче 9).

¹⁾ Статистики и астрономы иногда задаются вопросом: содержит ли данное распределение нормальную компоненту? Этот вопрос осмыслен, так как характеристическая функция нормального распределения $\mathfrak{N}_a$ не имеет нулей и, следовательно, из $\mathfrak{N}_a \star F_1 = \mathfrak{N}_a \star F_2$ вытекает $\varphi_1 = \varphi_2$, и, следовательно, по теореме единственности $F_1 = F_2$.

§ 3. ЕДИНСТВЕННОСТЬ. ФОРМУЛЫ ОБРАЩЕНИЯ

Пусть F и G — два распределения с характеристическими функциями φ и γ соответственно. Тогда

$$e^{-i\zeta t} \varphi(\zeta) = \int_{-\infty}^{+\infty} e^{i\zeta(x-t)} F\{dx\}. \quad (3.1)$$

Интегрируя по $G\{d\zeta\}$, имеем

$$\int_{-\infty}^{+\infty} e^{-i\zeta t} \varphi(\zeta) G\{d\zeta\} = \int_{-\infty}^{+\infty} \gamma(x-t) F\{dx\}. \quad (3.2)$$

Это соотношение известно под названием *равенства Парсеваля* (которое, впрочем, может быть записано во многих эквивалентных формах; мы вернемся к этому в гл. XIX).

Будем рассматривать только специальный случай, когда $G = \mathfrak{N}_a$ есть нормальное распределение с плотностью $a\pi(a\zeta)$. Соответствующая характеристическая функция равна $\gamma(\zeta) = \sqrt{2\pi} \pi(\zeta/a)$, и поэтому (3.2) принимает вид

$$\int_{-\infty}^{+\infty} e^{-i\zeta t} \varphi(\zeta) a\pi(a\zeta) d\zeta = \sqrt{2\pi} \int_{-\infty}^{+\infty} \pi\left(\frac{x-t}{a}\right) F\{dx\}, \quad (3.3)$$

или, что то же самое,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\zeta t} \varphi(\zeta) e^{-\frac{1}{2} a^2 \zeta^2} d\zeta = \int_{-\infty}^{+\infty} \frac{1}{a} \pi\left(\frac{x-t}{a}\right) F\{dx\}. \quad (3.4)$$

Удивительно много выводов может быть сделано из этого тождества. Прежде всего правая часть является плотностью $\mathfrak{N}_a \star F$ — свертки F с нормальным распределением с нулевым математическим ожиданием и дисперсией a^2 . Таким образом, знание φ позволяет в принципе вычислять распределения $\mathfrak{N}_a \star F$ для всех a . Но $\mathfrak{N}_a$ имеет дисперсию a^2 , и, следовательно, $\mathfrak{N}_a \star F \rightarrow F$ при $a \rightarrow 0$. Отсюда следует, что φ однозначно определяет распределение F . Таким образом, доказана важная

Теорема 1. *Различным распределениям вероятностей соответствуют различные характеристические функции.*

Предположим далее, что дана последовательность распределений вероятностей F_n с характеристическими функциями φ_n , такими, что $\varphi_n(\zeta) \rightarrow \varphi(\zeta)$ для всех ζ . По теореме о выборе из гл. VIII, 6, существует последовательность $\{n_k\}$ и (возможно, несобственное)

распределение F , такие, что $F_{n_k} \rightarrow F$. Применяя (3.4) к паре (φ_{n_k}, F_{n_k}) и устремляя $k \rightarrow \infty$ в пределе, мы снова получим тождество (3.4) (в левой части используется теорема об ограниченной сходимости, в правой части — то, что подынтегральное выражение $\pi((x-t)/a)$ обращается в нуль на бесконечности). Но мы видим, что для данной φ тождеством (3.4) F определяется однозначно и, следовательно, предел F одинаков для всех сходящихся подпоследовательностей $\{F_{n_k}\}$. Таким образом, справедлива

Лемма. Пусть F_n — распределение вероятностей с характеристической функцией φ_n . Если $\varphi_n(\xi) \rightarrow \varphi(\xi)$ при всех ξ , то существует (возможно, несобственное) распределение F , такое, что $F_n \rightarrow F$.

Пример. а) Пусть распределение вероятностей U имеет действительную неотрицательную характеристическую функцию ω . Пусть $F_n = U^{n*}$, так что $\varphi_n(\xi) = \omega^n(\xi)$. Тогда $\varphi_n(\xi) \rightarrow 0$, за исключением множества точек, где $\omega(\xi) = 1$, а по лемме 4 § 1 это множество состоит из всех точек вида $\pm n\lambda$, где λ — некоторое неотрицательное фиксированное число. Отсюда следует, что левая часть в (3.4) равна тождественно нулю, и поэтому $U^{n*} \rightarrow 0$. Используя симметризацию, заключаем, что $G^{n*} \rightarrow 0$ для любого распределения вероятностей G , которое не сосредоточено в нуле. ►

В следующей теореме, по существу, утверждается, что предельное распределение F является несобственным тогда и только тогда, когда предельная функция φ разрывна в нуле.

Теорема 2. (Теорема непрерывности.) Последовательность $\{F_n\}$ распределений вероятностей сходится в собственном смысле к распределению вероятностей F тогда и только тогда, когда последовательность $\{\varphi_n\}$ соответствующих характеристических функций сходится поточечно к предельной функции φ и эта функция φ является непрерывной в некоторой окрестности нуля.

В этом случае φ является характеристической функцией распределения F . (Отсюда следует, что φ непрерывна всюду и сходимость $\varphi_n \rightarrow \varphi$ равномерна в каждом конечном интервале.)

Доказательство. а) Предположим, что $F_n \rightarrow F$, где F — собственное распределение вероятностей. Согласно следствию из гл. VIII, 1, характеристические функции φ_n сходятся к характеристической функции φ распределения F , и сходимость эта равномерна на конечных интервалах.

б) Предположим, что $\varphi_n(\xi) \rightarrow \varphi(\xi)$ для всех ξ . По предыдущей лемме предел $F = \lim F_n$ существует и справедливо тождество (3.3). Левая часть есть математическое ожидание ограниченной функции $e^{-t\varphi(\xi)}$, вычисленное относительно нормального распределения

с нулевым математическим ожиданием и дисперсией a^{-2} . При $a \rightarrow \infty$ это распределение стремится сосредоточиться вблизи нуля, и поэтому левая часть стремится к $\varphi(0)$, если φ непрерывна в некоторой окрестности нуля. Но так как $\varphi_n(0) = 1$, то $\varphi(0) = 1$. С другой стороны, $\sqrt{2\pi} \pi(x) \leq 1$ для всех x и поэтому правая часть не превосходит $F\{-\infty, \infty\}$. Таким образом, $F\{-\infty, \infty\} \geq 1$, и, следовательно, распределение F — собственное. ►

Следствие. Непрерывная функция, являющаяся пределом (в смысле поточечной сходимости) последовательности характеристических функций, сама является характеристической.

Пример. б) Критерий Поля. Пусть ω — действительная непрерывная четная функция, равная единице в нуле и нулю на бесконечности. Пусть график ее является выпуклым на $[0, \infty)$. Тогда ω — характеристическая функция. Действительно, мы видели в примере 2, б), что это утверждение верно для случая, когда график функции ω — выпуклый полигон. Так как полигоны, вписанные в вогнутую линию, сами выпуклы, то общее утверждение вытекает из приведенного выше следствия. Этот критерий (вместе со сложным доказательством) в свое время сыграл интересную роль. Поля с его помощью доказал в 1920 г., что $e^{-|\xi|^\alpha}$ при $0 < \alpha \leq 1$ является характеристической функцией устойчивого распределения (говорят, что Коши был уверен в этом, но не имел доказательства). В действительности $e^{-|\xi|^\alpha}$ есть характеристическая функция даже при $1 < \alpha \leq 2$, но критерий уже неприменим. ►

Мы отложим до гл. XIX систематическое применение метода, развитого при доказательстве теоремы 1. Мы, однако, используем этот метод здесь для вывода важной теоремы, которая уже была использована при составлении таблицы в § 2 (формулы 5 и 8). Условимся для краткости писать $\varphi \in L$ в том и только том случае, если $|\varphi|$ интегрируема на $-\infty, \infty$.

Теорема 3. (Обращение интеграла Фурье.) Пусть φ — характеристическая функция распределения F и $\varphi \in L$. Тогда F имеет ограниченную непрерывную плотность f , определенную формулой

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\xi x} \varphi(\xi) d\xi. \quad (3.5)$$

Доказательство. Обозначим правую часть (3.4) через $f_a(t)$. Тогда f_a — плотность свертки $F_a = \mathfrak{N}_a \star F$ распределения F с нормальным распределением $\mathfrak{N}_a$ с нулевым математическим ожиданием и дисперсией a^2 . Отсюда, как уже упоминалось, следует, что $F_a \rightarrow F$ при $a \rightarrow 0$. Из представления левой части (3.4) видно, что $f_a(t) \rightarrow f(t)$ ограничено, где f — ограниченная непрерывная

функция, определенная в (3.5). Таким образом, для каждого ограниченного интервала I

$$F_a\{I\} = \int_I f_a(t) dt \rightarrow \int_I f(x) dx. \quad (3.6)$$

Но если I — интервал непрерывности для F , то $F_a\{I\}$ стремится к $F\{I\}$, так что f действительно является плотностью распределения F . ►

Следствие. Если $\varphi \geq 0$, то $\varphi \in L$ тогда и только тогда, когда соответствующее распределение F имеет ограниченную плотность.

Доказательство. По теореме 3 интегрируемость функции φ влечет за собой существование у F непрерывной ограниченной плотности. Обратно, если F имеет плотность $f < M$, то мы получаем из (3.4) при $t=0$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(\xi) e^{-\frac{1}{2}a^2\xi^2} d\xi = \frac{1}{\sqrt{2\pi} \cdot a} \int_{-\infty}^{+\infty} e^{-x^2/(2a^2)} f(x) dx < M. \quad (3.7)$$

Подынтегральное выражение слева неотрицательно, и, если бы φ не была интегрируема, левый интеграл стремился бы к ∞ при $a \rightarrow 0$. ►

Примеры. в) *Тождество Планшереля.* Пусть распределение F имеет плотность f и характеристическую функцию φ . Имеем $|\varphi|^2 \in L$ тогда и только тогда, когда $f^2 \in L$, и в этом случае

$$\int_{-\infty}^{+\infty} f^2(y) dy = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\varphi(\xi)|^2 d\xi. \quad (3.8)$$

Действительно, $|\varphi|^2$ — это характеристическая функция симметризованного распределения ${}^0F$. Если $|\varphi|^2 \in L$, то плотность

$${}^0f(x) = \int_{-\infty}^{+\infty} f(y+x) f(y) dy \quad (3.9)$$

распределения ${}^0F$ ограничена и непрерывна. Левая часть в (3.8) равна ${}^0f(0)$, а применяя формулу обращения (3.5) к ${}^0f$, убеждаемся, что то же самое справедливо и для правой стороны. Обратно, если $f^2 \in L$, то применение неравенства Шварца к (3.9) показывает, что ${}^0f$ ограничена, и в силу приведенного выше следствия $|\varphi|^2 \in L$. Мы вернемся к соотношению (3.8) в гл. XIX, 7.

г) *Теорема непрерывности для плотностей.* Пусть φ_n и φ — интегрируемые характеристические функции, для которых

$$\int_{-\infty}^{+\infty} |\varphi_n(\xi) - \varphi(\xi)| d\xi \rightarrow 0, \quad (3.10)$$

По следствию теоремы 3 соответствующие распределения F_n и F имеют ограниченные и непрерывные плотности f_n и f . Из формулы обращения (3.5) ясно, что

$$|f_n(x) - f(x)| \leq (2\pi)^{-1} \int_{-\infty}^{+\infty} |\varphi_n(\xi) - \varphi(\xi)| d\xi.$$

При этом сходимость $f_n \rightarrow f$ равномерна. (См. задачу 12.)

д) *Формула обращения для функций распределения.* Пусть F — распределение с характеристической функцией φ , и пусть $h > 0$ — произвольно, но фиксировано. Докажем, что

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(\xi) \frac{1 - e^{-i\xi h}}{i\xi h} e^{-i\xi x} d\xi, \quad (3.11)$$

как только подынтегральное выражение в правой части интегрируемо (например, если подынтегральная функция есть $O(1/\xi^2)$, т. е. если $|\varphi(\xi)| = O(1/\xi)$ при $\xi \rightarrow \infty$). В самом деле, левая часть представляет собой плотность свертки F с распределением, равномерным на $[-h, 0]$; множитель при $e^{-i\xi x}$ под интегралом есть характеристическая функция этой свертки. Поэтому (3.11) представляет собой не что иное, как частный случай общей формулы обращения (3.5). ►

Замечание о так называемой формуле обращения. Формула (3.11) применима только в случае, когда функция $|\varphi(\xi)/\xi|$ интегрируема на бесконечности. В общем же случае применяются формулы, получаемые ее видоизменением. Например, пусть F_a снова обозначает свертку распределения F с симметричным относительно нуля нормальным распределением с дисперсией a^2 . Тогда по (3.11)

$$\frac{F_a(x+h) - F_a(x)}{h} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(\xi) e^{-\frac{1}{2}a^2\xi^2} \frac{1 - e^{-i\xi h}}{i\xi h} e^{-i\xi x} d\xi. \quad (3.12)$$

Если x и $x+h$ — точки непрерывности распределения F , то правая часть стремится к $[F(x+h) - F(x)]/h$ при $a \rightarrow 0$. Это — типичный пример «теоремы обращения». Можно предложить бесчисленное множество эквивалентных формулировок. Традиционная форма получается заменой в (3.12) нормального распределения распределением, равномерным на $[-t, t]$, и переходом к пределу при $t \rightarrow \infty$. В силу традиции формулы обращения остаются популярной темой, хотя они и потеряли в значительной степени свою важность. Заметим, что вывод их из интеграла Дирихле выпадает из логической структуры всей теории.

От распределений с интегрируемыми характеристическими функциями мы перейдем к *решетчатым распределениям*. Пусть F приписывает массу p_k точке $b + kh$, где $p_k \geq 0$ и $\sum p_k = 1$. Характеристическая функция φ имеет вид

$$\varphi(\xi) = \sum_{-\infty}^{+\infty} p_k e^{i(b+kh)\xi}. \quad (3.13)$$

Предположим, что $h > 0$.

Теорема 4. Если φ — характеристическая функция вида (3.13), то

$$\rho_r = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} \varphi(\xi) e^{-i(b+rh)\xi} d\xi. \quad (3.14)$$

Доказательство. Подынтегральное выражение может быть представлено в виде ряда, где множитель при ρ_k равен $e^{i(k-r)h\xi}$. Интеграл от него равен 0, если $k \neq r$, и $2\pi/h$, если $k=r$, так что (3.14) верно. ►

§ 4. СВОЙСТВА РЕГУЛЯРНОСТИ

Основной результат этого параграфа может быть грубо резюмирован так: чем меньше «хвосты» распределения F , тем глаже его характеристическая функция φ ; обратно, чем глаже F , тем лучше ведет себя φ на бесконечности (леммы 2 и 4). Многие оценки, связанные с характеристическими функциями, опираются на оценку ошибки, с которой функция e^{it} аппроксимируется конечным отрезком своего ряда Тейлора. Следующая ниже лемма показывает, что эта ошибка мажорируется первым отброшенным членом.

Лемма 1¹⁾. Для $n=1, 2, \dots$ и $t > 0$

$$\left| e^{it} - 1 - \frac{it}{1!} - \dots - \frac{(it)^{n-1}}{(n-1)!} \right| \leq \frac{t^n}{n!}. \quad (4.1)$$

Доказательство. Обозначим выражение, стоящее под знаком абсолютной величины, через $\rho_n(t)$. Тогда

$$\rho_1(t) = i \int_0^t e^{ix} dx, \quad (4.2)$$

откуда $|\rho_1(t)| \leq t$. Далее, при $n > 1$

$$\rho_n(t) = i \int_0^t \rho_{n-1}(x) dx, \quad (4.3)$$

и (4.1) выводится теперь по индукции. ►

В дальнейшем F будет обозначать произвольное распределение, а φ — его характеристическую функцию. Моменты и абсолютные моменты распределения F (коль скоро они существуют) обозначим

$$m_n = \int_{-\infty}^{+\infty} x^n F\{dx\}, \quad M_n = \int_{-\infty}^{+\infty} |x|^n F\{dx\}. \quad (4.4)$$

¹⁾ То же самое доказательство показывает, что при оставлении в ряде Тейлора для $\sin t$ или $\cos t$ конечного числа членов ошибка имеет тот же знак, что и первый отброшенный член, и меньше его по абсолютной величине. Например, $1 - \cos t \leq t^2/2$.

Лемма 2. Если $M_n < \infty$, то n -я производная от φ существует, непрерывна и задается формулой

$$\varphi^{(n)}(\xi) = i^n \int_{-\infty}^{+\infty} e^{i\xi x} x^n F\{dx\}. \quad (4.5)$$

Доказательство. Разностное отношение для φ равно

$$\frac{\varphi(\xi+h) - \varphi(\xi)}{h} = \int_{-\infty}^{+\infty} e^{i\xi x} \frac{e^{ihx} - 1}{h} F\{dx\}. \quad (4.6)$$

В соответствии с последней леммой подынтегральное выражение мажорируется функцией $|x|$, и потому для $n=1$ утверждение (4.5) вытекает из теоремы о мажорированной сходимости. Общий случай доказывается по индукции. ►

Следствие. Если $m_2 < \infty$, то

$$\varphi'(0) = im_1, \quad \varphi''(0) = -m_2. \quad (4.7)$$

Утверждение, обратное к последнему¹⁾, также верно: если $\varphi''(0)$ существует, то $m_2 < \infty$.

Доказательство. Обозначая действительную часть функции φ через u , имеем

$$\frac{1 - u(h)}{h^2} = \int_{-\infty}^{+\infty} \frac{1 - \cos hx}{h^2 x^2} \cdot x^2 F\{dx\}. \quad (4.8)$$

Так как $u''(0)$ существует, то производная u' определена в окрестности нуля и в нуле непрерывна. В частности, $u'(0) = 0$, так как u — четная функция. По теореме о среднем значении существует такое число θ , что $0 < \theta < 1$ и

$$\left| \frac{u(h) - 1}{h^2} \right| = \left| \frac{u'(\theta h)}{h} \right| \leq \left| \frac{u'(\theta h)}{\theta h} \right|. \quad (4.9)$$

При $h \rightarrow 0$ правая часть стремится к $u''(0)$. Но отношение под интегралом в (4.8) стремится к $1/2$, так что интеграл стремится к ∞ , если $m_2 = \infty$. См. обобщение в задаче 15. ►

Примеры. а) Функция φ , отличная от постоянной и такая, что $\varphi''(0) = 0$, не может быть характеристической, так как соответствующее распределение должно было бы иметь второй момент, равный нулю. Например, функция $e^{-|\xi|^\alpha}$ при $\alpha > 2$ не будет характеристической.

¹⁾ Метод неприменим к первой производной. Проблема отыскания условий существования $\varphi'(0)$, привлекавшая долгое время внимание, решена в гл. XVII, 2а.

б) *Слабый закон больших чисел.* Пусть $X_1, X_2, \dots$ — независимые случайные величины с $E(X_j) = 0$ и одной и той же характеристической функцией φ . Положим $S_n = X_1 + \dots + X_n$. Среднее арифметическое S_n/n имеет характеристическую функцию $\varphi^n(\xi/n)$. В окрестности нуля $\varphi(h) = 1 + o(h)$, и потому $\varphi(\xi/n) = 1 + o(1/n)$ при $n \rightarrow \infty$. Переходя к логарифмам, мы видим, что $\varphi^n(\xi/n) \rightarrow 1$. По теореме непрерывности (теорема 2 из § 3) отсюда вытекает, что *распределение* величины S_n/n сходится к распределению, сосредоточенному в нуле. Это и есть слабый закон больших чисел. Такая простота и ясность доказательства вообще типичны для метода характеристических функций. Некоторое видоизменение этого рассуждения приведет нас к доказательству центральной предельной теоремы¹⁾. ►

Лемма 3. (Риман — Лебег.) Если g интегрируема и

$$\gamma(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} g(x) dx, \quad (4.10)$$

то $\gamma(\xi) \rightarrow 0$ при $\xi \rightarrow \pm \infty$.

Доказательство. Утверждение проверяется непосредственно для ступенчатых функций с конечным числом ступенек. Для произвольной интегрируемой функции g и любого $\varepsilon > 0$ найдется по теореме об аппроксимации в среднем из гл. IV, 2, такая ступенчатая функция g_1 с конечным числом ступенек, что

$$\int_{-\infty}^{+\infty} |g(x) - g_1(x)| dx < \varepsilon. \quad (4.11)$$

Соответствующее преобразование g_1 функции γ_1 из (4.10) обращается в нуль на бесконечности, и вследствие (4.10) и (4.11) для всех ξ имеет место неравенство $|\gamma(\xi) - \gamma_1(\xi)| < \varepsilon$. Соответственно $|\gamma_1(\xi)| < 2\varepsilon$ для всех достаточно больших $|\xi|$, а поскольку ε произвольно, то $\gamma_1(\xi) \rightarrow 0$ при $\xi \rightarrow \pm \infty$. ►

В качестве простого следствия получается

Лемма 4. Если F имеет плотность f , то $\varphi(\xi) \rightarrow 0$ при $\xi \rightarrow \pm \infty$. Если f имеет интегрируемые производные $f', \dots, f^{(n)}$, то $|\varphi(\xi)| = o(|\xi|^{-n})$ при $|\xi| \rightarrow \infty$.

Доказательство. Первое утверждение содержится в лемме 3.

¹⁾ В гл. VII, 7, было показано, что слабый закон больших чисел может выполняться даже в том случае, когда величины X_j не имеют математических ожиданий. Доказательство в тексте показывает, что существование производной $\varphi'(0)$ есть достаточное условие. На самом деле это условие также необходимо (см. гл. XVII, 2а).

Если f' интегрируема, то интегрирование по частям показывает, что

$$\varphi(\xi) = \frac{1}{i\xi} \int_{-\infty}^{+\infty} e^{i\xi x} f'(x) dx \quad (4.12)$$

и, следовательно, $|\varphi(\xi)| = o(|\xi|^{-1})$ и т. д. ►

Дополнение. Разложение Тейлора для характеристических функций.

Неравенство (4.1) можно переписать в виде

$$\left| e^{it\xi} \left(e^{it\xi} - 1 - \frac{it\xi}{1!} - \dots - \frac{(it\xi)^{n-1}}{(n-1)!} \right) \right| \leq \frac{|t\xi|^n}{n!}. \quad (4.13)$$

Отсюда, используя (4.5), получаем

$$\left| \varphi(\xi + t) - \varphi(\xi) - \frac{t}{1!} \varphi'(\xi) - \dots - \frac{t^{n-1}}{(n-1)!} \varphi^{(n-1)}(\xi) \right| < M_n \frac{|t|^n}{n!}. \quad (4.14)$$

Если $M_n < \infty$, то это неравенство верно при любых ξ и t и дает верхнюю границу для модуля разности между φ и первыми членами ряда Тейлора. В частном случае, когда F сосредоточено в точке 1, неравенство (4.14) превращается в (4.1).

Предположим теперь, что все моменты существуют и что

$$\lim_{n \rightarrow \infty} \frac{1}{n} M_n^{1/n} = \lambda < \infty. \quad (4.15)$$

Из формулы Стирлинга для $n!$ тривиально вытекает, что при $|t| < 1/(3\lambda)$ правая часть неравенства (4.14) стремится к нулю при $n \rightarrow \infty$, так что ряд Тейлора для φ сходится в некотором интервале вокруг ξ . Отсюда следует, что φ аналитична в некоторой окрестности любой точки действительной оси и, стало быть, однозначно определяется своим разложением в степенной ряд в окрестности нуля. Но $\varphi^{(n)}(0) = (i)^n m_n$, так что φ однозначно определяется моментами m_n распределения F . Соответственно если (4.15) выполняется, то распределение F однозначно определяется своими моментами и φ аналитична в некоторой окрестности действительной оси. Этот критерий единственности слабее, чем указанное в гл. VII достаточное условие (3.14), принадлежащее Карлеману: $\sum \bar{M}_n^{1/n} = \infty$.

Но оба критерия не так уж далеки один от другого (пример распределения, не определяемого однозначно своими моментами, приведен в гл. VII, 3).

§ 5. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ ОДИНАКОВО РАСПРЕДЕЛЕННЫХ СЛАГАЕМЫХ

Исследования, связанные с центральной предельной теоремой, сильно влияли на развитие и совершенствование методов, применяемых теперь повсюду в теории вероятностей. Поэтому представляется поучительным сравнение различных доказательств. До последнего времени метод характеристических функций (впервые употребленный П. Леви) был несравненно проще, чем прямой подход, примененный Линдбергом (мы не говорим здесь о других доказательствах). Освобожденный от усложнений современный вариант последнего метода (изложенный в гл. VIII, 4) не труднее метода характеристических функций и, кроме того, имеет некото-

рые преимущества. С другой стороны, метод характеристических функций приводит к уточнениям, которые в настоящий момент не удастся получить прямыми методами. К их числу относятся локальные предельные теоремы этого параграфа, а также оценки ошибок и асимптотические разложения, изучаемые в следующей главе. Мы выделяем случай слагаемых, имеющих одно и то же распределение отчасти из-за важности и отчасти потому, что желаем на примере простейшего случая показать существование метода.

Всюду в этом параграфе $X_1, X_2, \dots$ предполагаются независимыми случайными величинами с одним и тем же распределением F , которому соответствует характеристическая функция φ . Мы предполагаем также, что

$$E(X_i) = 0, \quad E(X_i^2) = 1, \quad (5.1)$$

и обозначаем $S_n = X_1 + \dots + X_n$.

Теорема 1¹⁾. *Распределение величины $S_n/\sqrt{n}$ сходится к нормальному распределению $\mathfrak{N}$.*

В силу теоремы непрерывности (теорема 2 из § 3) это утверждение эквивалентно следующему: при $n \rightarrow \infty$ и всех ξ

$$\varphi^n\left(\frac{\xi}{\sqrt{n}}\right) \rightarrow e^{-\frac{1}{2}\xi^2}. \quad (5.2)$$

Доказательство. По лемме 2 предыдущего параграфа существует непрерывная вторая производная φ'' , и, следовательно, по формуле Тейлора

$$\varphi(x) = \varphi(0) + x\varphi'(0) + \frac{1}{2}x^2\varphi''(0) + o(x^2), \quad x \rightarrow 0. \quad (5.3)$$

Задаваясь произвольным ξ и полагая $x = \xi/n$, получаем

$$\varphi\left(\frac{\xi}{\sqrt{n}}\right) = 1 - \frac{1}{2n}\xi^2 + o\left(\frac{1}{n}\right), \quad n \rightarrow \infty. \quad (5.4)$$

Возводя (5.4) в n -ю степень, получаем (5.2). ►

Естественно ожидать, что, когда F обладает плотностью f , плотность величины $S_n/\sqrt{n}$ должна сходиться к нормальной плотности π . Это не всегда так, но, к счастью, исключения носят «патологический» характер. Сформулированная ниже теорема охватывает множество практически интересных случаев.

Теорема 2. *Если функция $|\varphi|$ интегрируема, то случайная величина $S_n/\sqrt{n}$ имеет плотность f_n , равномерно сходящуюся к нормальной плотности π .*

¹⁾ Существование дисперсии вовсе не обязательно для асимптотической нормальности величины S_n . Необходимые и достаточные условия даны в следствии 1 в гл. XVII, 5.

Доказательство. Формула обращения интегралов Фурье (3.5) применима как к f_n , так и к n , и поэтому

$$|f_n(x) - n(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \varphi^n\left(\frac{\xi}{\sqrt{n}}\right) - e^{(-1/2)\xi^2} \right| d\xi. \quad (5.5)$$

Правая часть здесь не зависит от x , и мы должны показать, что она стремится к нулю при $n \rightarrow \infty$. Ввиду (5.3) можно выбрать $\delta > 0$ так, чтобы

$$|\varphi(\xi)| \leq e^{(-1/4)\xi^2} \text{ при } |\xi| < \delta. \quad (5.6)$$

Мы разобьем теперь интеграл (5.5) на три части и покажем, что каждую часть можно сделать меньше ε при достаточно большом n . 1) Как мы видели в последнем доказательстве, внутри фиксированного интервала $-a \leq \xi \leq a$ подынтегральное выражение равномерно стремится к нулю, так что интеграл по этому промежутку стремится к нулю. 2) Для $a < |\xi| < \delta\sqrt{n}$ подынтегральное выражение меньше, чем $2e^{(-1/4)\xi^2}$, так что величина соответствующего интеграла меньше ε , если только a взято достаточно большим. 3) Мы знаем из леммы 4 из § 1, что $|\varphi(\xi)| < 1$ для $\xi \neq 0$, и из леммы 3 § 4, что $\varphi(\xi) \rightarrow 0$ при $|\xi| \rightarrow \infty$. Отсюда следует, что максимум функции $|\varphi(\xi)|$ при $|\xi| \geq \delta$ равен числу $\eta < 1$. Поэтому часть интеграла (5.5), взятая из области $|\xi| > \delta\sqrt{n}$, меньше, чем

$$\eta^{n-1} \int_{-\infty}^{+\infty} \left| \varphi\left(\frac{\xi}{\sqrt{n}}\right) \right| d\xi + \int_{|\xi| > \delta\sqrt{n}} e^{(-1/2)\xi^2} d\xi. \quad (5.7)$$

Первый интеграл равен интегралу от $\sqrt{n}|\varphi|$, так что выражение (5.7) стремится к нулю. ►

На самом деле из доказательства вытекает ¹⁾ более сильный результат, а именно: если $|\varphi|^r \in L$ при некотором целом r , то $f_n \rightarrow n$ равномерно для всех x . В то же время следствие теоремы 3 § 3 показывает, что если $|\varphi|^r$ не интегрируема ни при каком r , то каждая из функций f_n не ограничена. Приведем любопытные примеры, подтверждающие, что такие патологические случаи действительно встречаются.

Примеры. а) Для $x > 0$ и $p \geq 1$ положим

$$u_p(x) = \frac{1}{x \log^2 p x}. \quad (5.8)$$

Пусть g — плотность распределения, сосредоточенного на интервале $\overline{0, 1}$, и пусть $g(x) > u_p(x)$ на некотором интервале $\overline{0, h}$. Существует интервал $\overline{0, \delta}$, в котором u_p монотонно убывает, и внутри этого интервала

$$g^{2*}(x) \geq \int_0^x u_p(x-y) u_p(y) dy > x u_p^2(x) = u_{2p}(x). \quad (5.9)$$

¹⁾ Единственное изменение состоит в том, что в (5.7) множитель η^{n-1} заменяется на η^{n-r} , а φ на φ^r .

По индукции устанавливаем, что при $n = 2^k$ существует интервал $\overline{0, h_n}$, в котором $g^{n*} \geq u_{np}$, и, следовательно, $g^{n*}(x) \rightarrow \infty$ при $x \rightarrow 0+$. Таким образом, каждая свертка g^{n*} не ограничена.

б) В варианте предыдущего примера та же самая патологическая особенность проявляется в еще более резкой форме. Пусть v — плотность, полученная симметризацией функции g . Положим

$$f(x) = \frac{1}{2} (v(x+1) + v(x-1)). \quad (5.10)$$

Тогда $f(x)$ будет четной плотностью, сосредоточенной на $\overline{-2, 2}$. Мы можем предположить, что соответствующая дисперсия равна 1. Анализируя предыдущий пример, мы видим, что v непрерывна всюду, кроме нуля, где она не ограничена. Это же верно и по отношению ко всем сверткам v^{n*} . Теперь $f^{2n*}(x)$ является линейной комбинацией функций $v^{2n*}(x+2k)$ с $k=0, \pm 1, \pm 2, \dots$, $\pm n$, следовательно, не ограничена во всех этих точках. Плотности нормированной суммы $S_{2n}/\sqrt{2n}$ случайных величин X_j с плотностью f определяется формулой $f_{2n}(x) = \sqrt{2n} f^{2n*}(x/\sqrt{2n})$. Она непрерывна всюду, кроме $2n+1$ точек вида $2k/\sqrt{2n}$ ($k=0, \pm 1, \dots, \pm n$), в которых она не ограничена. Так как каждой рациональной точке t соответствует бесконечно много пар (k, n) целых чисел, таких, что $2k/\sqrt{2n} = t$, то мы приходим к выводу: *распределение случайной величины $S_n/\sqrt{n}$ стремится к $\mathfrak{N}$, но плотности f_n не сходятся ни в одной рациональной точке и последовательность $\{f_n\}$ не будет ограниченной ни в одном интервале.*

в) См. задачу 20. ▶

Чтобы сделать картину полной, перейдем теперь к *решетчатым распределениям*. Предположим, что случайные величины X_j принимают значения вида $b, b \pm h, b \pm 2h, \dots$. Предположим, что h является *шагом* распределения, т. е. максимальным положительным числом с указанным свойством. Лемма 4 из § 1 утверждает, что $|\varphi|$ имеет период $2\pi/h$ и, следовательно, функция $|\varphi|$ не интегрируема. Однако теореме 2 соответствует аналогичная теорема, касающаяся «масс» возможных значений (атомов) распределения величины $S_n/\sqrt{n}$. Все эти атомы находятся среди точек вида $x = (nb + kh)/\sqrt{n}$, где $k=0, \pm 1, \pm 2, \dots$. Положим для таких x

$$p_n(x) = P \left\{ \frac{S_n}{\sqrt{n}} = x \right\}. \quad (5.11)$$

Для всех других x функция $p_n(x)$ не будет определена. Таким образом, значения x в соотношении (5.12) принадлежат наименьшей решетке, содержащей все атомы случайной величины $S_n/\sqrt{n}$.

Теорема 3¹). Если F — решетчатое распределение с шагом h , то при $n \rightarrow \infty$

$$\frac{\sqrt{n}}{h} p_n(x) - n(x) \rightarrow 0 \quad (5.12)$$

равномерно относительно x .

¹) Предполагается, как и раньше, что выполнены условия (5.1). — Прим. перев.

Доказательство. В силу (3.14) имеем

$$\frac{\sqrt{n}}{h} p_n(x) = \frac{1}{2\pi} \int_{-\sqrt{n}\pi/h}^{\sqrt{n}\pi/h} \varphi^n\left(\frac{\xi}{\sqrt{n}}\right) e^{-ix\xi} d\xi. \quad (5.13)$$

Используя снова формулу обращения (3.5) для нормальной плотности π , получаем, что левая часть в (5.12) не превосходит

$$\begin{aligned} \frac{1}{2\pi} \int_{-\sqrt{n}\pi/h}^{\sqrt{n}\pi/h} \left| \varphi^n\left(\frac{\xi}{\sqrt{n}}\right) - e^{(-1/2)\xi^2} \right| d\xi + \\ + \frac{1}{2\pi} \int_{|\xi| > \sqrt{n}\pi/h} e^{(-1/2)\xi^2} d\xi. \end{aligned} \quad (5.14)$$

Повторяя в точности доказательство теоремы 2, видим, что первый интеграл стремится к нулю. Ясно, что и второй интеграл стремится к нулю. Этим завершается доказательство. ►

§ 6. УСЛОВИЕ ЛИНДЕБЕРГА

Рассмотрим теперь последовательность таких независимых случайных величин X_k , что

$$E(X_k) = 0, \quad E(X_k^2) = \sigma_k^2. \quad (6.1)$$

Мы обозначим распределение величины X_k через F_k , характеристическую функцию через φ_k , а также, как обычно, обозначим $S_n = X_1 + \dots + X_n$ и $s_n^2 = \text{Var}(S_n)$. Таким образом,

$$s_n^2 = \sigma_1^2 + \dots + \sigma_n^2. \quad (6.2)$$

Мы будем говорить, что выполняется *условие Линдеберга*, если при любом фиксированном $t > 0$

$$\frac{1}{s_n^2} \sum_{k=1}^n \int_{|x| > ts_n} x^2 F_k\{dx\} \rightarrow 0, \quad n \rightarrow \infty. \quad (6.3)$$

Грубо говоря, это условие требует, чтобы дисперсии σ_k^2 образовывались в основном за счет масс, сосредоточенных в интервалах длины, малой по сравнению с s_n . Далее ясно, что σ_k^2/s_n^2 меньше, чем t^2 плюс левая часть в (6.3). Так как t произвольно, то из (6.3) следует, что при любом $\varepsilon > 0$ и достаточно большом n

$$\frac{\sigma_k}{s_n} \leq \varepsilon, \quad k = 1, 2, \dots, n. \quad (6.4)$$

Разумеется, отсюда следует, что $s_n \rightarrow \infty$.

Отношение σ_n/s_n можно принять за меру того вклада, который вносит слагаемое X_n в нормированную сумму S_n/s_n . Тогда условие (6.4) можно описать следующим образом: асимптотически S_n/s_n является суммой «многих индивидуально пренебрежимых слагаемых».

Условие Линдеберга было введено в г. VIII (соотношение (4.15)), и следующая ниже теорема совпадает с теоремой 3 из гл. VIII, 4. Каждое из доказательств имеет свои преимущества. Приводимое здесь доказательство позволяет установить, что условие Линдеберга в некотором смысле и необходимо; это доказательство применимо для вывода асимптотических разложений (гл. XVI) и теорем о сходимости плотностей (задача 28).

Теорема 1. Если выполняется условие Линдеберга (6.3), то распределения нормированных сумм S_n/s_n сходятся к стандартному нормальному распределению $\mathfrak{N}$.

Доказательство. Выберем $\xi > 0$ произвольное, но фиксированное. Мы должны показать, что

$$\varphi_1(\xi/s_n) \dots \varphi_n(\xi/s_n) \rightarrow e^{-\frac{1}{2}\xi^2}. \quad (6.5)$$

Поскольку $\varphi'_k(0) = 0$ и $|\varphi''_k(x)| \leq \sigma_k^2$ для всех x , из формулы Тейлора и (6.4) следует, что для достаточно больших n

$$|\varphi_k(\xi/s_n) - 1| \leq \frac{1}{2} \xi^2 \sigma_k^2 / s_n^2 \leq \varepsilon \xi^2. \quad (6.6)$$

Мы покажем, что если (6.6) выполнено, то (6.5) эквивалентно соотношению

$$\sum_{k=1}^n [\varphi_k(\xi/s_n) - 1] + \frac{1}{2} \xi^2 \rightarrow 0. \quad (6.7)$$

В самом деле, как показано в (2.4), $e^{\varphi_k^{-1}}$ является характеристической функцией обобщенного распределения Пуассона и поэтому $|e^{\varphi_k^{-1}}| \leq 1$. Далее, для любых комплексных чисел, таких, что $|a_k| \leq 1$ и $|b_k| \leq 1$, справедливо неравенство

$$|a_1 \dots a_n - b_1 \dots b_n| \leq \sum_{k=1}^n |a_k - b_k|, \quad (6.8)$$

которое может быть получено по индукции из тождества

$$x_1 x_2 - y_1 y_2 = (x_1 - y_1) x_2 + (x_2 - y_2) y_1.$$

Для произвольного $\delta > 0$ имеем $|e^z - 1 - z| < \delta |z|$, если $|z|$ достаточно мало, и, следовательно, для достаточно больших n из (6.6)

имеем

$$\begin{aligned} & |e^{\sum [\varphi_k(\xi/s_n) - 1]} - \varphi_1(\xi/s_n) \dots \varphi_n(\xi/s_n)| \leq \\ & \leq \sum_{k=1}^n |e^{\varphi_k(\xi/s_n) - 1} - \varphi_k(\xi/s_n)| \leq \\ & \leq \delta \sum_{k=1}^n |\varphi_k(\xi/s_n) - 1| \leq \delta (\xi^2/s_n^2) \sum_{k=1}^n \sigma_k^2 = \delta \xi^2. \quad (6.9) \end{aligned}$$

Поскольку δ произвольно, то левая часть стремится к нулю и, следовательно, (6.5) справедливо тогда и только тогда, когда выполняется (6.7).

Теперь (6.7) можно переписать в виде

$$\sum_{k=1}^n \int_{-\infty}^{+\infty} \left[e^{ix\xi/s_n} - 1 - \frac{ix\xi}{s_n} + \frac{x^2\xi^2}{2s_n^2} \right] F_k\{dx\} \rightarrow 0. \quad (6.10)$$

Из основного неравенства (4.1) следует, что при $|x| \leq ts_n$ подынтегральная функция мажорируется выражением $|x\xi/s_n|^3 < t\xi^3 x^2/s_n^2$, а при $|x| > ts_n$ — выражением $x^2\xi^2/s_n^2$. Поэтому левая часть в (6.10) по абсолютной величине меньше, чем

$$t\xi^3 + \xi^2 s_n^{-2} \sum_{k=1}^n \int_{|x| > ts_n} x^2 F_k\{dx\}. \quad (6.11)$$

Вследствие условия Линдеберга (6.3) второй член стремится к нулю, а поскольку t можно выбрать произвольно малым, то отсюда следует, что (6.10) справедливо. ►

Пояснительные примеры даны в гл. VIII, 4, в задачах 17—20 из гл. VIII, 10, и в задачах 26, 27 настоящей главы.

Следующая теорема содержит утверждение, в определенном смысле обратное к теореме 1.

Теорема 2. *Предположим, что $s_n \rightarrow \infty$ и $\sigma_n/s_n \rightarrow 0$. Тогда условие Линдеберга (6.3) является необходимым для сходимости распределения S_n/s_n к $\mathfrak{N}$.*

Предостережение. Мы скоро увидим, что даже в том случае, когда условие Линдеберга не выполнено, распределение S_n/s_n может стремиться к нормальному распределению (с дисперсией < 1).

Доказательство. Сначала покажем, что (6.4) выполнено. По предположению существует v , такое, что $\sigma_n/s_n < \varepsilon$ для $n > v$. Для $v < k \leq n$ имеем $\sigma_k/s_n \leq \sigma_k/s_k < \varepsilon$, а отношение σ_k/s_n с $k \leq v$ стремится в 0, так как $s_n \rightarrow \infty$.

Предположим далее, что распределение S_n/s_n стремится к $\mathfrak{N}$, т. е. предположим (6.5). В предыдущем доказательстве мы видели,

что выполнение соотношений (6.4) влечет за собой (7.10). Поскольку $\cos z - 1 + \frac{1}{2} z^2 \geq 0$, то действительная часть подынтегрального выражения в (6.10) неотрицательна и поэтому действительная часть всего выражения, стоящего слева в (6.10), не меньше, чем

$$\sum_{k=1}^n \int_{|x| > ts_n} \left(\frac{x^2 \zeta^2}{2s_n^2} - 2 \right) F_k \{dx\} \geq \left(\frac{1}{2} \zeta^2 - 2t^{-2} \right) \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x| > ts_n} x^2 F_k \{dx\}. \quad (6.12)$$

Таким образом, для произвольных ζ и t правая часть стремится к нулю и, следовательно, (6.3) выполняется. ►

Условие $\sigma_n/s_n \rightarrow 0$ не является строго говоря необходимым. Например, в частном случае, когда все распределения F_k нормальны, σ_k можно выбрать произвольными и тем не менее распределение S_n/s_n совпадает с $\mathfrak{N}$ (см. также задачу 27). Однако условие $\sigma_n/s_n \rightarrow 0$ естественным образом гарантирует, что влияние каждого из слагаемых X_k является предельно пренебрежимым, а без этого условия радикальным образом меняется характер этой проблемы. Но даже если $\sigma_n/s_n \rightarrow 0$ и $s_n \rightarrow \infty$, условие Линдеберга не является необходимым для существования таких нормирующих постоянных a_n , чтобы распределение S_n/a_n стремилось к $\mathfrak{N}$. Следующий пример разъясняет эту ситуацию.

Пример. Пусть $\{X_n\}$ — последовательность величин, удовлетворяющих условиям теоремы 1, включая нормировку (6.1). Пусть величины X'_n взаимно независимы, независимы от X_k и такие, что

$$\sum_{n=1}^{\infty} P\{X'_n \neq 0\} < \infty. \quad (6.13)$$

Положим $\bar{X}_n = X_n + X'_n$ и обозначим частные суммы $\{X'_n\}$ и $\{\bar{X}_n\}$ через S'_n и S_n соответственно. По первой лемме Бореля — Кантелли с вероятностью единица только конечное число X'_n будут отличаться от нуля, и, следовательно, с вероятностью единица $S'_n = O(s_n)$. Отсюда легко сделать вывод, что асимптотическое поведение распределений $\bar{S}_n/s_n$ и S_n/s_n одно и то же. Таким образом, *распределение $\bar{S}_n/s_n$ стремится к $\mathfrak{N}$ даже тогда, когда s_n^2 не является дисперсией $\bar{S}_n$; на самом деле, не требуется даже, чтобы $\bar{X}_n$ имела конечное математическое ожидание. Если $E(\bar{S}_n) = 0$ и $E(\bar{S}_n^2) = \bar{s}_n^2 < \infty$, то распределение $\bar{S}_n/\bar{s}_n$ сходится только тогда, когда $s_n/\bar{s}_n$ стремится к пределу p . В этом случае предельное распределение является нормальным с дисперсией $p \leq 1$.* ►

Этот пример показывает, что частные суммы S_n могут быть

асимптотически нормально распределены даже в том случае, когда $\mathbf{X}_n$ не имеют математических ожиданий, а также то, что дисперсии не всегда являются подходящими нормирующими константами. Мы не будем здесь развивать эту тему по двум причинам. Во-первых, общая теория будет рассмотрена в гл. XVII. Во-вторых, и это более важно, обобщение теорем, доказанных выше, может служить превосходным упражнением, а задачи 29—32 предназначены для того, чтобы с помощью простых методов получить необходимые и достаточные условия для центральной предельной теоремы.

§ 7. ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ МНОГОМЕРНЫХ РАСПРЕДЕЛЕНИЙ

Теория характеристических функций в пространствах нескольких измерений так близка к теории для $\mathcal{R}^1$, что ее систематическое изложение представляется излишним. Для того чтобы описать основные идеи и обозначения, достаточно рассмотреть случай двух измерений. Мы будем через $\mathbf{X}$ обозначать в этом параграфе пару действительных случайных величин X_1 и X_2 с заданным совместным распределением вероятностей F . Мы будем рассматривать $\mathbf{X}$ как *вектор-строку* с компонентами X_1 и X_2 ; аналогично аргумент x в $F(x)$ следует понимать как *вектор-строку* с компонентами x_1 и x_2 . С другой стороны, аргумент ζ соответствующей характеристической функции будет пониматься как *вектор-столбец* с компонентами ζ_1 и ζ_2 . Это соглашение удобно тем, что скалярное произведение записывается в виде $x\zeta = x_1\zeta_1 + x_2\zeta_2$. *Характеристическая функция φ для $\mathbf{X}$ (или для F) определяется равенством*

$$\varphi(\zeta) = E(e^{i\zeta\mathbf{X}}). \quad (7.1)$$

Это определение по форме такое же, как и в одномерном случае, однако показатель степени имеет новый смысл и интегрирование производится по отношению к двумерному распределению.

Основные свойства двумерных характеристических функций очевидны. Например, при $\zeta_2 = 0$ скалярное произведение $x\zeta$ превращается в $x_1\zeta_1$ и, следовательно, $\varphi(\zeta_1, 0)$ представляет собой *характеристическую функцию (маргинального) распределения X_1* . При любых фиксированных значениях параметров ζ_1, ζ_2 линейная комбинация $\zeta_1 X_1 + \zeta_2 X_2$ является (одномерной) случайной величиной с характеристической функцией

$$E(e^{i\lambda(\zeta_1 X_1 + \zeta_2 X_2)}) = \varphi(\lambda\zeta_1, \lambda\zeta_2). \quad (7.2)$$

Здесь ζ_1 и ζ_2 фиксированы, а независимой переменной служит λ . В частности, характеристическая функция суммы $X_1 + X_2$ равна $\varphi(\lambda, \lambda)$. Таким путем двумерная характеристическая функция порождает одномерные характеристические функции всех линейных комбинаций $\zeta_1 X_1 + \zeta_2 X_2$. Обратно, если известны распределения

всех таких комбинаций, мы можем вычислить все выражения типа $\varphi(\lambda\zeta_1, \lambda\zeta_2)$, а вместе с ними и двумерную характеристическую функцию¹⁾. Следующий пример показывает пользу и гибкость этого подхода. В примере используются обозначения, введенные в гл. III, 5.

Примеры. а) *Многомерные нормальные характеристические функции.* Пусть $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ (который надо считать вектором-строкой!) имеет невырожденное нормальное распределение. Предположим для простоты, что $\mathbf{E}(\mathbf{X}) = 0$. Обозначим матрицу ковариаций $\mathbf{E}(\mathbf{X}^T \mathbf{X})$ буквой $\mathbf{C}$. Ее элементами будут $c_{kk} = \text{Var}(\mathbf{X}_k)$ и $c_{12} = c_{21} = \text{Cov}(\mathbf{X}_1, \mathbf{X}_2)$. При фиксированных ζ_1 и ζ_2 линейная комбинация $\mathbf{X}\zeta = \zeta_1 \mathbf{X}_1 + \zeta_2 \mathbf{X}_2$ имеет математическое ожидание 0 и дисперсию

$$\sigma^2 = \zeta^T \mathbf{C} \zeta = c_{11} \zeta_1^2 + 2c_{12} \zeta_1 \zeta_2 + c_{22} \zeta_2^2. \quad (7.3)$$

Поэтому характеристическая функция для $\mathbf{X}\zeta = \zeta_1 \mathbf{X}_1 + \zeta_2 \mathbf{X}_2$ (ее аргумент обозначим через λ) имеет вид $e^{(-1/2)\sigma^2 \lambda^2}$, следовательно, двумерная характеристическая функция для $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ равна

$$\varphi(\zeta) = e^{-(1/2)\zeta^T \mathbf{C} \zeta}. \quad (7.4)$$

Эта формула применима и в r -мерном случае с той лишь разницей, что $\zeta^T \mathbf{C} \zeta$ является квадратичной формой r переменных $\zeta_1, \dots, \zeta_r$. Таким образом, (7.4) представляет собой характеристическую функцию r -мерного нормального распределения с нулевым математическим ожиданием и матрицей ковариаций $\mathbf{C}$.

Иногда желательно перейти от переменных $(\mathbf{X}_1, \mathbf{X}_2)$ и (ζ_1, ζ_2) к полярным координатам. Введем новые величины

$$\mathbf{X}_1 = R \cos \Theta, \quad \mathbf{X}_2 = R \sin \Theta, \quad \zeta_1 = \rho \cos \alpha, \quad \zeta_2 = \rho \sin \alpha. \quad (7.5)$$

(О таких преобразованиях см. гл. III, 1.) Тогда

$$\varphi(\zeta) = \mathbf{E}(e^{i\rho R \cos(\Theta - \alpha)}). \quad (7.6)$$

Следует иметь в виду, однако, что это *не есть* характеристическая функция пары случайных величин R, Θ ; последняя равна $\mathbf{E}(e^{i(\zeta_1 R + \zeta_2 \Theta)})$.

Примеры. б) *Круговая симметрия.* Если пара $(\mathbf{X}_1, \mathbf{X}_2)$ описывает «случайно направленный вектор» (см. гл. I, 10), то совместное распределение для (R, Θ) является произведением распределения G для R и равномерного распределения на $-\pi < \theta < \pi$. Тогда математическое ожидание в (7.6) не зависит от α и принимает форму

$$\varphi(\zeta_1, \zeta_2) = \frac{1}{2\pi} \int_0^\infty G\{dr\} \int_{-\pi}^\pi e^{i\rho r \cos \theta} d\theta. \quad (7.7)$$

¹⁾ Этим мы доказали попутно, что *распределение вероятностей в $\mathcal{R}^2$ однозначно определяется по вероятностям всех полуплоскостей*. Этот результат (отмеченный Крамером и Вальдом) пока не доказан элементарными методами. Его применение к проблеме моментов указано в задаче 21.

Производя замену переменной $\cos \theta = x$ и преобразуя внутренний интеграл (см. задачу 6), получаем

$$\varphi(\xi_1, \xi_2) = \int_0^{\infty} J_0(\rho r) G\{dr\}, \quad (\rho = \sqrt{\xi_1^2 + \xi_2^2}), \quad (7.8)$$

где

$$J_0(x) = J_0(ix) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!k!} \left(\frac{x}{2}\right)^{2k}. \quad (7.9)$$

(Бесселева функция J_0 была введена в гл. II, 7.)

Единичный вектор в случайном направлении имеет распределение G , сосредоточенное в точке 1. Таким образом, $J_0(\sqrt{\xi_1^2 + \xi_2^2})$ является *характеристической функцией суммы n независимых единичных случайно направленных векторов*. Этот результат был получен Рэлеем в одном специальном случае.

в) Мы рассмотрим частный случай, когда (X_1, X_2) имеет двумерную плотность

$$f(x_1, x_2) = (2\pi)^{-1} a^2 e^{-ar}, \quad r = \sqrt{x_1^2 + x_2^2}, \quad (7.10)$$

где a — положительная постоянная. Тогда (7.8) принимает вид¹⁾

$$\varphi(\xi_1, \xi_2) = a^2 \int_0^{\infty} e^{-ar} J_0(\rho r) r dr = (1 + \rho^2/a^2)^{-3/2}. \quad (7.11)$$

г) *Сферическая симметрия в $\mathcal{R}^3$* . Пример б) может быть распространен на трехмерный случай. Отличие состоит в том, что теперь мы имеем два полярных угла: географическую долготу ϕ и полярное расстояние θ . Внутренний интеграл в (7.7) принимает вид

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} d\omega \int_0^{\pi} e^{i\rho r \cos \theta} \sin \theta d\theta = \frac{1}{2} \int_0^{\pi/2} (e^{i\rho r \cos \theta} + e^{-i\rho r \cos \theta}) \sin \theta d\theta, \quad (7.12)$$

где $\rho^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$. Подстановка $\cos \theta = x$ преобразует это выражение к $(r\rho)^{-1} \sin r\rho$, а поэтому равенство (7.8) получает трехмерный аналог

$$\varphi(\xi_1, \xi_2, \xi_3) = \int_0^{\infty} \frac{\sin r\rho}{r\rho} G\{dr\}. \quad (7.13)$$

¹⁾ Подставляя вместо J_0 ее разложение (7.9), получаем

$$\varphi(\xi_1, \xi_2) = \sum \frac{(2k+1)!}{k!k!} (-1)^k \left(\frac{\rho}{2a}\right)^{2k} = \sum \binom{-\frac{3}{2}}{k} \left(\frac{\rho^2}{a^2}\right)^k,$$

что представляет собой ряд для $(1 + \rho^2/a^2)^{-3/2}$.

В частности, для *единичного* случайного вектора интеграл сводится к виду $\rho^{-1} \sin \rho$. Полагая $\xi_2 = \xi_3$, видим, что характеристическая функция компоненты X_1 единичного вектора записывается в виде $\xi_1^{-1} \sin \xi_1$. Таким образом, мы получаем новое доказательство факта (установленного в гл. I, 10), что эта компонента имеет равномерное распределение в интервале $-1, 1$. ►

Мы предоставляем читателю проверить, что основные теоремы, касающиеся характеристических функций в одномерном случае, обобщаются без существенных изменений. Так, *теорема об обращении интегралов Фурье в $\mathcal{R}^2$ утверждает, что если φ (абсолютно) интегрируема на всей плоскости, то X имеет ограниченную и непрерывную плотность, задаваемую формулой*

$$f(x_1, x_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-i(x_1 \xi_1 + x_2 \xi_2)} \varphi(\xi_1, \xi_2) d\xi_1 d\xi_2. \quad (7.14)$$

Пример. д) Двумерное распределение Коши. Если формула обращения (7.14) применяется к плотности f из примера в), то, как легко видеть, после деления на $f(0, 0) = (2\pi)^{-1} a^2$ имеем

$$\gamma(\xi_1, \xi_2) = e^{-a\sqrt{\xi_1^2 + \xi_2^2}}, \quad (7.15)$$

что представляет собой характеристическую функцию двумерной плотности g , определяемой формулой

$$g(x_1, x_2) = \frac{a}{2\pi(a^2 + x_1^2 + x_2^2)^{3/2}}. \quad (7.16)$$

Отсюда следует, что эта плотность обладает основными свойствами обычной плотности Коши. В частности, она *строго устойчива*: если $X^{(1)}, \dots, X^{(n)}$ — взаимно независимые векторы с плотностью (7.16), то их среднее арифметическое $(X^{(1)} + \dots + X^{(n)})/n$ имеет ту же самую плотность. ►

§ 8*). ДВЕ ХАРАКТЕРИЗАЦИИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Мы начнем с известной теоремы, высказанной П. Леви и доказанной в 1936 г. Г. Крамером. К сожалению, ее доказательство опирается на теорию аналитических функций и потому несколько выходит за рамки нашей трактовки теории характеристических функций.

Теорема 1. Пусть X_1 и X_2 — независимые случайные величины, сумма которых распределена нормально. Тогда X_1 и X_2 имеют нормальное распределение.

* В этом параграфе разбирается специальная тема, Результаты используются только в задаче 27.

Другими словами, нормальное распределение может быть разложено только на нормальные компоненты. Доказательство будет опираться на лемму, представляющую самостоятельный интерес.

Лемма. Пусть F — такое распределение вероятностей, что

$$f(\eta) = \int_{-\infty}^{+\infty} e^{\eta^2 x^2} F\{dx\} < \infty \quad (8.1)$$

при некотором $\eta > 0$. Тогда его характеристическая функция (рассматриваемая как функция комплексного аргумента ξ) является целой функцией. Если $\varphi(\xi) \neq 0$ для всех комплексных ξ , то F — нормальное распределение.

Доказательство леммы. При любых комплексных ξ и действительных x, η имеем $|x\xi| \leq \eta^2 x^2 + \eta^{-2} |\xi|^2$. Поэтому интеграл, определяющий φ , сходится при всех комплексных ξ и

$$|\varphi(\xi)| \leq e^{\eta^{-2} |\xi|^2} f(\eta). \quad (8.2)$$

Это значит, что φ есть целая функция порядка не более 2, и если такая функция не имеет нулей, то $\log \varphi(\xi)$ является многочленом не выше второй степени¹⁾. Следовательно, $\varphi(\xi) = e^{-\frac{1}{2} a \xi^2 + i b \xi}$, где a и b — некоторые, возможно, комплексные числа. Но φ — характеристическая функция, и поэтому $-i\varphi'(0)$ равно математическому ожиданию, а $-\varphi''(0)$ — второму моменту соответствующего распределения. Отсюда видим, что b действительно и $a \geq 0$, так что распределение F нормально. ►

Доказательство теоремы 1. Не ограничивая общности, мы можем предположить, что случайные величины X_1 и X_2 центрированы своими медианами. Тогда

$$P\{|X_1 + X_2| > t\} \geq \frac{1}{2} P\{|X_1| > t\}. \quad (8.3)$$

Теперь с помощью обычного интегрирования по частям [см. гл. V, 6] находим, что

$$f(\eta) \leq \eta^2 \int_0^{\infty} x \cdot e^{\eta^2 x^2} [1 - F(x) + F(-x)] dx. \quad (8.4)$$

Поэтому функции f_k , соответствующие X_k , удовлетворяют неравенству $f_k(\eta) \leq 2f(\eta) < \infty$. Так как $\varphi_1(\xi)\varphi_2(\xi) = e^{(-1/2) a \xi^2 + i b \xi}$, то ни φ_1 , ни φ_2 не могут обращаться в нуль. Поэтому X_1 и X_2 нормальны. ►

Мы переходим к доказательству другого характеристического

¹⁾ См., например: Hille E., *Analytic Function Theory*, Boston, 1962, т. 11, стр. 199 (теорема Адамара о факторизации).

свойства нормального распределения, о котором уже шла речь в гл. III, 4.

Теорема 2. Пусть X_1 и X_2 — независимые случайные величины, и пусть

$$Y_1 = a_{11}X_1 + a_{12}X_2, \quad Y_2 = a_{21}X_1 + a_{22}X_2. \quad (8.5)$$

Если Y_1 и Y_2 взаимно независимы, то или все четыре случайные величины распределены нормально, или же преобразование (8.5) тривиально в том смысле, что либо $Y_1 = aX_1$, $Y_2 = bX_2$, либо $Y_1 = aX_2$, $Y_2 = bX_1$.

Доказательство. При специальном предположении, что случайные величины X_j имеют непрерывные плотности, теорема была доказана в гл. III, 4. Доказательство опиралось на решение функционального уравнения (4.4) гл. III. Мы покажем теперь, что характеристические функции φ_j случайных величин X_j удовлетворяют функциональному уравнению того же типа. Мы покажем сначала, что достаточно рассматривать действительные характеристические функции. При этом мы продемонстрируем пользу теоремы 1.

Примеры. а) *Редукция к случаю симметричных распределений.* Введем случайные величины X_1^- и X_2^- , которые не зависят друг от друга и от X_j и которые распределены как $-X_1$ и $-X_2$ соответственно. Линейное преобразование (8.5) переводит симметризованные случайные величины ${}^0X_j = X_j + X_j^-$ в пару $({}^0Y_1, {}^0Y_2)$ симметричных и независимых случайных величин. Если теорема верна для таких случайных величин, то величины ${}^0X_j$ нормальны и по теореме 1 величины X_j также нормальны.

б) *Функциональные уравнения.* В силу независимости Y_1 и Y_2 двумерная характеристическая функция для (Y_1, Y_2) должна представляться в форме произведения соответствующих характеристических функций, т. е.

$$E(e^{i(\zeta_1 Y_1 + \zeta_2 Y_2)}) = E(e^{i\zeta_1 Y_1}) E(e^{i\zeta_2 Y_2}). \quad (8.6)$$

Подставляя сюда выражение (8.5), мы получаем следующее соотношение между характеристическими функциями для X_1 и X_2 :

$$\begin{aligned} \varphi_1(a_{11}\zeta_1 + a_{21}\zeta_2) \varphi_2(a_{12}\zeta_1 + a_{22}\zeta_2) = \\ = \varphi_1(a_{11}\zeta_1) \varphi_2(a_{12}\zeta_1) \varphi_1(a_{21}\zeta_2) \varphi_2(a_{22}\zeta_2). \end{aligned} \quad (8.7)$$

Теперь (8.7) совпадает с соотношением (4.4) из гл. III (с тем лишь исключением, что a_{12} и a_{21} меняются местами). По предположению φ_j действительны и непрерывны и так же, как в гл. III, 4, можно считать, что все a_{jk} отличны от нуля. Поэтому по лемме из гл. III, 4, $\varphi_j(\zeta) = e^{-a_j \zeta^2}$. Отсюда следует, что X_j нормальны. ►

§ 9. ЗАДАЧИ

1. Выведите из неравенства (1.7) (без всяких вычислений), что для любой характеристической функции φ

$$|\varphi(\xi)|^2 \leq 1 - \frac{1 - |\varphi(2\xi)|}{4} \leq e^{-\frac{1}{4}(1 - |\varphi(2\xi)|)}. \quad (9.1)$$

2. Пусть $\varphi = u + iv$ — характеристическая функция. Покажите, что

$$u^2(\xi) \leq \frac{1}{2}(1 + u(2\xi)). \quad (9.2)$$

Последнее в свою очередь влечет за собой неравенство

$$|\varphi(\xi)|^2 \leq \frac{1}{2}(1 + |\varphi(2\xi)|). \quad (9.3)$$

Указание. Для (9.2) используйте неравенство Шварца, для (9.3) рассмотрите характеристические функции вида $e^{i\alpha\xi}\varphi(\xi)$.

3. Докажите (обозначения те же, что и выше), что

$$|\varphi(\xi_2)| - |\varphi(\xi_1)|^2 \leq 2[1 - u(\xi_2 - \xi_1)]. \quad (9.4)$$

При $\xi_2 = -\xi_1$ отсюда следует (1.7).

4. Из элементарных формул выведите (без явного использования интегрирования), что характеристическая функция φ плотности $\frac{1}{\pi} \frac{1 - \cos x}{x^2}$ отличается лишь постоянным множителем от функции $2|\xi| - |\xi + 1| - |\xi - 1|$. Отсюда вытекает, что $\varphi(\xi) = 1 - |\xi|$ для $|\xi| \leq 1$.

5. Из характеристической функции для плотности $\frac{1}{2}ae^{-a|x|}$ получите дифференцированием по a новую характеристическую функцию. Используйте этот результат — покажите, что свертка данного распределения с собой имеет плотность $\frac{1}{4}ae^{-a|x|}(1 + a|x|)$.

6. Пусть плотность f распределения, сосредоточенного на $[-1, 1]$, имеет вид

$$f(x) = \frac{1}{\pi \sqrt{1 - x^2}}. \quad (9.5)$$

Покажите, что ее характеристическая функция равна

$$\varphi(\xi) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!k!} \left(\frac{1}{2}\xi\right)^{2k} = J_0(\xi). \quad (9.6)$$

Отметим, что $J_0(\xi) = I_0(i\xi)$, где I_0 — функция Бесселя, определенная формулой (7.1) в гл. II. **Указание.** Разложите $e^{i\xi x}$ в степенной ряд. Коэффициент при ξ^n выражается интегралом, а (9.6) может быть доказано индукцией по n , используя интегрирование по частям.

7. Докажите, что распределение *арксинуса* с плотностью $1/[\pi \sqrt{x(1-x)}]$, сосредоточенной на $[0, 1]$, имеет характеристическую функцию $e^{i\xi/2}J_0(\xi/2)$. **Указание.** Сведите к предыдущей задаче.

8. Исходя из формулы 10 таблицы на стр. 563, покажите, что $2\pi^2x(\operatorname{sh} x)^{-2}$ есть плотность с характеристической функцией $\frac{2}{1 + \operatorname{ch}(\pi\xi)}$. **Указание.** Используйте задачу 6 из гл. II, 9.

9. Пусть γ_L обозначает характеристическую функцию с периодом $2L$, описанную в примере 2, в). Покажите, что $2\gamma_{2L} - \gamma_L$ является снова характеристической функцией арифметического распределения. Ее график получается

из графика на рис. 2 (см. стр. 567) отражением каждого второго треугольника относительно оси ξ .

10. ¹⁾ Пусть X и Y — независимые случайные величины с распределениями F и G и характеристическими функциями φ и γ соответственно. Покажите, что произведение XY имеет характеристическую функцию

$$\int_{-\infty}^{+\infty} \gamma(\xi x) F\{dx\} = \int_{-\infty}^{+\infty} \varphi(\xi x) G\{dx\}. \quad (9.7)$$

11. Если $\{\varphi_n\}$ — такая последовательность характеристических функций, что $\varphi_n(\xi) \rightarrow 1$ при $-\delta < \xi < \delta$, то $\varphi_n(\xi) \rightarrow 1$ при всех ξ .

12. Пусть g — четная плотность со строгой положительной характеристической функцией γ . Тогда

$$g_a(x) = \frac{g(x) [1 - \cos ax]}{1 - \gamma(a)} \quad (9.8)$$

есть плотность вероятности с характеристической функцией

$$\gamma_a(\xi) = \frac{2\gamma(\xi) - \gamma(\xi + a) - \gamma(\xi - a)}{2[1 - \gamma(a)]}. \quad (9.9)$$

При $a \rightarrow \infty$ имеем $\gamma_a \rightarrow \gamma$, но g_a не стремится к g . Это показывает, что в теореме непрерывности для плотностей условие (3.10) существенно.

13. Если γ — действительная неотрицательная характеристическая функция, то существуют четные плотности g_n со строго положительными характеристическими функциями γ_n , такими, что $\gamma_n \rightarrow \gamma$. Указание. Рассмотрите смеси $(1-\varepsilon)G + \varepsilon F$ и свертки.

14. Если γ — характеристическая функция, такая, что $\gamma \geq 0$ и $\gamma(a) \neq 1$, тогда (9.9) определяет характеристическую функцию. Указание. Воспользуйтесь предыдущими двумя задачами.

15. Обобщение утверждения, обратного к (4.7) Рассматривая распределение $\frac{1}{m_{2k}} x^{2k} F\{dx\}$ (если они существуют), докажите по индукции, что распределение F имеет конечный момент порядка m_{2r} , тогда и только тогда, когда существует $2r$ -я производная характеристической функции φ в нуле.

16. Пусть f — плотность вероятности с положительной и интегрируемой характеристической функцией. Тогда f имеет единственный максимум в нуле. Если вторая производная f'' существует, то

$$f(0) > f(x) > f(0) - \frac{x^2}{2} f''(0); \quad (9.10)$$

аналогичное разложение верно и для первых $2r$ членов формулы Тейлора [заметьте, что f — четная функция и, следовательно, $f^{(2k+1)}(0) = 0$].

17. Пусть φ — действительная характеристическая функция с непрерывной второй производной φ'' . Тогда [исключая случай $\varphi(\xi) = 1$ для всех ξ]

$$\psi(\xi) = \frac{1 - \varphi(\xi)}{\xi^2} \cdot \frac{2}{|\varphi''(0)|} \quad (9.11)$$

будет характеристической функцией. Ей соответствует четная плотность f_2 ,

¹⁾ Комбинируя (9.7) с теоремой, указанной в сноске на стр. 186, получаем следующий критерий (принадлежащий Хинчину): функция ω является характеристической функцией унимодального распределения тогда и только

тогда, когда $\omega(\xi) = \int_0^1 \varphi(\xi x) dx$, где φ — характеристическая функция.

определяемая для $x > 0$ формулой

$$\frac{2}{|\varphi''(0)|} \int_x^{\infty} [1 - F(t)] dt. \quad (9.12)$$

Распространите это утверждение на моменты высших порядков.

18. Пусть f — четная плотность с характеристической функцией φ . Для $x > 0$ положим

$$g(x) = \int_x^{\infty} \frac{f(s) ds}{s}, \quad g(-x) = g(x). \quad (9.13)$$

Тогда g — снова четная плотность, и ее характеристической функцией будет

$$\gamma(\xi) = \frac{1}{\xi} \int_0^{\xi} \varphi(s) ds. \quad (9.14)$$

19. Пусть γ — характеристическая функция, такая, что $\overline{\lim} |\gamma(\xi)| = 1$ при $\xi \rightarrow \infty$. Соответствующее распределение F является чисто сингулярным (относительно меры Лебега).

20. Предположим, что $c_k > 0$, $\sum c_k = 1$, но $\sum c_k 2^k = \infty$. Пусть u — четная, непрерывная плотность, сосредоточенная на $-1, 1$, и пусть ω — соответствующая характеристическая функция. Тогда

$$f(x) = \sum c_k 2^k u(2^k x) \quad (9.15)$$

определяет плотность, непрерывную всюду, за исключением нуля, и имеющую характеристическую функцию

$$\varphi(\xi) = \sum c_k \omega(2^{-k} \xi). \quad (9.16)$$

Покажите, что функция $|\varphi|^n$ не интегрируема ни при каком n . *Указание.* Для $x \neq 0$ ряд в (9.15) конечен. Воспользуйтесь тривиальным неравенством $(\sum c_k p_k)^n \geq \sum c_k^n p_k^n$, которое выполняется при $p_k \geq 0$.

21. Проблема моментов в $\mathcal{R}^2$. Пусть X_1 и X_2 — две случайные величины с совместным распределением F . Положим $A_k = E(|X_1|^k) + E(|X_2|^k)$. Покажите, что F однозначно определяется своими моментами, если $\overline{\lim} k^{-1} A_k^{1/k} < \infty$. *Указание.* Как отмечено в сноске на стр. 585 достаточно доказать, что распределения всех линейных комбинаций $a_1 X_1 + a_2 X_2$ определены единственным образом. Используйте критерий (4.15).

22. В вырожденное двумерное распределение. Пусть φ — одномерная характеристическая функция и a_1, a_2 — произвольные постоянные. Покажите, что $\varphi(c_1 \xi_1 + a_2 \xi_2)$ как функция от ξ_1 и ξ_2 является двумерной характеристической функцией пары (X_1, X_2) для которой тождественно $a_2 X_1 = a_1 X_2$. Сформулируйте обратное утверждение. Рассмотрите специальный случай $a_2 = 0$.

23. Пусть X, Y, U — взаимно независимые случайные величины с характеристическими функциями φ, γ, ω . Покажите, что произведение $\varphi(\xi_1) \gamma(\xi_2) \times \times \omega(\xi_1 + \xi_2)$ представляет собой двумерную характеристическую функцию пары $(U + X, U + Y)$. *Указание.* Используйте трехмерную характеристическую функцию.

Примеры и дополнения к центральной предельной теореме

24. Докажите центральную предельную теорему 4 из гл. VIII, 4, для сумм случайного числа слагаемых методом характеристических функций.

25. Допустим, что X_k имеет плотность $e^{-x} x^{a_k-1} / \Gamma(a_k)$, где $a_k \rightarrow \infty$. Дисперсия S_n равна $s_n^2 = (a_1 + \dots + a_n)$. Покажите, что условие Линдберга выполнено, если

$$s_n^{-2} \sum_{k=1}^n a_k^2 \rightarrow 0.$$

26. Пусть $P\{X_k = \pm 1\} = (k-1)/2k$ и $P\{X_k = \pm \sqrt{k}\} = \frac{1}{2}k$. Покажите, что ни при каких нормирующих постоянных a_n распределение S_n/a_n не сходится к $\mathfrak{N}$. *Указание.* Перейдите к логарифмам и используйте неравенство

$$1 - \frac{\xi^2}{a_n^2} \leq \varphi_k\left(\frac{\xi}{a_n}\right) \leq 1 - \frac{k-1}{2k} \frac{\xi^2}{a_n^2}.$$

27. Если распределение S_n/s_n стремится к $\mathfrak{N}$, но $\sigma_n/s_n \rightarrow p > 0$, то распределение X_n/s_n стремится к нормальному распределению с дисперсией p^2 . *Указание.* По теореме Крамера—Леви из § 8 из равенства $\mathfrak{N} = U \star V$ следует, что как U , так и V — нормальные распределения. Используйте сходящиеся подпоследовательности последовательности распределений X_n/s_n и S_{n-1}/s_n .

28. Центральная предельная теорема для плотностей. Покажите, что теорема 2 § 5 обобщается на последовательности с различными плотностями f_k в предположении, что эти плотности удовлетворяют некоторым «равномерным» по k условиям гладкости. Например, достаточно, чтобы третьи абсолютные моменты были ограничены и чтобы f_k имели производные, причем $|f'_k| < M$ для всех k .

29. Центральная предельная теорема для схемы серий. При каждом n рассмотрим n независимых случайных величин $X_{1,n}, \dots, X_{n,n}$ с функциями распределения $F_{k,n}$. Пусть $T_n = X_{1,n} + \dots + X_{n,n}$. Предположим, что $E(X_{k,n}) = 0$, $E(T_n^2) = 1$ и что

$$\sum_{k=1}^n \int_{|x|>t} x^2 F_{k,n} \{dx\} \rightarrow 0 \quad (9.17)$$

для каждого $t > 0$. Покажите, что распределение T_n стремится к $\mathfrak{N}$. *Указание.* Используйте доказательство теоремы 1 из § 6.

Замечание. В частном случае, когда $X_{k,n} = X_k/s_n$ и $T_n = S_n/s_n$, получаем теорему Линдберга. При этом (9.17) превращается в условие Линдберга (6.3). О схемах серий см. в гл. VI, 3.

30. *Урезание.* Допустим, что $\{X_k\}$ — последовательность независимых случайных величин с симметричными распределениями. Пусть для каждого n и $k \leq n$ $X_{k,n}$ обозначает случайную величину, полученную урезанием величины

X_k на уровне $\pm a_n$. Предположим, что $\sum_{k=1}^n P\{|X_k| > a_n\} \rightarrow 0$ и выполнено (9.17). Покажите, что распределение S_n/a_n стремится к $\mathfrak{N}$.

31. *Обобщенная центральная предельная теорема.* Предположим, что распределения F_k симметричны и что для любого $t > 0$

$$\sum_{k=1}^n \int_{|x|>ta_n} F_k \{dx\} \rightarrow 0, \quad a_n^{-2} \sum_{k=1}^n \int_{|x|<ta_n} x^2 F_k \{dx\} \rightarrow 1. \quad (9.18)$$

Докажите, что распределение S_n/a_n стремится к $\mathfrak{N}$: а) используя две последние задачи; б) непосредственно приспособив доказательство теоремы 1 § 6¹⁾.

32. *Продолжение.* Условие симметричности распределений может быть за-

¹⁾ Теорема 2 обобщается подобным же образом, но требует другого доказательства.

менено более слабым условием:

$$\sum_{k=1}^n \left| \int_{|x| < a_n} x F_k \{dx\} \right| \rightarrow 0, \quad (9.19)$$

33. Для того чтобы существовали нормирующие постоянные a_n , удовлетворяющие условиям (9.18), необходимо и достаточно, чтобы для некоторой последовательности $t_n \rightarrow \infty$

$$\sum_{k=1}^n \int_{|x| > t_n} F_k \{dx\} \rightarrow 0, \quad \frac{1}{t_n^2} \sum_{k=1}^n \int_{|x| < t_n} x^2 F_k \{dx\} \rightarrow \infty.$$

В последнем случае можно взять

$$a_n^2 = \sum_{k=1}^n \int_{|x| < t_n} x^2 F_k \{dx\}.$$

(Обычно этот критерий применяется без затруднений.)

Вопросы, рассматриваемые в настоящей главе, технически довольно трудны. Они группируются вокруг двух проблем. Одна — вывод оценок для точности центральной предельной теоремы и улучшение этих результатов за счет построения асимптотических разложений. Другая проблема — уточнение центральной предельной теоремы для далеких от математического ожидания значений аргумента, для которых классические формулировки мало содержательны.

Чтобы облегчить подход к важным теоремам и объяснить основные идеи, мы выделяем сначала случай одинаково распределенных слагаемых. Методы, развитые в § 7 для больших отклонений, не опираются на содержание первых пяти параграфов, в которых излагается теория, существенно использующая два приема: прямую оценку абсолютно сходящихся интегралов Фурье и сглаживание. Мы различаем две главные идеи, рассматривая сначала асимптотическое разложение для плотностей, хотя это и приводит к некоторым повторениям и уменьшает изящество построений.

Кульминационный пункт главы — теорема Берри — Эссеена в § 5. Метод сглаживания, описанный в § 3, был впервые использован Берри при доказательстве этой теоремы. Сейчас используется бесчисленное множество приемов сглаживания. Заметим, к сожалению, что долгая и примечательная история предмета этой главы привела к печальному обстоятельству: случайности исторического развития все еще влияют на подход к отдельным вопросам. Возникшее в результате разнообразие средств и избыток созданных *ad hoc* методов привели к тому, что эта область прославилась своей хаотичностью.

Систематическое использование метода Берри и современных неравенств позволяет, к счастью, придать всей теории удивительную цельность и простоту²⁾.

¹⁾ Эта глава посвящена специальным вопросам и может быть опущена при первом чтении.

²⁾ Наилучшим введением в асимптотические разложения может служить книга Г. Крамера (1962). В ней содержатся теоремы из § 2 и § 4 для одинаково распределенных слагаемых и несколько более сильный вариант теорем из § 7. Первое строгое изложение теорем о асимптотических разложениях принадлежит Крамеру, однако его методы больше не используются. Материал § 1—5 изложен в монографии Гнеденко и Колмогорова (1954).

§ 1. ОБОЗНАЧЕНИЯ

Через F мы будем обозначать всюду (кроме § 6, где рассматривается случай неодинаково распределенных величин) одномерное распределение вероятностей с характеристической функцией φ . Момент порядка k (если он существует) будет обозначаться μ_k ;

$$\mu_k = \int_{-\infty}^{+\infty} x^k F \{dx\}. \quad (1.1)$$

Предположим, что $\mu_1 = 0$. Положим, как обычно, $\mu_2 = \sigma^2$. Распределение, полученное после нормировки n -кратной свертки F , обозначим F_n , т. е.

$$F_n(x) = F^{n*}(x\sigma\sqrt{n}). \quad (1.2)$$

Обозначим плотность F_n (если она существует) через f_n .

За исключением § 7 (посвященного большим отклонениям), нам придется иметь дело с функциями вида

$$u(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\zeta x} v(\zeta) d\zeta, \quad (1.3)$$

к которым будем применять очевидную оценку

$$|u(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |v(\zeta)| d\zeta. \quad (1.4)$$

Как u , так и v предполагаются интегрируемыми. Если u — плотность, то v будет соответствующей характеристической функцией. Для упрощения формулировок примем

Соглашение. Функция v в (1.3) будет называться преобразованием Фурье для u , а правая часть неравенства (1.4) — нормой Фурье для u .

Как всегда, нормальная плотность обозначается

$$n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \quad (1.5)$$

Ее преобразование Фурье — характеристическая функция $e^{(-1/2)\zeta^2}$. Повторным дифференцированием получаем тождество

$$\frac{d^k}{dx^k} n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\zeta x} (-i\zeta)^k e^{-\frac{1}{2}\zeta^2} d\zeta, \quad (1.6)$$

верное для $k=1, 2, \dots$. Очевидно, что левая часть имеет вид

$$\frac{d^k}{dx^k} \Pi(x) = (-1)^k H_k(x) \Pi(x), \quad (1.7)$$

где H_k — полином степени k . Полиномы H_k называют *полиномами Эрмита*¹⁾.

В частности,

$$H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x. \quad (1.8)$$

Характеристическое свойство полиномов H_k состоит, следовательно, в том, что *преобразование Фурье для $H_k(x) \Pi(x)$ равно $(i\xi)^k e^{(-1/2)\xi^2}$* .

§ 2. АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ДЛЯ ПЛОТНОСТЕЙ

Центральную предельную теорему для плотностей (теорема 2 из гл. XV, 5) можно значительно усилить, если допустить конечность моментов высшего порядка. Важную роль играет предположение²⁾ о том, что

$$\int_{-\infty}^{+\infty} |\varphi(\xi)|^v d\xi < \infty \quad (2.1)$$

при некотором $v \geq 1$. Доказательство, данное в гл. XV, 5, можно резюмировать, грубо говоря, следующим образом: разность $u_n = f_n - \Pi$ имеет преобразование Фурье

$$v_n(\xi) = \varphi^n \left(\frac{\xi}{\sigma \sqrt{n}} \right) - e^{(-1/2)\xi^2}. \quad (2.2)$$

Интеграл от $|v_n(\xi)|$ стремится к нулю в силу двух причин. При сколь угодно малом, но фиксированном $\delta > 0$ интеграл по области $|\xi| > \delta \sigma \sqrt{n}$ в силу (2.1) сходится к нулю. На интервале $|\xi| < \delta \sigma \sqrt{n}$ подынтегральное выражение $|v_n|$ мало из-за поведения φ вблизи нуля. Последнее связано лишь с тем, что $\mu_1 = 0$ и $\mu_2 = \sigma^2$. Если существуют моменты более высокого порядка, то мы можем использовать большее число членов разложения функции φ по формуле Тейлора и можем получить таким путем более

¹⁾ Иногда полиномами Чебышева — Эрмита. Единой терминологии нет. Встречаются различные нормирующие множители, а вместо $e^{-x^2/2}$ берут e^{-x^2} .

²⁾ В связи с этим условием см. примеры 5 а), б) из гл. XV и задачу 20 гл. XV, 9.

точную информацию относительно скорости сходимости $f_n \rightarrow \pi$. К сожалению, все вычисления заметно усложняются, если взять более чем три члена асимптотического разложения. Поэтому мы рассмотрим отдельно простейший и наиболее важный частный случай.

Теорема 1. Допустим, что момент μ_3 существует и что для некоторого $\nu \geq 1$ функция $|\varphi|^{\nu}$ интегрируема. Тогда при $n \geq \nu$ существует плотность f_n и при $n \rightarrow \infty$

$$f_n(x) - \pi(x) - \frac{\mu_3}{6\sigma^3 \sqrt{n}} (x^3 - 3x) \pi(x) = o\left(\frac{1}{\sqrt{n}}\right) \quad (2.3)$$

равномерно относительно x .

Доказательство. По теореме обращения для интегралов Фурье [гл. XV, 3] при $n \geq \nu$ левая часть равенства (2.3) определена и имеет норму Фурье

$$N_n = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \varphi^n\left(\frac{\xi}{\sigma\sqrt{n}}\right) - e^{(-1/2)\xi^2} - \frac{\mu_3}{6\sigma^3 \sqrt{n}} (i\xi)^3 e^{(-1/2)\xi^2} \right| d\xi \quad (2.4)$$

Фиксируем произвольное $\delta > 0$. Так как φ^n — характеристическая функция распределения, имеющего плотность, то $|\varphi(\xi)| < 1$ при $\xi \neq 0$ и $\varphi(\xi) \rightarrow 0$ при $\xi \rightarrow \infty$ (лемма 4 из гл. XV, 1, и лемма 3 из гл. XV, 4). Поэтому существует $q_\delta < 1$ с тем свойством, что $|\varphi(\xi)| < q_\delta$ для $|\xi| \geq \delta$. Часть интеграла (2.4), взятая по области $|\xi| > \delta\sigma\sqrt{n}$, меньше

$$q_\delta^{n-\nu} \int_{-\infty}^{+\infty} \left| \varphi\left(\frac{\xi}{\sigma\sqrt{n}}\right) \right|^\nu d\xi + \int_{|\xi| > \delta\sigma\sqrt{n}} e^{(-1/2)\xi^2} \left(1 + \left|\frac{\mu_3 \xi^3}{\sigma^3}\right|\right) d\xi \quad (2.5)$$

и, таким образом, стремится к нулю быстрее любой степени числа $1/n$.

Полагая для краткости¹⁾

$$\psi(\xi) = \ln \varphi(\xi) + \frac{1}{2} \sigma^2 \xi^2, \quad (2.6)$$

имеем

$$N_n = \frac{1}{2\pi} \int_{|\xi| < \delta\sigma\sqrt{n}} e^{(-1/2)\xi^2} \left| \exp\left(n\psi\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right) - 1 - \frac{\mu_3}{6\sigma^3 \sqrt{n}} (i\xi)^3 \right| d\xi + o\left(\frac{1}{n}\right). \quad (2.7)$$

¹⁾ Используемые в дальнейшем логарифмы комплексных чисел определяются разложением Тейлора $\log(1+z) = \sum (-z)^n/n$, которое имеет место при $|z| < 1$. Никакие другие значения z не рассматриваются.

Оценим подынтегральное выражение, используя неравенство $|e^\alpha - 1 - \beta| = |(e^\alpha - e^\beta) + (e^\beta - 1 - \beta)| \leq (|\alpha - \beta| + \frac{1}{2}|\beta|^2) e^\gamma$, (2.8)

где $\gamma \geq \max(|\alpha|, |\beta|)$ (справедливость этого неравенства для любых действительных или комплексных α и β очевидным образом доказывается заменой e^α и e^β их разложениями в степенной ряд).

Функция ψ трижды непрерывно дифференцируема и $\psi(0) = \psi'(0) = \psi''(0) = 0$, тогда как $\psi'''(0) = i^3 \mu_3$. Поскольку ψ''' непрерывна, можно найти окрестность $|\xi| < \delta$ точки $\xi = 0$, в которой изменение ψ''' менее ε . Из разложения в ряд Тейлора выводим, что при $|\xi| < \delta$

$$|\psi(\xi) - \frac{1}{6} \mu_3 (i\xi)^3| < \varepsilon \sigma^3 |\xi|^3. \quad (2.9)$$

Теперь мы выберем δ столь малым, чтобы, кроме того, при $|\xi| < \delta$

$$|\psi(\xi)| < \frac{1}{4} \sigma^2 \xi^2, \quad \left| \frac{1}{6} \mu_3 (i\xi)^3 \right| \leq \frac{1}{4} \sigma^2 \xi^2. \quad (2.10)$$

В самом деле, благодаря существованию μ_3 найдется окрестность нуля, в которой ψ трижды дифференцируема и (2.9) оказывается лишь формой записи разложения по формуле Тейлора.

При таком выборе δ из неравенства (2.8) следует, что подынтегральное выражение в (2.7) меньше, чем

$$e^{-\frac{1}{2} \xi^2} \left(\frac{\varepsilon}{\sqrt{n}} |\xi|^3 + \frac{\mu_3^2}{72n} \xi^6 \right). \quad (2.11)$$

В силу произвольности ε имеем $N_n = o(1/\sqrt{n})$, так что (2.3) действительно верно. ►

Точно такие же рассуждения приводят к асимптотическим разложениям, включающим большее число моментов. Однако их члены нельзя выразить простыми явными формулами. Поэтому мы пока отложим точное описание полиномов, входящих в эти разложения.

Теорема 2. Предположим, что моменты $\mu_3, \dots, \mu_r$ существуют и что для некоторого $\nu \geq 1$ функция $|\varphi|^{\nu}$ интегрируема. Тогда при $n \geq \nu$ существует f_n и при $n \rightarrow \infty$

$$f_n(x) - n(x) - n(x) \sum_{k=3}^r n^{(-1/2)k+1} P_k(x) = o(n^{(-1/2)r+1}) \quad (2.12)$$

равномерно относительно x . Здесь P_k — полином с действительными коэффициентами, зависящими только от $\mu_1, \dots, \mu_k$ (но не от n и r или других характеристик распределения F).

Первые два члена имеют вид

$$P_3 = \frac{\mu_3}{6\sigma^3} H_3, \quad P_4 = \frac{\mu_4}{72\sigma^6} H_3 + \frac{\mu_4 - 3\sigma^4}{24\sigma^4} H_4, \quad (2.13)$$

где H_k — полиномы Эрмита, определенные в (1.7). Разложение (2.12) называется (точнее, называлось) разложением Эджворта для f_n .

Доказательство. Ниже мы используем обозначение (2.6). Если p — поли-

ном с действительными коэффициентами $p_1, p_2, \dots$, то

$$f_n - p - n \sum p_k H_k \quad (2.14)$$

имеет норму Фурье

$$N_n = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\xi^2/2} \left| \exp \left(n\psi \left(\frac{\xi}{\sigma\sqrt{n}} \right) \right) - 1 - p(i\xi) \right| d\xi. \quad (2.15)$$

Теорема будет доказана построением подходящих полиномов p (чтобы не делать обозначения громоздкими, мы не будем писать индекс n).

Начнем с оценки подынтегрального выражения. Процедура будет такой же, как и в предыдущем доказательстве, но с той лишь разницей, что для аппроксимации ψ будут взяты члены формулы Тейлора со степенями аргумента до r -й включительно. Это аппроксимирующее выражение мы обозначим $\xi^2 \psi_r(\xi)$. Таким образом, ψ_r — полином степени $r-2$ с $\psi_r(0) = 0$. Он однозначно определяется требованием

$$\psi(\xi) - \xi^2 \psi_r(\xi) = o(|\xi|^r), \quad \xi \rightarrow 0.$$

Положим теперь

$$p(\xi) = \sum_{k=1}^{r-2} \frac{1}{k!} \left[\xi^2 \psi_r \left(\frac{\xi}{\sigma\sqrt{n}} \right) \right]^k. \quad (2.16)$$

Тогда $p(i\xi)$ будет полиномом с действительными коэффициентами, зависящими от n . При фиксированном ξ с другой стороны p будет полиномом относительно $1/\sqrt{n}$, коэффициенты которого можно вычислить в явном виде как полиномы относительно $\mu_1, \mu_2, \dots, \mu_r$. Как и в предыдущем доказательстве, очевидно, что при фиксированном $\delta > 0$ часть интеграла (2.15), взятая по области $|\xi| > \delta\sigma\sqrt{n}$, стремится к нулю быстрее любой степени числа $1/n$. Рассмотрим поэтому только подынтегральное выражение при $|\xi| < \delta\sigma\sqrt{n}$. Чтобы оценить его, мы воспользуемся [вместо (2.8)] неравенством

$$\begin{aligned} \left| e^\alpha - 1 - \sum_1^{r-2} \beta^k/k! \right| &\leq |e^\alpha - e^\beta| + \left| e^\beta - 1 - \sum_1^{r-2} \beta^k/k! \right| \leq \\ &\leq e^\gamma \left(|\alpha - \beta| + \frac{1}{(r-1)!} |\beta|^{r-1} \right), \end{aligned} \quad (2.17)$$

верным при $|\alpha| < \gamma$ и $|\beta| < \gamma$.

Подберем δ по аналогии с (2.9) таким образом, чтобы при $|\xi| < \delta$ выполнялось неравенство

$$|\psi(\xi) - \xi^2 \psi_r(\xi)| \leq \varepsilon \sigma^r |\xi|^r. \quad (2.18)$$

Мы можем предположить также (учитывая, что коэффициент при ξ в ψ_r равен $i^2 \mu_3/6$), что при $|\xi| < \delta$ и $a > 1 + |\mu_3|$

$$|\psi_r(\xi)| < a |\xi| < \frac{1}{4} \sigma^2. \quad (2.19)$$

Наконец, мы потребуем, чтобы при $|\xi| < \delta$

$$|\psi(\xi)| < \frac{1}{4} \sigma^2 \xi^2. \quad (2.20)$$

Тогда при $|\xi| < \delta \sigma \sqrt{n}$ подынтегральное выражение в (2.15) меньше, чем

$$e^{-\frac{1}{4}\xi^2} \left(\frac{\varepsilon |\xi|^r}{n^{\frac{1}{2}r-1}} + \frac{\sigma^{r-1}}{(r-1)!} \cdot \frac{|\xi|^{3(r-1)}}{(\sigma \sqrt{n})^{r-1}} \right). \quad (2.21)$$

В силу произвольности ε имеем $N_n = o\left(n^{-\frac{1}{2}r+1}\right)$.

Теперь мы должны определить действительные коэффициенты p_k , зависящие от n , такие, чтобы левая часть (2.14) была $o(n^{(-1/2)r+1})$ равномерно по x . При фиксированном ξ правая часть (2.16) является полиномом относительно $1/\sqrt{n}$. Производя упорядочение по возрастающим степеням $1/\sqrt{n}$, мы получаем разложение типа (2.12) с той лишь особенностью, что суммирование может идти и по индексам k , большим r . Но слагаемые, содержащие степени $1/n^k$ с $k > \frac{1}{2}r-1$, можно отбросить, и мы приходим к желаемому разложению (2.12). ►

Таким образом, явное вычисление полиномов P_k можно провести следующим образом. Полином ψ_r степени $r-2$ однозначно определяется из формулы Тейлора

$$\ln \varphi(\xi) = \xi^2 \left[-\frac{1}{2} \sigma^2 + \psi_r(\xi) \right] + o(|\xi|^r), \quad (2.22)$$

справедливой в окрестности нуля. Упорядочим слагаемые в (2.16) соответственно степеням $1/\sqrt{n}$. Обозначим коэффициент при $n^{(-1/2)k+1}$ через $q_k(i\xi)$. Тогда P_k — такой полином, что преобразование Фурье для $n(x) P_k(x)$ равно $e^{(-1/2)\xi^2} q_k(i\xi)$.

§ 3. СГЛАЖИВАНИЕ

Интегрируя асимптотическое разложение для плотностей f_n , получаем аналогичное асимптотическое разложение для распределений F_n , но эта простая процедура не подходит, если не выполняется условие (2.1). Чтобы справиться с этим случаем, мы должны будем идти обходным путем (следуя Берри). Для оценки расхождения $F_n - \mathfrak{N}$ или какой-либо аналогичной функции Δ мы сначала применим основанные на анализе преобразований Фурье методы предыдущего параграфа к некоторой аппроксимации функции Δ (обозначим ее ${}^T\Delta$). Затем мы прямыми методами оценим разность ${}^T\Delta - \Delta$. В этом параграфе мы дадим основные способы такой оценки.

Пусть V_T — распределение вероятностей с плотностью

$$v_T(x) = \frac{1}{\pi} \frac{1 - \cos Tx}{Tx^2} \quad (3.1)$$

и характеристической функцией ω_T . Для $|\xi| \leq T$ имеем

$$\omega_T(\xi) = 1 - \frac{|\xi|}{T}, \quad (3.2)$$

но эта явная форма не играет существенной роли. Важно лишь то, что $\omega_T(\xi)$ обращается в нуль при $|\xi| \geq T$. Последнее обстоятельство снимает все вопросы о сходимости.

Нас будут интересовать границы для $F_n - \mathfrak{N}$ или, более общим образом, для функций вида $\Delta_n = F_n - G_n$. Эти функции мы будем аппроксимировать их свертками с V_T . Положим вообще ${}^T\Delta = V_T \star \Delta$. Другими словами, для данной функции Δ мы определяем

$${}^T\Delta(t) = \int_{-\infty}^{+\infty} \Delta(t-x) v_T(x) dx. \quad (3.3)$$

Если функция Δ ограничена и непрерывна, то ${}^T\Delta \rightarrow \Delta$ при $T \rightarrow \infty$. Наша основная задача состоит в том, чтобы оценить максимум функции $|\Delta|$ через максимум функции $|{}^T\Delta|$.

Лемма 1. Пусть F — распределение вероятностей и G — такая функция, что $G(-\infty) = 0$, $G(\infty) = 1$ и $|G'(x)| \leq m < \infty$.

Положим

$$\Delta(x) = F(x) - G(x) \quad (3.4)$$

и

$$\eta = \sup_x |\Delta(x)|, \quad \eta_T = \sup_x |{}^T\Delta(x)|. \quad (3.5)$$

Тогда

$$\eta_T \geq \frac{\eta}{2} - \frac{12m}{\pi T}. \quad (3.6)$$

Доказательство. Функция Δ обращается в нуль на бесконечности, и пределы справа $\Delta(x_0 +)$ и слева $\Delta(x_0 -)$ существуют в каждой точке. Отсюда ясно видно, что найдется точка x_0 , такая, что или $|\Delta(x_0 +)| = \eta$ или $|\Delta(x_0 -)| = \eta$. Допустим, что $\Delta(x_0) = \eta$. Так как F не убывает, а модуль производной от G не превосходит m , то

$$\Delta(x_0 + s) \geq \eta - ms \text{ для } s > 0. \quad (3.7)$$

Полагая

$$h = \frac{\eta}{2m}, \quad t = x_0 + h, \quad x = h - s, \quad (3.8)$$

имеем

$$\Delta(t-x) \geq \frac{\eta}{2} + mx \text{ при } |x| \leq h. \quad (3.9)$$

Оценим интеграл (3.3), используя (3.9) и неравенство $\Delta(t-x) \geq -\eta$ при $|x| > h$. Интеграл от линейного члена обращается в нуль в силу симметрии. Далее плотность v_T приписывает области $|x| > h$ массу, которая не превосходит $4/(\pi Th)$. Поэтому

$$\eta_T \geq {}^T\Delta(x_0) \geq \frac{\eta}{2} \left[1 - \frac{4}{\pi Th} \right] - \eta \frac{4}{\pi Th} = \frac{\eta}{2} - \frac{6\eta}{\pi Th} = \frac{\eta}{2} - \frac{12m}{\pi T}. \quad (3.10)$$



В наших приложениях G будет иметь производную g , совпа-

дающую или с нормальной плотностью Π или с начальным отрезком асимптотического разложения, указанного в предыдущем параграфе. В любом случае преобразование Фурье γ , соответствующее g , имеет две непрерывные производные и $\gamma(0)=1$, $\gamma'(0)=0$. Очевидно, что преобразование Фурье свертки ${}^Tg = V_T \star g$ равно $\gamma \omega_T$. Аналогично по теореме обращения преобразований Фурье из гл. XV, 3, произведение $\varphi \omega_T$ является преобразованием Фурье плотности ${}^Tf$ распределения $V_T \star F$. Иными словами,

$${}^Tf(x) - {}^Tg(x) = \frac{1}{2\pi} \int_{-T}^T e^{-i\xi x} [\varphi(\xi) - \gamma(\xi)] \omega_T(\xi) d\xi. \quad (3.11)$$

Интегрируя по x , получим

$${}^T\Delta(x) = \frac{1}{2\pi} \int_{-T}^T e^{-i\xi x} \frac{\varphi(\xi) - \gamma(\xi)}{-i\xi} \omega_T(\xi) d\xi. \quad (3.12)$$

Постоянная интегрирования здесь отсутствует, так как при $|x| \rightarrow \infty$ обе части стремятся к 0, причем левая часть — в силу $F(x) - G(x) \rightarrow 0$, а правая часть — по лемме Римана — Лебега (лемма 4 из гл. XV, 4). Заметим, что $\varphi(0) = \gamma(0) = 1$ и $\varphi'(0) = \gamma'(0) = 0$, следовательно, выражение под знаком интеграла является непрерывной функцией, обращающейся в нуль в нуле, так что вопрос о сходимости не возникает.

Комбинируя верхнюю границу для η_T , содержащуюся в (3.12), с (3.6), получаем верхнюю границу для η , а именно

$$|F(x) - G(x)| \leq \frac{1}{\pi} \int_{-T}^T \left| \frac{\varphi(\xi) - \gamma(\xi)}{\xi} \right| d\xi + \frac{24m}{\pi T}. \quad (3.13)$$

Поскольку это неравенство служит основой всех оценок в последующих двух параграфах, суммируем условия его применимости.

Лемма 2. Пусть F — распределение вероятностей с нулевым математическим ожиданием и характеристической функцией φ . Предположим, что $F - G$ обращается в нуль на бесконечности и что G имеет производную g с $|g| \leq m$. Допустим, что g имеет преобразование Фурье γ , которое дважды непрерывно дифференцируемо и при этом $\gamma(0) = 1$, $\gamma'(0) = 0$. Тогда для всех x и $T > 0$ выполняется (3.13).

Мы дадим два независимых друг от друга применения этого неравенства. В следующем параграфе мы выведем интегральный вариант теорем § 2. В § 5 мы выведем знаменитое неравенство Берри — Эссеена для разности $F_n - \mathfrak{N}$.

§ 4. АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ДЛЯ РАСПРЕДЕЛЕНИЙ

Интегрируя асимптотическое разложение (2.3) для плотностей, получаем, что

$$F_n(x) - \mathfrak{N}(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}}(1-x^2)\Pi(x) = o\left(\frac{1}{\sqrt{n}}\right). \quad (4.1)$$

Однако для того, чтобы это асимптотическое разложение имело место, не обязательно, чтобы распределение F имело плотность. Действительно, сейчас мы докажем, что (4.1) выполняется для всех распределений, за единственным исключением решетчатых распределений (т. е. когда F сосредоточено на множестве точек вида $b \pm nh$). Из формулы обращения для решетчатых распределений следует, что наибольший скачок распределения F_n имеет порядок $1/\sqrt{n}$ [см. (5.12) из гл. XV], и, следовательно, (4.1) не может быть верным для решетчатых распределений. Оказывается, однако, что следующая теорема с небольшими поправками применима даже для решетчатых распределений. Для удобства мы эти два случая разделяем.

Теорема 1. Если F — нерешетчатое распределение и если момент третьего порядка μ_3 существует, то (4.1) выполняется равномерно для всех x .

Доказательство. Положим

$$G(x) = \mathfrak{N}(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}}(x^2 - 1)\Pi(x). \quad (4.2)$$

Тогда G удовлетворяет условиям последней леммы 2 с

$$\gamma(\xi) = e^{(-1/2)\xi^2} \left[1 + \frac{\mu_3}{6\sigma^3\sqrt{n}}(i\xi)^3 \right]. \quad (4.3)$$

Мы применим неравенство (3.13) с $T = a\sqrt{n}$, где постоянная a выбирается столь большой, что $24|G'(x)| < \varepsilon a$ для всех x . Тогда

$$|F_n(x) - G(x)| \leq \int_{-a\sqrt{n}}^{a\sqrt{n}} \left| \frac{\varphi^n\left(\frac{\xi}{\sigma\sqrt{n}}\right) - \gamma(\xi)}{\xi} \right| d\xi + \frac{\varepsilon}{\sqrt{n}}. \quad (4.4)$$

Так как область интегрирования конечна, то мы можем использовать рассуждения из § 2 даже в том случае, когда $|\varphi|$ не интегрируема на всей прямой. Разобьем интервал интегрирования на две части. Так как распределение F не является решетчатым, то в соответствии с леммой 4 из гл. XV, 1, максимум $|\varphi(\xi)|$ при $\delta \leq |\xi| \leq a\sigma$ строго меньше единицы. Как и в § 2, отсюда следует, что интеграл по области $|\xi| > \delta\sigma\sqrt{n}$ стремится к нулю быстрее, чем любая степень числа $1/n$. Далее, при $|\xi| \leq \delta\sigma\sqrt{n}$

подынтегральное выражение в (4.4) в силу оценки (2.11) меньше

$$e^{-\frac{1}{4}\zeta^2} \left(\frac{\varepsilon}{\sqrt{n}} |\zeta| + \frac{\mu_3^2}{72n} |\zeta|^3 \right),$$

и поэтому при больших n правая часть в (4.4) меньше $1000\varepsilon/\sqrt{n}$. В силу произвольности ε это завершает доказательство теоремы. ►

Это рассуждение неприменимо к решетчатым распределениям, потому что их характеристические функции периодичны (и, стало быть, интеграл по области $|\zeta| > \delta\sigma\sqrt{n}$ не стремится к нулю). Но утверждения теоремы можно спасти естественным изменением формулировки с учетом решетчатого характера распределения. Функция распределения F является ступенчатой функцией, но мы будем аппроксимировать ее непрерывной функцией распределения F^* с полигональным графиком.

Определение. Пусть F сосредоточено на решетке точек $b \pm nh$, но не сосредоточено ни на какой подрешетке (т. е. h есть шаг F).

Будем называть распределение F^* полигональным приближением к F , если график функции распределения F^* представляет собой полигон с вершинами в точках $b \pm \left(n + \frac{1}{2}\right)h$, лежащих на графике F (эти точки называются средними точками решетки).

Таким образом,

$$F^*(x) = F(x), \text{ если } x = b \pm \left(n + \frac{1}{2}\right)h, \quad (4.5)$$

$$F^*(x) = \frac{1}{2} [F(x) + F(x-)], \text{ если } x = b \pm nh. \quad (4.6)$$

Пусть теперь F_n есть решетчатое распределение с шагом

$$h_n = \frac{h}{\sigma\sqrt{n}}, \quad (4.7)$$

и, следовательно, при больших n полигональное приближение F_n^* очень близко к F_n .

Теорема 2¹. Для решетчатых распределений асимптотическая формула (4.1) остается верной, если заменить F_n ее полигональным приближением F_n^* .

В частности, (4.1) выполняется во всех средних точках $b \pm \left(n + \frac{1}{2}\right)h_n$ решетки для F_n (с шагом h_n), тогда как в точках решетки (4.1) выполняется с $\frac{1}{2} [F_n(x) + F_n(x-)]$ вместо $F_n(x)$.

¹) Вместо того чтобы заменять F_n на F_n^* , можно разложить $F_n^* - F_n$ в ряд Фурье и прибавить это к правой части (4.1). Таким образом, мы формально приходим к теореме, имеющей вид той, которую доказал Эсseen с помощью длинных вычислений. См., например, книгу Гнеденко и Колмогорова,

Доказательство. Нетрудно установить, что приближение $F^\#$ представляет собой свертку F с равномерным распределением на интервале $-\frac{1}{2}h < x < \frac{1}{2}h$. Соответственно $F_n^\#$ есть свертка F_n с равномерным распределением на интервале $-\frac{1}{2}h_n < x < \frac{1}{2}h_n$. Обозначим через $G^\#$ свертку последнего равномерного распределения с G , т. е.

$$G^\#(x) = h_n^{-1} \int_{-h_n/2}^{h_n/2} G(x-y) dy. \quad (4.8)$$

Пусть максимум $|G''|$ равен M , [тогда из разложения G в ряд Тейлора в точке x следует, что

$$|G^\#(x) - G(x)| < \frac{1}{3} M h_n^3 = O\left(\frac{1}{n}\right). \quad (4.9)$$

Для доказательства теоремы достаточно поэтому показать, что

$$|F_n^\#(x) - G^\#(x)| = o(1/\sqrt{n}). \quad (4.10)$$

Поскольку вычисление сверток соответствует перемножению преобразований, мы выводим из (4.4), что

$$|F_n^\#(x) - G^\#(x)| \leq \int_{-a\sqrt{n}}^{a\sqrt{n}} \left| \frac{\varphi^n\left(\frac{\xi}{\sigma n}\right) - \gamma(\xi)}{\xi} \right| \cdot |\omega_n(\xi)| d\xi + \frac{\varepsilon}{\sqrt{n}}, \quad (4.11)$$

где $\omega_n(\xi) = \left(\sin \frac{1}{2} h_n \xi\right) / \left(\frac{1}{2} h_n \xi\right)$ есть характеристическая функция равномерного распределения. Все оценки, используемые для вывода (4.4), применимы и здесь с тем исключением, что требуются новые аргументы для обоснования соотношения

$$\int_{\delta\sigma\sqrt{n}}^{a\sqrt{n}} |\varphi^n(\xi/\sigma n) \omega_n(\xi)| \xi^{-1} d\xi = \frac{2}{h} \int_0^{a/\sigma} \left| \varphi^n(y) \sin \frac{hy}{2} \right| y^{-2} dy = o\left(\frac{1}{n}\right). \quad (4.12)$$

По лемме 4 из гл. XV, 1, характеристическая функция φ имеет период $2\pi/h$ и то же самое, очевидно, верно для $|\sin \frac{1}{2} hy|$.

Поэтому достаточно доказать, что

$$\int_0^{\pi/h} |\varphi^n(y)| y dy = o\left(\frac{1}{\sqrt{n}}\right). \quad (4.13)$$

Но это соотношение справедливо, так как в некоторой окрестности нуля $|\varphi(y)| < e^{-\frac{1}{4}\sigma^2 y^2}$, а за пределами этой окрестности $|\varphi(y)|$ отделено от 1, и, следовательно, подынтегральное выраже-

ние в (4.13) убывает быстрее любой степени числа $1/n$. Таким образом, интеграл на самом деле имеет порядок $O(1/n)$. ►

Перейдем теперь к асимптотическим разложениям более высокого порядка. Доказательство соотношения (4.1) отличается от доказательства (2.3) только применением сглаживания, которое делает пределы интегрирования в (4.4) конечными. Точно такое же сглаживание можно применить и к разложениям, указанным в (2.12), но при этом, очевидно, чтобы получился остаток $O(n^{(-1/2)r+1})$, мы должны взять $T \sim an^{(1/2)r-1}$. Здесь возникает следующая трудность. Доказательство соотношения (4.1) использует тот факт, что максимум функции $|\varphi(\xi/\sqrt{n})|$ на интервале $\delta\sqrt{n} < |\xi| < T$ меньше единицы. Для нерешетчатых распределений это всегда верно при $T = a\sqrt{n}$, но может быть неверно при T , растущем как более высокая степень числа n . Поэтому для вывода асимптотических разложений более высокого порядка мы вынуждены предположить, что

$$\lim_{|\xi| \rightarrow \infty} |\varphi(\xi)| < 1, \quad (4.14)$$

откуда для нерешетчатых распределений следует, что максимум q_δ функции $|\varphi(\xi)|$ для $|\xi| > \delta$ меньше 1. При дополнительном условии (4.14) метод, детально объясненный для (4.1), применим без всяких изменений к разложениям (2.12), что приводит к следующей теореме.

Теорема 3. Если существуют моменты $\mu_2, \dots, \mu_r$ и выполняется (4.14), то при $n \rightarrow \infty$

$$F_n(x) - \mathfrak{M}(x) - n(x) \sum_{k=3}^r n^{(-1/2)k+1} R_k(x) = o(n^{(-1/2)r+1}) \quad (4.15)$$

равномерно по x . Здесь R_k — полином, зависящий только от $\mu_1, \dots, \mu_r$, но не от n или r (или других характеристик распределения F).

Разложение (4.15) — это лишь интегральный вариант формулы (2.12), и полиномы R_k связаны с P_k из (2.12) соотношением

$$n(x) P_k(x) = \frac{d}{dx} n(x) R_k(x). \quad (4.16)$$

Следовательно, нет нужды повторять их конструкцию. Заметим, что (4.14) выполняется для любого несингулярного распределения F .

Формула (4.15) называется *разложением Эджворта* для F . Если F имеет конечные моменты всех порядков, возникает искушение перейти к пределу при $r \rightarrow \infty$. Однако соответствующий бесконечный ряд не сходится ни при каком n . (Г. Крамер показал, что он сходится при любом n тогда и только тогда, когда функция $e^{(1/4)x^2}$ интегрируема относительно F .) Формальное разложение Эджворта не следует смешивать с разложением по полиномам Эрмита

$$F(x) - \mathfrak{M}(x) = \sum_{k=1}^{\infty} c_k H_k(x) e^{-x^2/2}, \quad (4.17)$$

которое сходится при условии, что F имеет конечное математическое ожидание, но без какого-либо вероятностного смысла. Например, даже если возможно разложить каждую функцию F_n в ряд типа (4.17), то коэффициенты не дают нам никакой информации относительно скорости сходимости $F_n \rightarrow \mathfrak{M}$.

§ 5. ТЕОРЕМА БЕРРИ — ЭССЕЕНА ¹⁾

Приводимая ниже важная теорема была доказана (совершенно различными путями) А. С. Берри (1941) и К. Г. Эссееном (1942).

Теорема 1. Пусть взаимно независимые случайные величины X_k с одинаковым распределением F таковы, что

$$E(X_k) = 0, \quad E(X_k^2) = \sigma^2 > 0, \quad E(|X_k|^3) = \rho < \infty. \quad (5.1)$$

Пусть F_n обозначает распределение нормированной суммы $(X_1 + \dots + X_n)/\sigma\sqrt{n}$. Тогда при всех x и n

$$|F_n(x) - \mathfrak{N}(x)| \leq \frac{3\rho}{\sigma^3 \sqrt{n}}. \quad (5.2)$$

Поразительное свойство этого неравенства заключается в том, что оно зависит только от первых трех моментов. Разложение (4.1) дает лучший асимптотический результат, но скорость сходимости определяется более тонкими свойствами исходного распределения. Множитель 3 в правой части (5.2) можно было бы заменить лучшей верхней границей C , но мы не ставим себе цели достичь в нашем изложении оптимальных результатов ²⁾.

Доказательство. Доказательство будет опираться на неравенство (3.13) с $F = F_n$ и $G = \mathfrak{N}$. В качестве T выберем

$$T = \frac{4}{3} \frac{\sigma^3}{\rho} \sqrt{n}; \quad (5.3)$$

тогда $T \leq \frac{4}{3} \sqrt{n}$ в силу неравенства между моментами $\sigma^3 < \rho$ (см. 8, в) из гл. V). Поскольку нормальная плотность имеет максимум, меньший $2/5$, мы получаем

$$\pi |F_n(x) - \mathfrak{N}(x)| \leq \int_{-T}^T \left| \varphi_n(\xi/\sigma\sqrt{n}) - e^{-\frac{1}{2}\xi^2} \right| \left| \frac{d\xi}{|\xi|} + \frac{9,6}{T} \right|. \quad (5.4)$$

Чтобы оценить подынтегральное выражение, воспользуемся тем, что для разности $\alpha^n - \beta^n$ имеет место неравенство

$$|\alpha^n - \beta^n| \leq n |\alpha - \beta| \cdot \gamma^{n-1}, \text{ если } |\alpha| \leq \gamma, \quad |\beta| \leq \gamma. \quad (5.5)$$

Пусть $\alpha = \varphi(\xi/\sigma\sqrt{n})$ и $\beta = e^{-\frac{1}{2}\xi^2/n}$. Из неравенства (4.1) для e^{it}

¹⁾ В этом параграфе используется неравенство сглаживания (3.13) (где G обозначает нормальное распределение), но в остальном этот параграф не зависит от предыдущих.

²⁾ Берри утверждал, что $C \leq 1,88$, но его вычисления содержали ошибку. Эссеен показал, что $C \leq 7,59$. Неопубликованные вычисления дают оценки $C \leq 2,9$ (Эссеен, 1956) и $C \leq 2,05$ (Д. Л. Уоллес, 1958). Наш прямолинейный метод дает удивительно хорошую границу, даже без обычных сложных численных расчетов. Никаких существенных улучшений нельзя ожидать без уточнения погрешности $24\pi/\pi$ в (3.13). [В. М. Золотарев (1966) показал, что $C \leq 0,9051$, — Прим. перев.].

из гл. XV получаем

$$\left| \varphi(t) - 1 + \frac{1}{2} \sigma^2 t^2 \right| = \left| \int_{-\infty}^{+\infty} \left(e^{itx} - 1 - itx + \frac{1}{2} t^2 x^2 \right) F\{dx\} \right| \leq \frac{1}{6} \rho |t|^3 \quad (5.6)$$

и, следовательно,

$$|\varphi(t)| < 1 - \frac{1}{2} \sigma^2 t^2 + \frac{1}{6} \rho |t|^3, \text{ если } \frac{1}{2} \sigma^2 t^2 \leq 1. \quad (5.7)$$

Тогда при $|\xi| \leq T$

$$|\varphi(\xi/\sigma\sqrt{n})| \leq 1 - \frac{1}{2n} \xi^2 + \frac{\rho}{6\sigma^3 n^{3/2}} |\xi|^3 \leq 1 - \frac{5}{18n} \xi^2 \leq e^{-\frac{5}{18} \xi^2/n}. \quad (5.8)$$

Так как $\sigma^3 < \rho$, то утверждение теоремы тривиально при $\sqrt{n} \leq 3$. Поэтому можно предположить, что $n \geq 10$. Тогда

$$|\varphi(\xi/\sigma\sqrt{n})|^{n-1} \leq e^{-\frac{1}{4} \xi^2}, \quad (5.9)$$

и оценку в правой части можно принять за γ^{n-1} в (5.5). Используя неравенство $e^{-x} - 1 + x \leq \frac{1}{2} x^2$ при $x > 0$, получаем из (5.6)

$$\begin{aligned} n \left| \varphi\left(\frac{\xi}{\sigma\sqrt{n}}\right) - e^{-\frac{1}{2} \xi^2/n} \right| &\leq n \left| \varphi\left(\frac{\xi}{\sigma\sqrt{n}}\right) - 1 + \frac{\xi^2}{2n} \right| + \\ &+ n \left| 1 - \frac{\xi^2}{2n} - e^{-\frac{1}{2} \xi^2/n} \right| \leq \frac{\xi}{6\sigma^3 \sqrt{n}} |\xi|^3 + \frac{1}{8n} \xi^4. \end{aligned} \quad (5.10)$$

Поскольку $\sqrt{n} > 3$, то из (5.5) и (5.9) следует, что подынтегральная функция в (5.4) не превосходит

$$\frac{1}{T} \left(\frac{2}{9} \xi^2 + \frac{1}{18} |\xi|^3 \right) e^{-\frac{1}{4} \xi^2}. \quad (5.11)$$

Эта последняя функция интегрируема на $-\infty < \xi < \infty$, и обычное интегрирование по частям приводит к неравенству

$$\pi T |F_n(x) - \mathfrak{N}(x)| \leq \frac{8}{9} \sqrt{\pi} + \frac{8}{9} + 10. \quad (5.12)$$

Поскольку $\sqrt{\pi} < \frac{8}{9}$, правая часть (5.12) меньше $\frac{113}{9} < 4\pi$, поэтому (5.2) выполняется. ►

Теорему и ее доказательство можно распространить на последовательности $\{X_k\}$ с разнораспределенными X_k .

Теорема 2¹⁾. Пусть X_k — взаимно независимые случайные величины и

$$E(X_k) = 0, \quad E(X_k^2) = \sigma_k^2, \quad E(|X_k|^3) = \rho_k. \quad (5.13)$$

¹⁾ Дана Эссееном (но с совершенно иным доказательством).

Положим

$$s_n^2 = \sigma_1^2 + \dots + \sigma_n^2, \quad r_n = \rho_1 + \dots + \rho_n \quad (5.14)$$

и обозначим через F_n распределение нормированной суммы $(X_1 + \dots + X_n)/s_n$. Тогда для всех x и n

$$|F_n(x) - \mathfrak{N}(x)| \leq 6 \frac{r_n}{s_n^3}. \quad (5.15)$$

Доказательство. Пусть ω_k — характеристическая функция X_k , тогда исходное неравенство (5.4) заменяется неравенством

$$\pi |F_n(x) - \mathfrak{N}(x)| \leq \int_{-T}^T \left| \omega_1\left(\frac{\xi}{s_n}\right) \dots \omega_n\left(\frac{\xi}{s_n}\right) - e^{-\frac{1}{2}\xi^2} \right| \frac{d\xi}{|\xi|} + \frac{9.6}{T}. \quad (5.16)$$

На этот раз выберем

$$T = \frac{8}{9} \cdot \frac{s_n^3}{r_n}. \quad (5.17)$$

Вместо (5.5) мы воспользуемся теперь неравенством

$$|\alpha_1 \dots \alpha_n - \beta_1 \dots \beta_n| \leq \sum_{k=1}^n (\gamma_1 \dots \gamma_{k-1} \alpha_k - \beta_k \gamma_{k+1} \dots \gamma_n), \quad (5.18)$$

которое имеет место при $|\alpha_k| \leq \gamma_k$ и $|\beta_k| \leq \gamma_k$. Применим это неравенство к

$$\alpha_k = \omega_k(\xi/s_n), \quad \beta_k = e^{-\frac{1}{2}(\sigma_k^2/s_n^2)\xi^2}, \quad |\xi| < T. \quad (5.19)$$

По аналогии с (5.8) получим

$$|\omega_k(\xi/s_n)| \leq 1 - \frac{1}{2} \frac{\sigma_k^2}{s_n^2} \xi^2 + \frac{\rho_k}{6s_n^3} |\xi|^3 \leq \exp\left(-\frac{\sigma_k^2}{2s_n^2} + \frac{\rho_k T}{6s_n^3}\right) \xi^2 \quad (5.20)$$

при условии, что $\sigma_k T < s_n \sqrt{2}$. Для того чтобы получить подходящее при всех k выражение для γ_k , изменим коэффициент $1/6$ на $3/8$ и положим

$$\gamma_k = \exp\left(-\frac{\sigma_k^2}{2s_n^2} + \frac{3}{8} \frac{\rho_k T}{s_n^3}\right) \xi^2. \quad (5.21)$$

Очевидно, что $|\beta_k| \leq \gamma_k$. Из (5.20) также следует, что $|\alpha_k| \leq \gamma_k$ для всех таких k , что $\sigma_k T \leq \frac{3}{4} s_n$. Но из неравенства для моментов $\rho_k \geq \sigma_k^3$ следует, что $\gamma_k > 1$, если $\sigma_k T > \frac{4}{3} s_n$. Поэтому $|\alpha_k| \leq \gamma_k$ при всех k .

Теорема тривиальна, если правая часть в (5.15) ≥ 1 , т. е. если $r_n/s_n^3 \geq 1/6$. Таким образом, мы предположим в дальнейшем, что $r_n/s_n^3 < 1/6$ или $T > 16/3$. Свое минимальное значение γ_k принимает

при некотором k , для которого верно неравенство $\sigma_k/s_n < \frac{4}{3}T > \frac{1}{4}$.

Следовательно, $\gamma_k \geq e^{-\zeta^{1/32}}$ при всех k . В итоге имеем

$$|\gamma_1 \cdots \gamma_{k-1} \beta_{k+1} \cdots \gamma_n| \leq \exp \zeta^2 \left(-\frac{1}{2} + \frac{3r_n T}{8s_n^3} + \frac{1}{32} \right) < e^{-\zeta^{1/8}}. \quad (5.22)$$

По аналогии с (5.10) получаем

$$\sum_{k=1}^n |\alpha_k - \beta_k| \leq \frac{r_n}{6s_n^3} |\zeta|^3 + \frac{\zeta^4}{8s_n^4} \sum_{k=1}^n \sigma_k^4. \quad (5.23)$$

Для того чтобы оценить последнюю сумму справа, вспомним, что $\sigma_k^4 \leq \rho_k^{4/3} \leq r_n^{1/3} \cdot \rho_k$, откуда

$$\frac{1}{s_n^4} \sum_{k=1}^n \sigma_k^4 \leq \left(\frac{r_n}{s_n^3} \right)^{4/3} \leq \frac{1}{6^{1/3}} \frac{r_n}{s_n^3} \leq \frac{5}{9} \frac{r_n}{s_n^3}. \quad (5.24)$$

Эти неравенства показывают, что подынтегральное выражение в (5.16) меньше

$$\frac{8}{9T} \left(\frac{1}{6} \zeta^2 + \frac{5}{72} |\zeta|^3 \right) e^{-\zeta^{1/8}}, \quad (5.25)$$

и поэтому

$$\pi T |F_n(x) - \mathfrak{N}(x)| \leq \frac{32}{27} \sqrt{2\pi} + \frac{5}{81} \cdot 64 + 9,6. \quad (5.26)$$

Правая часть здесь меньше $16\pi/3$, и из (5.26) следует (5.15). ►

В последнее время большое внимание уделялось обобщениям теоремы Берри — Эссеена на величины, не имеющие момента третьего порядка. В этом случае верхняя граница в неравенстве выражается через момент дробного порядка или какую-либо родственную величину. Первый шаг в этом направлении был сделан М. Л. Кацем (1963). Обычные вычисления в этом случае достаточно запутаны, а попытки разработать универсальные методы, применимые ко многим случаям, не предпринимались. Наше доказательство предназначалось для этой цели и может быть приспособлено для значительно более широкого круга задач. В самом деле, необходимость в моменте третьего порядка появляется в доказательстве теоремы только из-за неравенства

$$\left| e^{itx} - 1 - itx + \frac{1}{2} t^2 x^2 \right| \leq \frac{1}{6} |tx|^3.$$

Фактически было бы достаточно использовать эту оценку в некотором конечном интервале $|x| < a$, а в ином случае удовлетвориться границей $t^2 x^2$. Имеет место следующая теорема, полученная разными методами Л. В. Осиповым и В. В. Петровым.

Теорема 3. *Предположим, что выполнены условия теоремы 2 с той разницей, что моменты третьего порядка не обязаны существо-*

возврат. Тогда для произвольного $\tau_k > 0$

$$|F_n(x) - \mathfrak{H}(x)| \leq 6 \left(s_n^{-3} \int_{|x| \leq \tau_k} |x|^3 F \{dx\} + s_n^{-2} \int_{|x| > \tau_k} x^2 F \{dx\} \right). \quad (5.27)$$

С помощью обычных приемов урезания можно распространить этот результат на случай, когда не делается никаких предположений о существовании моментов¹⁾.

§ 6. АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ В СЛУЧАЕ РАЗЛИЧНО РАСПРЕДЕЛЕННЫХ СЛАГАЕМЫХ

Теория, развитая в § 2 и 4, легко переносится на последовательности $\{X_k\}$ независимых случайных величин с различными распределениями U_k . В действительности все наши обозначения и рассуждения были предназначены для подготовки к решению этой задачи и потому не всегда были простейшими из возможных.

Мы предполагаем, что $E(X_k) = 0$ и $E(X_k^2) = \sigma_k^2$. Как обычно, положим $s_n^2 = \sigma_1^2 + \dots + \sigma_n^2$. Чтобы сохранить единство, мы по-прежнему будем обозначать через F_n распределение нормированной суммы $(X_1 + \dots + X_n)/s_n$.

Для определенности рассмотрим асимптотическое разложение (4.1). Очевидным аналогом левой части (4.1) является выражение

$$D_n(x) = F_n(x) - \mathfrak{H}(x) - \frac{\mu_3^{(n)}}{6s_n^3} n(x), \quad (6.1)$$

где

$$\mu_3^{(n)} = \sum_{k=1}^n E(X_k^3). \quad (6.2)$$

Как было показано, в случае одинаково распределенных X_k $D_n(x) = o(1/\sqrt{n})$. Теперь D_n представляет собой сумму различных остаточных членов, которые в данной ситуации не обязаны быть сравнимыми по величине. Действительно, если X_k имеют четыре момента, то можно показать, что при слабых дополнительных условиях

$$|D_n(x)| = O(n^2 s_n^{-6}) + O(n s_n^{-4}). \quad (6.3)$$

Здесь любое из слагаемых может быть главным в зависимости от поведения последовательности $n s_n^{-2}$, которая может колебаться между 0 и ∞ . Теоретически возможно найти универсальные оценки

¹⁾ Подробнее см: W. Feller, On the Berry—Esseen theorem, Zs. Wahrscheinlichkeitstheorie verw. Gebiete, vol. 10 (1968), 261—268. Удивительно то, что универсальный общий метод фактически упрощает рассуждения даже в классическом случае и, более того, без излишних усилий приводит к лучшим числовым оценкам.

для суммарной ошибки¹⁾, но они по своему характеру не отражали бы существа дела, а результаты применения их в отдельных случаях, возникающих на практике, были бы разочаровывающими. Поэтому более осмотрительно ограничиться только последовательностями $\{X_k\}$, в некотором смысле типичными, но при этом доказательства сделать столь гибкими, чтобы они были пригодны в различных ситуациях.

В качестве типичной схемы мы рассмотрим такие последовательности, что отношения s_n^2/n остаются ограниченными сверху и снизу (положительными постоянными)²⁾. Мы покажем, что при слабых дополнительных ограничениях асимптотическое разложение (4.1) остается справедливым и доказательство не требует изменений. В других ситуациях остаточный член может принимать другой вид. Например, если $s_n^2 = o(n)$, то можно сказать только то, что $|D_n(x)| = o(n/s_n^2)$. Однако доказательство можно приспособить к такой ситуации.

Доказательство (4.1) опирается на преобразование Фурье функции $D_n(x)$. Если ω_k обозначает характеристическую функцию X_k , это преобразование можно записать в виде

$$e^{nv_n(\zeta/s_n)} = e^{-\frac{1}{2}\zeta^2} - \frac{nv_n'''(0)}{6s_n^3} \zeta^3 e^{-\frac{1}{2}\zeta^2}, \quad (6.4)$$

где

$$v_n(\zeta) = n^{-1} \sum_{k=1}^n \log \omega_k(\zeta). \quad (6.5)$$

Мы получили в точности ту же самую формулу, которая была использована в доказательстве (4.1), с той лишь разницей, что там функция $v_n(\zeta) = \log \varphi(\zeta)$ была независима от n . Посмотрим теперь, как эта зависимость влияет на доказательство. При доказательстве были использованы только два свойства v .

а) Мы использовали непрерывность производной третьего порядка φ''' для нахождения интервала $|\zeta| < \delta$, внутри которого изменение v_n''' меньше ε . Чтобы быть уверенным в том, что такое δ можно выбрать независимо от n , мы должны теперь ввести некоторые условия «равномерности» относительно поведения производных ω_k''' вблизи нуля. Чтобы обойти малоинтересные технические

¹⁾ Например, у Крамера оценка имеет вид

$$D_n(x) = O\left(n^{1/2} s_n^{-6} \left(\sum_{k=1}^n E(X_k^4)\right)^{3/2}\right),$$

что может быть хуже (6.3).

²⁾ Тогда (6.3) дает $|D_n(x)| = O(1/n)$, что точнее, чем оценка $o[1/\sqrt{n}]$, полученная в (4.1). Возможно уточнение, если предположить существование четырех моментов.

проблемы, мы предположим, что моменты $E(X_k^i)$ существуют и остаются ограниченными. Тогда производные ω_k'''' равномерно ограничены и то же самое справедливо для v_n'''' .

б) Доказательство (4.1) было основано на том, что $|\varphi^n(\xi)| = o(1/\sqrt{n})$ равномерно для всех $\xi > \delta$. Аналогом этого здесь служит условие

$$|\omega_1(\xi) \dots \omega_n(\xi)| = o(1/\sqrt{n}) \text{ равномерно по } \xi > \delta > 0. \quad (6.6)$$

Это условие исключает ту возможность, когда все X_k имеют решетчатые распределения с одним и тем же шагом; в последнем случае произведение в (6.6) было бы периодической функцией ξ . В остальных отношениях это условие достаточно умеренно, чтобы тривиально выполняться в большинстве случаев. Например, если X_k имеют плотности, то каждый множитель $|\omega_k|$ отделен от единицы, и левая часть в (6.6) убывает быстрее, чем любая степень числа $\frac{1}{n}$, за исключением того случая, когда $|\omega_n(\xi)|$ стремится к 1, т. е. случая, когда X_n стремится стать постоянной. Таким образом, условие (более сильное, чем (6.6))

$$|\omega_1(\xi) \dots \omega_n(\xi)| = o(n^{-a}) \text{ равномерно при } \xi > \delta. \quad (6.7)$$

при всех $a > 0$ часто выполнено и легко проверяется.

При двух указанных дополнительных условиях доказательство (4.1) проходит без изменений, и мы получаем теорему.

Теорема 1. *Предположим, что существуют такие положительные постоянные c, C, M , что при всех n*

$$cn < s_n^2 < Cn, \quad E(X_n^2) < M \quad (6.8)$$

и что выполняется (6.6). Тогда $|D_n(x)| = o(1/\sqrt{n})$ равномерно для всех x .

Как указывалось, доказательство применимо и к другим ситуациям. Например, предположим, что

$$s_n^2/n \rightarrow 0, \quad \text{но } s_n^3/n \rightarrow \infty. \quad (6.9)$$

Доказательство (4.1) проводится с $T = as_n^3/n$, а поскольку $T = o(s_n)$, условие (6.6) становится излишним. Таким образом, мы приходим к следующему варианту теоремы.

Теорема 1а. *Если выполняется (6.9) и $E(X_k^4)$ равномерно ограничено, то $|D_n(x)| = o(n/s_n^3)$ равномерно относительно x .*

Подобным же образом обобщаются другие теоремы из § 2 и 4. Например, доказательство теоремы 3 из § 4 приводит без существенных изменений к следующей общей теореме об асимптотических разложениях¹⁾.

¹⁾ Эта теорема в несколько более слабыми условиями равномерности содержится в основополагающей работе Крамера. Сами методы Крамера, однако, теперь устарели.

Теорема 2. Допустим, что

$$0 < c < E(|X|^v) < C < \infty, \quad v = 1, \dots, r+1, \quad (6.10)$$

и что (6.7) выполняется с $a = r+1$. Тогда асимптотическое разложение (4.15) выполняется равномерно относительно x .

Полиномы R_j зависят от моментов, входящих в условия (6.10), но при фиксированном x последовательность $\{R_j(x)\}$ ограничена.

§ 7. БОЛЬШИЕ ОТКЛОНЕНИЯ¹⁾

Мы снова начнем с рассмотрения общей проблемы в частном случае одинаково распределенных случайных величин X_k с $E(X_k) = 0$ и $E(X_k^2) = \sigma^2$. Как и раньше, F_n обозначает распределение нормированных сумм $(X_1 + \dots + X_n)/\sigma\sqrt{n}$. Информация о том, что F_n стремится к нормальному распределению $\mathfrak{N}$, является очень важной при умеренных значениях x , но при больших x как $F_n(x)$, так и $\mathfrak{N}(x)$ близки к единице и утверждение центральной предельной теоремы становится бессодержательным. Аналогично большая часть наших приближенных формул и асимптотических разложений становятся бесполезными. В действительности нам нужна относительная ошибка аппроксимации $1 - F_n$ величиной $1 - \mathfrak{N}$. Часто желательно бывает использовать соотношение

$$\frac{1 - F_n(x)}{1 - \mathfrak{N}(x)} \rightarrow 1, \quad (7.1)$$

когда и x , и n стремятся к бесконечности. Это соотношение не может быть верно всегда. Так, для симметричного биномиального распределения числитель (7.1) обращается в нуль для всех $x > \sqrt{n}$. Мы покажем, однако, что (7.1) верно, если x меняется вместе с n таким способом, что $xn^{-1/6} \rightarrow 0$. При этом предположим, что интеграл

$$f(\xi) = \int_{-\infty}^{+\infty} e^{\xi x} F\{dx\} \quad (7.2)$$

конечен при всех ξ из некоторого интервала $|\xi| < \xi_0$. [Это равносильно тому, что характеристическая функция $\varphi(\xi) = f(i\xi)$ аналитична в некоторой окрестности нуля; но мы предпочитаем иметь дело с действительной функцией f .]

Теорема 1. Если интеграл (7.2) сходится в некоторой окрестности точки $\xi = 0$ и x меняется вместе с n так, что $x \rightarrow \infty$ и $x = o(n^{1/6})$, то верно (7.1).

Заменяя x на $-x$, мы получаем двойственную теорему для левого конца. Теорема представляется достаточно общей, чтобы

¹⁾ Этот параграф полностью не зависит от предыдущих параграфов этой главы.

охватить все «практически интересные ситуации», но метод доказательства приводит к значительно более сильным результатам.

Для доказательства перейдем от функции f к ее логарифму. Равенство

$$\psi(\zeta) = \ln f(\zeta) = \sum \frac{\psi_k}{k!} \zeta^k \quad (7.3)$$

определяет функцию, аналитическую в некоторой окрестности нуля. Коэффициент ψ_k зависит только от моментов $\mu_1, \dots, \mu_k$ распределения F и носит название *семиинварианта порядка k* распределения F . В общем случае $\psi_1 = \mu_1, \psi_2 = \sigma^2, \dots$. Мы предполагаем $\mu_1 = 0$, и потому $\psi_1 = 0, \psi_2 = \sigma^2, \psi_3 = \mu_3, \dots$.

Доказательство основано на применении сопряженных распределений¹⁾. Распределение V , сопряженное к F , задается равенством

$$V\{dx\} = e^{-\psi(s)} e^{sx} F\{dx\}, \quad (7.4)$$

где параметр s изменяется в интервале сходимости функции ψ . Функция

$$v(\zeta) = \frac{f(\zeta + s)}{f(s)} \quad (7.5)$$

играет по отношению к V ту же роль, что и f по отношению к исходному распределению F . В частности, дифференцируя (7.5), видим, что V имеет математическое ожидание $\psi'(s)$ и дисперсию $\psi''(s)$.

Идею доказательства можно объяснить следующим образом. Как легко видеть или из (7.4), или из (7.5), распределения F^{n*} и V^{n*} также связаны соотношением (7.4). Лишь нормирующая константа $e^{-\psi(s)}$ заменяется на $e^{-n\psi(s)}$.

Обращая соотношение (7.4) относительно F , получаем

$$1 - F_n(x) = 1 - F^{n*}(x\sigma\sqrt{n}) = e^{n\psi(s)} \int_{x\sigma\sqrt{n}}^{\infty} e^{-s \cdot y} V^{n*}\{dy\}. \quad (7.6)$$

Имея в виду центральную предельную теорему, представляется естественным заменить V^{n*} соответствующим нормальным распределением с математическим ожиданием $n\psi'(s)$ и дисперсией $n\psi''(s)$. Относительная ошибка такой аппроксимации будет малой, если нижний предел интеграла в (7.6) близок к математическому ожиданию V^{n*} , т. е. если x близко к $\psi'(s)\sqrt{n}/\sigma$. На этом пути можно получить при достаточно больших значениях x хорошие аппроксимации для величины $1 - F_n(x)$ и в том числе (7.1).

Доказательство. Функция ψ в окрестности нуля является ана-

¹⁾ Они уже были использованы в теории восстановления (гл. XI, 6) и при изучении случайных блужданий (гл. XII, 4).

литической функцией со степенным рядом

$$\psi(s) = 1 + \frac{1}{2} \sigma^2 s^2 + \frac{1}{6} \mu_3 s^3 + \dots \quad (7.7)$$

Кроме того, ψ — выпуклая функция с $\psi'(0) = 0$, и поэтому она возрастает при $s > 0$. Соотношение

$$\sqrt{n} \psi'(s) = \sigma x, \quad s > 0, \quad x > 0, \quad (7.8)$$

устанавливает взаимно однозначное соответствие между переменными s и x , когда s и $x/\sqrt{n}$ принимают значения из достаточно малой окрестности нуля. Каждая из этих величин может быть рассмотрена как аналитическая функция другой, и, очевидно,

$$s \sim \frac{x}{\sigma \sqrt{n}}, \quad \text{если} \quad \frac{x}{\sqrt{n}} \rightarrow 0. \quad (7.9)$$

Разобьем доказательство на две части.

а) Сначала мы вычислим величину A_s , получаемую заменой V^{n*} в (7.6) нормальным распределением с тем же математическим ожиданием $n\psi'(s)$ и с той же самой дисперсией $n\psi''(s)$. Стандартная подстановка $y = n\psi'(s) + t\sqrt{n\psi''(s)}$ дает

$$A_s = e^{n[\psi(s) - s\psi'(s)]} \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-ts\sqrt{n\psi''(s)} - (1/2)t^2} dt. \quad (7.10)$$

Дополняя показатель под интегралом до полного квадрата, получаем

$$A_s = \exp \left(n \left[\psi(s) - s\psi'(s) + \frac{1}{2} s^2 \psi''(s) \right] \right) \cdot [1 - \mathfrak{N}(s\sqrt{n\psi''(s)})]. \quad (7.11)$$

Показатель и первые две его производные в нуле обращаются в нуль, и поэтому соответствующий ряд по степеням s начинается с члена третьего порядка. Таким образом,

$$A_s = [1 - \mathfrak{N}(s\sqrt{n\psi''(s)})][1 + O(ns^3)], \quad s \rightarrow 0. \quad (7.12)$$

Если $ns^3 \rightarrow 0$ или, что то же самое, если $x = o(n^{1/6})$, то (7.12) можно переписать в виде

$$A_s = [1 - \mathfrak{N}(\bar{x})][1 + O(x^3/\sqrt{n})], \quad (7.13)$$

где

$$\bar{x} = s\sqrt{n\psi''(s)}. \quad (7.14)$$

Осталось показать, что в (7.13) можно заменить $\bar{x}$ на x . Разложение $(\bar{x} - x)/n$ по степеням s не зависит от n и, как показывают обычные вычисления, начинается с члена третьего порядка. Таким образом,

$$|\bar{x} - x| = O(\sqrt{ns^3}) = O(x^3/n). \quad (7.15)$$

Из 1, гл. VII, (1.8), мы знаем, что при $t \rightarrow \infty$

$$\frac{n(t)}{1-\mathfrak{N}(t)} \sim t. \quad (7.16)$$

Интегрируя в пределах от x до $\bar{x}$, получаем при $x \rightarrow \infty$

$$\left| \log \frac{1-\mathfrak{N}(\bar{x})}{1-\mathfrak{N}(x)} \right| = O(x \cdot |\bar{x} - x|) = O(x^4/n). \quad (7.17)$$

Следовательно,

$$\frac{1-\mathfrak{N}(\bar{x})}{1-\mathfrak{N}(x)} = 1 + O(x^4/n). \quad (7.18)$$

Отсюда и из (7.13) получаем при $x \rightarrow \infty$ и $x = o(n^{1/6})$

$$A_s = [1 - \mathfrak{N}(x)] [1 + O(x^3/\sqrt{n})]. \quad (7.19)$$

б) Пусть $\mathfrak{N}$ обозначает нормальное распределение с математическим ожиданием $n\psi'(s)$ и дисперсией $n\psi''(s)$. Тогда A_s по определению есть правая часть в (7.6), где распределение V^{n*} заменено на $\mathfrak{N}_s$. Мы приступим теперь к оценке ошибки, совершаемой при такой замене. По теореме Берри—Эссеена из § 5 для всех y

$$|V^{n*}(y) - \mathfrak{N}_s(y)| < 3M_s/\sigma^3\sqrt{n}, \quad (7.20)$$

где M_s обозначает абсолютный момент третьего порядка распределения V . Производя интегрирование в (7.20), после интегрирования по частям находим, что

$$\begin{aligned} |1 - F_n(x) - A_s| &< \frac{3M_s}{\sigma^3\sqrt{n}} e^{n\psi(s)} \left[e^{-s\psi'(s)} + s \int_{n\psi'(s)}^{\infty} e^{-sy} dy \right] = \\ &= \frac{6M_s}{\sigma^3\sqrt{n}} e^{n[\psi(s) - s\psi'(s)]}. \end{aligned} \quad (7.21)$$

В силу (7.11)

$$A_s = e^{n[\psi(s) - s\psi'(s)]} \cdot e^{\frac{1}{2}} \bar{x}^2 [1 - \mathfrak{N}(\bar{x})] \sim \frac{1}{x} e^{n[\psi(s) - s\psi'(s)]}, \quad (7.22)$$

и, следовательно, правая часть в (7.21) равна $A_s \cdot O(x/\sqrt{n})$. Таким образом,

$$1 - F_n(x) = A_s [1 + O(x/\sqrt{n})]. \quad (7.23)$$

Вместе с (7.19) это доказывает не только теорему, но и более сильное утверждение.

Следствие. Если $x \rightarrow \infty$, так что $x = o(n^{1/6})$, то

$$\frac{1 - F_n(x)}{1 - \mathfrak{N}(x)} = 1 + O\left(\frac{x^3}{\sqrt{n}}\right). \quad (7.24)$$

Мы попутно получили более общее соотношение, которое применимо

тогда, когда x меняется вместе с n так, что $x \rightarrow \infty$, но $x = O(\sqrt{n})$. Действительно, в этом случае в силу (7.23) $1 - F_n(x) \sim A_s$, где A_s определено в (7.11). Аргументом функции распределения $\mathfrak{H}$ в (7.11) служит $\bar{x}$, но из (7.18) следует, что $\bar{x}$ можно заменить на x . Мы получаем, следовательно, общую формулу

$$1 - F_n(x) = \exp \{n[\psi(s) - s\psi'(s) + (1/2)\psi''(s)]\} [1 - \mathfrak{H}(x)] [1 + O(x/\sqrt{n})]. \quad (7.25)$$

Показатель можно представить степенным рядом относительно s , начинающимся с члена третьего порядка. Как и в (7.8), мы определяем теперь аналитическую функцию переменного z равенством $\psi'(s) = \sigma z$. Затем мы определяем степенной ряд λ соотношением

$$z^2 \lambda(z) = \lambda_1 z^3 + \lambda_2 z^4 + \dots = \psi(s) - s\psi'(s) + \frac{1}{2} \psi''(s). \quad (7.26)$$

Тогда верна

Теорема 2¹⁾. Если в теореме 1 условие $x = o(n^{1/6})$ заменить условием $x = o(\sqrt{n})$, то

$$\frac{1 - F_n(x)}{1 - \mathfrak{H}(x)} = \exp \{x^2 \lambda(x/\sqrt{n})\} \left[1 + O\left(\frac{x}{\sqrt{n}}\right) \right]. \quad (7.27)$$

В частности, если $x = o(n^{1/4})$, то достаточно учитывать только первый член степенного ряда λ . Мы получаем

$$\frac{1 - F_n(x)}{1 - \mathfrak{H}(x)} \sim \exp \{\lambda_1 x^3 / \sqrt{n}\}, \quad \lambda_1 = \frac{\mu_3}{6\sigma^3}. \quad (7.28)$$

Если x возрастает как $o(n^{3/10})$, имеем

$$\frac{1 - F_n(x)}{1 - \mathfrak{H}(x)} \sim \exp \left(\lambda_1 \frac{x^3}{\sqrt{n}} + \lambda_2 \frac{x^4}{n} \right), \quad \lambda_2 = \frac{\sigma^2 \psi_4 - 3\psi_3^2}{24\sigma^6} \quad (7.29)$$

и так далее.

Заметим, что правые части в приведенных соотношениях могут стремиться или к 0, или к ∞ (и, следовательно, из этих формул не вытекает асимптотическая эквивалентность $1 - F_n(x)$ и $1 - \mathfrak{H}(x)$). Такая эквивалентность существует только в случае, когда $x = o(n^{1/6})$ [или в случае, когда $x = o(n^{1/4})$ и третий момент равен нулю]. При любых условиях мы имеем следующее интересное

Следствие. Если $x = o(\sqrt{n})$, то при любом $\varepsilon > 0$

$$\exp \{-(1+\varepsilon)x^2/2\} < 1 - F_n(x) < \exp \{-(1-\varepsilon)x^2/2\}. \quad (7.30)$$

Развитую здесь теорию можно распространить на суммы случайных величин X_k с различными распределениями и характеристическими функциями ω_k . Эту процедуру можно иллюстрировать

¹⁾ Использование преобразования (7.4) в связи с центральной предельной теоремой принадлежит, по-видимому, Эшсеру (1932). Теорема 2 доказана Г. Крамером (1938) и была затем распространена на разнораспределенные слагаемые Феллером (1943). Более новые результаты указаны в статьях В. В. Петрова [УМН, 9 (1954)] и В. Рихтера [Локальные теоремы для больших отклонений, Теория вероятностей и ее применения, 2 (1957), стр. 206—220]. Последний рассматривает, правда, не функции распределения, а плотности.

О другом подходе, приводящем к аппроксимации вида $1 - F_n(x) \approx \exp[\nu(x) + o(\nu(x))]$, см. в работе Феллера, Zs. Wahrscheinlichkeitstheorie verw. Gebiete, v. 14 (1969), 1—20.

следующим ниже обобщением теоремы 1, где условия равномерности являются излишне строгими. Распределение нормированной суммы $(X_1 + \dots + X_n)/s_n$ снова обозначается через F_n .

Теорема 3. *Предположим, что существует интервал $\overline{-a, a}$, на котором все характеристические функции ω_k являются аналитическими, и что*

$$E(|X_n|^3) \leq M\sigma_n^2, \quad (7.31)$$

где постоянная M не зависит от n . Если s_n и x стремятся к ∞ так, что $x = o(s_n^{1/3})$, то

$$\frac{1 - F_n(x)}{1 - \mathfrak{F}(x)} \rightarrow 1 \quad (7.32)$$

с ошибкой $O(x^3/s_n)$.

Доказательство остается тем же с той лишь разницей, что ψ заменяется теперь определенной при $-a < s < a$ аналитической функцией ψ_n , принимающей на этом интервале действительные значения:

$$\psi_n(s) = \frac{1}{n} \sum_{k=1}^n \ln \omega_k(-is). \quad (7.33)$$

Во всех выкладках место $x\sigma\sqrt{n}$ займет теперь xs_n . Основное соотношение (7.8) примет вид $\psi'_n(s) = xs_n/n$.

В этой главе представлено ядро ставшей теперь классической области предельных теорем теории вероятностей. Эта область подобна водохранилищу, образовавшемуся из бесчисленных отдельных потоков и течений. Наиболее экономное изложение следовало бы начинать со схемы серий, представленной в § 7. Однако мы и на этот раз начнем с обсуждения простых частных случаев для того, чтобы облегчить восприятие разнообразных важных вопросов.

Понятия безграничной делимости, устойчивости и т. д. и их интуитивный смысл обсуждались в гл. VI. Основные результаты этой главы были получены в другой форме и другими методами в гл. IX, но эта глава содержит более детальную информацию. Настоящая глава автономна и ее можно изучать как непосредственное продолжение гл. XV (о характеристических функциях) независимо от других глав.

§ 1. БЕЗГРАНИЧНО ДЕЛИМЫЕ РАСПРЕДЕЛЕНИЯ

Мы по-прежнему будем использовать определенную терминологию, в равной степени относящуюся как к распределениям, так и к характеристическим функциям. Имея это в виду, переформулируем определение безграничной делимости, данное в гл. VI,3, следующим образом.

Определение. *Характеристическую функцию ω называют безгранично делимой тогда и только тогда, когда при каждом n существует характеристическая функция ω_n , такая, что*

$$\omega_n^n = \omega. \quad (1.1)$$

Ниже будет показано, что безграничная делимость может характеризоваться и другими замечательными свойствами, которые объясняют, почему это понятие играет такую важную роль в теории вероятностей.

Замечание о корнях и логарифмах характеристических функций. Заманчиво рассматривать ω_n в (1.1) как корень n -й степени из ω . Однако, чтобы этим воспользоваться, мы должны показать, что этот корень почти всюду определяется однозначно. Для обсуждения вопроса о неопределенности корней и логарифмов в комплексной области удобно тригонометрическое представление $a = re^{i\theta}$ комплексного числа $a \neq 0$. Положительное число r однозначно опре-

делено, однако аргумент θ определен только с точностью до кратных 2π . По существу, эта неопределенность распространяется на представления $\log a = \log r + i\theta$ и $a^{1/n} = r^{1/n} e^{i\theta/n}$ (где $r^{1/n}$ обозначает арифметический корень, а $\log r$ — обычный логарифм действительного числа). Тем не менее в любом интервале $|\xi| < \xi_0$, в котором $\omega(\xi) \neq 0$, характеристическая функция ω допускает *единственное* представление вида $\omega(\xi) = r(\xi) e^{i\theta(\xi)}$, такое, что θ непрерывна и $\theta(0) = 0$. В таком интервале мы можем записать однозначным образом $\log \omega(\xi) = \log r(\xi) + i\theta(\xi)$ и $\omega^{1/n}(\xi) = r^{1/n}(\xi) e^{i\theta(\xi)/n}$. Наши определения — это единственный способ сделать $\log \omega$ и $\omega^{1/n}$ непрерывными функциями, которые действительны в точке $\xi = 0$. В этом смысле $\log \omega$ и $\omega^{1/n}$ *однозначно определены в любом интервале $|\xi| < \xi_0$, не содержащем нулей характеристической функции ω* . Мы будем использовать обозначения $\log \omega$ и $\omega^{1/n}$ только в указанном смысле. Однако надо иметь в виду, что это определение перестает действовать¹⁾, как только $\omega(\xi_0) = 0$. ►

Пусть F — произвольное вероятностное распределение и φ — его характеристическая функция. Напомним, что, согласно (2.4) из гл. XV, F порождает семейство обобщенных *распределений Пуассона*

$$e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} F^{k*} \quad (1.2)$$

с характеристическими функциями $e^{c(\varphi-1)}$, где $c > 0$ — произвольный параметр. Очевидно, что функция $\omega = e^{c(\varphi-1)}$ — безгранично делимая (корень $\omega^{1/n}$ имеет ту же самую форму с заменой c на c/n). Нормальное распределение и распределение Коши свидетельствуют, что безгранично делимое распределение не обязано принадлежать обобщенному пуассоновскому типу. Однако мы покажем, что любое безгранично делимое распределение является пределом последовательности обобщенных пуассоновских распределений. Основой всей теории является

Теорема 1. Пусть $\{\varphi_n\}$ — последовательность характеристических функций. Для того чтобы существовал непрерывный предел

$$\omega(\xi) = \lim \varphi_n^n(\xi), \quad (1.3)$$

необходимо и достаточно, чтобы

$$n[\varphi_n(\xi) - 1] \rightarrow \psi(\xi), \quad (1.4)$$

где функция ψ непрерывна. В этом случае

$$\omega(\xi) = e^{\psi(\xi)}. \quad (1.5)$$

¹⁾ В конце гл. XV, 2 (а также в задаче 9, гл. XV, 9), указаны пары действительных характеристических функций, таких, что $\varphi_1^2 = \varphi_2^2$. Это показывает, что при наличии нулей даже действительные характеристические функции могут иметь два действительных корня, которые также будут характеристическими функциями.

Доказательство. Напомним, что согласно теореме непрерывности из гл. XV,3, если последовательность характеристических функций сходится к непрерывной функции, то последняя является характеристической функцией, а сходимость автоматически будет равномерной на любом конечном интервале.

а) Начнем с более простой части теоремы. Пусть выполнено (1.4), где ψ непрерывна. Тогда $\varphi_n(\xi) \rightarrow 1$ при каждом фиксированном ξ и сходимость равномерна на конечных интервалах. Это означает, что в любом интервале $|\xi| < \xi_1$ для всех достаточно больших n выполняется неравенство $|1 - \varphi_n(\xi)| < 1$. Тогда для таких больших n из разложения Тейлора функции $\log(1-z)$ мы заключаем, что

$$\begin{aligned} n \log \varphi_n(\xi) &= n \log [1 - [1 - \varphi_n(\xi)]] = \\ &= -n[1 - \varphi_n(\xi)] - \frac{n}{2}[1 - \varphi_n(\xi)]^2 - \dots \end{aligned} \quad (1.6)$$

В соответствии с (1.4) первый член справа стремится к $\psi(\xi)$ и так как $\varphi_n(s) \rightarrow 1$, то получаем, что все остальные члены стремятся к нулю. Таким образом, $n \log \varphi_n \rightarrow \psi$ или $\varphi_n^n \rightarrow e^\psi$, что и утверждалось.

в) Обратное утверждение также легко доказать, если известно, что в (1.3) предельная функция ω не имеет нулей. Действительно, рассмотрим произвольный конечный интервал $|\xi| \leq \xi_1$. В нем определяемая соотношением (1.3) сходимость равномерная, и для всех достаточно больших n отсутствие нулей у функции ω приводит к тому, что также и $\varphi_n(\xi) \neq 0$ при $|\xi| \leq \xi_1$. Следовательно, мы можем перейти к логарифмам и сделать вывод, что $n \log \varphi_n \rightarrow \log \omega$ и поэтому $\log \varphi_n \rightarrow 0$. Это приводит нас к тому, что $\varphi_n(\xi) \rightarrow 1$ для каждого фиксированного ξ , причем на каждом конечном интервале эта сходимость равномерная. Так же как и в а), мы делаем вывод, что разложение (1.6) справедливо, и, так как $1 - \varphi_n(\xi) \rightarrow 0$, приходим к тому, что

$$n \log \varphi_n(\xi) = -n[1 - \varphi_n(\xi)][1 + O(1)], \quad (1.7)$$

где $O(1)$ обозначает величину, стремящуюся к нулю при $n \rightarrow \infty$. По предположению левая сторона имеет пределом $\log \omega(\xi)$, и поэтому очевидно, что $n[1 - \varphi_n] \rightarrow -\log \psi$, что и утверждалось.

Для завершения доказательства мы должны показать, что $\omega(\xi)$ не может обращаться в нуль ни при каком значении ξ . Для этого заменим функции ω и φ_n характеристическими функциями $|\omega|^2$ и $|\varphi_n|^2$ соответственно и нам достаточно рассмотреть частный случай (1.3), когда все φ_n — действительные и $\varphi_n \geq 0$. Пусть теперь $|\xi| \leq \xi_1$ — интервал, в котором $\omega(\xi) > 0$. Внутри этого интервала $-n \log \varphi_n(\xi)$ — положительная и ограниченная функция. С другой стороны, для $|\xi| \leq \xi_1$ справедливо разложение (1.6) и, так как все члены имеют одинаковый знак, заключаем, что и $n[1 - \varphi_n(\xi)]$ ограничена для всех $|\xi| \leq \xi_1$. Однако на основании основного

неравенства гл. XV, (1.7), для характеристических функций

$$n[1 - \varphi_n(2\xi)] \leq 4n[1 - \varphi_n(\xi)],$$

мы заключаем, что $n(1 - \varphi_n(\xi))$ остается ограниченной для всех $|\xi| \leq 2\xi_1$. Следовательно, этот интервал не содержит нулей функции ω . Теперь, применяя наши рассуждения к этому интервалу, сделаем вывод, что $\omega(\xi) > 0$ для всех $|\xi| \leq 4\xi_1$. Этот процесс показывает, что $\omega(\xi) > 0$ для всех ξ , что и завершает доказательство. ►

Теорема 1 имеет много следствий. Умножив соотношение (1.4) на $t > 0$, отметим, что оно эквивалентно следующему:

$$e^{tn[\varphi_n(\xi) - 1]} \rightarrow e^{t\psi(\xi)} = \omega^t(\xi). \quad (1.8)$$

Левая часть представляет собой характеристическую функцию обобщенного распределения Пуассона, и, следовательно, $e^{t\psi(\xi)}$ — характеристическая функция для всех $t > 0$. В частности, мы отмечаем, что функция $\omega = e^\psi$ безгранично делима. Другими словами, каждая характеристическая функция ω , представимая в виде предела последовательности $\{\varphi_n^n\}$ характеристических функций, безгранично делимая. Это можно было бы принять за обобщение определения безграничной делимости, заменив тождество (1.1) более общим предельным соотношением (1.3). В § 7 будет показано, что этот результат можно дальше продолжить на более общую схему серий. Зафиксируем наши результаты в виде теоремы 2.

Теорема 2. *Характеристическая функция ω безгранично делима тогда и только тогда, когда существует последовательность $\{\varphi_n\}$ характеристических функций $\{\varphi_n^n\}$, таких, что $\varphi_n^n \rightarrow \omega$.*

В этом случае для любого $t > 0$ ω^t — характеристическая функция и $\omega(\xi) \neq 0$ для всех ξ .

Следствие. *Непрерывный предел последовательности $\{\omega_n\}$ безгранично делимых характеристических функций — также безгранично делимая функция.*

Доказательство. По предположению $\varphi_n = \omega_n^{1/n}$ снова характеристическая функция, и поэтому соотношение $\omega_n = \omega$ можно переписать в виде $\varphi_n^n \rightarrow \omega$. ►

Каждое обобщенное распределение Пуассона безгранично делимо, и теорема 1 указывает, что любое безгранично делимое распределение можно представить как предел последовательности обобщенных распределений Пуассона (см. (1.8) при $t = 1$).

Таким образом, мы получаем новую характеристику безгранично делимых распределений.

Теорема 3. *Класс безгранично делимых распределений совпадает с классом предельных распределений для обобщенных пуассоновских распределений.*

Применение к процессам с независимыми приращениями. Согласно гл. VI,3, такие процессы могут быть описаны посредством семейств $\{X(t)\}$ случайных величин, обладающих тем свойством, что для любого набора $t_0 < t_1 < \dots < t_n$ приращения $X(t_k) - X(t_{k-1})$ представляют собой n взаимно независимых случайных величин. Приращения *стационарны*, если распределение $X(s+t) - X(s)$ зависит только от длины t интервала и не зависит от его положения на временной оси. В этом случае $X(s+t) - X(s)$ есть сумма n независимых случайных величин, распределенных так же, как $X(s+t/n) - X(s)$. Следовательно, *распределение $X(s+t) - X(s)$ безгранично делимо*. Обратно, любое семейство безгранично делимых распределений с характеристическими функциями вида $e^{t\psi}$ определяет процесс с независимыми стационарными приращениями. В § 7 этот результат будет обобщен для схемы серий на процессы с *нестационарными* независимыми приращениями. Приращение $X(t+s) - X(s)$ будет тогда суммой приращений $X(t_k+1) - X(t_k)$, которые являются взаимно независимыми случайными величинами. Тогда теорема из § 7 применима при условии, что процесс непрерывен, т. е. $X(t+h) - X(t)$ стремится по вероятности к нулю при $h \rightarrow 0$. Для таких процессов *распределение приращений $X(t+s) - X(s)$ безгранично делимо*. (Существуют разрывные процессы такого типа, однако разрывы имеют простую природу и в определенном смысле устранимы. См. обсуждение в гл. IX,5а, и IX,9.)

Обобщенные пуассоновские процессы допускают очень простую вероятностную интерпретацию (см. гл. VI,3, и гл. IX,5), и тот факт, что любое безгранично делимое распределение представимо как предел обобщенных пуассоновских распределений, помогает понять природу более общих процессов с независимыми приращениями.

§ 2. КАНОНИЧЕСКИЕ ФОРМЫ. ОСНОВНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Мы видели, что для отыскания наиболее общего вида безгранично делимых характеристических функций $\omega = e^\psi$ достаточно определить общий вид возможных пределов последовательностей характеристических функций $\exp c_n(\xi_n - 1)$ обобщенных пуассоновских распределений. Для различных приложений желательно поставить более общую задачу, используя произвольное центрирование. Таким образом, мы ищем возможные пределы характеристических функций вида $\omega_n = e^{\psi_n}$, где мы полагаем

$$\psi_n(\xi) = c_n[\varphi_n(\xi) - 1 - i\beta_n\xi]. \quad (2.1)$$

Так как ω_n безгранично делимы, то же самое верно и для их непрерывных пределов.

Наша цель — найти условия, при которых существует *непрерывный* предел

$$\psi(\xi) = \lim \psi_n(\xi). \quad (2.2)$$

При этом φ_n — характеристическая функция вероятностного распределения F_n , c_n — положительные постоянные, β_n — действительные центрирующие постоянные.

Для распределений, имеющих математическое ожидание, естественно центрирование математическим ожиданием, и всякий раз,

когда это возможно, мы будем выбирать β_n соответствующим образом. Однако нам нужно подходящее ко всем случаям центрирование с подобными свойствами. Как оказывается, в простейшем случае такое центрирование получается, если потребовать, чтобы при $\zeta = 1$ значение ψ_n было действительным. Если u_n и v_n обозначают действительную и мнимую части φ_n , то наше условие требует, чтобы

$$\beta_n = v_n(1) = \int_{-\infty}^{\infty} \sin x F_n \{dx\}. \quad (2.3)$$

Отсюда видно, что такое центрирование всегда возможно. При этом

$$\psi_n(\zeta) = c_n \int_{-\infty}^{\infty} [e^{i\zeta x} - 1 - i\zeta \sin x] F_n \{dx\}. \quad (2.4)$$

Подынтегральное выражение в (2.4) ведет себя вблизи нуля подобно $-\frac{1}{2}\zeta^2 x^2$, т. е. так же, как при более привычном центрировании математическим ожиданием. Полезность центрирования с постоянной (2.3) в большой степени обусловлена следующим утверждением.

Лемма. Пусть заданы последовательности $\{c_n\}$ и $\{\varphi_n\}$. Если существуют центрирующие постоянные β_n , такие, что ψ_n стремятся к непрерывному пределу ψ , тогда и для центрирующих постоянных (2.3) ψ_n стремятся к непрерывному пределу.

Доказательство. Определим ψ_n формулой (2.1) с произвольными β_n и предположим, что $\psi_n \rightarrow \psi$. Если b обозначает мнимую часть $\psi(1)$, то

$$c_n(v_n(1) - \beta_n) \rightarrow b. \quad (2.5)$$

Умножая на $i\zeta$ и вычитая левую часть из ψ_n , а правую часть из ψ , мы получаем

$$c_n[\varphi_n(\zeta) - 1 - iv_n(1)\zeta] \rightarrow \psi(\zeta) - ib\zeta, \quad (2.6)$$

что завершает доказательство. ►

Мы начнем исследование проблемы сходимости в специальном случае, когда ее решение особенно простое. Предположим, что функции ψ_n и ψ дважды непрерывно дифференцируемы (это означает, что соответствующие распределения имеют дисперсии; см. гл. XV, 4). Предположим также, что не только $\psi_n \rightarrow \psi$, но и $\psi_n'' \rightarrow \psi''$. Это означает, согласно (2.1), что

$$c_n \int_{-\infty}^{+\infty} e^{i\zeta x} x^2 F_n \{dx\} \rightarrow -\psi''(\zeta). \quad (2.7)$$

Предполагая, что $c_n x^2 F_n \{dx\}$ определяет конечную меру, обозна-

чим ее полную массу через μ_n . Полагая $\xi = 0$ в (2.7), мы заключаем, что $\mu_n \rightarrow -\psi''(0)$. После деления обеих частей (2.7) на μ_n мы получим слева характеристическую функцию некоторого собственного распределения вероятностей, которая при $n \rightarrow \infty$ стремится к $\psi''(\xi)/\psi''(0)$. Отсюда следует, что $\psi''(\xi)/\psi''(0)$ есть характеристическая функция некоторого распределения вероятностей и, следовательно,

$$-\psi''(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} M\{dx\}, \quad (2.8)$$

где M — конечная мера. Находим теперь ψ повторным интегрированием равенства (2.8). Принимая во внимание, что $\psi(0) = 0$ и при нашем центрировании $\psi(1)$ должно быть действительным, мы получаем

$$\psi(\xi) = \int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi \sin x}{x^2} M\{dx\}. \quad (2.9)$$

Этот интеграл имеет смысл, поскольку подынтегральное выражение — ограниченная непрерывная функция, принимающая в нуле значение $-\frac{1}{2}\xi^2$.

При наших предположениях относительно дифференцируемости предел ψ необходимо имеет вид (2.9). Теперь мы покажем, что для произвольно выбранной конечной меры M интеграл (2.9) определяет безгранично делимую характеристическую функцию e^ψ . Однако мы можем сделать еще один шаг. Для того чтобы интеграл (2.9) был определен, нет необходимости требовать конечности меры M . Достаточно того, что M принимает конечные значения на конечных интервалах и что $M\{\overline{-x, x}\}$ возрастает достаточно медленно, чтобы интегралы

$$M^+(x) = \int_x^{+\infty} y^{-2} M\{dy\}, \quad M^-(-x) = \int_{-\infty}^{-x} y^{-2} M\{dy\} \quad (2.10)$$

сходились для всех $x > 0$. (Для определенности мы считаем интервалы интегрирования замкнутыми.) Типичные примеры представляют собой меры, определяемые плотностями $|x|^p dx$ при $0 < p < 1$. Мы покажем, что если M обладает указанными свойствами, то (2.9) определяет безгранично делимую характеристическую функцию, и что все безгранично делимые характеристические функции получаются таким способом. В связи с этим удобно ввести специальное название для таких мер.

Определение 1. Назовем меру M канонической, если она принимает конечные значения на конечных интервалах и интегралы (2.10) сходятся для некоторых (и, следовательно, всех) $x > 0$.

Лемма 2. Если M — каноническая мера и ψ задается выраже-

нием (2.9), то e^ψ — безгранично делимая характеристическая функция.

Доказательство. Мы рассмотрим два важных частных случая.

а) Предположим, что мера M сосредоточена в нуле и принимает там значение $m > 0$. Тогда $\psi(\xi) = -m\xi^2/2$ и e^ψ — характеристическая функция нормального распределения с дисперсией m^{-1} .

б) Предположим, что мера M сосредоточена на множестве $|x| > \eta$, где $\eta > 0$. В этом случае можно переписать (2.9) в более простом виде. В самом деле, теперь $x^{-2}M\{dx\}$ определяет конечную меру с полной массой $\mu = M^+(\eta) + M^-(-\eta)$. Соответственно $x^{-2}M\{dx\}/\mu = F\{dx\}$ определяет вероятностную меру с характеристической функцией φ и очевидно, что $\psi(\xi) = \mu[\varphi(\xi) - 1 - ib\xi]$, где b — действительная константа. Таким образом, в этом случае e^ψ — характеристическая функция обобщенного пуассоновского распределения и, следовательно, является безгранично делимой.

в) В общем случае, пусть мера M задает массу $m \geq 0$ в нуле и пусть

$$\psi_\eta(\xi) = \int_{|x| > \eta} \frac{e^{i\xi x} - 1 - i\xi \sin x}{x^2} M\{dx\}. \quad (2.11)$$

Тогда

$$\psi(\xi) = -\frac{m}{2}\xi^2 + \lim_{\eta \rightarrow 0} \psi_\eta(\xi). \quad (2.12)$$

Мы видели, что $e^{\psi_\eta(\xi)}$ — характеристическая функция безгранично делимого распределения U_η . Если $m > 0$, то сложение величины $(-m\xi^2/2)$ с $\psi_\eta(\xi)$ соответствует свертке U_η с нормальным распределением. Таким образом, с учетом (2.12) e^ψ предстает как предел последовательности безгранично делимых характеристических функций, и, следовательно, e^ψ безгранично делима, как и утверждалось. ►

Мы теперь покажем, что представление (2.9) единственно в том смысле, что различные канонические меры приводят к различным интегралам.

Лемма 3. *Представление (2.9) для ψ единственно.*

Доказательство. В частном случае конечной меры M ясно, что существует вторая производная ψ'' и что $-\psi''(\xi)$ совпадает с математическим ожиданием $e^{i\xi x}$ относительно меры M . Из теоремы единственности для характеристических функций следует, что M однозначно определяется по ψ'' и, следовательно, по ψ .

Это рассуждение может быть приспособлено для случая неограниченных канонических мер. При этом необходимо вторую производную заменить оператором с теми же свойствами, но применимым к произвольным непрерывным функциям. Такие операторы можно

выбрать различными способами (см. задачи 1—3). Мы выберем оператор, который преобразует ψ в функцию ψ^* по формуле

$$\psi^*(\xi) = \psi(\xi) - \frac{1}{2h} \int_{-h}^h \psi(\xi + s) ds, \quad (2.13)$$

где $h > 0$ — произвольное, но фиксированное число. Для функции ψ , определяемой равенством (2.9), мы получим

$$\psi^*(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} K(x) M\{dx\}, \quad (2.14)$$

где мы для краткости обозначили

$$K(x) = x^{-2} \left[1 - \frac{\sin xh}{xh} \right]. \quad (2.15)$$

Это строго положительная непрерывная функция, принимающая в нуле значение $h^2/6$; $K(x) \sim x^{-2}$ при $x \rightarrow \pm\infty$. Поэтому мера M^* , определяемая соотношением $M^*\{dx\} = K(x) M\{dx\}$, конечна, а равенство (2.14) задает ψ^* как преобразование Фурье меры M^* . По теореме единственности для характеристических функций значения ψ^* однозначно определяют меру M^* . Но тогда и $M\{dx\} = K^{-1}(x) M^*\{dx\}$ однозначно определено, а поэтому значение ψ позволяет нам вычислять соответствующую каноническую меру (ср. с задачей 3). ►

Наша следующая цель состоит в том, чтобы доказать, что лемма 2 описывает совокупность всех безгранично делимых характеристических функций. Однако для этого мы должны сначала решить проблему сходимости, поставленную в начале этого параграфа. Теперь мы сформулируем ее в несколько более общей форме. Пусть $\{M_n\}$ — последовательность канонических мер и

$$\psi_n(\xi) = \int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi \sin x}{x^2} M_n\{dx\} + ib_n \xi, \quad (2.16)$$

где b_n — действительные числа. Нас интересуют необходимые и достаточные условия сходимости $\psi_n \rightarrow \psi$, где ψ — непрерывная функция. Заметим, что функции ψ_n , определенные формулами (2.1) или (2.4), представляют собой частный случай (2.16), когда

$$M_n\{dx\} = c_n x^2 F_n\{dx\}. \quad (2.17)$$

Предположим, что $\psi_n \rightarrow \psi$ и предельная функция ψ непрерывна. Тогда для преобразований, определяемых формулой (2.13), выполнено $\psi_n^* \rightarrow \psi^*$, т. е.

$$\int_{-\infty}^{+\infty} e^{i\xi x} K(x) M_n\{dx\} \rightarrow \psi^*(\xi), \quad (2.18)$$

где K — строго положительная непрерывная функция из (2.15). Слева в (2.18) мы имеем преобразование Фурье конечной меры с полной массой

$$\mu_n = \int_{-\infty}^{+\infty} K(x) M_n \{dx\}. \quad (2.19)$$

Очевидно, что $\mu_n \rightarrow \psi^*(0)$. Легко видеть, что $\mu_n \rightarrow 0$ должно влечь за собой $\psi(\xi) = 0$ для всех ξ . Следовательно, мы можем положить $\psi^*(0) = \mu > 0$. Тогда меры M_n^* , определяемые равенством

$$M_n^* \{dx\} = \frac{1}{\mu_n} K(x) M_n \{dx\}, \quad (2.20)$$

являются вероятностными мерами, и, согласно (2.18), их характеристические функции стремятся к непрерывной функции $\psi^*(\xi)/\psi^*(0)$. Поэтому

$$M_n^* \rightarrow M^*, \quad (2.21)$$

где M^* — распределение вероятностей с характеристической функцией $\psi^*(\xi)/\psi^*(0)$. Однако ψ_n можно переписать в виде

$$\psi_n(\xi) = \mu_n \int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi \sin x}{x^2} K^{-1}(x) M_n^* \{dx\} + ib_n \xi. \quad (2.22)$$

Поскольку подынтегральное выражение представляет собой ограниченную непрерывную функцию аргумента x , то ввиду (2.21) интегралы сходятся при $n \rightarrow \infty$. Отсюда следует, что $b_n \rightarrow b$ и предельная функция ψ имеет вид

$$\psi(\xi) = \mu \int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi \sin x}{x^2} K^{-1}(x) M^* \{dx\} + ib\xi. \quad (2.23)$$

Так как M^* — вероятностная мера, то мера M , определяемая соотношением

$$M \{dx\} = \mu K^{-1}(x) M^* \{dx\}, \quad (2.24)$$

является канонической и

$$\psi(\xi) = \int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi \sin x}{x^2} M \{dx\} + ib\xi. \quad (2.25)$$

Теперь видно, что, если отвлечься от центрирующего слагаемого $ib\xi$, все предельные функции имеют вид, установленный в лемме 2. Как уже отмечалось, функции ψ_n из (2.1) являются собой частный случай (2.16), и, следовательно, мы решили проблему сходимости, сформулированную в начале этого параграфа. Зафиксируем этот результат в форме теоремы.

Теорема 1. Класс безгранично делимых характеристических функций совпадает с классом функций вида e^{ψ} , где ψ определяется равенством (2.25) через каноническую меру M и действительное число b .

Другими словами, с точностью до произвольного способа центрирования существует взаимно однозначное соответствие между каноническими мерами и безгранично делимыми распределениями.

В предыдущих рассуждениях мы подчеркивали, что условия $M_n^* \rightarrow M^*$ и $b_n \rightarrow b$ являются необходимыми для сходимости $\psi_n \rightarrow \psi$. На самом деле мы также показали и достаточность этих условий, так как ψ_n можно записать в виде (2.22), откуда с очевидностью следует, что ψ_n стремится к предельной функции, определяемой формулой (2.23). Таким образом, нами получена полезная предельная теорема, но было бы желательно условие $M_n^* \rightarrow M^*$ выразить в терминах канонических мер M_n и M . Связь между M_n и M_n^* задается соотношением (2.22). Поскольку на конечных интервалах значения функции $K(x)$ отделены от 0 и ∞ , то для любого конечного интервала I соотношения $M_n^* \{I\} \rightarrow M^* \{I\}$ и $M_n \{I\} \rightarrow M \{I\}$ равносильны. При $x \rightarrow \infty$ функция $K(x)$ приближенно ведет себя как x^{-2} , и, следовательно, $M_n^* \{x, \infty\} \sim M_n^+(x)$, где M_n^+ обозначает один из интегралов (2.10), которые были введены при определении канонических мер. Таким образом, собственная сходимость $M_n^* \rightarrow M^*$ полностью эквивалентна соотношению

$$M_n \{I\} \rightarrow M \{I\} \quad (2.26)$$

для всех конечных интервалов непрерывности меры M или соотношениям

$$M_n^+(x) \rightarrow M^+(x), \quad M_n^-(x) \rightarrow M^-(x) \quad (2.27)$$

во всех точках $x > 0$ непрерывности предельных функций. В случае канонических мер вида $M_n \{dx\} = c_n x^2 F_n \{dx\}$ (где F_n — вероятностное распределение) эти соотношения принимают вид

$$c_n \int_I x^2 F_n \{dx\} \rightarrow M \{I\} \quad (2.28)$$

и

$$c_n [1 - F_n(x)] \rightarrow M^+(x), \quad c_n F_n(-x) \rightarrow M^-(-x). \quad (2.29)$$

Если $I = \overline{a, b}$ — конечный интервал непрерывности, который не содержит точку нуль, то из (2.28), очевидно, вытекает $c_n F_n \{I\} \rightarrow M^+(a) - M^+(b)$ и поэтому можно рассматривать (2.29) как распространение (2.28) на полупрямые. Это условие равносильно требованию, чтобы не было «утечки массы на бесконечности» в том смысле, что для любого $\varepsilon > 0$ найдется такое τ , что

$$c_n [1 - F_n(\tau) + F_n(-\tau)] < \varepsilon \quad (2.30)$$

по крайней мере для всех достаточно больших n . При наличии (2.28) условия (2.29) и (2.30) эквивалентны. [Заметим, что левая часть (2.30) есть убывающая функция τ .] Последовательности канонических мер $M_n \{dx\} = c_n x^2 F_n \{dx\}$ будут встречаться настолько часто, что желательно ввести удобное понятие для ссылок.

Определение 2. Последовательность $\{M_n\}$ канонических мер называют собственно сходящейся в собственном смысле к канонической мере M , если выполняются условия (2.26) и (2.27). В этом и только в этом случае мы пишем $M_n \rightarrow M$.

Используя новую терминологию можно сформулировать полученный результат о сходимости $\psi_n \rightarrow \psi$ в следующем виде.

Теорема 2. Пусть M_n — каноническая мера, а ψ_n определяется по формуле (2.16). Для того чтобы ψ_n сходилась к непрерывной предельной функции ψ , необходимо и достаточно, чтобы существовала такая каноническая мера M , что $M_n \rightarrow M$, и чтобы $b_n \rightarrow b$. В этом случае ψ определяется из (2.25).

В дальнейшем мы будем пользоваться этой теоремой только в частном случае, когда

$$\psi_n(\xi) = c_n [\varphi_n(\xi) - 1 - ib_n \xi], \quad (2.31)$$

где φ_n есть характеристическая функция вероятностного распределения F_n . Тогда наши условия принимают вид

$$c_n x^2 F_n \{dx\} \rightarrow M \{dx\}, \quad c_n (\beta_n - b_n) \rightarrow b, \quad (2.32)$$

где

$$\beta_n = \int_{-\infty}^{+\infty} \sin x \cdot F \{dx\}. \quad (2.33)$$

В силу теоремы 1 § 1 наши условия применимы не только к последовательностям обобщенных пуассоновских распределений, но и к более общим последовательностям вида $\{\varphi_n^n\}$.

Замечание о других канонических представлениях. Представление с мерой M не единственное, которое можно встретить в литературе. В первоначальных исследованиях П. Леви использовалась мера Λ , определяемая для $x \neq 0$ равенством $\Lambda \{dx\} = x^{-2} M \{dx\}$. Эта мера служит пределом для $n F_n \{dx\}$. Она конечна на интервалах $|x| > \delta > 0$, но может быть неограниченной в окрестности нуля. Атом M в нуле, если он существует, не учитывается мерой Λ . В терминах этой меры (2.9) принимает вид

$$\psi(\xi) = -\frac{1}{2} \sigma^2 \xi^2 + ib\xi + \lim_{\delta \rightarrow 0} \int_{|x| > \delta} [e^{i\xi x} - 1 + i\xi \sin x] \Lambda \{dx\}. \quad (2.34)$$

Это есть (с точностью до выбора центрирующей функции) первоначальное каноническое представление, данное П. Леви. Его основной недостаток в том, что полное описание свойств меры Λ оказывается слишком длинным.

Хинчин ввел ограниченную меру K , определяемую равенством $K \{dx\} = (1+x^2)^{-1} M \{dx\}$. Эта ограниченная мера может быть выбрана произвольно.

Поэтому каноническое представление по Хинчину задается выражением

$$x(\zeta) = i b \zeta + \int_{-\infty}^{+\infty} \left[e^{i \zeta x} - 1 - \frac{i \zeta x}{1+x^2} \right] \frac{1+x^2}{x^2} K\{dx\}. \quad (2.35)$$

Его легче всего описать, поскольку оно имеет дело только с ограниченными мерами. Это преимущество нейтрализуется тем обстоятельством, что искусственный выбор меры K излишне усложняет многие рассуждения. Устойчивые распределения и пример 3, е) показывают характер этих трудностей.

2а. Производные характеристических функций

Пусть F — распределение вероятностей с характеристической функцией φ . Как было показано в гл. XV, 4, если F имеет математическое ожидание μ , то φ имеет производную φ' и $\varphi'(0) = i\mu$. Обратное утверждение неверно. Дифференцируемость функции φ тесно связана с законом больших чисел для последовательности $\{X_n\}$ независимых случайных величин с общей функцией распределения F . Поэтому во многих работах изучались налагаемые на F условия, при которых существует φ' . Эта проблема была решена Питмэном в 1956 г. вслед за А. Зигмундом (1947), который получил частичный ответ, налагая на φ некоторые условия гладкости. Поскольку прямой подход к проблеме грозит большими трудностями, интересно узнать, что ее решение есть простое следствие последней теоремы 2.

Теорема. Каждое из трех следующих условий влечет за собой два других.

- (i) $\varphi'(0) = i\mu$.
- (ii) При $t \rightarrow \infty$

$$t[1 - F(t) + F(-t)] \rightarrow 0, \quad \int_{-t}^t xF\{dx\} \rightarrow \mu. \quad (2.36)$$

- (iii) Арифметическое среднее $(X_1 + \dots + X_n)/n$ сходится по вероятности к μ .

Доказательство. Так как действительная часть φ является четной функцией, то производная $\varphi'(0)$ обязательно будет чисто мнимой. Чтобы проследить связь между нашей предельной теоремой и соотношением $\varphi'(0) = i\mu$, последнее лучше записать в виде

$$t[\varphi(\zeta/t) - 1] \rightarrow i\mu\zeta, \quad t \rightarrow \infty. \quad (2.37)$$

Когда t пробегает последовательность $\{c_n\}$, то (2.37) становится частным случаем (2.31), где $\varphi_n(\zeta) = \varphi(\zeta/c_n)$ и $F_n(x) = F(c_n x)$. Стало быть, теорема 2 утверждает, что (2.37) справедливо тогда и только тогда, когда

$$tx^2 F\{tx\} \rightarrow 0, \quad t \int_{-\infty}^{+\infty} \sin x F\{tx\} dx \rightarrow \mu. \quad (2.38)$$

а) Предположим, что выполнено (2.36). Производя интегрирование по частям, получаем, что для произвольного $a > 0$

$$t \int_{-a}^a x^2 F\{tx\} dx \leq 4 \int_0^a tx[1 - F(tx) + F(-tx)] dx. \quad (2.39)$$

Подынтегральное выражение правой части при $t \rightarrow \infty$ стремится к нулю, следовательно, то же самое верно и для интеграла в левой части. Так как $|t \sin x/t - x| < Cx^2$, легко показать, что верно (2.38), а это влечет (2.37). Обратно, очевидно, что из (2.38) следует (2.36). Таким образом, (i) и (ii) эквивалентны.

б) Согласно теореме 1 из § 1, $\varphi^n(\zeta/n) \rightarrow e^{i\mu\zeta}$ тогда и только тогда, когда при $n \rightarrow \infty$

$$n[\varphi(\zeta/n) - 1] \rightarrow i\mu\zeta. \quad (2.40)$$

Другими словами, для применимости закона больших чисел необходимо и достаточно выполнение (2.37) при том условии, что t пробегает последовательность целых положительных чисел. Так как сходимость характеристических функций автоматически равномерная на конечных интервалах, то ясно, что из (2.40) следует (2.37) и поэтому условия (i) и (ii) эквивалентны.

То, что (2.36) представляет собой необходимое и достаточное условие для закона больших чисел (iii), было показано другими методами в теореме 1 гл. VII, 7.

§ 3. ПРИМЕРЫ И СПЕЦИАЛЬНЫЕ СВОЙСТВА

Мы укажем несколько специальных распределений, а затем рассмотрим такие свойства, как существование моментов или положительность. Эти свойства помещены в разделе «примеры» частично для того, чтобы сделать изложение более ясным, частично для того, чтобы подчеркнуть отсутствие связи между отдельными пунктами. Никакая часть материала этого параграфа не используется в дальнейшем. Дополнительные примеры можно найти в задачах 6, 7 и 19.

Примеры. а) *Нормальное распределение.* Если мера M сосредоточена в нуле и приписывает ему массу σ^2 , то (2.25) превращается в равенство $\psi(\zeta) = -\frac{1}{2}\sigma^2\zeta^2$ и e^ψ будет нормальным распределением с нулевым математическим ожиданием и дисперсией σ^2 .

б) *Распределение Пуассона.* Стандартное распределение Пуассона с математическим ожиданием α имеет характеристическую функцию $\omega = e^\psi$ с $\psi(\zeta) = \alpha(e^{i\zeta} - 1)$. Изменим параметры так, чтобы получить распределение, сосредоточенное в точках вида $-b + nh$. При этом показатель степени превращается в $\rho(\zeta) = \alpha(e^{i\zeta h} - 1) - ib\zeta$, что представляет собой частный случай (2.25) с мерой M , сосредоточенной в одной точке h . Свойство, состоящее в том, что мера M сосредоточена в одной точке, характеризует класс нормальных распределений и класс пуассоновских распределений с произвольными параметрами. Свертка конечного числа подобных распределений приводит к каноническим мерам с конечным числом атомов. В общем случае мера M может быть представлена как предел подобных мер. Поэтому все безгранично делимые распределения могут быть представлены как пределы свертки конечного числа распределений Пуассона с нормальными распределениями.

в) *Рандомизированное случайное блуждание.* В (7.7) гл. II мы встретились с семейством арифметических распределений, приписывающих точкам $r = 0, \pm 1, \pm 2, \dots$ вероятности

$$a_r(t) = \sqrt{\left(\frac{p}{q}\right)^r} e^{-t} I_r(2\sqrt{pq}t), \quad (3.1)$$

где параметры p, q, t положительны, $p+q=1$ и I_r — функция Бесселя, определенная в (7.1) гл. II. Из того факта, что распределение $\{a_r\}$ удовлетворяет уравнению Чепмена—Колмогорова, вытекает его безграничная делимость. Соответствующую характеристическую функцию $\omega = e^\psi$ легко вычислить, так как она отличается от разложения Шлёмилха [(7.8) гл. II] только заменой переменных $u = \sqrt{p/q}e^{-i\zeta}$. Окончательный результат

$$\psi(\zeta) = -t + t(pe^{i\zeta} + qe^{-i\zeta}) \quad (3.2)$$

показывает, что $\{a_r(t)\}$ можно рассматривать как *распределение разности двух независимых случайных величин, распределенных по закону Пуассона* с математическими ожиданиями pt и qt соответственно. Каноническая мера сосредоточена в точках ± 1 .

г) *Гамма-распределение*. Распределение с плотностью

$$g_t(x) = e^{-x} x^{t-1} / \Gamma(t)$$

при $x > 0$ имеет характеристическую функцию $\gamma_t(\zeta) = (1 - i\zeta)^{-t}$, которая, очевидно, безгранично делима. Чтобы представить ее в каноническом виде, заметим, что

$$(\log \gamma_t(\zeta))' = it(1 - i\zeta)^{-1} = it \int_0^\infty e^{i\zeta x - x} dx. \quad (3.3)$$

Проинтегрировав, получаем

$$\log \gamma_t(\zeta) = t \int_0^\infty \frac{e^{i\zeta x} - 1}{x} e^{-x} dx. \quad (3.4)$$

Таким образом, каноническая мера M задается плотностью $t x e^{-x}$, $x > 0$. При этом нет необходимости в центрирующем члене, так как интеграл сходится и без него.

д) *Плотность гиперболического косинуса*. В гл. XV, 2, было показано, что плотности $f(x) = \frac{1}{\pi \operatorname{ch} x}$ соответствует характеристическая функция $\omega(\zeta) = 1/\operatorname{ch}(\pi\zeta/2)$. Для того чтобы показать ее безграничную делимость, заметим, что $(\log \omega)'' = -(\pi^2/4)\omega^2$. Здесь ω^2 есть характеристическая функция плотности f^{2*} , которая была вычислена в задаче 6 из гл. II, 9. Таким образом,

$$\frac{d^2}{d\zeta^2} \log \omega(\zeta) = - \int_{-\infty}^{+\infty} e^{i\zeta x} \frac{x}{e^x - e^{-x}} dx. \quad (3.5)$$

Так как в нуле $(\log \omega)'$ обращается в нуль, то мы получаем

$$\log \omega(\zeta) = \int_{-\infty}^{+\infty} \frac{e^{i\zeta x} - 1 - i\zeta x}{x^2} \frac{x}{e^x - e^{-x}} dx. \quad (3.6)$$

Следовательно, каноническая мера имеет плотность $x/(e^x - e^{-x})$. По соображениям симметрии вклад слагаемого $i\zeta x$ в интеграл будет равен нулю, а так как интеграл сходится и без него, то $i\zeta x$ в числителе можно опустить.

е) *Пример П. Леви. Функция*

$$\psi(\zeta) = 2 \sum_{k=-\infty}^{+\infty} 2^{-k} [\cos 2^k \zeta - 1] \quad (3.7)$$

имеет вид (2.9) с симметричной мерой M , приписывающей массы 2^k точкам $\pm 2^k$, $k=0, \pm 1, \pm 2, \dots$. (Ряд сходится, так как $1 - \cos 2^k \zeta \sim 2^{2k-1} \zeta^2$ при $k \rightarrow -\infty$.) Характеристическая функция $\omega = e^\psi$ обладает любопытным свойством: $\omega^2(\zeta) = \omega(2\zeta)$, и, следовательно, $\omega^{2^k}(\zeta) = \omega(2^k \zeta)$. Эта функция была бы устойчивой (как это определено в гл. VI, 1), если бы при каждом n $\omega^n(\zeta) = \omega(a_n \zeta)$, но последнее соотношение выполняется только для $n=2, 4, 8, \dots$. В терминологии § 9 функция ω принадлежит своей собственной области *частичного притяжения*, но она не имеет области притяжения (см. задачу 10).

ж) *Односторонние устойчивые плотности.* Перейдем к вычислению характеристических функций, которые соответствуют сосредоточенным на $0, \infty$ каноническим мерам, таким, что

$$M\{0, x\} = Cx^{\alpha-1}, \quad 0 < \alpha < 2, \quad C > 0. \quad (3.8)$$

Этот пример очень важен, так как с его помощью мы получим общую форму устойчивых характеристических функций.

(i) При $0 < \alpha < 1$ рассмотрим характеристическую функцию $\omega_\alpha = e^{\psi_\alpha}$, где

$$\psi_\alpha(\zeta) = C(2-\alpha) \int_0^\infty \frac{e^{i\zeta x} - 1}{x^{\alpha+1}} dx. \quad (3.9)$$

По сравнению с каноническим видом (2.9) здесь отсутствует центрирующее слагаемое, которое несущественно, поскольку интеграл сходится и без него. Для вычисления интеграла в (3.9) мы предположим, что $\zeta > 0$, и рассмотрим его как предел при $\lambda \rightarrow 0+$ интегралов

$$\begin{aligned} \int_0^\infty \frac{e^{-(\lambda - i\zeta)x} - 1}{x^{\alpha+1}} dx &= \frac{1}{\alpha} (\lambda - i\zeta) \int_0^\infty e^{-(\lambda - i\zeta)x} x^{-\alpha} dx = \\ &= -\frac{1}{\alpha} \Gamma(1-\alpha) (\lambda - i\zeta)^\alpha \end{aligned} \quad (3.10)$$

(характеристическая функция гамма-распределения рассматривалась в гл. XV, 2). Заметим, что

$$(\lambda - i\zeta)^\alpha = (\lambda^2 + \zeta^2)^{\alpha/2} e^{i\theta\alpha},$$

где θ — аргумент $\lambda - i\zeta$, т. е. $\operatorname{tg} \theta = -\zeta/\lambda$. Очевидно, что $\theta \rightarrow -\pi/2$ при $\lambda \rightarrow 0+$ и, следовательно, $(\lambda - i\zeta)^\alpha \rightarrow \zeta^\alpha e^{-i\alpha\pi/2}$. Окончательный результат мы запишем в виде

$$\psi_\alpha(\zeta) = \zeta^\alpha \cdot C \cdot \frac{\Gamma(3-\alpha)}{(\alpha-1)\alpha} e^{-i\pi\alpha/2}, \quad \zeta > 0. \quad (3.11)$$

Для $\zeta < 0$ вычислим $\psi_\alpha(\zeta)$ как функцию, сопряженную к $\psi_\alpha(-\zeta)$.

(ii) При $1 < \alpha < 2$ положим

$$\psi_\alpha(\zeta) = C \int_0^\infty \frac{e^{i\zeta x} - 1 - i\zeta x}{x^{\alpha+1}} dx. \quad (3.12)$$

Отличие от канонического вида (2.9) состоит в более подходящем центрировании при котором математические ожидания становятся равными нулю. Интегрирование по частям приводит к понижению показателя степени в знаменателе и позволяет нам воспользоваться предыдущим результатом. Обычные вычисления показывают, что ψ_α опять задается выражением (3.11). (Действительная часть снова отрицательна, так как теперь $\cos \alpha\pi/2 < 0$.)

(iii) При $\alpha = 1$ используем канонический вид

$$\psi_1(\zeta) = C \int_0^\infty \frac{e^{i\zeta x} - 1 - i\zeta \sin x}{x^2} dx. \quad (3.13)$$

Из гл. XV, 2, известно, что $(1 - \cos x)/(\pi x^2)$ есть плотность вероятности и, следовательно, действительная часть $\psi_1(\zeta)$ равна $-\frac{1}{2}\pi\zeta$. Для мнимой части получаем

$$\int_0^\infty \frac{\sin \zeta x - \zeta \sin x}{x^2} dx = \lim_{\varepsilon \rightarrow 0} \left[\int_\varepsilon^\infty \frac{\sin \zeta x}{x^2} dx - \zeta \int_\varepsilon^\infty \frac{\sin x}{x^2} dx \right]. \quad (3.14)$$

Когда $\zeta > 0$, замена переменных $\zeta x = y$ приводит первый интеграл к тому же виду, что и второй. В целом получаем

$$-\zeta \lim_{\varepsilon \rightarrow 0} \int_\varepsilon^\zeta \frac{\sin x}{x^2} dx = -\zeta \lim_{\varepsilon \rightarrow 0} \int_1^\zeta \frac{\sin ey}{ey} \cdot \frac{dy}{y} = -\zeta \log \zeta. \quad (3.15)$$

Окончательно

$$\psi_1(\zeta) = C \left(-\frac{1}{2}\pi\zeta - i\zeta \log \zeta \right), \quad \zeta > 0. \quad (3.16)$$

Естественно, что функция $\psi_1(-\zeta)$ сопряжена с $\psi_1(\zeta)$.

При $\alpha \neq 1$ характеристическая функция $\omega = e^{\psi_\alpha}$ обладает тем свойством, что $\omega^n(\zeta) = \omega(n^{1/\alpha}\zeta)$. Это в соответствии с определением гл. VI, 1, означает, что ω строго устойчива, т. е. сумма независимых случайных величин $X_1, \dots, X_n$ с характеристической функцией ω имеет такое же распределение, что и $n^{1/\alpha}X_1$.

Если $\alpha = 1$, то $\omega^n(\zeta) = \omega(n\zeta) e^{-i\zeta \log n}$, и, следовательно, распределение суммы отличается от распределения $n^{1/\alpha} X_1$ центрированием. Таким образом, характеристическая функция ψ_1 *устойчива* в широком смысле. [Свойства и примеры устойчивых распределений даны в гл. XV, 1, 2. Дополнительные свойства будут получены в § 5. В § 4 (см. пример 4, в)) мы покажем, что при $\alpha < 1$ распределение сосредоточено на положительной полуоси. При $\alpha \neq 1$ это неверно.]

з) *Произвольные устойчивые плотности.* Каждой функции ψ_k предыдущего примера можно поставить в соответствие аналогичную характеристическую функцию, которая порождается канонической мерой с точно такой же, но сосредоточенной на отрицательной полуоси плотностью. Чтобы получить такие характеристические функции, мы должны только заменить в наших формулах i на $-i$. Можно получить более общие устойчивые характеристические функции, составляя линейные комбинации из характеристических функций, относящихся к двум экстремальным случаям, т. е. используя каноническую меру M , такую, что при $x > 0$

$$M\{\overline{0, x}\} = Cpx^{2-\alpha}, \quad M\{\overline{-x, 0}\} = Cqx^{2-\alpha}, \quad (3.17)$$

где $p \geq 0$, $q \geq 0$ и $p+q=1$. Из сказанного ясно, что соответствующая характеристическая функция $\omega = e^\psi$ задается выражением

$$\psi(\zeta) = |\zeta|^\alpha C \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)} \left[\cos \frac{\pi\alpha}{2} \pm i(p-q) \sin \frac{\pi\alpha}{2} \right], \quad (3.18)$$

если $0 < \alpha < 1$ или $1 < \alpha \leq 2$, и выражением

$$\psi(\zeta) = -|\zeta| \cdot C \left[\frac{1}{2} \pi \pm i(p-q) \log |\zeta| \right],$$

если $\alpha = 1$ (здесь везде берется верхний знак для $\zeta > 0$ и нижний для $\zeta < 0$). Заметим, что при $\alpha = 2$ мы получаем $\psi(\zeta) = -\frac{1}{2} \zeta^2$, т. е. *нормальное распределение*. Это соответствует мере M , сосредоточенной в нуле.

В § 5 будет показано, что (с точностью до произвольного центрирования) эти формулы задают *самый общий вид устойчивых характеристических функций*. В частности, все симметричные устойчивые распределения имеют характеристические функции вида $e^{-\alpha|\zeta|^\alpha}$, $\alpha > 0$. ►

§ 4. СПЕЦИАЛЬНЫЕ СВОЙСТВА

Везде в этом параграфе $\omega = e^\psi$ обозначает безгранично делимую характеристическую функцию с ψ , имеющей стандартный вид

$$\psi(\zeta) = \int_{-\infty}^{+\infty} \frac{e^{i\zeta x} - 1 - i\zeta \sin x}{x^2} M\{dx\} + i b \zeta, \quad (4.1)$$

где M — каноническая мера, а b — действительное число. В соответствии с определением канонической меры интеграл

$$M^+(x) = \int_x^{\infty} y^{-2} M \{dy\} \quad (4.2)$$

сходится при всех $x > 0$, и аналогичное утверждение имеет место для $M^-(x)$ при $x < 0$.

Будем обозначать через U вероятностное распределение с характеристической функцией ω .

а) *Существование моментов.* В гл. XV, 4, было показано, что второй момент U конечен тогда и только тогда, когда ω дважды дифференцируема, т. е. когда существует ψ'' . Точно так же доказывается, что это имеет место тогда и только тогда, когда мера M конечна. Другими словами, для существования второго момента U необходимо и достаточно, чтобы мера M была конечна.

Аналогичное рассуждение (см. задачу 15 из гл. XV, 9) позволяет доказать более общее утверждение: для любого целого $k \geq 1$ $2k$ -й момент распределения U существует тогда и только тогда, когда M имеет момент порядка $2k-2$.

б) *Разложения.* Каждое представление меры M в виде суммы $M = M_1 + M_2$ двух мер порождает разложение $\omega = e^{\psi_1} e^{\psi_2}$ характеристической функции ω на два множителя, каждый из которых является безгранично делимой характеристической функцией. Если мера M сосредоточена в одной точке, то же самое справедливо для M_1 и M_2 . Другими словами, если ω — характеристическая функция нормального распределения или распределения Пуассона¹⁾, то же самое будет верно и для множителей e^{ψ_1} и e^{ψ_2} . Однако любая другая безгранично делимая ω может быть представлена произведением двух совершенно различных компонент. В частности, любая отличная от нормальной устойчивая характеристическая функция может быть разложена в произведение не являющихся устойчивыми безгранично делимых характеристических функций.

Особенно полезно разложение $\omega = e^{\psi_1} e^{\psi_2}$, получаемое представлением M как суммы двух мер, сосредоточенных на интервалах $|x| \leq \eta$ и $|x| > \eta$ соответственно. Для второго интервала выразим M через меру N , определяемую соотношением $N\{dx\} = x^{-2} M\{dx\}$. Тогда мы запишем

$$\psi(\xi) = \psi_1(\xi) + \psi_2(\xi) + i\beta\xi, \quad (4.3)$$

$$\psi_1(\xi) = \int_{|x| \leq \eta} \frac{e^{i\xi x} - 1 - i\xi x}{x^2} M\{dx\}, \quad (4.4)$$

¹⁾ По теореме 1 гл. XV, 8, для характеристической функции нормального распределения не допускается произвольное разложение через произведение характеристических функций, не являющихся нормальными. Аналогичное утверждение справедливо и для распределения Пуассона (теорема Райкова).

$$\psi_2(\zeta) = \int_{|x| > \eta} (e^{i\zeta x} - 1) N \{dx\}, \quad (4.5)$$

а разность $b - \beta$ образуется за счет изменения центрирующих слагаемых в (4.4) и (4.5).

Заметим, что e^{ψ_2} является характеристической функцией обобщенного распределения Пуассона, порожденного таким распределением вероятностей F , что $F \{dx\} = cN \{dx\}$ или

$$1 - F(x) = cM^+(x), \quad x > 0. \quad (4.6)$$

Характеристическая функция e^{ψ_2} бесконечно дифференцируема. Таким образом, *каждое безгранично делимое распределение U представимо в виде свертки распределения U_1 , имеющего конечные моменты всех порядков, и обобщенного распределения Пуассона U_2 , порожденного распределением вероятностей F , хвосты которого пропорциональны M^+ и M^- . Отсюда, в частности, следует, что U имеет момент порядка k тогда и только тогда, когда существует момент порядка k распределения F .*

в) *Положительные случайные величины.* Мы покажем, что распределение U сосредоточено на $0, \infty$ тогда и только тогда, когда ¹⁾

$$\psi(\zeta) = \int_0^\infty \frac{e^{i\zeta x} - 1}{x} P \{dx\} + ib\zeta, \quad (4.7)$$

где $b \geq 0$ и P — такая мера, что функция $(1+x)^{-1}$ интегрируема относительно P . (В соответствии с (4.1) имеем $P \{dx\} = = x^{-1}M \{dx\}$.)

Предположим, что мера U сосредоточена на $0, \infty$, и рассмотрим разложение, описываемое соотношениями (4.3) — (4.5). Для обобщенного распределения Пуассона U_2 нуль является точкой роста. Распределение U_1 имеет нулевое математическое ожидание и, следовательно, некоторую точку роста $s \leq 0$. Отсюда следует, что $s + \beta$ — точка роста U и, следовательно, $\beta \geq 0$. Аналогичное рассуждение показывает, что U_1 не может иметь нормальной компоненты, и поэтому вклад U_1 в $\psi(\zeta)$ должен стремиться к нулю при $\eta \rightarrow 0$. Наконец, если t — точка роста распределения вероятностей F , порождающего U_2 , тогда nt есть точка роста самого распределения U_2 . Поэтому F и, следовательно, N сосредоточены на положительной полуоси. Таким образом, интеграл в (4.5) на самом деле сводится к интегралу по области $x > \eta$. Подынтегральная функция в нуле обращается в нуль, и поэтому при $\eta \rightarrow 0$ мера N не обязана быть ограни-

¹⁾ Это было отмечено П. Леви и непосредственно следует из варианта преобразования Лапласа (7.2) гл. XIII. Интересно отметить, что формальная проверка утверждения без вероятностных рассуждений достаточно громоздка (см. Бакстер и Шапиро, Sannhya, vol. 22).

ченной. Однако при $x > \eta$ можно перейти от меры N к мере $P\{dx\} = xN\{dx\}$ (что совпадает с $x^{-1}M\{dx\}$). При этом новое подынтегральное выражение $(e^{ix} - 1)x^{-1}$ при приближении x к 0 само к нулю не стремится, и, следовательно, мера P должна приписывать конечные значения различным окрестностям нуля. Таким путем мы получим представление (4.7).

Обратно, если ψ определено формулой (4.7), то проведенные нами рассуждения показывают, что e^ψ будет пределом характеристических функций обобщенных распределений Пуассона, сосредоточенных на $0, \infty$. Следовательно, то же самое верно для предельного распределения U .

г) *Асимптотическое поведение.* Полученное нами утверждение относительно существования моментов, по-видимому, означает, что асимптотическое поведение функции распределения U при $x \rightarrow \pm \infty$ зависит только от поведения канонической меры M вблизи $\pm \infty$ или, что равносильно, от асимптотического поведения функций M^+ и M^- . Прежде чем приступить к доказательству этого утверждения в наиболее общей форме, мы рассмотрим типичный случай.

Предположим, что M^+ правильно меняется на бесконечности, т. е.

$$M^+(x) = x^{-\xi} L(x), \quad (4.8)$$

где $\xi \geq 0$ и L — медленно меняющаяся функция. Тогда

$$1 - U(x) \sim M^+(x), \quad x \rightarrow \infty. \quad (4.9)$$

Доказательство. Пусть S — случайная величина с функцией распределения U . Рассмотрим каноническую меру M как сумму $M_1 + M_2 + M_3$ трех мер, сосредоточенных на интервалах $\overline{1, \infty}$, $\overline{-1, 1}$ и $\overline{-\infty, -1}$ соответственно. Как показано в п. б), это порождает представление $S = X_1 + X_2 + X_3 + \beta$ в виде суммы трех независимых случайных величин, таких, что X_1 имеет обобщенное распределение Пуассона U_1 , которое порождено распределением F , сосредоточенным на $\overline{1, \infty}$ и определенным формулой (4.6); каноническая мера, соответствующая X_2 , сосредоточена на $\overline{-1, 1}$; наконец, X_3 определяется так же, как и X_1 , но только интервал $\overline{1, \infty}$ заменяется теперь на $\overline{-\infty, -1}$. Нетрудно показать, что

$$P\{X_1 > x\} \sim M^+(x), \quad x \rightarrow \infty \quad (4.10)$$

(см. теорему 2 из гл. VIII, 9). Поэтому, чтобы доказать утверждение (4.9), достаточно показать, что

$$P\{S > x\} \sim P\{X_1 > x\}, \quad x \rightarrow \infty. \quad (4.11)$$

В этой связи центрирующая постоянная β не играет никакой роли, и поэтому мы полагаем $\beta = 0$. Тогда для любого $\varepsilon > 0$

$$P\{S > x\} \geq P\{X_1 > (1 + \varepsilon)x\} \cdot P\{X_2 + X_3 > -\varepsilon x\}. \quad (4.12)$$

С другой стороны, поскольку $X_2 \leq 0$,

$$P\{S > x\} \leq P\{X_1 > (1-\varepsilon)x\} + P\{X_2 > \varepsilon x\}. \quad (4.13)$$

При $x \rightarrow \infty$ последняя вероятность в (4.12) стремится к 1, тогда как последняя вероятность в (4.13) убывает быстрее, чем любая степень $x^{-\alpha}$, так как X_2 имеет моменты всех порядков. Таким образом, (4.12) и (4.13) подтверждают справедливость (4.11).

д) *Подчиненность*. Если e^ψ безгранично делима, то при любом $s > 0$ $e^{s\psi}$ также безгранично делима. Каково бы ни было сосредоточенное на $0, \infty$ распределение G , можно, рандомизируя параметр s , построить новую характеристическую функцию

$$\varphi(s) = \int_0^\infty e^{s\psi(t)} G\{ds\}. \quad (4.14)$$

Характеристическая функция φ не обязана быть безгранично делимой, однако легко проверить, что если $G = G_1 \star G_2$ — свертка двух вероятностных распределений, то (обозначения очевидны) $\varphi = \varphi_1 \varphi_2$. Отсюда следует, что *если распределение G безгранично делимо, то (4.14) определяет безгранично делимую характеристическую функцию*.

Этот результат допускает простую вероятностную интерпретацию. Пусть $\{X(t)\}$ — случайный процесс с независимыми приращениями, причем случайная величина $X(t)$ имеет характеристическую функцию $e^{t\psi}$. Если T — положительная случайная величина с распределением G , то φ из (4.14) можно интерпретировать как характеристическую функцию сложной случайной величины $X(T)$. Теперь предположим, что распределение G безгранично делимо с характеристической функцией e^γ . Мы можем представить себе другой процесс $\{T(t)\}$ с независимыми приращениями, такой, что $e^{t\gamma}$ будет характеристической функцией $T(t)$. При каждом $t > 0$ мы получим новые случайные величины $X(T(t))$, которые снова являются значениями процесса с независимыми приращениями¹⁾. Таким образом, случайная величина $T(t)$ служит *операционным временем*. В терминологии гл. X, 7, новый процесс $\{X(T(t))\}$ получается *подчинением процессу $\{X(t)\}$ с направляющим процессом $\{T(t)\}$* . Мы нашли чисто аналитическое доказательство того, что процедура подчинения всегда приводит к безгранично делимым распределениям.

¹⁾ Если G имеет преобразование Лапласа $e^{-\rho(\lambda)}$, то $T(t)$ соответствует преобразованию Лапласа $e^{-t\rho(\lambda)}$, и легко проверить, что характеристическая функция $X(T(t))$ имеет вид $e^{-t\rho(-\psi)}$.

§ 5. УСТОЙЧИВЫЕ РАСПРЕДЕЛЕНИЯ И ИХ ОБЛАСТИ ПРИТЯЖЕНИЯ

Пусть $\{X_n\}$ — последовательность взаимно независимых случайных величин с общим распределением F . Положим $S_n = X_1 + \dots + X_n$. Пусть U — некоторое распределение, не сосредоточенное в одной точке. В соответствии с терминологией гл. VI, 1, мы говорим, что F принадлежит области притяжения распределения U , если существуют такие постоянные $a_n > 0$ и b_n , что распределение случайной величины $a_n^{-1} S_n - nb_n$ сходится к U . Исключение предельных, служит цели устранить тривиальный результат, когда $b_n \rightarrow b$, а a_n возрастают настолько быстро, что $a_n^{-1} S_n$ сходится по вероятности к нулю.

Мы хотим перефразировать определение в терминах характеристических функций φ и ω распределений F и U соответственно.

Согласно лемме 4 гл. XV, 1, распределение U сосредоточено в одной точке тогда и только тогда, когда $|\omega(\xi)| = 1$ для всех ξ . Таким образом, φ принадлежит области притяжения характеристической функции ω , если $|\omega|$ не равно тождественно единице и существуют такие постоянные $a_n > 0$ и b_n , что

$$(\varphi(\xi/a_n) e^{-ib_n \xi})^n \rightarrow \omega(\xi). \quad (5.1)$$

В гл. VI, 1, было показано, что предельная функция ω необходимо устойчива. Однако мы построим всю теорию заново, как простое следствие основной предельной теоремы § 2. Для согласия с принятыми там обозначениями положим

$$\varphi_n(\xi) = \varphi(\xi/a_n) e^{-ib_n \xi}, \quad F_n(x) = F(a_n(x + b_n)). \quad (5.2)$$

Согласно теореме 1 из § 1, соотношение (5.1) выполняется тогда и только тогда, когда для всех ξ

$$n[\varphi_n(\xi) - 1] \rightarrow \psi(\xi), \quad (5.3)$$

где $\omega = e^\psi$.

Рассмотрим сначала случай симметричного распределения F . Тогда $b_n = 0$. Из теоремы 1 § 2 мы знаем, что из (5.3) следует существование такой канонической меры M , что $nx^2 F_n\{dx\} \rightarrow M\{dx\}$. Чтобы выразить это, мы введем урезанные моменты

$$\mu(x) = \int_{-x}^x y^2 F\{dy\}, \quad x > 0. \quad (5.4)$$

Тогда во всех точках непрерывности x

$$\frac{n}{a_n^2} \mu(a_n x) \rightarrow M\{\overline{-x, x}\} \quad (5.5)$$

и

$$n[1 - F(a_n x)] \rightarrow M^+(x), \quad (5.6)$$

где

$$M^+(x) = \int_x^\infty y^{-2} M \{dy\}. \quad (5.7)$$

Из соотношения $\varphi(\xi/a_n) \rightarrow 1$ следует, что $a_n \rightarrow \infty$, и, следовательно, S_n/a_n и S_n/a_{n+1} имеют одинаковое предельное распределение U . Отсюда заключаем, что отношение $\frac{a_{n+1}}{a_n}$ стремится к 1, и, следовательно, лемма 3 из гл. VIII, 8, применима к (5.5). Мы приходим к выводу, что μ *меняется правильно*, а каноническая мера M имеет вид

$$M\{\overline{-x, x}\} = Cx^{2-\alpha}, \quad x > 0, \quad (5.8)$$

где $\alpha \leq 2$ (обозначение $2-\alpha$ показателя степени соответствует введенному П. Леви). Если $\alpha = 2$, то мера M сосредоточена в нуле. Для сходимости интеграла (5.7) нужно, чтобы $\alpha > 0$. При $0 < \alpha < 2$ мы находим

$$M^+(x) = C \frac{2-\alpha}{\alpha} x^{-\alpha}, \quad x > 0. \quad (5.9)$$

Аналогичное рассуждение применимо и к несимметричному распределению F , но вместо (5.6) мы получим менее привлекательные соотношения

$$n[1 - F(a_n(x + b_n))] \rightarrow M^+(x), \quad nF(a_n(-x + b_n)) \rightarrow M^-(-x); \quad (5.10)$$

подобное изменение касается и (5.5). Тем не менее из $\varphi_n(\epsilon) \rightarrow 1$ и $a_n \rightarrow \infty$ следует $b_n \rightarrow 0$, и поэтому (5.10) и в самом деле полностью эквивалентно (5.6) и аналогичному соотношению для левого хвоста.

Таким образом, мы видим, что (5.6) выполняется всякий раз, когда F принадлежит области притяжения. Ввиду леммы 3 гл. VIII, 8, это означает, что или M^+ тождественно равно нулю, или же хвост $1 - F$ *правильно меняется* и $M^+(x) = Ax^{-\alpha}$. Тогда (5.7) показывает, что на положительной полуоси мера M имеет плотность $A\alpha x^{1-\alpha}$. То же самое рассуждение применимо для левого хвоста, а также для суммы хвостов. Следовательно, показатель α должен быть общим для того и другого хвоста.

Если оба хвоста тождественно равны нулю, то мера M сосредоточена в нуле. Ни в каком другом случае у меры M не может быть атома в нуле. Это подтверждается тем, что каноническая мера ${}^0M$, соответствующая симметризованному распределению ${}^0U$, представляет собой сумму M и ее зеркального отражения относительно нуля, а мы видели, что ${}^0M$ либо вовсе не имеет атомов, либо сосредоточена в нуле. Таким образом, при $\alpha < 2$ каноническая мера M однозначно определяется своими плотностями на

каждой из полуосей, и эти плотности пропорциональны $|x|^{1-\alpha}$. Следовательно, для интервалов $-y, x$, содержащих нуль, мы имеем

$$M\{-y, x\} = C(qy^{2-\alpha} + px^{2-\alpha}), \quad (5.11)$$

причем в последнем равенстве можно считать, что $0 < \alpha \leq 2$ (при $\alpha = 2$ мера M сосредоточена в нуле); кроме того, $C > 0$ и $p + q = 1$. В силу (5.7) все это равносильно

$$M^+(x) = Cp \frac{2-\alpha}{\alpha} x^{-\alpha}, \quad M^-(-x) = Cq \frac{2-\alpha}{\alpha} x^{-\alpha}. \quad (5.12)$$

Соответствующие этим мерам характеристические функции определяются формулами (3.18) и (3.19). Они ясно показывают, что наши распределения *устойчивы* в том смысле, что распределение U^{n*} отличается от распределения U только параметрами сдвига. Это означает, что каждое устойчивое распределение принадлежит своей собственной области притяжения. Таким образом, мы решили задачу отыскания *всех* распределений, обладающих областью притяжения. Суммируем наши заключения в виде теоремы.

Теорема 1. *Распределение обладает областью притяжения тогда и только тогда, когда устойчиво.*

(i) *Класс устойчивых распределений совпадает с классом безгранично делимых распределений, канонические меры которых заданы в (5.11).*

(ii) *Соответствующие характеристические функции имеют вид $\phi(\xi) = e^{\psi(\xi) + i b \xi}$, где ψ определяется формулами (3.18), (3.19) с $0 < \alpha \leq 2$.*

(iii) *При $x \rightarrow \infty$ хвосты соответствующего распределения U удовлетворяют соотношениям*

$$x^\alpha [1 - U(x)] \rightarrow Cp \frac{2-\alpha}{\alpha}, \quad x^\alpha U(-x) \rightarrow Cq \frac{2-\alpha}{\alpha}. \quad (5.13)$$

Если вспомнить, что U принадлежит своей собственной области притяжения с нормирующими постоянными $a_n = n^{1/\alpha}$, то последнее утверждение теоремы является прямым следствием (5.6). [С другой стороны, (5.13) представляет собой частный случай результата, полученного в 4, г.)]

Заметим, что каждая из трех характеристик, указанных в теореме, определяет распределение U однозначно с точностью до произвольного сдвига.

Прежде чем вернуться к исследованию условий, при которых распределение F принадлежит области притяжения устойчивого распределения, мы вспомним основные факты о правильном изменении функций. В согласии с определением гл. VIII, 8, функция L меняется *медленно* на бесконечности, если для каждого фиксированного $x > 0$

$$\frac{L(tx)}{L(t)} \rightarrow 1, \quad t \rightarrow \infty. \quad (5.14)$$

В этом случае для произвольного $\delta > 0$ и всех достаточно больших x мы имеем

$$x^{-\delta} < L(x) < x^{\delta}. \quad (5.15)$$

Функция μ изменяется правильно, если она имеет вид $\mu(x) = x^{\alpha} L(x)$. Мы рассматриваем, в частности, функцию $\mu(x)$ — усеченный момент (5.4). Применим теорему 2 гл. VIII, 9, с $\xi = 2$, $\eta = 0$ к функции распределения $F(x) - F(-x)$, сосредоточенной на $0, \infty$, и получим следующий важный результат.

Если μ изменяется правильно с показателем $2 - \alpha$ (где $0 < \alpha \leq 2$), то

$$\frac{x^2 [1 - F(x) + F(-x)]}{\mu(x)} \rightarrow \frac{2 - \alpha}{\alpha}. \quad (5.16)$$

Обратно, если (5.16) справедливо при $\alpha < 2$, то $\mu(x)$ и сумма хвостов

$$1 - F(x) + F(-x)$$

изменяются правильно с показателями $2 - \alpha$ и $-\alpha$ соответственно. Если (5.16) выполнено при $\alpha = 2$, то μ изменяется медленно.

При выводе (5.8) мы видели, что для того, чтобы симметричное распределение F принадлежало некоторой области притяжения, необходимо, чтобы функция μ изменялась правильно:

$$\mu(x) \sim x^{2-\alpha} L(x), \text{ при } x \rightarrow \infty, \quad (5.17)$$

где L — медленно меняющаяся функция. Мы сейчас увидим, что это верно и для несимметричных распределений. При $\alpha = 2$ это условие оказывается также и достаточным. Однако при $\alpha < 2$ каноническая мера (5.11) приписывает положительной и отрицательной полуосям массы в отношении $p:q$ и оказывается, что оба хвоста распределения F должны быть сбалансированы подобным же образом.

Мы теперь в состоянии доказать основную теорему.

Теорема 2. а) Для того чтобы распределение F принадлежало некоторой области притяжения, необходимо, чтобы функция (урезанный момент) μ менялась правильно с показателем $2 - \alpha$, $0 < \alpha \leq 2$ (т. е. выполнялось (5.17)).

б) При $\alpha = 2$ это условие также достаточно, если только F не сосредоточено в одной точке.

в) Если при $0 < \alpha < 2$ выполняется (5.17), то распределение F принадлежит некоторой области притяжения в том и только том случае, когда хвосты F сбалансированы так, что при $x \rightarrow \infty$

$$\frac{1 - F(x)}{1 - F(x) + F(-x)} \rightarrow p, \quad \frac{F(-x)}{1 - F(x) + F(-x)} \rightarrow q. \quad (5.18)$$

Отметим, что никаких предположений относительно центрирования F не делается. Таким образом, теорема приводит к выводу,

что (5.17) либо верно при любом центрировании, либо не выполняется ни при каком. Справедливость этого заключения легко установить прямой проверкой. Исключением является лишь тот случай, когда F сосредоточена в единственной точке t ; при этом левая часть соотношения (5.17) тождественно равна нулю при центрировании величиной t и меняется правильно при всех других способах центрирования.

Теорема была сформулирована так, чтобы охватить и сходимость к нормальным распределениям. При $\alpha < 2$, по-видимому, более естественно выразить основное условие через сумму хвостов распределения F . В приводимых ниже следствиях заключены равносильные варианты теоремы 2.

Следствие 1. *Распределение F , не сосредоточенное в одной точке, принадлежит области притяжения нормального распределения тогда и только тогда, когда μ — медленно меняющаяся функция.*

Это имеет место в том и только том случае, когда условие (5.16) выполнено при $\alpha = 2$.

Излишне говорить, что функция μ меняется медленно, если распределение F имеет дисперсию.

Следствие 2. *Распределение F принадлежит области притяжения устойчивого распределения с показателем $\alpha < 2$ тогда и только тогда, когда хвосты F удовлетворяют условию (5.18), а сумма хвостов правильно меняется с показателем α .*

Последнее условие полностью равносильно (5.16).

Доказательство. а) *Необходимость.* Предположим, что для предельного распределения U каноническая мера задается равенством (5.11). При выводе этого соотношения мы убедились, что распределение, принадлежащее области притяжения U , удовлетворяет условию (5.6) и аналогичному условию для левого хвоста. Поэтому

$$n[1 - F(a_n x) + F(-a_n x)] \rightarrow M^+(x) + M^-(-x). \quad (5.19)$$

Сначала предположим, что $\alpha < 2$, поэтому правая часть не равна тождественно нулю. Уже говорилось о том, что лемма 3 гл. VIII, 8, гарантирует в этом случае правильное изменение с показателем $-\alpha$ суммы хвостов $1 - F(x) + F(-x)$. Тогда выполняется (5.16) и поэтому μ меняется правильно с показателем $2 - \alpha$. Условие (5.18) теперь немедленно следует из (5.6).

Остается рассмотреть случай $\alpha = 2$. Левая часть (5.19) в этом случае стремится к нулю. Поэтому вероятность того, что $|X_k| > a_n$ при некотором $k \leq n$, стремится к нулю. Отсюда нетрудно вывести следующее утверждение: если S_n/n не стремится по вероятности к нулю, то и сумма урезанных вторых моментов величин

$X_k a_n^{-1}$ не стремится к нулю. Но

$$\frac{\mu(a_n)}{1 - F(a_n) + F(-a_n)} \rightarrow \infty, \quad (5.20)$$

и, следовательно, при $\alpha = 2$ (5.16) выполняется. Отсюда следует медленное изменение функции μ , и поэтому наши условия необходимы.

б) *Достаточность.* Мы не только докажем, что наши условия являются достаточными, но одновременно определим нормирующие постоянные a_n и b_n , которые будут обеспечивать сходимость к предписанному устойчивому распределению. Это сделано в теореме 3. ►

Формулировка теоремы 3 предполагает известным то, что распределения, принадлежащие любой области притяжения при $\alpha > 1$, имеют математическое ожидание. При доказательстве нам потребуются дополнительная информация относительно урезанных первых моментов. Естественно сформулировать эти результаты в обстановке наибольшей общности, хотя нам будет нужен только частный случай $\beta = 1$.

Лемма. *Распределение F , принадлежащее некоторой области притяжения с показателем α , обладает абсолютными моментами t_β всех порядков β , $\beta < \alpha$. При $\alpha < 2$ не существуют моменты порядка $\beta > \alpha$.*

Более точно, если $\beta < \alpha$, то при $t \rightarrow \infty$

$$\frac{t^{2-\beta}}{\mu(t)} \int_{|x|>t} |x|^\beta F\{dx\} \sim \frac{2-\alpha}{\alpha-\beta}, \quad (5.21)$$

тогда как при $\alpha < 2$ и $\beta > \alpha$

$$\int_{|x|<t} |x|^\beta F\{dx\} \sim \frac{\alpha}{\beta-\alpha} t^\beta [1 - F(t) + F(-t)]. \quad (5.22)$$

(Заметим, что в каждом из этих случаев интеграл является правильно изменяющейся функцией с показателем $\beta - \alpha$.)

Доказательство. Соотношения (5.21) и (5.22) обобщают (5.16) и являются прямым следствием теоремы 2 гл. VIII, 9, если применить ее к распределению, заданному функцией $F(x) + F(-x)$ на $0, \infty$ (для вывода достаточно положить $\xi = 2$, $\eta = \beta$ и $\zeta = 2$, $\eta = 0$). ►

При доказательстве теоремы 2 подразумевалось, что нормирующие постоянные a_n должны удовлетворять условию

$$\frac{n\mu(a_n)}{a_n^2} \rightarrow C. \quad (5.23)$$

Если μ меняется правильно [удовлетворяет (5.17)], такие a_n су-

ществуют. Можно определить a_n как нижнюю границу всех x , для которых $nx^{-2}\mu(x) \leq C$. Вследствие правильного регулярного изменения μ мы имеем при $x > 0$

$$\frac{n\mu(a_n x)}{a_n^2} \rightarrow Cx^{2-\alpha}. \quad (5.24)$$

Это означает, что масса, которую мера $nx^2 F\{a_n dx\}$ приписывает любому симметричному интервалу $-x, x$, стремится к $M\{-x, x\}$. Ввиду (5.16) соотношение (5.24) автоматически влечет за собой соотношение (5.19) для сумм хвостов распределения F . При $\alpha = 2$ правая часть (5.19) тождественно равна нулю. При $\alpha < 2$ условие (5.18) обеспечивает подчинение каждого отдельного хвоста условиям

$$n[1 - F(a_n x)] \rightarrow Cp \frac{2-\alpha}{\alpha} x^{-\alpha}, \quad nF(-a_n x) \rightarrow Cq \frac{2-\alpha}{\alpha} x^{-\alpha}, \quad (5.25)$$

где правые части совпадают соответственно с $M^+(x)$ и $M^-(x)$. [Между прочим, при $\alpha < 2$ из соотношений (5.25) в свою очередь вытекает (5.24).]

Таким образом, мы показали, что меры $nx^2 F\{a_n dx\}$ сходятся в собственном смысле к канонической мере M . По теореме 2 из § 2 имеем тогда

$$\int_{-\infty}^{\infty} \frac{e^{i\zeta x} - 1 - i\zeta \sin x}{x^2} nx^2 F\{a_n dx\} \rightarrow \int_{-\infty}^{\infty} \frac{e^{i\zeta x} - 1 - i\zeta \sin x}{x^2} M\{dx\}. \quad (5.26)$$

Теперь легко получить теорему 3.

Теорема 3. Пусть устойчивое распределение U определяется (включая и центрирование) характеристической функцией (3.18), если $\alpha \neq 1$, или (3.19), если $\alpha = 1$.

Пусть распределение F удовлетворяет условиям теоремы 2, а последовательность $\{a_n\}$ — соотношению (5.23).

- (i) Если $0 < \alpha < 1$, то $\varphi^n(\zeta/a_n) \rightarrow \omega(\zeta) = e^{\psi(\zeta)}$.
- (ii) Если $1 < \alpha \leq 2$, то справедливо то же самое при условии, что распределение F имеет нулевое математическое ожидание.
- (iii) Если $\alpha = 1$, то

$$(\varphi(\zeta/a_n) e^{-ib_n \zeta})^n \rightarrow \omega(\zeta) = e^{\psi(\zeta)}, \quad (5.27)$$

где

$$b_n = \int_{-\infty}^{\infty} \sin \frac{x}{a_n} F\{dx\}. \quad (5.28)$$

Таким образом, мы имеем приятный результат, который состоит в том, что при $\alpha < 1$ вовсе не требуется центрирование,

тогда как при $\alpha > 1$ нам достаточно естественное центрирование математическим ожиданием.

Доказательство. (i) Пусть $\alpha < 1$. В стандартном интеграле для $\psi(\xi)$, который приводит к (3.18), отсутствует слагаемое $i\xi \sin x$, и этим он отличается от правой части (5.26). Мы покажем, что слагаемые $i\xi \sin x$ могут быть опущены и в (5.26), так что

$$\int_{-\infty}^{\infty} \frac{e^{i\xi x} - 1}{x^2} n x^2 F\{a_n dx\} \rightarrow \int_{-\infty}^{\infty} \frac{e^{i\xi x} - 1}{x^2} \cdot M\{dx\}. \quad (5.29)$$

Вне некоторой окрестности точки нуль подынтегральное выражение непрерывно, а так как $n x^2 F\{a_n dx\} \rightarrow M\{dx\}$, то соотношение (5.29) выполняется, если исключить интервал $|x| < \delta$ из области интегрирования. Достаточно показать, что, выбирая достаточно малое δ , можно сделать интеграл по области $|x| < \delta$ как часть интеграла слева в (5.29) сколь угодно малым. Указанный интеграл мажорируется выражением

$$n \int_{|x| < \delta} |x| F\{a_n dx\} = \frac{n}{a_n} \int_{|y| < a_n \delta} |y| F\{dy\}. \quad (5.30)$$

Теперь (5.22) при $\beta = 1$ и (5.25) показывают, что правая часть (5.30) $\sim \frac{2-\alpha}{1-\alpha} C \delta^{1-\alpha}$, $0 < \alpha < 1$, и поэтому стремится к нулю вместе с δ .

Таким образом, (5.29) справедливо. Это соотношение можно переписать в виде $n[\varphi(\xi/a_n) - 1] \rightarrow \psi(\xi)$, что, по теореме 1 § 1, равносильно искомому соотношению $\varphi^n(\xi/a_n) \rightarrow \log \psi(\xi)$.

(ii) Пусть $\alpha > 1$. На этот случай переносится предыдущее рассуждение, с той лишь разницей, что (5.26) заменяется соотношением вида

$$\int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi x}{x^2} n x^2 F\{a_n dx\} \rightarrow \int_{-\infty}^{+\infty} \frac{e^{i\xi x} - 1 - i\xi x}{x^2} M\{dx\}. \quad (5.31)$$

Для обоснования (5.31) мы должны показать, что, выбирая t достаточно большим, можно сделать часть левого интеграла в (5.31), взятую по области $|x| > t$, сколь угодно малой. Это непосредственно вытекает из (5.21).

(iii) Пусть $\alpha = 1$. В этом случае соотношение (5.26) сохраняется без изменений, но для того, чтобы доказать, что это соотношение равносильно искомому (5.27), необходимо показать, что при фиксированных ξ

$$\varphi^n(\xi/a_n) \sim e^n [\varphi(\xi/a_n) - 1], \quad (5.32)$$

или, что то же самое,

$$n |\varphi(\xi/a_n) - 1|^2 \rightarrow 0. \quad (5.33)$$

При $\beta < 1$ абсолютный момент m_β распределения F конечен. Из очевидного неравенства $|e^u - 1| < 2|u|^\beta$ мы заключаем, что $|\varphi(\xi/a_n) - 1| < 2m_\beta |\xi|^\beta a_n^{-\beta}$, и поэтому левая часть (5.33) есть $O(n a_n^{-2\beta})$. Используя определяющее a_n соотношение (5.23), получаем, что $n = O(a_n^{1+\varepsilon})$ при любом $\varepsilon > 0$, и поэтому (5.33) справедливо.

Заключительное замечание. Область притяжения нормального распределения не следует смешивать с введенным Б. В. Гнеденко понятием «области нормального притяжения устойчивого распределения U с показателем α ». Говорят, что распределение F принадлежит указанной области, если оно принадлежит области притяжения U при нормирующих коэффициентах $a_n = n^{1/\alpha}$. Выделение этой области представляло собой первоначально серьезную задачу, однако в нашей обстановке решение дается условием (5.23), определяющим нормирующие постоянные. Распределение F принадлежит области «нормального» притяжения U в том и только том случае, когда $x^2 [1 - F(x)] \rightarrow C\rho$ и $x^2 F(-x) \rightarrow Cq$ при $x \rightarrow \infty$. Здесь $C > 0$ — некоторая постоянная. (Отметим попутно, используя введенную терминологию, что нормальное распределение обладает областью ненормального притяжения.)

§ 6*). УСТОЙЧИВЫЕ ПЛОТНОСТИ

Невозможно, по-видимому, выразить устойчивые плотности в «замкнутой» форме. Разложения в ряды для них были даны независимо друг от друга Феллером (1952) и Бергстромом (1953). В этих работах неявно содержатся результаты, установленные позже более сложными методами. Вывод этих разложений представляет собой хороший пример использования формулы обращения Фурье (хотя при этом применяется интегрирование в комплексной области). Мы не будем рассматривать случай $\alpha = 1$.

При $\xi > 0$ мы можем представить устойчивые характеристические функции в форме $e^{-a^2 \xi^2}$, где a — некоторое комплексное число. Его абсолютная величина определяется только масштабным параметром, поэтому мы можем допустить, что a по модулю равно единице, именно $a = e^{i\pi\gamma/2}$, где γ — действительное число. Таким образом, мы полагаем

$$\psi(\xi) = -|\xi|^2 \cdot e^{\pm i\pi\gamma/2}, \quad (6.1)$$

где в символе « $\pm$ » верхний знак выбирается при $\xi > 0$, а ниж-

*) Этот параграф посвящен специальным вопросам, Его следует пропустить при первом чтении.

ний—при $\zeta < 0$. [См. каноническую форму (3.18) для $\psi(\zeta)$.] Отношение действительной и мнимой частей подчиняется неравенствам, очевидным из (3.18). Тогда в принятых здесь обозначениях для того, чтобы функция e^ψ была устойчивой характеристической функцией, необходимы и достаточны условия

$$|\gamma| \leq \begin{cases} \alpha, & \text{если } 0 < \alpha < 1, \\ 2 - \alpha, & \text{если } 1 < \alpha < 2. \end{cases} \quad (6.2)$$

Так как e^ψ абсолютно интегрируема, то соответствующее распределение имеет плотность, которую мы обозначим $p(x; \alpha, \gamma)$. Мы вычислим ее, используя формулу обращения для интегралов Фурье (3.5) из гл. XV. Так как p —действительная функция и величина $\psi(-\zeta)$ сопряжена с $\psi(\zeta)$, мы получаем

$$p(x; \alpha, \gamma) = \frac{1}{\pi} \operatorname{Re} \int_0^\infty e^{-ix\zeta - \zeta^\alpha e^{i\pi\gamma/2}} d\zeta. \quad (6.3)$$

Так как

$$p(-x; \alpha, \gamma) = p(x; \alpha, -\gamma), \quad (6.4)$$

то эту функцию достаточно вычислить лишь для $x > 0$.

а) *Случай $\alpha < 1$.* Рассмотрим подынтегральное выражение как функцию комплексного переменного ζ . При $x > 0$ и $\operatorname{Im} \zeta \rightarrow -\infty$ подынтегральное выражение стремится к нулю благодаря тому, что главную роль в показателе играет линейный член. Мы можем заменить путь интегрирования отрицательной частью мнимой оси, что сводится к подстановке $\zeta = \frac{t}{x} e^{-i\pi/2}$ и проведению дальнейших выкладок так, как если бы все коэффициенты были вещественны. Новое подынтегральное выражение имеет вид $e^{-t - ct^\alpha}$. Используя разложение e^{-ct^α} в ряд и формулу для гамма-функции, мы получаем сразу

$$p(-x; \alpha, \gamma) = \operatorname{Re} \frac{-i}{\pi x} \sum_{k=0}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} \left(-x^{-\alpha} \exp \left[i \frac{\pi}{2} (\gamma - \alpha) \right] \right)^k. \quad (6.5)$$

б) *Случай $1 < \alpha < 2$.* Формальную подстановку

$$\zeta = t^{\alpha-1} \exp \left(-i \frac{\pi\gamma}{2\alpha} \right)$$

легко обосновать так же, как это было сделано при $\alpha < 1$. Новое подынтегральное выражение имеет вид $e^{-t - ct^{\alpha-1}} t^{\alpha-1-1}$. Разлагая $e^{-ct^{\alpha-1}}$ в степенной ряд, получаем

$$p(-x; \alpha, \gamma) = \frac{1}{\alpha\pi} \operatorname{Re} \exp \left(-i \frac{\pi\gamma}{2\alpha} \right) \sum_{n=0}^{\infty} \frac{\Gamma\left(\frac{n+1}{\alpha}\right)}{n!} \left(-ix \exp \left[-i \frac{\pi\gamma}{2\alpha} \right] \right)^n. \quad (6.6)$$

Заменяя индекс суммирования n на $k-1$ и применяя известную рекуррентную формулу $\Gamma(s+1)=s\Gamma(s)$, имеем

$$p(-x; \alpha, \gamma) =$$

$$= \frac{1}{\pi x} \operatorname{Re} i \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha^{-1}+1)}{k!} \left(-x \exp \left[-i \frac{\pi}{2\alpha} (\gamma - \alpha) \right] \right)^k. \quad (6.7)$$

Таким образом, нами доказана

Лемма 1. Для $x > 0$ и $0 < \alpha < 1$

$$p(x; \alpha, \gamma) = \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{k!} (-x^{-\alpha})^k \sin \frac{k\pi}{2} (\gamma - \alpha). \quad (6.8)$$

Для $x > 0$ и $1 < \alpha < 2$

$$p(x; \alpha, \gamma) = \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha^{-1}+1)}{k!} (-x)^k \sin \frac{k\pi}{2\alpha} (\gamma - \alpha). \quad (6.9)$$

Для $x < 0$ значения p определяются в соответствии с (6.4).

Отметим, что (6.8) дает асимптотическое разложение для $x \rightarrow \infty$. Попутно из доказанных формул вытекает следующий любопытный результат.

Лемма 2. Если $\frac{1}{2} < \alpha < 1$ и $x > 0$, то

$$\frac{1}{x^{\alpha+1}} p \left(\frac{1}{x^{\alpha}}; \frac{1}{\alpha}, \gamma \right) = p(x; \alpha, \gamma^*), \quad (6.10)$$

где $\gamma^* = \alpha(\gamma + 1) - 1$.

Простая проверка показывает, что значения γ^* попадают в интервал, указанный в (6.2). Тождество (6.10) было впервые отмечено (с другим доказательством) В. М. Золотаревым.

§ 7. СХЕМА СЕРИЙ

Понятие схемы серий было описано в гл. VI, 3, следующим образом. Каждому n соответствует конечное число, скажем r_n , независимых случайных величин $X_{k,n}$ ($k=1, 2, \dots, r_n$) с распределениями $F_{k,n}$ и характеристическими функциями $\varphi_{k,n}$. Мы образуем «суммы по строкам» $S_n = X_{1,n} + \dots + X_{r_n,n}$. Обозначим их распределения и характеристические функции через U_n и ω_n соответственно. По причинам, объясненным в гл. VI, 3, мы интересуемся в первую очередь схемами, где влияние отдельных компонент асимптотически пренебрежимо. Чтобы обеспечить это свойство, мы ввели условие (3.2) гл. VI: величины $X_{k,n}$ должны стремиться по вероятности к нулю равномерно относительно $k=1, \dots, r_n$. В терминах характеристических функций это означает, что при любых числах $\varepsilon > 0$ и $\zeta_0 > 0$ для всех достаточно

больших n :

$$|1 - \varphi_{k,n}(\xi)| < \varepsilon \text{ при } |\xi| < \xi_0, \quad k=1, \dots, r_n. \quad (7.1)$$

Такие схемы мы назовем *нулевыми схемами*.

В § 1 и 2 мы имели дело со схемами серий, в которых распределения $F_{k,n}$ не зависели от k . Такие схемы автоматически оказываются нулевыми. Условие (7.1) дает возможность применить построенную в первых двух параграфах теорию. В частности, будет показано, что основной результат переносится на произвольные нулевые схемы: *если распределения сумм по строкам S_n сходятся к некоторому пределу, то последний безгранично делим¹⁾*. Нами будет найден критерий сходимости к заданному безгранично делимому распределению.

Для удобства читателей напомним, что мера M считается *канонической*, если она приписывает конечные массы $M\{I\}$ конечным интервалам и если при любом $x > 0$ существуют интегралы

$$M^+(x) = \int_x^\infty y^{-2} M\{dy\}, \quad M^-(-x) = \int_{-\infty}^{-x} y^{-2} M\{dy\} \quad (7.2)$$

(см. определение 1 из § 1).

Для упрощения обозначений введем меру M_n следующим образом:

$$M_n\{dx\} = \sum_{k=1}^{r_n} x^2 F_{k,n}\{dx\}. \quad (7.3)$$

Эта мера есть аналог мер $n x^2 F_n\{dx\}$ из предыдущих параграфов. При $x > 0$ положим по аналогии с (7.2)

$$M_n^+(x) = \sum_{k=1}^{r_n} [1 - F_{k,n}(x)], \quad M_n^-(x) = \sum_{k=1}^{r_n} F_{k,n}(-x). \quad (7.4)$$

Мы будем широко использовать урезанные случайные величины, однако стандартная процедура урезания будет несколько видоизменена для того, чтобы избежать обычных затруднений, возникающих при использовании разрывных функций. Новая процедура урезания заменяет случайную величину X величиной $\tau(X)$, где $\tau(x)$ — *такая непрерывная монотонная функция $\tau(x)$, что*

$$\tau(x) = x \text{ при } |x| \leq a, \quad \tau(x) = \pm a \text{ при } |x| \geq a \quad (7.5)$$

[очевидно, $\tau(-x) = -\tau(x)$]. Введем обозначения для математических ожиданий усеченных случайных величин:

$$\begin{aligned} \beta_{k,n} &= E(\tau(X_{k,n})), \\ b_n &= \sum_{k=1}^{r_n} \beta_{k,n}, \quad B_n = \sum_{k=1}^{r_n} \beta_{k,n}^2. \end{aligned} \quad (7.6)$$

¹⁾ Относительно применения этого результата к процессам с независимыми, но нестационарными приращениями см. замечание в конце § 1.

Существует теоретическая возможность центрировать $X_{k,n}$ таким образом, чтобы все $\beta_{k,n}$ обратились в нуль. Это упростит рассуждения, но полученный критерий нельзя будет применять непосредственно во многих конкретных ситуациях. Однако возможно так центрировать $X_{k,n}$, чтобы сделать $\beta_{k,n}$ настолько малыми, чтобы $B_n \rightarrow 0$. В этом случае условия нижеследующей теоремы сводятся к условию $M_n \rightarrow M$, хорошо известному по предыдущим параграфам. В общем случае $M_n\{I\} \rightarrow M\{I\}$ для интервалов I , расположенных на положительном расстоянии от нуля. В окрестности же нуля будет сказываться влияние B_n . (Выбор точки урезания a не имеет значения.)

Теорема. Пусть $\{X_{k,n}\}$ — нулевая схема. Если можно найти такие постоянные b_n , что распределения $S_n - b_n$ сходятся к предельному распределению U , то они сходятся и при постоянных, удовлетворяющих (7.6)¹⁾. Предельное распределение U безгранично делимо²⁾.

Для сходимости к предельному распределению U с канонической мерой M необходимо и достаточно, чтобы во всех точках непрерывности $x > 0$

$$M_n^+(x) \rightarrow M^+(x), \quad M_n^-(-x) \rightarrow M^-(-x) \quad (7.7)$$

и чтобы для некоторого $s > 0$

$$M_n\{\overline{-s}, s\} - B_n \rightarrow M\{\overline{-s}, s\}. \quad (7.8)$$

В этом случае распределение $S_n - b_n$ стремится к распределению с характеристической функцией $\omega = e^\psi$, где

$$\psi(\zeta) = \int_{-\infty}^{\infty} \frac{e^{i\zeta x} - 1 - i\zeta\tau(x)}{x^2} M\{dx\}. \quad (7.9)$$

[Условие (7.8) будет автоматически выполняться во всех точках непрерывности.]

Доказательство. Разобьем доказательство на отдельные этапы.

а) Предположим сначала, что все случайные величины $X_{k,n}$ симметричны, так что распределения S_n должны сходиться без предварительного центрирования. Характеристические функции $\varphi_{k,n}$ действительны, и ввиду (7.1) для произвольного ζ справед-

¹⁾ Теорема остается также в силе и при обычной процедуре урезания, т. е. когда τ заменяется функцией, обращающейся в нуль при $|x| > a$. Чтобы избежать усложнения обозначений, необходимо при этом предположить, что у меры M нет атомов в точках $\pm a$. [Часть б) доказательства становится более сложной, так как можно не найти такое θ , чтобы $E(\tau(X + \theta)) = 0$.]

²⁾ Распределение, сосредоточенное в одной точке, является безгранично делимым, соответствующая ему каноническая мера тождественно равна нулю.

ливо разложение Тейлора

$$-\log \varphi_{k,n}(\zeta) = [1 - \varphi_{k,n}(\zeta)] + \frac{1}{2} [1 - \varphi_{k,n}(\zeta)]^2 + \dots \quad (7.10)$$

при одном условии, что n достаточно велико. Вопрос состоит в том, справедливо ли соотношение

$$\sum_{k=1}^{r_n} \log \varphi_{k,n}(\zeta) \rightarrow \psi(\zeta). \quad (7.11)$$

Все члены разложения (7.10) положительны и, следовательно, (7.11) требует, чтобы сумма линейных членов была ограниченной. Ввиду (7.1) это означает, что влияние членов более высокого порядка асимптотически незначимо. Поэтому (7.11) справедливо тогда и только тогда, когда

$$\sum_{k=1}^{r_n} [\varphi_{k,n}(\zeta) - 1] \rightarrow \psi(\zeta). \quad (7.12)$$

Левая часть может быть переписана в виде $r_n[\varphi_n - 1]$, где φ_n — характеристическая функция арифметического среднего распределений $F_{k,n}$. Таким образом, мы имеем дело с частным случаем теоремы 2 из § 2 и можем сделать вывод, что соотношение (7.12) выполняется тогда и только тогда, если существует каноническая мера M , такая, что $M_n \rightarrow M$. Если $B_n = 0$, то условия (7.7), (7.8) равносильны $M_n \rightarrow M$, так как для интервала I , находящегося на положительном расстоянии от нуля, соотношение $M_n\{I\} \rightarrow M\{I\}$ следует из (7.7). Это доказывает теорему для симметричных распределений.

б) Предположим теперь, что $\beta_{k,n} = 0$ для всех k и n . Мы докажем, что (7.11) верно, если верно

$$\sum_{k=1}^{r_n} |\varphi_{k,n}(\zeta) - 1| < C(\zeta), \quad (7.13)$$

т. е. если сумма слева ограничена. В таком случае (7.11) и (7.12) будут снова равносильны, и, используя заключительное рассуждение предыдущего пункта доказательства а), снова сводим утверждение к теореме 2 из § 2. Правда, в этой теореме центрирующие постоянные b'_n вычисляются по функции $\tau(x)$ вместо $\sin x$; однако это легко компенсируется сдвигом предельного распределения, поскольку

$$b_n - b'_n = \int_{-\infty}^{\infty} \frac{\tau(x) - \sin x}{x^2} M_n\{dx\} \rightarrow \int_{-\infty}^{\infty} \frac{\tau(x) - \sin x}{x^2} M\{dx\}. \quad (7.14)$$

Чтобы вывести (7.13) из (7.11), мы начнем с тождества

$$\varphi_{k,n}(\zeta) - 1 = \int_{-\infty}^{+\infty} [e^{i\zeta x} - 1 - i\zeta \tau(x)] F_{k,n}\{dx\}, \quad (7.15)$$

которое справедливо, так как $\beta_{k,n} = 0$. Подынтегральное выражение (7.15) при $|x| < a$ равно $e^{i\zeta x} - 1 - i\zeta x$ и мажорируется функцией $\frac{1}{2}\zeta^2 x^2$. Так как $|\tau(x)| \leq a$, мы заключаем, что

$$\sum_{k=1}^{r_n} |\varphi_{k,n}(\zeta) - 1| \leq \frac{1}{2} \zeta^2 M_n \{ -\overline{a}, a \} + (2+a|\zeta|) (M_n^+(a) + M_n^-(-a)). \quad (7.16)$$

Чтобы показать, что $M_n^+(a)$ должно быть ограничено, рассмотрим схему $\{^0X_{k,n}\}$, полученную симметризацией схемы $\{X_{k,n}\}$. Из условия (7.1) для нулевой схемы следует, что при достаточно большом n вероятность события $^0X_{k,n} > X_{k,n} - \varepsilon$ превосходит $\frac{1}{2}$ для всех $k \leq r_n$. Таким образом, $\frac{1}{2} M_n^+(a) < ^0M_n^+(a)$, и мы знаем, что последняя величина остается ограниченной, если имеет место сходимость. Мы заключаем, что в случае сходимости сумма $M_n^+(a) + M_n^-(-a)$ остается ограниченной и, следовательно,

$$\sum_{k=1}^{r_n} |\varphi_{k,n}(\zeta) - 1|^2 = M_n \{ -\overline{a}, a \} \cdot \varepsilon_n(\zeta), \quad (7.17)$$

где ε_n обозначает величину, стремящуюся к нулю. С другой стороны, действительная часть подынтегрального выражения в (7.15) не меняет знака. При $|x| < a$ и достаточно малых ζ ее абсолютная величина больше $\frac{1}{4} \cdot \zeta^2 x^2$, и, следовательно,

$$-\operatorname{Re} \sum_{k=1}^{r_n} (\varphi_{k,n}(\zeta) - 1) \geq \frac{1}{4} \zeta^2 M_n \{ -\overline{a}, a \}. \quad (7.18)$$

Последние два неравенства показывают, что левая часть (7.11) не может быть ограничена, если не ограничена $M_n \{ -\overline{a}, a \}$. Поэтому в нашем случае (7.13) является следствием (7.16).

в) Наконец, мы рассмотрим произвольную нулевую схему $\{X_{k,n}\}$. Так как $E(\tau(X_{k,n} - \theta))$ — непрерывная монотонная функция параметра θ , которая изменяется от $+a$ до $-a$, то существует такое единственное значение $\theta_{k,n}$, при котором случайная величина $Y_{k,n} = X_{k,n} - \theta_{k,n}$ удовлетворяет условию $E(\tau(Y_{k,n})) = 0$. Ясно, что $\{Y_{k,n}\}$ образует нулевую схему и, следовательно, к ней применима наша теорема.

Таким образом, мы нашли общий вид возможных предельных распределений. Однако условия сходимости выражены в терминах меры $N_n \{dx\} = \sum x^2 F_{k,n} \{\theta_{k,n} \mp dx\}$ для искусственно центрированных распределений $Y_{k,n}$. Другими словами, мы доказали теорему с N_n вместо M_n в (7.7) и (7.8) и заменой B_n на 0.

Чтобы исключить центрирующие постоянные $\theta_{k,n}$, вспомним, что они равномерно стремятся к 0 и поэтому в конечном счете

$$M_n^+(x + \varepsilon) \leq N_n^+(x) \leq M_n^+(x - \varepsilon).$$

Это позволяет заключить, что условие (7.7) применимо в равной степени к обеим схемам.

Прежде чем обратиться к условию (7.8), мы замечаем, что схемы $\{Y_{k,n}\}$ и $\{X_{k,n} - \beta_{k,n}\}$ имеют одно и то же предельное распределение, т. е.

$$b_n - \sum_{k=1}^{r_n} \theta_{k,n} \rightarrow 0. \quad (7.19)$$

Положим $Z_{k,n} = \tau(Y_{k,n}) - \tau(X_{k,n}) + \theta_{k,n}$. Из определения τ ясно, что $Z_{k,n}$ обращается в нуль, если $\|X_{k,n} - a\| < \theta_{k,n}$, а в противном случае $|Z_{k,n}| \leq |\theta_{k,n}| \rightarrow 0$. Следовательно, условие (7.7) гарантирует, что

$$\sum_{k=1}^{r_n} |E(Z_{k,n})| = \sum_{k=1}^{r_n} |\beta_{k,n} - \theta_{k,n}| \rightarrow 0, \quad (7.20)$$

и это сильнее, чем (7.19).

В заключение мы обратимся к условию (7.8). Для того чтобы показать, что при $n \rightarrow \infty$ разность двух частей (7.8) стремится к нулю, мы используем знак $\approx$. Легко увидеть, что если справедливы (7.7) и (7.20), то

$$\begin{aligned} M_n \{-a, a\} - B_n &\approx \sum_{k=1}^{r_n} \int_{|x| < a} (x - \theta_{k,n})^2 F_{k,n} \{dx\} \approx \\ &\approx \sum_{k=1}^{r_n} \int_{|y| < a} y^2 F_{k,n} \{dy + \theta_{k,n}\} \approx N_n \{-a, a\}. \end{aligned} \quad (7.21)$$

Таким образом, (7.8) эквивалентно соответствующему условию для схемы $\{Y_{k,n}\}$.

Пример. Роль центрирования. Пусть величины $X_{k,n}$ ($k=1, 2, \dots, n$) распределены нормально с математическим ожиданием $n^{-1/4}$ и дисперсией n^{-1} . При центрировании математическими ожиданиями предельное распределение существует и является нормальным. Однако, когда центрирующие постоянные $\beta_{k,n} = n^{-1/4}$, мы имеем $B_n \sim 2\sqrt{n} \rightarrow \infty$. Отсюда следует, что $M_n \{-a, a\} \rightarrow \infty$. Этот пример показывает, что при произвольном центрировании «нелинейная» форма теоремы неизбежна. Пример показывает также, что в общем случае в разложении (7.10) для $\log \varphi_{k,n}$ уже недостаточно рассматривать только линейные члены. Другие результаты см. в задачах 17 и 18.

§ 8*). КЛАСС L

Мы проиллюстрируем силу последней теоремы, приведя простое доказательство одной теоремы, установленной П. Леви. Мы еще раз вернемся к частным суммам $S_n = X_1 + \dots + X_n$ последовательности взаимно независимых случайных величин. Однако здесь в отличие от § 5 распределение F_n случайной величины X_n может зависеть от n . Положим $S_n^* = (S_n - b_n)/a_n$. Мы желаем описать все возможные предельные распределения для $\{S_n^*\}$ при условиях

$$a_n \rightarrow \infty, \quad \frac{a_{n+1}}{a_n} \rightarrow 1. \quad (8.1)$$

Первое условие исключает сходящиеся ряды $\sum X_k$, которые рассматриваются в § 10. Ситуации, которые исключает второе условие, легче всего поможет уяснить следующий пример.

Пример. Пусть X_n имеет показательное распределение с математическим ожиданием $n!$. Положим, $a_n = n!$ и $b_n = 0$. Тогда очевидно, что распределение S_n^* сходится к показательному распределению с математическим ожиданием 1, но эта сходимость имеет место исключительно благодаря подавляющей роли слагаемого X_n . ▶

Следуя Хинчину, обычно говорят, что *распределение принадлежит классу L , если оно является предельным распределением для последовательности $\{S_n^*\}$, удовлетворяющей условиям (8.1).*

При таком определении не ясно, что все распределения класса L безгранично делимы. Мы выведем этот факт из следующей леммы.

Лемма. *Характеристическая функция ω принадлежит классу L тогда и только тогда, когда для любого s ($0 < s < 1$) отношение $\omega(\xi)/\omega(s\xi)$ есть характеристическая функция.*

Доказательство. а) *Необходимость.* Обозначим характеристическую функцию S_n^* через ω_n . Пусть $n > m$. Случайная величина S_n^* может быть представлена в виде суммы $(a_m/a_n) S_m^*$ и случайной величины, зависящей только от $X_{m+1}, \dots, X_n$. Следовательно,

$$\omega_n(\xi) = \omega_m\left(\frac{a_m}{a_n}\xi\right)\varphi_{m,n}(\xi), \quad (8.2)$$

где $\varphi_{m,n}$ — характеристическая функция. Пусть теперь $n \rightarrow \infty$ и $m \rightarrow \infty$ так, что $a_m/a_n \rightarrow s < 1$ [этого возможно добиться в силу (8.1)]. Левая часть (8.2) сходится к $\omega(\xi)$, а первый множитель справа — к $\omega(s\xi)$, так как сходимость характеристических функций равномерна в каждом конечном интервале (теорема 2, гл. XV, 3). Из этого мы выводим прежде всего, что ω не имеет нулей. В самом деле, функции $\varphi_{m,n}$ ограничены, поэтому равенство

*) Здесь рассматриваются специальные вопросы.

$\omega(\xi_0) = 0$ влекло бы $\omega(s\xi_0) = 0$ и вообще $\omega(s^k\xi_0) = 0$ при всех $k > 0$, в то время как $\omega(s^k\xi_0) \rightarrow 1$. Соответственно отношение $\omega(\xi)/\omega(s\xi)$ будет непрерывной функцией, предельной для характеристических функций $\varphi_{m,n}$, а потому оно будет характеристической функцией.

б) *Достаточность.* Рассуждения выше показывают, что ω не имеет нулей и, следовательно, справедливо тождество

$$\omega(n\xi) = \omega(\xi) \cdot \frac{\omega(2\xi)}{\omega(\xi)} \cdots \frac{\omega(n\xi)}{\omega((n-1)\xi)}. \quad (8.3)$$

В условиях леммы множитель $\frac{\omega(k\xi)}{\omega((k-1)\xi)}$ будет характеристической функцией некоторой случайной величины X_k . Поэтому $\omega(\xi)$ — характеристическая функция для $(X_1 + \dots + X_n)/n$. ►

Мы не только доказали теорему, но и установили, что ω является характеристической функцией n -й суммы в схеме серий. Условие (7.1), выделяющее нулевые схемы, тривиальным образом выполняется. Следовательно, функция ω будет безгранично делимой. Чтобы отыскать каноническую меру M , определяющую ω , отметим, что отношение $\omega(\xi)/\omega(s\xi)$ также безгранично делимо. Это видно из формулы (8.3). Каноническая мера N , соответствующая $\omega(\xi)/\omega(s\xi)$, связана с M тождеством

$$N\{dx\} = M\{dx\} - s^2 M\{s^{-1}dx\}. \quad (8.4)$$

В терминах функций M^+ и M^- это соотношение имеет вид

$$N^+(x) = M^+(x) - M^+(x/s), \quad N^-(-x) = M^-(-x) - M^-(-x/s). \quad (8.5)$$

Мы показали, что если каноническая мера M определяет характеристическую функцию ω класса L , то функции N^+ и N^- , определенные в (8.5), должны быть монотонны при каждом $0 < s < 1$. Обратно, если это верно, то (8.4) определяет каноническую меру, соответствующую $\omega(\xi)/\omega(s\xi)$. Нами доказана, таким образом,

Теорема. *Характеристическая функция ω принадлежит классу L тогда и только тогда, когда она безгранично делима и ее каноническая мера M такова, что две функции (8.5) монотонны при любом фиксированном $0 < s < 1$.*

Примечание. Легко проверить, что указанные функции монотонны тогда и только тогда, когда $M^+(e^x)$ и $M^-(-e^x)$ суть выпуклые функции.

§ 9*). ЧАСТИЧНОЕ ПРИТЯЖЕНИЕ. «УНИВЕРСАЛЬНЫЕ ЗАКОНЫ»

Как мы уже видели, распределение F не обязано принадлежать какой-либо области притяжения. Поэтому возникает вопрос о том, существуют ли какие-нибудь общие законы асимптотического поведения последовательности $\{F^{n*}\}$ свертки F с собой. К сожалению, можно встретить практически любое мыслимое поведение. Мы опишем здесь несколько возможностей главным образом по причине их неожиданного характера.

Говорят, что характеристическая функция φ принадлежит области частичного притяжения γ , если существуют такие нормирующие постоянные a_r и b_r и такая последовательность натуральных чисел $n_r \rightarrow \infty$, что

$$\left[\varphi\left(\frac{\zeta}{a_r}\right) e^{-ib_r \zeta} \right]^{n_r} \rightarrow \gamma(\zeta). \quad (9.1)$$

Здесь предполагается, что $|\gamma|$ не равен тождественно единице, т. е. что соответствующее распределение не сосредоточено в одной точке. Таким образом, соотношение (9.1) обобщает понятие области притяжения, позволяя рассматривать пределы подпоследовательностей.

По теореме 2 § 1 предел γ будет обязательно безгранично делимым. Приводимые ниже примеры показывают, что возможны оба крайних случая: существуют как распределения, не принадлежащие никакой области частичного притяжения, так и распределения, принадлежащие области частичного притяжения любого безгранично делимого закона.

Примеры. а) В примере 3, е) указана характеристическая функция φ , которая не является устойчивой, но принадлежит своей собственной области частичного притяжения.

б) Симметричное распределение с медленно меняющимися хвостами не принадлежит ни одной области частичного притяжения. Допустим, что функция $L(x) = 1 - F(x) + F(-x)$ медленно меняется на бесконечности. По теореме 2 гл. VIII, 9, в этом случае

$$U(x) = \int_{-x}^x y^2 F\{dy\} = o(x^2 L(x)), \quad x \rightarrow \infty. \quad (9.2)$$

Согласно теореме § 7, для того чтобы распределение F принадлежало какой-либо области частичного притяжения, необходимо условие: для n , пробегающих некоторую подпоследовательность,

*) Здесь рассматриваются специальные вопросы,

функции $n[1 - F(a_n x) + F(-a_n x)]$ и $na_n^{-2}U(a_n x)$ сходятся во всех точках непрерывности. Из первого условия следует $nL(a_n) \sim 1$, а из второго — $nL(a_n) \rightarrow \infty$.

в) *Безгранично делимое распределение не обязано принадлежать своей собственной области частичного притяжения.* В самом деле, из теоремы 1 из § 1 следует, что если φ принадлежит области притяжения γ , то тем же свойством обладает безгранично делимая характеристическая функция $e^{\varphi^{-1}}$. Последний пример показывает, что функция $e^{\varphi^{-1}}$ может не принадлежать ни одной области частичного притяжения.

г) Мы докажем сейчас одно утверждение, которое подготовит нас к странностям последующих примеров. Рассмотрим произвольную последовательность безгранично делимых характеристических функций $\omega_r = e^{\psi_r}$ с ограниченными показателями. Положим

$$\lambda(\xi) = \sum_{k=1}^{\infty} \frac{1}{n_k} \psi_k(a_k \xi). \quad (9.3)$$

Тогда возможно подобрать постоянные $a_k > 0$ и целые n_k так, что при всех ξ при $r \rightarrow \infty$

$$n_r \lambda\left(\frac{\xi}{a_r}\right) - \psi_r(\xi) \rightarrow 0. \quad (9.4)$$

Доказательство. Выберем в качестве $\{n_k\}$ монотонную последовательность целых чисел, возрастающую столь быстро, что $n_k/n_{k-1} > 2^k \max |\psi_k|$. Левая часть (9.4) не превосходит в этом случае

$$n_r \sum_{k=1}^{r-1} \left| \psi_k\left(\frac{a_k \xi}{a_r}\right) \right| + \sum_{k=r+1}^{\infty} \frac{1}{2^k}. \quad (9.5)$$

Коэффициенты a_r мы будем выбирать один за другим по следующему правилу. Положим $a_1 = 1$. При данных $a_1, \dots, a_{r-1}$ выберем a_r столь большим, чтобы величина (9.5) была меньше $1/r$ для всех $|\xi| < r$. Это можно сделать, так как первая сумма непрерывно зависит от ξ и равна нулю при $\xi = 0$.

д) *Каждая безгранично делимая характеристическая функция $\omega = e^{\psi}$ имеет область частичного притяжения.* В самом деле, мы знаем, что функция ω может быть представлена как предел последовательности характеристических функций $\omega_k = e^{\psi_k}$ обобщенного пуассоновского типа. Определим λ по (9.3) и положим $\varphi = e^{\lambda}$. Тогда φ будет характеристической функцией и, согласно (9.4),

$$\lim \varphi^{n_r}\left(\frac{\xi}{a_r}\right) = \lim e^{\psi_r(\xi)} = \omega(\xi). \quad (9.6)$$

е) *Другие возможности.* Возьмем две безгранично делимые

характеристические функции e^α и e^β . Выберем в (9.3) слагаемые ψ_k так, что $\psi_{2k} \rightarrow \alpha$, а $\psi_{2k+1} \rightarrow \beta$. Из (9.4) легко выводим, что если последовательность $n_r \lambda(\xi/a_{k_r})$ сходится, то ее предел необходимо является линейной комбинацией α и β . Другими словами, e^λ принадлежит области частичного притяжения каждого распределения с характеристической функцией вида $e^{p\alpha+q\beta}$ ($p+q=1$) и не принадлежит никаким другим областям частичного притяжения. Этот пример легко обобщить. В терминах выпуклых множеств можно утверждать, что существует распределение F , принадлежащее области частичного притяжения любого распределения, расположенного в выпуклой оболочке n заданных безгранично делимых распределений.

ж) Пусть дана последовательность безгранично делимых характеристических функций $e^{\alpha_1}, e^{\alpha_2}, \dots$. Тогда существует функция $\varphi = e^\lambda$, принадлежащая области частичного притяжения каждой из этих функций. Разобьем последовательность целых чисел на бесконечное число подпоследовательностей (например, включая в n -ю подпоследовательность все те целые числа, которые делятся на 2^{n-1} , но не делятся на 2^n). Выберем ψ_r в примере г) так, что $\psi_r \rightarrow \alpha_n$ при r , пробегающем n -ю подпоследовательность. В этом случае, как показывает (9.4), функция $\varphi = e^\lambda$ обладает требуемым свойством.

з) «Универсальные законы» Деблина. Существуют функции φ , принадлежащие области частичного притяжения любой безгранично делимой функции ω . Действительно, очевидно, что если φ принадлежит области частичного притяжения каждой из функций $\omega_1, \omega_2, \dots$ и $\omega_n \rightarrow \omega$, то φ принадлежит и области частичного притяжения ω . Далее рассмотрим безгранично делимые характеристические функции, у которых соответствующие канонические меры сосредоточены в конечном числе рациональных точек и приписывают каждой из них только рациональные массы. Множество этих функций счетно. Расположим их в последовательность $e^{\alpha_1}, e^{\alpha_2}, \dots$. Тогда каждая безгранично делимая функция ω будет пределом некоторой подпоследовательности последовательности $\{e^{\alpha_k}\}$. Характеристическая функция φ , построенная по методу предыдущего примера, принадлежит области частичного притяжения каждой из функций α_k , а потому и области частичного притяжения ω .

[Примечание. Последний результат был получен в замечательной статье Деблина (1940), продолжившей работу Хинчина (1937). Технические трудности, связанные с этой задачей, в то время были весьма значительными. Явление, описанное в примере б), было в частных случаях отмечено Гнеденко, Хинчиным и Леви. Интересно отметить те трудности, с которыми сталкиваются в отдельных примерах, если не принимают во внимание основной роли, которую в этой теории играют правильно меняющиеся функции.

§ 10*). БЕСКОНЕЧНЫЕ СВЕРТКИ

Пусть $X_1, X_2, \dots$ — независимые случайные величины с характеристическими функциями $\varphi_1, \varphi_2, \dots$. Как и в (7.5), мы определим «монотонную непрерывную урезающую функцию» τ равенствами: $\tau(x) = x$ при $|x| \leq a$ и $\tau(x) = \pm 1$ при $|x| \geq a$. Основная теорема о бесконечных свертках утверждает: *распределения частных сумм $X_1 + \dots + X_n$ сходятся к некоторому распределению вероятностей U тогда и только тогда, когда*

$$\sum_{k=1}^{\infty} \text{Var}(\tau(X_k)) < \infty, \quad \sum_{k=1}^{\infty} P\{|X_k| > a\} < \infty \quad (10.1)$$

и

$$\sum_{k=1}^n E(\tau(X_k)) \rightarrow b, \quad (10.2)$$

где b — некоторое число.

Частный случай конечных дисперсий был разобран в гл. VIII, 5, с указанием примеров и применений. В общей форме теорема появилась в гл. IX, 9, где сформулированный результат был обобщен (была доказана сходимость ряда $\sum X_n$ — «теорема о трех рядах»). Было показано, что упомянутая теорема является простым следствием основных теорем, касающихся схемы серий, и поэтому нет необходимости повторять сделанные рассуждения¹⁾. Поэтому мы удовлетворимся примерами, иллюстрирующими использование характеристических функций.

Примеры. а) *Факторизация равномерного распределения.* Пусть $X_k = \pm 2^{-k}$ с вероятностью $1/2$. В примере 11, в) гл. I было объяснено, что $\sum X_k$ можно интерпретировать как «число, выбранное наудачу на отрезке от -1 до 1 ». Это равносильно утверждению, что характеристическая функция $(\sin \xi)/\xi$ равномерного распределения представима в виде бесконечного произведения характеристических функций $\cos(\xi/2^k)$. Аналитическое доказательство основано на тождестве

$$\frac{\sin \xi}{\xi} = \cos \frac{\xi}{2} \cdot \cos \frac{\xi}{4} \dots \cos \frac{\xi}{2^n} \cdot \frac{\sin(\xi/2^n)}{\xi/2^n}, \quad (10.3)$$

которое доказывается по индукции, исходя из формулы $\sin 2\alpha =$

*) Здесь рассматриваются специальные вопросы.

¹⁾ Прямая проверка того, что (10.1) и (10.2) обеспечивают равномерную в каждом конечном интервале сходимость произведений $\varphi_1 \dots \varphi_n$, представляет собой хорошее упражнение. Необходимость этих условий менее очевидна. Однако она легко вытекает из того факта, что схема серий с n -й строкой $X_n, X_{n+1}, \dots, X_{n+n}$ должна удовлетворять условиям теоремы из § 7 с $M=0$.

$= 2 \sin \alpha \cos \alpha$. При $n \rightarrow \infty$ последний множитель сходится к 1 равномерно в каждом конечном интервале.

Отметим, что произведение четных членов снова соответствует сумме независимых случайных величин. Мы знаем из примера 11, г) гл. I, что эта сумма имеет *сингулярное распределение* канторовского типа¹⁾ [см. задачи 5, 7 и 19].

б) Пусть Y_k имеет плотность $\frac{1}{2} e^{-|x|}$ с характеристической функцией $1/(1+\zeta^2)$. Ряд $\sum Y_k/k$ сходится. Для соответствующей характеристической функции $\frac{\pi \zeta}{\operatorname{sh} \pi \zeta}$ мы получаем каноническое разложение на множители (sh — гиперболический синус). Используя задачу 8 из гл. XV, 9, находим, что *плотность* $\sum Y_k/k$ равна

$$\frac{1}{2 + e^x + e^{-x}} = \frac{1}{4 (\operatorname{ch} (x/2))^2}.$$

§ 11. МНОГОМЕРНЫЙ СЛУЧАЙ

Развитая в этой главе теория без существенных изменений переносится на многомерный случай. Мы не будем останавливаться на деталях. В канонической форме для безгранично делимых распределений лучше выделять нормальную компоненту и рассматривать только канонические меры, не имеющие атома в начале координат. Прежние формулы не потребуют никаких изменений, если интерпретировать ζx как скалярное произведение (см. гл. XV, 7). Для определенности мы выпишем формулу для двумерного случая.

Мера M , не имеющая атома в 0, является канонической, если она приписывает конечные массы конечным интервалам и если по отношению к ней интегрируема функция $1/(1+x_1^2+x_2^2)$. Выберем подходящую *одномерную* центрирующую функцию, скажем $\tau(x) = \sin x$, или функцию, указанную в (7.5). Положим

$$\psi(\zeta_1, \zeta_2) = \int \frac{e^{i(\zeta_1 x_1 + \zeta_2 x_2)} - 1 - i\zeta_1 \tau(x_1) - i\zeta_2 \tau(x_2)}{x_1^2 + x_2^2} M\{dx\}, \quad (11.1)$$

где интеграл распространяется на всю плоскость. Тогда $\omega = e^\psi$ будет двумерной безгранично делимой характеристической функцией. Наиболее общая безгранично делимая характеристическая функция получается умножением на нормальную характеристическую функцию.

Переход к полярным координатам может сделать ситуацию легче воспринимаемой. Положим

$$\zeta_1 = \rho \cos \varphi, \quad \zeta_2 = \rho \sin \varphi, \quad x = r \cos \theta, \quad y = r \sin \theta. \quad (11.2)$$

Определим каноническую меру в полярных координатах следую-

¹⁾ Шокé дал привлекательное геометрическое доказательство, применимое и к более общим бесконечным сверткам. Оно приведено в статье: Tortrat A., J. Math. Pures. Appl., 39 (1960), 231—273.

щим образом. Выберем при каждом θ , $-\pi \leq \theta \leq \pi$, одномерную каноническую меру Λ_θ , сосредоточенную на $[0, \infty)$; выберем далее конечную меру W на $-\pi < \theta \leq \pi$ (на окружности). Тогда M можно определить с помощью рандомизации по параметру θ , и (с тривиальными изменениями центрирующей функции) мы можем переписать (11.1) в форме

$$\psi(\xi_1, \xi_2) = \int_{-\pi}^{\pi} W\{d\theta\} \int_{0+}^{\infty} \frac{e^{i\rho r \cos(\varphi-\theta)} - 1 - i\rho\tau(r) \cos(\varphi-\theta)}{r^2} \Lambda_\theta\{dr\}. \quad (11.3)$$

(В такой форме, добавляя к Λ_θ атом в нуле, можно охватить и случай нормальной компоненты.)

Пример. Устойчивые распределения. Положим по аналогии с одномерным случаем $\Lambda_\theta\{dr\} = r^{-\alpha+1} dr$. Можно было бы добавить произвольный множитель C_θ , но это свелось бы к замене меры W . Как мы видели в примере 3, ж), при такой мере Λ_θ функция (11.3) принимает вид

$$\psi(\xi_1, \xi_2) = -C\rho^\alpha \int_{-\pi}^{+\pi} |\cos(\varphi-\theta)|^\alpha \left(1 \pm \operatorname{tg} \frac{\pi}{2\alpha}\right) W\{d\theta\}, \quad (11.4)$$

где верхний (нижний) знак выбирается при $\varphi-\theta > 0$ ($\varphi-\theta < 0$). Отсюда ясно, что e^ψ — строго устойчивая характеристическая функция и так же, как и в § 5, можно установить, что других строго устойчивых функций нет. Тем не менее точно так же, как и в одномерном случае при $\alpha=1$ существуют характеристические функции, которые устойчивы только в широком смысле и имеют логарифмический член в показателе.

При $\alpha=1$ и равномерном распределении W мы получаем характеристическую функцию $e^{-\alpha\rho}$ симметричного распределения Коши в $\mathbb{R}^2$ [см. пример 7, д) гл. XV и задачи 21—23]. ►

§ 12. ЗАДАЧИ

1. В § 2 было показано, что если ψ представляет собой логарифм безгранично делимой характеристической функции, то функция

$$\psi(\xi) - \frac{1}{2h} \int_{-h}^{+h} \psi(\xi+s) ds = \chi(\xi) \quad (12.1)$$

отличается лишь действительным множителем от некоторой характеристической функции. Докажите обратное: если ψ — непрерывная функция и $\chi(\xi)/\chi(0)$ — характеристическая функция при любом выборе $h > 0$, то ψ отличается от логарифма безгранично делимой характеристической функции только на линейную функцию. Более того, ψ есть тот самый логарифм, если дополнительно $\psi(0)=0$ и $\psi(-\xi)=\overline{\psi(\xi)}$. *Указание.* Докажите, что все решения однородного уравнения (12.1) (при $\chi=0$) линейны.

2. Показать, что задача 1 и рассуждения в § 2 остаются верными, если

в (12.1) или (2.13) вместо равномерного распределения используется распределение, сосредоточенное в точках $\pm h$:

$$\psi(\xi) - \frac{1}{2} [\psi(\xi+h) + \psi(\xi-h)] = \chi(\xi). \quad (12.2)$$

Однако здесь появляются небольшие осложнения из-за того, что плотность, соответствующая χ , не является строго положительной.

3. *Обобщение.* Пусть R — произвольное симметричное распределение вероятностей с конечной дисперсией. Если e^ψ — безгранично делимая характеристическая функция и

$$\chi = \psi - R \star \psi, \quad (12.3)$$

то $\chi(\xi)/\chi(0)$ — характеристическая функция. При этом применимо рассуждение § 2 с заменой (2.13) на (12.3).

В частности, если R имеет плотность $\frac{1}{2} e^{-|x|}$, то приходим непосредственно к нормальной форме Хинчина для ψ (см. заключительное замечание к § 2). Однако необходима некоторая осторожность в связи с тем, что ψ не ограничена.

4. Пусть ω — безгранично делимая характеристическая функция. Тогда существуют такие постоянные a и b , что $|\log \omega(\xi)| < a + b\xi^2$ при всех ξ .

5. *Дробовой эффект в вакуумных лампах.* В примере 3, 3) гл. VI мы имели дело со схемой серий, в которой $X_{k,n}$ обладает характеристической функцией

$$\varphi_{k,n}(\xi) = 1 + \alpha h [e^{i\xi/(kh)} - 1],$$

где $h = 1/\sqrt{n}$. Покажите, что характеристические функции сумм $S_n = X_{1,n} + \dots + X_{n,n}$ сходятся к пределу e^ψ , где

$$\psi(\xi) = \alpha \int_0^\infty [e^{i\xi/(x)} - 1] dx;$$

e^ψ является характеристической функцией случайной величины $X(t)$. Дифференцированием получаем *теорему Кемпбелла*, см. (3.4) в гл. VI.

6. Пусть $U = \sum X_n/n$, где случайные величины X_k независимы и имеют одну и ту же плотность $\frac{1}{2}e^{-|x|}$. Покажите¹⁾, что случайная величина U безгранично делима и что соответствующая каноническая мера равна

$$M\{dx\} = |x| \frac{e^{-|x|}}{1 - e^{-|x|}} dx.$$

[Не требуется никаких вычислений, кроме суммирования геометрической прогрессии.]

7. Пусть $P(s) = \sum p_k s^k$, где $p_k \geq 0$ и $\sum p_k = 1$. Допустим, что $P(0) > 0$ и что $\log \frac{P(s)}{P(0)}$ разлагается в степенной ряд с положительными коэффициентами. Если φ — характеристическая функция какого-либо распределения F , то $P(\varphi)$ — безгранично делимая характеристическая функция. Опишите соответствующую каноническую меру M в терминах $F^{n\star}$.

¹⁾ Характеристическая функция ω равна бесконечному произведению, которое совпадает с хорошо известным каноническим произведением для функции $\frac{2\pi|\xi|}{e^{\pi|\xi|} - e^{-\pi|\xi|}}$.

Интересный частный случай: если $0 \leq a < b < 1$, то функция

$$\frac{1-b}{1-a} \cdot \frac{1-af}{1-bf}$$

будет безгранично делимой характеристической функцией. (См. также задачу 19.)

8. *Продолжение.* Интерпретируйте $P(\varphi)$ в терминах рандомизации и подчиненных процессов, принимая во внимание тот факт, что P — производящая функция безгранично делимой целочисленной случайной величины.

9. Пусть случайная величина X имеет устойчивое распределение с характеристической функцией $e^{-|\zeta|^\alpha}$ ($0 < \alpha \leq 2$), и пусть Y не зависит от X . Если Y — положительная случайная величина с распределением G , сосредоточенным на $0, \infty$, то покажите, что характеристическая функция $XY^{1/\alpha}$ имеет вид

$$\int_0^\infty e^{-|\zeta|^\alpha y} G(dy).$$

Вывод: если X и Y — независимые строго устойчивые случайные величины с показателями α и β и если $Y > 0$, то $XY^{1/\alpha}$ — строго устойчивая величина с показателем $\alpha\beta$.

10. Пусть ω — такая характеристическая функция, что $\omega^2(\zeta) = \omega(a\zeta)$ и $\omega^3(\zeta) = \omega(b\zeta)$. Тогда ω устойчива.

[Пример 3, е) показывает, что одного первого условия недостаточно. Показатели степени 2 и 3 могут быть заменены любыми двумя взаимно простыми числами.]

11. Покажите, что простая лемма 3 из гл. VIII, 8, применима не только к монотонным функциям, но и к логарифмам характеристических функций. Выведите отсюда, что если $\omega^n(\zeta) = \omega(a_n \zeta)$ при всех n , то $\log \omega(\zeta) = A\zeta^\alpha$ при $\zeta > 0$, где A — комплексная постоянная.

12. *Продолжение.* Используя результат задачи 28 из гл. VIII, 10, покажите непосредственно, что если ω — устойчивая характеристическая функция, то при $\zeta > 0$ или $\log \omega(\zeta) = A\zeta^\alpha + ib\zeta$, или $\log \omega(\zeta) = A\zeta^\alpha + ib\zeta \log \zeta$, где b — действительное число.

13. Пусть F сосредоточено на $0, \infty$ и $1 - F(x) \sim x^{-\alpha} L(x)$ с $0 < \alpha < 1$, где L — медленно меняющаяся на бесконечности функция. Докажите, что $1 - \varphi(\zeta) \sim A\zeta^\alpha L(1/\zeta)$ при $\zeta \rightarrow 0+$.

14. *Продолжение.* Опираясь на результаты § 5, докажите обратное утверждение, а также докажите, что $A = \Gamma(1-\alpha)e^{-i\pi\alpha/2}$.

15. *Продолжение.* Докажите индукцией по k следующее утверждение: для того чтобы при $x \rightarrow \infty$ было $1 - F(x) \sim ax^{-\alpha} L(x)$, где L — медленно меняющаяся функция и $k < \alpha < k+1$, необходимо и достаточно, чтобы при $\zeta \rightarrow 0+$ было

$$\varphi(\zeta) - 1 - \frac{\mu_1(i\zeta)}{1!} - \dots - \frac{\mu_k(i\zeta)^k}{k!} \sim A\zeta^\alpha L\left(\frac{1}{\zeta}\right). \quad (*)$$

Тогда автоматически $A = -a\Gamma(k-\alpha)e^{-i\pi\alpha/2}$.

16. Сформулируйте слабый закон больших чисел для схемы серий как частный случай общей теоремы § 7.

17. Пусть $\{X_{k,n}\}$ — нулевая схема, в которой суммы по строкам имеют предельное распределение с канонической мерой M . Покажите, что при $x > 0$

$$P\{\max[X_{1,n}, \dots, X_{r,n}] \leq x\} \rightarrow e^{-M^+(x)}.$$

Сформулируйте обратное утверждение.

18. Пусть $\{X_{k,n}\}$ — нулевая схема из симметричных случайных величин, в которой суммы по строкам имеют предельное распределение с канонической

мерой M , имеющей атом массы σ^2 в нуле. Покажите, что распределение случайной величины $S_n^* = \sum X_{k,n}^2 - \sigma^2$ сходится к распределению, определяемому мерой M_* без атома в нуле и такой, что $M_*^+(x) = 2M^+(\sqrt{x})$ при $x > 0$.

19. Пусть $0 < r_j < 1$ и $\sum r_j < \infty$. При любых действительных a_j бесконечное произведение

$$\frac{1-r_1}{1-r_1 e^{ia_1 \zeta}} \cdot \frac{1-r_2}{1-r_2 e^{ia_2 \zeta}} \cdots$$

сходится и представляет собой безгранично делимую характеристическую функцию. Указание. В соответствии с задачей 7 каждый множитель безгранично делим.

20. Используйте метод примера 9, г) для построения распределения F , такого, что $\limsup F^{n*}(x) = 1$ и $\liminf F^{n*}(x) = 0$ во всех точках.

21. Возьмем в (11.4) в качестве W равномерное распределение. Тогда

$$\psi(\xi_1, \xi_2) = -c [\xi_1^2 + \xi_2^2]^{\alpha/2},$$

так что e^{ψ} — симметричное устойчивое распределение.

22. Возьмем в (11.4) в качестве W распределение с массами $1/4$ в четырех точках $0, \pi, \frac{1}{2}\pi, -\frac{1}{2}\pi$. Тогда (11.4) определяет двумерную характеристическую функцию пары независимых одномерных устойчивых случайных величин.

23. Пусть распределение W в (11.4) сосредоточено в двух точках σ и $\sigma + \pi$. Тогда (11.4) задает вырожденную характеристическую функцию такой пары, что

$$X_1 \sin \sigma - X_2 \cos \sigma = 0.$$

Вообще любой дискретной мере W соответствует свертка вырожденных распределений. Объясните (11.4) с помощью предельного перехода.

В этой главе в значительной степени повторяются вопросы, уже разобранные в гл. XII, и по этой причине применения сведены до минимума. Была предпринята серьезная попытка построить эту главу независимо от других и сделать ее доступной при минимальных предварительных сведениях (исключая анализ Фурье из гл. XV). Излагаемая теория совсем не связана с содержанием двух предшествующих глав. Содержание § 6 не зависит от § 1—5.

§ 1. ОСНОВНОЕ ТОЖДЕСТВО

Всюду в этой главе $X_1, X_2, \dots$ обозначают взаимно независимые случайные величины с одним и тем же распределением F и характеристической функцией φ . Как обычно, мы полагаем $S_0 = 0$ и $S_n = X_1 + \dots + X_n$; последовательность $\{S_n\}$ образует случайное блуждание, порожденное F .

Пусть A — произвольное множество на действительной прямой и A' — его дополнение (в большинстве применений A' будет конечным или бесконечным интервалом). Если I — подмножество (интервал) из A' и если

$$S_1 \in A, \dots, S_{n-1} \in A, \quad S_n \in I \quad (I \subset A'), \quad (1.1)$$

то мы скажем, что (*первое*) *вхождение*¹⁾ в A' происходит в момент n в точке из I .

Так как вхождение в A' не обязано быть достоверным событием, то *момент вхождения* N может оказаться несобственной случайной величиной, равно как и *точка* S_N *первого вхождения*. Для совместного распределения пары (N, S_N) мы примем обозначение

$$P\{N=n, S_n \in I\} = H_n\{I\}, \quad n=1, 2, \dots \quad (1.2)$$

Таким образом, $H_n\{I\}$ есть вероятность события (1.1). Дополнительное соглашение, что $H_n\{I\} = 0$ при $I \subset A$, позволяет определить распределение (1.2) для всех I на числовой прямой. Вероятности (1.2) будут называться *вероятностями достижения*. Их изучение неотделимо от изучения случайного блуждания до момента первого вхождения в A' , т. е., иными словами, *случайного блуждания, ограниченного множеством A* . Положим для $I \subset A$ и

¹⁾ Иногда мы будем говорить «достижение A' ». — Прим. перев.

$n = 1, 2, \dots$

$$G_n \{I\} = P \{S_1 \in A, \dots, S_{n-1} \in A, S_n \in I\}, \quad (1.3)$$

т. е. $G_n(I)$ есть вероятность того, что в момент n блуждание приводит в $I \subset A$ и что до момента n не произошло вхождения в A' . Мы распространим это определение на все множества I , считая $G_n \{I\} = 0$ при $I \subset A'$.

Отметим, что, согласно нашему определению,

$$G_n \{A\} = 1 - P \{N \leq n\}. \quad (1.4)$$

Случайная величина N будет собственной тогда и только тогда, когда эта величина стремится к нулю при $n \rightarrow \infty$.

Рассматривая положение S_n блуждающей точки в моменты $n = 1, 2, \dots$, видим, что при $I \subset A'$

$$H_{n+1} \{I\} = \int_A G_n \{dy\} F \{I - y\}, \quad (1.5a)$$

а при $I \subset A$

$$G_{n+1} \{I\} = \int_A G_n \{dy\} F \{I - y\}. \quad (1.5b)$$

Условимся теперь обозначать буквой G_0 *распределение вероятностей, сосредоточенное в нуле*. Тогда соотношения (1.5) верны при $n = 0, 1, 2, \dots$, и с их помощью можно последовательно вычислить все вероятности H_n и G_n . Эти два соотношения можно объединить в одно. Разобьем произвольное множество I на числовой прямой на части IA' и IA и применим к ним (1.5). Вспоминая, что H_n и G_n сосредоточены на A' и A соответственно, получаем для $n = 0, 1, \dots$ и произвольного I

$$H_{n+1} \{I\} + G_{n+1} \{I\} = \int_A G_n \{dy\} F \{I - y\}. \quad (1.6)$$

Частный случай $A = \overline{0, \infty}$ разбирался в гл. XII, 3. Соотношение (3.5) гл. XII играло там ту же роль, что (1.5) здесь. Мы могли бы повторить здесь прежний путь и вывести интегральное уравнение типа Винера—Хопфа, аналогичное (3.9) гл. XII. Оно снова имело бы только одно вероятностное решение (хотя эта единственность решения не абсолютна). Однако на этот раз предпочтительнее опираться на мощный метод анализа Фурье.

Нас интересует распределение пары (N, S_N) . Так как N — целочисленная случайная величина, то мы будем использовать производящую функцию для N и характеристическую для S_N . Соответственно этому положим

$$\chi(s, \zeta) = \sum_{n=1}^{\infty} s^n \int_{A'} e^{i\zeta x} H_n \{dx\}, \quad \gamma(s, \zeta) = \sum_{n=0}^{\infty} s^n \int_A e^{i\zeta x} G_n \{dx\} \quad (1.7)$$

(члены с $n=0$ в этих двух рядах равны соответственно 0 и 1.) Оба ряда сходятся по крайней мере при $|s| < 1$.

Для ясности здесь указана существенная часть области интегрирования; можно было бы взять интеграл в пределах от $-\infty$ до ∞ . В частности, интеграл в (1.6) есть обычная свертка. После применения преобразования Фурье—Стилтьеса соотношение (1.6) принимает вид

$$\chi_{n+1}(\zeta) + \gamma_{n+1}(\zeta) = \gamma_n(\zeta) \varphi(\zeta). \quad (1.8)$$

Умножая на s^{n+1} и суммируя по $n=0, 1, \dots$, получаем

$$\chi(s, \zeta) + \gamma(s, \zeta) - 1 = s\gamma(s, \zeta) \varphi(\zeta)$$

для всех s , при которых ряд (1.7) сходится. Таким образом, мы получили *основное тождество*

$$1 - \chi = \gamma [1 - s\varphi]. \quad (1.9)$$

(Другое доказательство см. в задаче 6.)

В принципе χ и γ могут быть вычислены, исходя из (1.5), и поэтому тождество (1.9) может на первый взгляд показаться избыточным. В действительности прямые вычисления редко можно довести до конца. В то же время непосредственно из (1.9) можно извлечь много ценной информации.

Пример. Пусть F обозначает двустороннее показательное распределение с плотностью $\frac{1}{2} e^{-|x|}$ и характеристической функцией

$\varphi(\zeta) = 1/(1 + \zeta^2)$, и пусть $A = -\frac{1}{a}, \frac{1}{a}$. При $x > a$ из (1.5а) получаем

$$H_{n+1}\{\overline{x, \infty}\} = H_{n+1}\{\overline{-\infty, -x}\} = \frac{1}{2} \int_{-a}^a G_n\{dy\} e^{-(x-y)} = c_n e^{-x}, \quad (1.10)$$

где c_n не зависит от x . Отсюда следует, что точка S_N первого вхождения в множество $|x| > a$ не зависит от момента этого вхождения и имеет плотность, пропорциональную $e^{-|x|}$ (при $|x| > a$). Этот результат, как подсказывает интуиция, согласуется со свойством отсутствия памяти у показательного распределения, описанного в гл. 1. Независимость означает, что совместная характеристическая функция χ может быть разложена на множители, и из вида плотности S_N заключаем, что

$$\chi(s, \zeta) = \frac{1}{2} P(s) \left[\frac{e^{ia\zeta}}{1 - i\zeta} + \frac{e^{-ia\zeta}}{1 + i\zeta} \right], \quad (1.11)$$

где P —производящая функция момента N первого вхождения в множество $|x| > a$. (Множитель $1/2$ появляется вследствие того, что $\chi(1, 0) = 1$.)

Прямое вычисление $P(s)$ громоздко, однако явное выражение для нее легко может быть получено из (1.9). Действительно, в нашем случае правая часть (1.9) равна нулю при $\zeta = \pm i\sqrt{1-s}$, поэтому при таких ζ величина $\chi(s, \zeta)$ должна быть равна единице.

Таким образом,

$$P(s) = 2 \left[\frac{e^{-a\sqrt{1-s}}}{1+\sqrt{1-s}} + \frac{e^{a\sqrt{1-s}}}{1-\sqrt{1-s}} \right]^{-1}. \quad (1.12)$$

Отсюда следует, что момент N первого вхождения в множество $|x| > a$ имеет математическое ожидание $1 + a + \frac{1}{2} a^2$.
(Другие примеры см. в задачах 1—5). ▶

§ 2*). КОНЕЧНЫЕ ИНТЕРВАЛЫ. ВАЛЬДОВСКАЯ АППРОКСИМАЦИЯ

Теорема. Пусть $A = \overline{-a, b}$ — конечный интервал, содержащий точку нуль, и пусть пара (N, S_N) соответствует достижению дополнения A' к A .

Тогда случайные величины N и S_N будут собственными. Ряд для производящей функции

$$\sum_{n=0}^{\infty} s^n P\{N > n\} = \sum_{n=0}^{\infty} s^n G_n\{A\} \quad (2.1)$$

сходится при некотором¹⁾ $s > 1$, и потому N имеет моменты всех порядков. Величина S_N имеет математическое ожидание тогда и только тогда, когда распределение F , определяющее случайное блуждание, имеет математическое ожидание μ . В последнем случае

$$E(S_N) = \mu \cdot E(N). \quad (2.2)$$

Тождество (2.2) впервые было получено А. Вальдом. В частном случае $A = \overline{0, \infty}$ оно сводится к (2.8) гл. XII.

Доказательство. Как уже отмечалось, $G_n\{A\}$ и $P\{N > n\}$ есть разные обозначения для вероятности того, что случайное блуждание не выйдет из множества A за время n , и поэтому выражения в обеих частях (2.1) идентичны.

Выберем целое число r так, что $P\{|S_r| < a + b\} = \eta < 1$. Событие $\{N > n + r\}$ не может осуществиться, если не выполняется хотя бы одно из неравенств

$$N > n \text{ и } |X_{n+1} + \dots + X_{n+r}| < a + b.$$

(Эти два события независимы, потому что $\{N > n\}$ зависит только от $X_1, \dots, X_n$.) Так как $X_{n+1} + \dots + X_{n+r}$ и S_r одинаково распределены, имеем

$$P\{N > n + r\} \leq P\{N > n\} \eta.$$

^{*}) Включение в книгу этого параграфа объясняется его важностью в статистике; он может быть опущен при первом чтении.

¹⁾ Это утверждение известно статистикам как лемма Стейна. Другое доказательство см. в задаче 8.

Отсюда по индукции

$$P\{N > kr\} \leq \eta^k. \quad (2.3)$$

Последнее неравенство показывает, что последовательность $P\{N > n\}$ убывает не медленнее геометрической прогрессии со знаменателем $\eta^{1/r}$. Отсюда следует, что N является собственной случайной величиной и что ряд (2.1) сходится по крайней мере при $|s| < \eta^{-1/r}$. Это доказывает первое утверждение.

Отсюда следует также, что соотношение (1.9) справедливо при $|s| < \eta^{-1/r}$. При $s = 1$ получаем

$$1 - \chi(1, \xi) = \gamma(1, \xi) [1 - \varphi(\xi)]. \quad (2.4)$$

Но $\chi(1, \xi)$ есть характеристическая функция S_N , и то, что $\chi(1, 0) = 1$, показывает, что S_N — собственная случайная величина.

Событие $|S_N| > t + a + b$ не может осуществиться, если ни при каком n не выполняются неравенства $N > n - 1$ и $|X_n| > t$. Как уже отмечалось, последние события независимы, и так как X_n одинаково распределены, то

$$\begin{aligned} P\{|S_N| > t + a + b\} &\leq \\ &\leq \sum_{n=1}^{\infty} P\{N > n - 1\} \cdot P\{|X_1| > t\} = E(N) \cdot P\{|X_1| > t\}. \end{aligned}$$

Математическое ожидание $\mu = E(X_1)$ существует тогда и только тогда, когда правая часть интегрируема на $0, \infty$. В этом случае интегрируема и левая часть, и, следовательно, $E(S_N)$ существует. С другой стороны,

$$P\{|S_N| > t\} \geq P\{|X_1| > t + a + b\},$$

потому что осуществление события в правой части влечет за собой $S_N = X_1$. Таким образом, из существования $E(S_N)$ следует существование $\mu = E(X_1)$. Когда эти математические ожидания существуют, можно, дифференцируя (2.4), получить

$$iE(S_N) = \frac{\partial \chi(1, 0)}{\partial \xi} = \varphi'(0) \gamma(1, 0) = i\mu E(N). \quad (2.5) \blacktriangleright$$

Перейдем теперь к выводу одного из вариантов основного тождества (1.9), известного как тождество Вальда. Чтобы избежать использования мнимого аргумента, введем

$$f(\lambda) = \int_{-\infty}^{+\infty} e^{-\lambda x} F\{dx\}. \quad (2.6)$$

Предположим, что этот интеграл сходится в некотором интервале $-\lambda_0 < \lambda < \lambda_1$, содержащем точку нуль. Тогда характеристическая функция определяется соотношением $\varphi(i\lambda) = f(\lambda)$, причем $\varphi(z)$ аналитична в полосе $-\lambda_0 < \operatorname{Im}(z) < \lambda_1$. Формально тождество Вальда получается подстановкой $\xi = i\lambda$ и $s = 1/\varphi(i\lambda)$ в (1.9). При

таких значениях ξ и s правая часть (1.9) обращается в нуль и, следовательно, $\chi(s, \xi) = 1$. Исходя из определения χ , этот результат может быть сформулирован на вероятностном языке следующим образом.

Тождество Вальда¹⁾. Если интеграл (2.6) сходится при $-\lambda_0 < \lambda < \lambda_1$, то в этом интервале

$$E(f^{-N}(\lambda) e^{-\lambda S_N}) = 1. \quad (2.7)$$

Доказательство. Повторим рассуждение, приведшее к формуле (1.9). Так как меры G_n сосредоточены на конечных интервалах, их преобразования Фурье χ_n определены при всех комплексных ξ . По предположению $\varphi(i\lambda) = f(\lambda)$ существует, и, следовательно, соотношение (1.8) справедливо при $\xi = i\lambda$. После умножения на $f^{-n-1}(\lambda)$ оно принимает вид

$$f^{-n-1}(\lambda) \chi_{n+1}(i\lambda) = f(\lambda)^{-n} \gamma_n(i\lambda) - f^{-n-1}(\lambda) \gamma_{n+1}(i\lambda). \quad (2.8)$$

Если $f^{-n}(\lambda) \gamma_n(i\lambda) \rightarrow 0$, сумма по n правых частей (2.8) равна 1. В этом случае суммирование (2.8) приводит к утверждению (2.7), и, следовательно, достаточно показать, что

$$f^{-n}(\lambda) G_n\{A\} \rightarrow 0. \quad (2.9)$$

Теперь если $f(\eta) < \infty$, то

$$\begin{aligned} G_n\{A\} &\leq P\{-a < S_n < b\} \leq \\ &\leq e^{(a+b)|\eta|} \int_{-a}^b e^{-nx} F_n^*(dx) \leq e^{(a+b)|\eta|} \cdot f^n(\eta). \end{aligned}$$

Поэтому (2.9) справедливо, если $f(\lambda) > f(\eta)$. Так как выбор η находится в нашем распоряжении, то тем самым (2.7) доказано для всех точек, кроме тех, где f достигает своего минимума. Но, будучи выпуклой функцией, f может иметь не более одной точки минимума. Для нее (2.7) верно по непрерывности. ►

Пример. Оценки для распределения N . Вальд пришел к тождеству (2.7) от задач последовательного анализа, в которых было необходимо получить приближения для распределения момента N первого выхода из интервала A , а также оценить вероятности того, что выход произойдет вправо или влево от этого интервала. Метод Вальда является обобщением процедуры для арифметических распределений с конечным числом скачков, описанной в I; XIV, 8. (Там же показано, как можно получить строгие неравенства.)

¹⁾ Вальд применял (2.7) в связи с последовательным анализом. Это было еще до 1945 г. и до того времени, когда началось систематическое изучение общих случайных блужданий. Естественно поэтому, что его условия были жесткими, а методы — трудными. Однако, к сожалению, их влияние на статистическую литературу все еще продолжается. Рассуждения в тексте используют одну идею Миллера (H. D. Miller, 1961).

Положим

$$p_k = P\{N = k, S_N \geq b\}, \quad q_k = P\{N = k, S_N \leq -a\} \quad (2.10)$$

и обозначим соответствующие производящие функции через $P(s)$ и $Q(s)$. (Тогда $P+Q$ будет производящей функцией для N .) Допустим теперь, что a и b велики по сравнению с математическим ожиданием и дисперсией F . Тогда S_N , по-видимому, будет близко к b или $-a$. Если бы только эти числа были возможными значениями S_N , то тождество (2.7) имело бы вид

$$P\left(\frac{1}{f(\lambda)}\right)e^{-\lambda b} + Q\left(\frac{1}{f(\lambda)}\right)e^{\lambda a} = 1. \quad (2.11)$$

Естественно ожидать, что при сделанных предположениях (2.11) будет приближенно выполняться. Функция f выпукла, и обычно можно найти такой интервал $s_0 < s < s_1$, что на нем уравнение

$$sf(\lambda) = 1 \quad (2.12)$$

имеет два корня $\lambda_1(s)$ и $\lambda_2(s)$, непрерывно зависящих от s . Подставляя их в (2.11), мы получаем два линейных уравнения для производящих функций P и Q и таким образом можем найти (по крайней мере, приближенно) распределение N и вероятности выхода вправо или влево. ►

§ 3. ФАКТОРИЗАЦИЯ ВИНЕРА — ХОПФА

В этом параграфе чисто аналитическими методами мы выведем различные следствия из основного тождества (1.9). Мы увидим, что они в более гибкой и отчетливой форме содержат в себе многие результаты, полученные в гл. XII комбинаторными методами. Это может создать ложное впечатление о превосходстве методов Фурье, однако в действительности именно взаимодействие этих двух методов характеризует последние успехи в данной теории. Каждый метод приводит к результатам, которые кажутся недоступными для другого (см., например, § 5; закон арксинуса для числа положительных частных сумм, а также теория симметрично зависимых случайных величин иллюстрируют преимущества комбинаторного подхода).

В дальнейшем N и S_N будут обозначать момент и точку *первого* вхождения в открытую полупрямую $\overline{0, \infty}$. Их совместное распределение

$$P\{N = n, S_N \in I\} = H_n\{I\}$$

задается соотношением

$$H_n\{I\} = P\{S_1 \leq 0, \dots, S_{n-1} \leq 0, S_n \in I\}, \quad I \subset \overline{0, \infty}, \quad (3.1)$$

где подразумевается, что H_n сосредоточено на $\overline{0, \infty}$, а $H_0 = 0$. Как и прежде, вместо двумерной характеристической функции

введем более удобную комбинацию производящей и характеристической функций

$$\chi(s, \zeta) = E(s^N e^{i\zeta S_N}), \quad (3.2)$$

то есть

$$\chi(s, \zeta) = \sum_{n=1}^{\infty} s^n \int_0^{\infty} e^{i\zeta x} H_n \{dx\}. \quad (3.3)$$

(Интегрирование производится по полупрямой, однако ничего не изменится, если нижний предел заменить на $-\infty$.) Для краткости мы будем называть χ «преобразованием» последовательности мер H_n .

Для момента и точки первого вхождения в *открытую отрицательную* полупрямую введем обозначения N^- и S_{N^-} ; тогда $\{H_n^-\}$ и χ^- — соответствующие распределение и преобразование.

Когда распределение F не является непрерывным, мы должны различать первые вхождения в открытую и замкнутую полупрямые. Поэтому необходимо рассматривать событие, заключающееся в «возвращении в нуль через отрицательные значения». Соответствующее распределение вероятностей $\{f_n\}$ задается соотношениями

$$\{f_n = P\{S_1 < 0, \dots, S_{n-1} < 0, S_n = 0\}, \quad n \geq 1. \quad (3.4)$$

Положим $f(s) = \sum_{n=1}^{\infty} f_n s^n$. Вскоре будет показано, что правая часть (3.4) не изменится, если все неравенства заменить на противоположные. Очевидно, $\sum f_n \leq P\{X_1 < 0\} < 1$.

Используя эти обозначения, можно теперь сформулировать основную теорему.

Теорема о факторизации Винера — Хопфа. При $|s| \leq 1$ справедливо тождество

$$1 - s\varphi(\zeta) = [1 - f(s)] \cdot [1 - \chi(s, \zeta)] \cdot [1 - \chi^-(s, \zeta)]. \quad (3.5)$$

Доказательство приведет к явным выражениям для f и χ , которые мы оформим в виде отдельных лемм¹⁾.

Лемма 1. При $0 \leq s < 1$

$$\log \frac{1}{1 - \chi(s, \zeta)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \int_0^{\infty} e^{i\zeta x} F_n^* \{dx\}. \quad (3.6)$$

¹⁾ При $\zeta=0$ лемма 1 сводится к теореме 1 из гл. XII, 7. Обобщенный вариант (9.3) из гл. XII эквивалентен лемме 1, однако менее удобен для использования. Лемма 2 представляет собой иную формулировку (9.6) из гл. XII. Она принадлежит Г. Бакстеру (G. Baxter). Значительно упрощенное (но все еще технически сложное) доказательство было дано Спизером, F. Spitzer, Trans. Amer. Math. Soc., vol. 94 (1960), p. 150—169.

Аналогичная формула для χ^- получается из соображений симметрии.

Лемма 2. При $0 \leq s \leq 1$

$$\log \frac{1}{1-f(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbf{P}\{S_n = 0\}. \quad (3.7)$$

Так как в правую часть (3.7) не входят неравенства, то (3.4) остается неизменным если все неравенства в нем заменены на противоположные. Этот результат был получен в примере 2, а) гл. XII как следствие принципа двойственности.

Примечательная черта факторизации (3.5) в том, что с ее помощью произвольная характеристическая функция φ представима через два (возможно, несобственных) распределения, сосредоточенных на двух непересекающихся полупрямых. Лемма 1 показывает, что это представление единственно.

Для непрерывных распределений доказательство простое. В общем случае нам понадобится аналог леммы 1 для вероятностей вхождения в замкнутую полупрямую $\overline{0, \infty}$. Они будут обозначаться $R_n\{I\}$ для любого интервала $I \subset \overline{0, \infty}$, т. е.

$$R_n\{I\} = \mathbf{P}\{S_1 < 0, \dots, S_{n-1} < 0, S_n \in I\}. \quad (3.8)$$

Конечно, $R_0 = 0$ и $R_n\{\overline{-\infty, 0}\} = 0$.

Лемма 3. При $0 \leq s < 1$ преобразование ρ последовательности $\{R_n\}$ определяется соотношением

$$\log \frac{1}{1-\rho(s, \zeta)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \int_{0-}^{\infty} e^{i\zeta x} F_n^* \{dx\}. \quad (3.9)$$

Доказательство. Применим основное тождество (1.9) к $A = \overline{0, \infty}$. Для вероятностей вхождения R_n (1.9) записывается в виде

$$1 - \rho(s, \zeta) = \gamma(s, \zeta) [1 - s\varphi(\zeta)]. \quad (3.10)$$

Здесь γ —преобразование последовательности вероятностей G_n , определяемых на $\overline{-\infty, 0}$ соотношениями

$$G_n\{I\} = \mathbf{P}\{S_1 < 0, \dots, S_{n-1} < 0, S_n < 0, S_n \in I\}, \quad (3.11)$$

т. е.

$$\gamma(s, \zeta) - 1 = \sum_{n=1}^{\infty} s^n \int_{-\infty}^{0-} e^{i\zeta x} G_n \{dx\}. \quad (3.12)$$

При фиксированном $|s| < 1$ функции $1 - s\varphi(\zeta)$ и $1 - \gamma(s, \zeta)$ не имеют нулей и, следовательно, (см. гл. XVII, 1) их логарифмы

однозначно определяются как непрерывные функции ζ , равные нулю в нуле. Поэтому (3.10) можно переписать в виде

$$\log \frac{1}{1-s\varphi(\zeta)} = \log \frac{1}{1-\rho(s, \zeta)} + \log \gamma(s, \zeta), \quad (3.13)$$

или

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{s^n}{n} \int_{-\infty}^{+\infty} e^{i\zeta x} F^{n*} \{dx\} = \\ = \sum_{n=1}^{\infty} \frac{s^n}{n} \rho^n(s, \zeta) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n} [\gamma(s, \zeta) - 1]^n. \end{aligned} \quad (3.14)$$

Рассмотрим это соотношение при фиксированном $0 < s < 1$. Тогда $\rho^n(s, \zeta)$ — характеристическая функция несобственного вероятностного распределения, сосредоточенного на $\overline{0, \infty}$, и, следовательно, первый ряд справа представляет собой преобразование Фурье —

Стилтьеса конечной меры, сосредоточенной на $\overline{0, \infty}$. Аналогично, (3.12) показывает, что $\rho(s, \zeta) - 1$ является преобразованием Фурье — Стилтьеса конечной меры, сосредоточенной на $\overline{-\infty, 0}$. То же самое справедливо для $[\rho(s, \zeta) - 1]^n$, и поэтому второй ряд в правой части представляет собой преобразование разности двух мер на $\overline{-\infty, 0}$. Отсюда следует, что для множеств из $\overline{0, \infty}$ первые два ряда в (3.14) представляют одну и ту же конечную меру, а именно $\sum (s^n/n) F^{n*}$. Соотношение (3.9) выражает это свойство в терминах соответствующих преобразований.

Доказательство леммы 2. Эта лемма фактически содержится в лемме 3, потому что обе части (3.7) суть массы атомов в нуле мер, чьи преобразования содержатся в (3.9). Это утверждение очевидно для правых частей. Рассмотрим левые части. По определению (3.8) атом R_n в нуле имеет массу f_n . Тогда $f(s)$ есть масса, приписываемая нулю мерой $\sum s^n R_n$, чье преобразование равно $\rho(s, \zeta)$. Поэтому мера с преобразованием $\sum \rho^n(s, \zeta)/n$ приписывает нулю массу $\sum f^n(s)/n = \log(1 - f(s))^{-1}$.

Доказательство леммы 1. Проведем доказательство двумя способами.

(i) Лемма 1 является аналогом леммы 3 для открытых полупрямых и приемлет точно такое же доказательство. Если считать обе леммы доказанными, то, вычитая (3.6) из (3.9), получим

$$\rho(s, \zeta) = f(s) + [1 - f(s)]\chi(s, \zeta). \quad (3.15)$$

(Это тождество утверждает, что первое вхождение в $\overline{0, \infty}$ может быть возвращением в нуль через отрицательные значения и что, если такого возвращения в нуль не было, (условное) распреде-

ление точки первого вхождения в $\overline{0, \infty}$ равно распределению $\{H_n\}$ первого вхождения в $\overline{0, \infty}$.)

(ii) Иначе, соотношение (3.15) можно доказать непосредственно, исходя из определений (3.1) и (3.8) мер H_n и R_n . [Для этого достаточно в (3.8) рассмотреть последний индекс $k \leq n$, при котором $S_k = 0$, и взять точку $(k, 0)$ в качестве нового начала координат.] Подставляя (3.15) в (3.9), получаем лемму 1 как следствие лемм 2 и 3. ►

Доказательство теоремы о факторизации. Складывая тождества лемм 1, 2 и аналога леммы 1 для $\overline{-\infty, 0}$, получаем соотношение, представляющее собой логарифмический эквивалент (3.5). То, что (3.5) выполняется при $z=1$, следует по непрерывности.

Следствие.

$$\gamma(s, \zeta) = \frac{1}{1 - \chi^-(s, \zeta)}. \quad (3.16)$$

Доказательство. Из леммы 3 и (3.13) имеем

$$\gamma(s, \zeta) = \exp \left(\sum_{n=1}^{\infty} \frac{s^n}{n} \int_{-\infty}^{0-} e^{i\zeta x} F^{n*} \{dx\} \right), \quad (3.17)$$

и в силу леммы 1 правые части в (3.16) и (3.17) совпадают. ►

Примеры. а) *Биномиальное случайное блуждание.* Пусть

$$P\{X_1 = 1\} = p \text{ и } P\{X_1 = -1\} = q.$$

Первое вхождение в положительную и отрицательную полупрямые обязательно происходит через точки $+1$ и -1 соответственно, и, следовательно,

$$\chi(s, \zeta) = P(s) e^{i\zeta}, \quad \chi^-(s, \zeta) = Q(s) e^{-i\zeta}, \quad (3.18)$$

где P и Q — производящие функции моментов первого вхождения. Поэтому обе части факторизационного соотношения (3.5) являются линейными комбинациями трех величин $e^{ik\zeta}$ при $k=0, \pm 1$. Приравнявая коэффициенты, получаем три уравнения

$$\begin{aligned} [1 - f(s)][1 + P(s)Q(s)] &= 1, & [1 - f(s)]P(s) &= sp, \\ [1 - f(s)]Q(s) &= sq. \end{aligned} \quad (3.19)$$

Эти уравнения сводятся к квадратному уравнению для $1 - f(s)$; его решение, удовлетворяющее условию $f(0) = 0$, имеет вид

$$f(s) = \frac{1}{2} (1 - \sqrt{1 - 4pq s^2}). \quad (3.20)$$

Производящие функции P и Q находятся теперь из (3.19). Если $p > q$, имеем $f(1) = q$ и, следовательно, $Q(1) < 1$. Таким образом,

теорема о факторизации непосредственно приводит к распределениям времен первого достижения и возвращения, найденным другими методами в 1; XI, и 1; XIV.

б) *Конечные арифметические распределения.* Тот же метод в принципе применим, когда F сосредоточено в целых точках между $-a$ и b . Преобразования χ , χ^- и f в этом случае определяются $a+b+1$ уравнениями, однако решение в явном виде получить трудно (см. пример 4, в) гл. XII).

в) Пусть F — свертка двух показательных распределений, одно из которых сосредоточено на положительной, а другое — на отрицательной полуосях, т. е. пусть

$$\varphi(\zeta) = \frac{a}{a+i\zeta} \cdot \frac{b}{b-i\zeta}, \quad a > 0, \quad b > 0. \quad (3.21)$$

В силу непрерывности F имеем $f(s)=0$ тождественно. Левая часть формулы (3.5) в теореме факторизации имеет полюс в точке $\zeta = -ib$, но $\chi^-(s, \zeta)$ регулярна в окрестности любой точки ζ с отрицательной мнимой частью (это следует из того, что χ^- является преобразованием меры, сосредоточенной на $-\infty, 0$). Отсюда следует, что χ должно иметь вид $\chi(s, \zeta) = (b-i\zeta)^{-1} U(s, \zeta)$, где U регулярна при всех ζ . Можно поэтому предполагать, что U не будет зависеть от ζ , т. е. χ и χ^- будут иметь вид

$$\chi(s, \zeta) = \frac{P(s)}{b-i\zeta}, \quad \chi^-(s, \zeta) = \frac{Q(s)}{a+i\zeta}. \quad (3.22)$$

Для этого необходимо, чтобы

$$1-s \frac{ab}{(a+i\zeta)(b-i\zeta)} = \left(1 - \frac{P(s)}{b-i\zeta}\right) \left(1 - \frac{Q(s)}{a+i\zeta}\right). \quad (3.23)$$

Освобождаясь от знаменателей и приравнивая коэффициенты при одинаковых степенях ζ , находим, что $P(s)=Q(s)$ и что $P(s)$ удовлетворяет квадратному уравнению. Условию $P(0)=0$ удовлетворяет только один из двух корней этого уравнения, и в результате мы получаем

$$P(s)=Q(s) = \frac{1}{2} [a+b - \sqrt{(a+b)^2 - 4abs}]. \quad (3.24)$$

Предположим, что $a > b$. Тогда $P(1)=b$, и, следовательно, $P(s)/b$ и $Q(s)/a$ — производящие функции соответственно собственного и несобственного распределений вероятностей. Следовательно, функция χ , определенная соотношением (3.22), является преобразованием пары (N, S_N) , такой, что S_N не зависит от N и имеет характеристическую функцию $b/(b-i\zeta)$. Такое же утверждение справедливо для χ^- , и в силу единственности факторизации $P(s)/b$ и $Q(s)/a$ на самом деле являются производящими функциями моментов N и N^- первого вхождения. [То, что S_N и S_{N^-} показательно распределены, было получено также в примере 4, а)

гл. XII. Напомним (см. пример 9, д), гл. VI), что распределения вида (3.21) играют важную роль в теории очередей.] ►

Другие примеры см. в задачах 9—11.

§ 4. ВЫВОДЫ И ПРИМЕНЕНИЯ

Перейдем к анализу результатов предыдущего параграфа с вероятностной точки зрения и установим их связь с некоторыми результатами, полученными в гл. XII.

(i) *Принцип двойственности.* Сначала покажем, что следствие (3.16) эквивалентно следующему утверждению.

Лемма 1. Для любого интервала $I \subset \overline{0, \infty}$

$$\begin{aligned} P\{S_1 < S_n, \dots, S_{n-1} < S_n, S_n \in I\} = \\ = P\{S_1 > 0, \dots, S_{n-1} > 0, S_n \in I\}. \end{aligned} \quad (4.1)$$

Это равенство мы получили в (2.1) гл. XII, рассматривая величины $X_1, \dots, X_n$ в обратном порядке. Лемма, будучи оцениваема с этой точки зрения, кажется почти очевидной, однако мы видели, что ее простыми следствиями являются многие важные соотношения. Она не играет никакой роли при исследовании с помощью преобразований Фурье, но примечательно, что она может быть получена как строгий аналитический результат¹⁾. (Чтобы вспомнить удивительные следствия леммы 1, касающиеся блужданий, читатель может обратиться к примеру 2, б) гл. XII.)

Доказательство. Следствие (3.16) относится к отрицательной полуоси, и для прямого сравнения все неравенства в (4.1) должны быть заменены на противоположные. Вероятность в правой части (4.1) совпадает тогда с вероятностью $G_n\{I\}$, введенной в (3.11), а $\gamma(s, \xi)$ — это просто соответствующее преобразование. Для доказательства леммы мы должны, следовательно, показать, что $[1 - \chi(s, \xi)]^{-1}$ является преобразованием последовательности вероятностей из левой части (4.1). Далее, $\chi(s, \xi)$ было определено как преобразование распределения точки (N, S_N) первого вхождения в $\overline{0, \infty}$, поэтому χ^r есть преобразование r -й лестничной точки (N_r, S_{N_r}) . Отсюда следует, что

$$\frac{1}{1 - \chi} - 1 = \chi + \chi^2 + \dots \quad (4.2)$$

¹⁾ Используемые нами аналитические методы довольно элементарны, однако исторически так сложилось, что первоначально отправным пунктом здесь служила теория Винера—Хопфа. Поэтому в большинстве работ используются глубокие методы теории функций комплексного переменного, которые на самом деле не уместны в теории вероятностей, так как даже оригинальная теория Винера—Хопфа значительно упрощается, если ограничиться положительными ядрами. См. обсуждение в гл. XII, За,

есть преобразование последовательности вероятностей того, что n есть лестничный момент и $S_n \in I$. Но именно эти вероятности стоят в левой части (4.1). Лемма доказана. ►

(ii) Момент N первого вхождения в $\overline{0, \infty}$ имеет производящую функцию τ , равную $\tau(s) = \chi(s, 0)$. Следовательно, из (3.6)

$$\log \frac{1}{1-\tau(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n > 0\}. \quad (4.3)$$

Эта формула была выведена комбинаторными методами в гл. XII, 7. Там же обсуждались различные ее следствия. В частности, устремляя в (4.3) $s \rightarrow 1$, видим, что случайная величина N будет собственной тогда и только тогда, когда ряд $\sum n^{-1} P\{S_n > 0\}$ расходится; в случае сходимости этого ряда имеет место снос в $-\infty$. Прибавляя $\log(1-s) = -\sum s^n/n$ к (4.3) и устремляя $s \rightarrow 1$, получаем

$$\log E(N) = \log \tau'(1) = \sum_{n=1}^{\infty} \frac{1}{n} P\{S_n \leq 0\} \quad (4.4)$$

при условии, что N — собственная случайная величина. Но мы уже видели, что ряд из (4.4) сходится тогда и только тогда, когда случайное блуждание уходит в ∞ . Следовательно, мы доказали следующее утверждение.

Лемма 2. Для того чтобы случайная величина N была собственной и $E(N) < \infty$, необходимо и достаточно, чтобы случайное блуждание уходило в ∞ .

Этот результат был получен в гл. XII, 2, другими методами. Другие свойства распределения N см. в гл. XII, 7.

(iii) О математическом ожидании точки S_N первого вхождения. Используя методы гл. XII, о распределении S_N можно было сказать немного. Теперь же, подставляя $s=1$ в (3.6), мы найдем характеристическую функцию величины S_N . Тем не менее предпочтительно получить некоторую существенную информацию о распределении S_N непосредственно из факторизационной формулы.

Лемма 3. Если S_N и S_{N-} являются собственными случайными величинами с конечными математическими ожиданиями, то F имеет нулевое математическое ожидание и дисперсию σ^2 , заданную равенством

$$\frac{1}{2} \sigma^2 = -[1-f(1)] \cdot E(S_N) \cdot E(S_{N-}). \quad (4.5)$$

Теорема 1 следующего параграфа показывает, что справедливо и обратное утверждение. Неожиданным следствием является то,

что для конечности математического ожидания S_N необходимо существование второго момента F .

Доказательство. При $s=1$ из (3.5) получаем

$$\frac{\varphi(\xi)-1}{\xi^2} = [1-f(1)] \frac{\chi(1, \xi)-1}{\xi} \cdot \frac{\chi^-(1, \xi)-1}{\xi}. \quad (4.6)$$

При $\xi \rightarrow 0$ дроби в правой части стремятся к производным характеристических функций χ и χ^- , т. е. к $iE(S_N)$ и $iE(S_{N-})$. Левая часть имеет поэтому конечный предел σ^2 , а это означает, что $\varphi'(0)=0$ и $\varphi''(0)=\frac{1}{2}\sigma^2$. Отсюда следует, что σ^2 — дисперсия F (см. следствие из гл. XV, 4). ►

Рассмотрим случай сноса блуждания в ∞ . Из лемм 1, 2 § 3 и соотношения (4.5) следует, что в этом случае при $s \rightarrow 1$ и $\xi \rightarrow 0$

$$\begin{aligned} [1-f(s)]^{-1} \cdot [1-\chi^-(s, \xi)]^{-1} &\rightarrow \\ &\rightarrow \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} P\{S_n \leq 0\}\right) = E(N) < \infty. \end{aligned} \quad (4.7)$$

По теореме о факторизации

$$\frac{\chi(1, \xi)-1}{\xi} = \frac{\varphi(\xi)-1}{\xi} \cdot \frac{1}{[1-f(1)][1-\chi^-(1, \xi)]}. \quad (4.8)$$

Устремляя $\xi \rightarrow 0$, получаем следующий важный результат:

$$E(S_N) = E(X_1) \cdot E(N), \quad (4.9)$$

при условии, что $E(S_N)$ и $E(X_1)$ существуют (последнее математическое ожидание положительно вследствие предполагаемого сноса в ∞).

Из приведенного доказательства можно получить и другие результаты. Наши рассуждения показывают, что левая часть в (4.8) сходится к конечному пределу тогда и только тогда, когда существует $\varphi'(0) = i\mu$. Как было показано в гл. XVII, 2а, это имеет место в том и только том случае, когда наше случайное блуждание подчиняется обобщенному слабому закону больших чисел, а именно

$$\frac{1}{n} S_n \xrightarrow{P} \mu \quad (4.10)$$

($\xrightarrow{P}$ означает сходимость по вероятности). Было показано также, что для положительных случайных величин μ совпадает с их математическим ожиданием. Таким образом, левая часть в (4.8) сходится к пределу тогда и только тогда, когда $E(N) < \infty$, а правая часть в (4.8) сходится тогда и только тогда, когда $\varphi'(0)$ существует. Таким образом, справедлива

Лемма 4. Если случайное блуждание уходит в ∞ , то $E(N) < \infty$ тогда и только тогда, когда существует такое число $\mu > 0$, что выполняется (4.10).

В гл. XII, 8, мы смогли показать только, что для указанного выше утверждения достаточно существования $E(X_1)$, но даже для этого более слабого результата нам потребовался усиленный закон больших чисел и обратная к нему теорема.

§ 5. ДВЕ БОЛЕЕ ОСНОВАТЕЛЬНЫЕ ТЕОРЕМЫ

Чтобы проиллюстрировать использование более тонких методов, мы докажем две теоремы, представляющие самостоятельный интерес. Первая уточняет лемму 3 предыдущего параграфа, вторая имеет приложения к теории очередей. Доказательства опираются на фундаментальные тауберовы теоремы; при доказательстве второй используются преобразования Лапласа.

Теорема 1. Если F имеет нулевое математическое ожидание и дисперсию σ^2 , то ряд

$$\sum_{n=1}^{\infty} \frac{1}{n} \left[P\{S_n > 0\} - \frac{1}{2} \right] = c \quad (5.1)$$

сходится по крайней мере условно и

$$E(S_N) = \frac{\sigma}{\sqrt{2}} e^{-c}. \quad (5.2)$$

Эта теорема принадлежит Спичеру. Сходимость ряда важна в связи с теоремами 1а из гл. XII, 7 и 8.

Доказательство. Дифференцируя (3.6) по ζ и полагая $\zeta=0$, получим

$$-i \frac{d\chi(s, 0)}{d\zeta} = \sum_{n=1}^{\infty} \frac{s^n}{n} \int_0^{\infty} x F^{n*} \{dx\} \exp \left[- \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n > 0\} \right]. \quad (5.3)$$

Оба ряда сходятся абсолютно при $|s| < 1$, так как коэффициенты при s^n ограничены. Действительно, в силу центральной предельной теоремы моменты порядка ≤ 2 случайных величин $S_n/\sigma\sqrt{n}$ сходятся к соответствующим моментам нормального распределения, а это означает, что при $n \rightarrow \infty$

$$\int_0^{\infty} x F^{n*} \{dx\} \sim \sigma \sqrt{\frac{n}{2\pi}}. \quad (5.4)$$

Следовательно, в соответствии с «абелевой» частью теоремы 5 из гл. XIII, 5, при $s \rightarrow 1$

$$\sum_{n=1}^{\infty} \frac{s^n}{n} \int_0^{\infty} x F^{n*} \{dx\} \sim \frac{\sigma}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{s^n}{\sqrt{n}} \sim \frac{\sigma}{\sqrt{2}} (1-s)^{-\frac{1}{2}}. \quad (5.5)$$

Левая часть в (5.3) стремится к математическому ожиданию $E(S_N)$, которое может быть конечным или бесконечным, но не нулевым. Комбинируя (5.3) и (5.5), получаем

$$E(S_N) = \frac{\sigma}{\sqrt{2}} \lim_{s \rightarrow 1} \exp \left[\sum_{n=1}^{\infty} \frac{s^n}{n} \left(\frac{1}{2} - P\{S_n > 0\} \right) \right]. \quad (5.6)$$

Выражение в показателе экспоненты стремится или к конечному пределу, или к $+\infty$. То же самое рассуждение применимо к N^- , т. е. к показателю экспоненты в (5.6), в котором $\{S_n > 0\}$ заменено на $\{S_n < 0\}$. Но сумма этих двух показателей равна $\sum (s^n/n) P\{S_n = 0\}$ и остается ограниченной при $s \rightarrow 1$. Отсюда следует, что показатель экспоненты в (5.6) остается ограниченным и, следовательно, стремится к конечному пределу $-c$. Так как коэффициенты ряда в показателе имеют порядок $o(n^{-1})$, отсюда следует¹⁾, что при $s=1$ ряд сходится к $-c$. Доказательство завершено. ►

Рассмотрим теперь случайное блуждание со сносом в $-\infty$ и положим

$$M_n = \max \{0, S_1, \dots, S_n\}. \quad (5.7)$$

Напомним (см. гл. VI, 9), что в приложениях к теории очередей M_n изображает время ожидания n -го требования. Метод доказательства следующей предельной теоремы, по-видимому, имеет больший интерес, чем сама теорема.

Теорема 2. Если случайное блуждание уходит в $-\infty$, то распределение U_n случайной величины M_n стремится к предельному распределению U , характеристическая функция которого ω задается соотношением

$$\omega(\xi) = \exp \left[\sum_{n=1}^{\infty} \frac{1}{n} \int_0^{\infty} (e^{i\xi x} - 1) F^{n*} \{dx\} \right]. \quad (5.8)$$

Заметим, что вследствие (4.3) $\sum n^{-1} P\{S_n > 0\} < \infty$, и поэтому ряд в (5.8) сходится абсолютно при всех ξ с положительной мнимой частью.

Доказательство. Пусть ω_n — характеристическая функция U_n . Покажем сначала, что при $|s| < 1$

$$\sum_{n=0}^{\infty} s^n \omega_n(\xi) = \frac{1}{1-s} \exp \left[\sum_{n=1}^{\infty} \frac{s^n}{n} \int_0^{\infty} (e^{i\xi x} - 1) F^{n*} \{dx\} \right]. \quad (5.9)$$

Событие $\{M_v \in I\}$ осуществляется тогда и только тогда, когда выполняются следующие два условия: первое — для некоторого $0 \leq n \leq v$ точка (n, S_n) является лестничной и $S_n \in I$, второе — $S_k - S_n \leq 0$ при всех $n < k \leq v$. Первое условие включает только $X_1, \dots, X_n$, второе — только $X_{n+1}, \dots, X_v$. Следовательно, эти два события независимы и поэтому

$$P\{M_v \in I\} = a_0 b_v + \dots + a_v b_0, \quad (5.10)$$

¹⁾ В силу элементарной оригинальной теоремы Таубера. См., например: Titchmarsh E. C. Theory of Functions, 2nd ed., Oxford 1939, p. 10.

где

$$a_n = P\{S_1 < S_n, \dots, S_{n-1} < S_n, S_n \in I\}, \quad b_n = P\{N > n\}. \quad (5.11)$$

Вероятности a_n появляются в левой части (4.1), и мы видели, что их преобразование равно $[1 - \chi(s, \zeta)]^{-1}$. Производящая функция последовательности $\{b_n\}$ равна $[1 - \tau(s)]/(1 - s)$, где τ определено в (4.3). Так как правая часть (5.10) представляет собой свертку $\{a_n\}$ и $\{b_n\}$, то произведение функций $[1 - \chi(s, \zeta)]^{-1}$ и $[1 - \tau(s)]/(1 - s)$ является преобразованием для вероятностей $P\{M_n \in I\}$, а соотношение (5.9) просто выражает этот факт.

Мы уже отмечали, что показатели экспонент в (5.8) и (5.9) регулярны при всех ζ , имеющих положительную мнимую часть. Поэтому, полагая $\zeta = i\lambda$ при $\lambda > 0$, мы можем ввести преобразование Лапласа

$$\omega_n(i\lambda) = \int_0^\infty e^{-\lambda x} U_n\{dx\} - \lambda \int_0^\infty e^{-\lambda x} U_n(x) dx. \quad (5.12)$$

Из монотонности последовательности максимумов M_n следует, что при фиксированном x последовательность $\{U_n(x)\}$ убывает, и, следовательно, при фиксированном λ преобразования Лапласа $\omega_n(i\lambda)$ образуют убывающую последовательность. Ввиду (5.9) при $s \rightarrow 1$ имеем

$$\sum_{n=0}^{\infty} s^n \omega_n(i\lambda) \sim \frac{1}{1-s} \omega(i\lambda), \quad (5.13)$$

и, используя последнюю часть тауберовой теоремы 5 из гл. XIII, 5, получаем $\omega_n(i\lambda) \rightarrow \omega(i\lambda)$. Отсюда вытекает сходимость $U_n \rightarrow U$. ►

§ 6. КРИТЕРИИ ВОЗВРАТНОСТИ

Материал этого параграфа не связан с предшествующей теорией; он посвящен методу, развитому Чжуном и Фуксом (1950) и позволяющему определить, является ли случайное блуждание возвратным или невозвратным. Несмотря на наличие указанных в гл. VI и XII критериев и методов, метод, основанный на анализе Фурье, сохраняет и методологический, и исторический интерес. Кроме того, в настоящее время это единственный метод, применимый в многомерном случае. В дальнейшем F обозначает одномерное распределение с характеристической функцией $\varphi(\zeta) = u(\zeta) + iv(\zeta)$.

Определим при $0 < s < 1$ конечную меру

$$U_s = \sum_{n=0}^{\infty} s^n F^{n*} \quad (6.1)$$

В соответствии с теорией, развитой в гл. VI, 10, распределение F

является невозвратным тогда и только тогда, когда для некоторого открытого интервала I , содержащего нуль, $U_s\{I\}$ ограничено при $s \rightarrow 1$; в этом случае $U_s\{I\}$ ограничено для любого открытого интервала I .

Критерий. Распределение F будет невозвратным в том и только том случае, когда при некотором $\alpha > 0$ интеграл

$$\int_0^n \frac{1-su}{(1-su)^2 + s^2 v^2} d\xi \quad (6.2)$$

остается ограниченным при стремлении $s \rightarrow 1$ слева.

(Мы увидим, что в противном случае интеграл стремится к ∞ .)

Доказательство. (i) Предположим, что интеграл (6.2) остается ограниченным для некоторого фиксированного $a > 0$. Равенство Парсеваля (3.2) гл. XV, примененное к F^{n*} и треугольной плотности (см. формулу (4) в гл. XV, 2), дает

$$2 \int_{-\infty}^{+\infty} \frac{1 - \cos ax}{a^2 x^2} F^{n*}\{dx\} = \frac{1}{a} \int_{-a}^a \left(1 - \frac{|\xi|}{a}\right) \varphi^n(\xi) d\xi. \quad (6.3)$$

Умножая на s^n и суммируя по n , получаем

$$\begin{aligned} 2 \int_{-\infty}^{+\infty} \frac{1 - \cos ax}{a^2 x^2} U_s\{dx\} &= \frac{1}{a} \int_{-a}^a \left(1 - \frac{|\xi|}{a}\right) \frac{d\xi}{1 - s\varphi(\xi)} = \\ &= \frac{2}{a} \int_0^a \left(1 - \frac{\xi}{a}\right) \cdot \frac{1-su}{(1-su)^2 + s^2 v^2} d\xi \end{aligned} \quad (6.4)$$

(так как действительная часть φ — четная функция, а мнимая часть — нечетная). Пусть I означает интервал $|x| < 2/a$. При $x \in I$ подынтегральное выражение в крайнем левом интеграле (6.4) будет $> 1/3$, так что величина $U_s\{I\}$ ограничена. Поэтому условия теоремы достаточны.

(ii) Необходимость условия устанавливается на основе равенства Парсеваля с распределением, которое имеет характеристическую функцию $1 - \frac{|\xi|}{a}$ при $|\xi| < a$ (формула (5) в гл. XV, 2). При этом (6.4) заменяется соотношением

$$\int_{-a}^a \left(1 - \frac{|x|}{a}\right) U_s\{dx\} = \frac{2}{\pi} \int_0^\infty \frac{1 - \cos a\xi}{a\xi^2} \cdot \frac{1-su}{(1-su)^2 + s^2 v^2} d\xi. \quad (6.5)$$

Для невозвратного распределения F левая часть остается ограниченной, так что интеграл (6.2) ограничен. ►

В качестве применения критерия докажем лемму, результат

которой был получен другими методами в теореме 4 гл. VI, 10. Другие примеры см. в задачах 13—16.

Лемма 1¹⁾. Вероятностное распределение с нулевым математическим ожиданием является возвратным.

Доказательство. Характеристическая функция дифференцируема в нуле, и соответствующая производная равна нулю, следовательно, можно выбрать столь малое a , что

$$0 \leq 1 - u(\xi) \leq \varepsilon \xi \text{ при } 0 \leq \xi < a.$$

Тогда $1 - su(\xi) \leq 1 - s + \varepsilon \xi$. Используя неравенство $2|xy| \leq x^2 + y^2$, можно оценить интеграл в (6.2) снизу величиной

$$\frac{1}{3} \int_0^a \frac{(1-s) d\xi}{(1-s)^2 + \varepsilon^2 \xi^2} = \frac{1}{3\varepsilon} \operatorname{arctg} \frac{\varepsilon \xi}{1-s} \rightarrow \frac{\pi}{6\varepsilon}.$$

Правая часть может быть сделана как угодно большой, и поэтому интеграл (6.2) стремится к ∞ .

Переход к пределу, который включается в критерий, довольно сложен, поэтому весьма полезны приводимые ниже более простые достаточные условия.

Следствие. Распределение вероятностей F возвратно, если для любого $a > 0$

$$\int_0^a \frac{1-u}{(1-u)^2 + v^2} d\xi = \infty, \quad (6.6)$$

и невозвратно, если при некотором $a > 0$

$$\int_0^a \frac{d\xi}{1-u} < \infty. \quad (6.7)$$

Доказательство. Подынтегральное выражение в (6.2) уменьшится, если $1 - su$ в числителе заменить на $1 - u$, а sv в знаменателе — на v . Тогда (6.6) следует из теоремы о монотонной сходимости. Аналогично подынтегральное выражение в (6.2) увеличится, если слагаемое $s^2 v^2$ в знаменателе отбросить. Отсюда в силу теоремы о монотонной сходимости вытекает (6.7). ►

Эти критерии практически без изменений переносятся на многомерный случай (отличие только в том, что u и v будут функциями многих переменных, а интегралы распространяются на шары с центрами в нуле). Таким образом, мы доказываем следующие критерии.

Лемма 2. Невырожденное двумерное распределение с нулевыми математическими ожиданиями и конечными дисперсиями является возвратным.

Доказательство. Характеристическая функция в этом случае является дважды непрерывно дифференцируемой. Используя разложение в ряд Тейло-

¹⁾ Тот факт, что условие $E(X_j) = 0$ влечет возвратность, был установлен Чжуном и Фуксом. Интересно отметить, что в 1950 г. эта задача доставляла серьезные затруднения и многие попытки решить ее закончились неудачей. Внимание к этой задаче было привлечено неожиданным открытием обманчиво «безобидного» случайного блуждания, в котором $P\{S_n > n/\log n\} \rightarrow 1$. (См. I; гл. X, 3, и задачу 15 в I; гл. X, 8.) О родственных явлениях см. сноску к задаче 13 на стр. 691.

ра, видим, что в некоторой окрестности нуля подынтегральное выражение в (6.6) $> \delta/(\xi_1^2 + \xi_2^2)$. Следовательно, интеграл, соответствующий (6.6), расходится. ►

Лемма 3. Каждое невырожденное трехмерное распределение невозвратно.

Доказательство. Рассматривая разложение в ряд Тейлора функции $\cos(x_1\xi_1 + x_2\xi_2 + x_3\xi_3)$ в некоторой окрестности нуля, можно показать, что для любой характеристической функции существует окрестность начала координат, в которой

$$1 - u(\xi_1, \xi_2, \xi_3) \geq \delta(\xi_1^2 + \xi_2^2 + \xi_3^2).$$

Поэтому трехмерный аналог для (6.7) оценивается сверху интегралом от $(\xi_1^2 + \xi_2^2 + \xi_3^2)^{-1}$, распространенным по окрестности нуля, а этот интеграл в случае трехмерного пространства сходится. ►

§ 7. ЗАДАЧИ

1. Распространите пример § 1 на случай несимметричного интервала $-a, b$. (Выведите два линейных уравнения для двух производящих функций, соответствующие двум границам. Явное решение затруднительно).

Задачи 2—5 касаются симметричного случайного блуждания Бернулли, т. е. $\varphi(\xi) = \cos \xi$. Используются обозначения § 1.

2. Пусть A состоит из двух точек 0 и 1. Покажите элементарным путем, что

$$\chi(s, \xi) = \frac{1}{1 - \frac{1}{4}s^2} \left(\frac{s}{2} e^{-i\xi} + \frac{s^2}{4} e^{2i\xi} \right)$$

и

$$\gamma(s, \xi) = \frac{1}{1 - \frac{1}{4}s^2} \left(1 + \frac{s}{2} e^{i\xi} \right).$$

3. Если в предыдущем примере поменять ролями A и A' , то

$$\chi(s, \xi) = \frac{s}{2} e^{i\xi} + \frac{1}{2} (1 - \sqrt{1-s^2})$$

и

$$\gamma(s, \xi) = \left[1 - \frac{1 - \sqrt{1-s^2}}{s} e^{-i\xi} \right]^{-1}.$$

Дайте вероятностное объяснение.

4. Пусть A состоит из одной точки 0. Тогда χ зависит только от s и γ будет суммой двух степенных рядов, расположенных по степеням $e^{i\xi}$ и $e^{-i\xi}$ соответственно. Используя эту информацию, найдите χ и γ непосредственно из (1.9).

5. Если A_1 состоит из одной точки 0, то $\chi = s\varphi$ и $\gamma = 1$.

6. Другое доказательство тождества (1.9). Покажите, что (в обозначениях § 1)

$$F^n \star \{I\} = \sum_{k=1}^n \int_{A'} H_k \{dy\} F^{(n-k)} \star \{I-y\} + G_n \{I\} \quad (*)$$

а) прямыми вероятностными рассуждениями;

б) по индукции.

Покажите, что (*) равносильно (1.9).

7. В случае (не обязательно симметричного) случайного блуждания Бернулли вальдовская аппроксимация из § 2 совпадает с точным решением. Покажите, что (2.12) сводится к квадратному уравнению для $\tau = e^{-\lambda}$, которое приводит к решению, известному из I; XIV, (4.11). В частности, $Q(S)$ совпадает, вообще говоря, с U_s с той лишь разницей, что последняя величина определялась для интервала $0, a$, а не для $-a, b$, и для начальной точки z .

8. Пусть, как и в § 2, $G_n\{I\}$ обозначает вероятность того, что $S_n \in I \subset A$ и что все это время не было выхода из $A = -a, b$. Покажите, что если два распределения F и F^* совпадают в интервале $|x| < a+b$, то они приводят к одним и тем же вероятностям G_n . Используя этот факт, а также метод урезания, получите другое доказательство сходимости ряда (2.1) при некотором $s > 1$.

9. Случайные блуждания, в которых распределение F сосредоточено на конечном множестве целых чисел, рассматривались в примере 4, в) гл. XII. Покажите, что полученные там формулы содержат неявно факторизацию Винера—Хопфа для $1-\varphi$.

10. (Формула Хинчина—Поллачека.) Пусть F будет сверткой показательного распределения с математическим ожиданием $1/a$, сосредоточенного на $0, \infty$, и распределения B , сосредоточенного на $-\infty, 0$. Обозначим через β и $-b < 0$ соответственно характеристическую функцию и математическое ожидание распределения B . Предположим, что математическое ожидание $a^{-1} - b$ распределения F положительно. Тогда

$$1 - \varphi(\xi) = 1 - \frac{a}{a - i\xi} \beta(\xi) = \left(1 - \frac{a}{a - i\xi}\right) \left(1 - a \frac{1 - \beta(\xi)}{i\xi}\right).$$

Замечание. Эта формула играет важную роль в теории очередей. Другие способы ее получения см. в примерах 5, а)–б) гл. XII и 2, б) гл. XIV.

11. *Продолжение.* Покажите, что при $ab > 1$ существует единственное положительное число κ между 0 и a , такое, что

$$a\beta(-i\kappa) = a - \kappa.$$

Докажите, что $\chi^-(1, \xi) = \frac{a - \kappa - a\beta(\xi)}{i\xi - \kappa}$. *Указание.* Примените результат задачи 10 к сопряженному случайному блужданию с характеристической функцией $a\varphi(\xi) = \varphi(\xi - i\kappa)$. Вспомните, что $\chi^-(1, \xi) = a\chi^-(1, \xi - i\kappa)$. [См. пример 4, б) гл. XII.]

12. Пусть $U_n = \max[0, S_1, \dots, S_n]$ и $V_n = S_n - U_n$. Слегка изменяя рассуждения, использованные при выводе (5.9), покажите, что двумерная характеристическая функция для пары (U_n, V_n) равна коэффициенту при s^n в 1)

$$\frac{1}{1-s} \exp \sum_{n=1}^{\infty} \frac{s^n}{n} \left[\int_0^{\infty} (e^{i\xi_1 x} - 1) F^n \star \{dx\} + \int_{-\infty}^0 (e^{i\xi_2 x} - 1) F^n \star \{dx\} \right].$$

13³⁾. Допустим, что в некоторой окрестности нуля $|1 - \varphi(\xi)| < A \cdot |\xi|$. Тогда F возвратно, за исключением случая, когда оно имеет математическое

¹⁾ Результат получен впервые аналитическим путем Спitzerом, Spitzer F., Trans. Amer. Math. Soc. 82 (1956), 323—339.

²⁾ Эта задача представляет теоретический интерес. Она применима, если φ имеет производную в нуле. Мы видели в гл. XVII, 2а, что это возможно даже в том случае, когда F не имеет математического ожидания, и что тем не менее слабый закон больших чисел применим в этом случае. Таким образом, мы получаем примеры случайных блужданий, в которых при каждом достаточно большом n с очень большой вероятностью $S_n > (1-\varepsilon)n$ при $\mu > 0$ и, однако же, нет сноса в ∞ : случайное блуждание возвратно.

ожидание $\mu \neq 0$. *Указание.* Интеграл (6.6) превосходит $\int_0^a d\xi \int_{-1/\xi}^{1/\xi} x^2 F\{dx\}$.

Подстановка $\xi = 1/t$ и перемена порядка интегрирования показывают, что этот интеграл расходится (за исключением случая, когда существует μ).

14. Используя критерий (6.7), покажите, что если при некотором $\rho > 0$ и $t \rightarrow \infty$

$$t^{-1-\rho} \int_{-t}^t x^2 F\{dx\} \rightarrow \infty,$$

то распределение F невозвратно ¹⁾.

15. Распределение с характеристической функцией $\varphi(\xi) = e^{-1} \sum \frac{1}{n!} \cos(n! \xi)$ невозвратно. *Указание.* Используйте (6.7) и замену переменных $\xi = \frac{1}{n!} t$.

16. Асимметричное устойчивое распределение с показателем $\alpha = 1$ невозвратно, а распределение Коши возвратно.

¹⁾ Отсюда видно, что при слабых условиях регулярности F невозвратно, если абсолютный момент какого-либо порядка $\rho < 1$ расходится. Затруднения, возникающие при отказе от условий регулярности, показаны в статье: Shepp L. A. Bull. Amer. Math. Soc., 70 (1964), 540—542.

В этой главе даны дополнения к изложенной в гл. XV теории характеристических функций, а также даны применения гармонического анализа к случайным процессам и стохастическим интегралам. Обсуждение формулы суммирования Пуассона в § 5 практически не зависит от остального материала. Глава как целое не зависит от гл. XVI—XVIII.

§ 1. РАВЕНСТВО ПАРСЕВАЛЯ

Пусть U — распределение вероятностей с характеристической функцией

$$\omega(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} U\{dx\}. \quad (1.1)$$

Интегрируя это выражение по какому-либо другому распределению F , получаем

$$\int_{-\infty}^{+\infty} \omega(\xi) F\{d\xi\} = \int_{-\infty}^{+\infty} \varphi(x) U\{dx\}, \quad (1.2)$$

где φ — характеристическая функция F . Это одна из форм *равенства Парсеваля*, из которого были выведены основные результаты гл. XV, 3. Удивительно, какое изобилие новой информации можно получить, переписывая равенство Парсеваля в эквивалентных формах и рассматривая специальные случаи. Этот метод, который мы постоянно будем использовать, можно проиллюстрировать простым примером (имеющим и самостоятельный интерес).

Пример. Формула

$$\int_{-\infty}^{+\infty} e^{-iat} \omega(\xi) F\{d\xi\} = \int_{-\infty}^{+\infty} \varphi(x) U\{a+dx\} \quad (1.3)$$

отличается от (1.2) только обозначениями. Рассмотрим частный случай, когда F — равномерное распределение на $[-t, t]$ и $\varphi(x) = \frac{\sin tx}{tx}$. Эта функция не превосходит по абсолютной величине единицы и стремится к нулю при $t \rightarrow \infty$ при всех $x \neq 0$. По

теореме об ограниченной сходимости получаем

$$U(a) - U(a-) = \lim_{t \rightarrow \infty} \frac{1}{2t} \int_{-t}^t e^{-ia\zeta} \omega(\zeta) d\zeta. \quad (1.4)$$

Эта формула позволяет узнать, будет ли a точкой непрерывности, и, если $a-$ атом, позволяет определить его массу. Наиболее интересный результат получается применением (1.4) к симметризованному распределению ${}^0U$ с характеристической функцией $|\omega|^2$. Если $p_1, p_2, \dots$ суть массы атомов U , то ${}^0U$ имеет атом массы $\sum p_k^2$ в нуле (см. задачу 11 в гл. V, 12). Поэтому

$$\frac{1}{2t} \int_{-t}^t |\omega(\zeta)|^2 d\zeta \rightarrow \sum p_k^2. \quad (1.5)$$

Эта формула показывает, в частности, что характеристические функции непрерывных распределений «в среднем» малы. ►

Приведем теперь полезный и «гибкий» вариант формулы Парсеваля (1.2). Пусть A и B — произвольные распределения вероятностей с характеристическими функциями α и β соответственно. Тогда

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \omega(s-t) A\{ds\} B\{dt\} = \int_{-\infty}^{+\infty} \alpha(x) \overline{\beta(x)} U\{dx\}, \quad (1.6)$$

где $\bar{\beta}$ — величина, комплексно сопряженная с β . Справедливость (1.6) проверяется непосредственно интегрированием

$$\omega(s-t) = \int_{-\infty}^{+\infty} e^{i(s-t)x} U\{dx\} \quad (1.7)$$

по A и B . Может создаться обманчивое впечатление, что соотношение (1.6) является более общим, чем (1.2). На самом деле (1.6) есть специальный случай равенства Парсеваля (1.2), соответствующий $F = A \star \bar{B}$, где $\bar{B}$ — распределение с характеристической функцией $\bar{\beta}$ (т. е. $\bar{B}(x) = 1 - B(-x)$ во всех точках непрерывности). Действительно, F имеет характеристическую функцию $\varphi = \alpha \bar{\beta}$, так что правые части (1.2) и (1.6) идентичны. Левые части отличаются лишь формой записи. Лучше всего это можно показать, вводя две независимые случайные величины X и Y с распределениями A и B соответственно. Тогда левая часть (1.6) представляет собою прямое определение математического ожидания $E(\omega(X-Y))$, а левая часть (1.2) выражает это математическое ожидание в терминах распределения F разности $X-Y$. (Мы еще вернемся к формуле Парсеваля в § 7.)

§ 2. ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫЕ ФУНКЦИИ

В 1932 г. С. Бохнер доказал важную теорему, которая делает возможным описание класса характеристических функций по их «внутренним» свойствам. Путь рассуждений указывается следующим простым критерием.

Лемма 1. Пусть ω — ограниченная непрерывная комплекснозначная функция, интегрируемая¹⁾ на $-\infty, \infty$. Определим u равенством

$$u(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\xi x} \omega(\xi) d\xi. \quad (2.1)$$

Для того чтобы функция ω была характеристической, необходимо и достаточно, чтобы было $\omega(0) = 1$ и $u(x) \geq 0$ при всех x . При этих условиях u является плотностью вероятности, соответствующей ω .

Доказательство. Теорема об обращении интеграла Фурье [гл. XV, (3.5)] показывает, что высказанные условия необходимы. Выберем теперь произвольную четную плотность f с интегрируемой характеристической функцией $\varphi \geq 0$. Умножим (2.1) на $\varphi(tx)e^{iax}$ и проинтегрируем по x . Применяя к паре f, φ формулу обращения (3.5) гл. XV, получаем

$$\int_{-\infty}^{+\infty} u(x) \varphi(tx) e^{iax} dx = \int_{-\infty}^{+\infty} \omega(\xi) f\left(\frac{\xi - a}{t}\right) \frac{d\xi}{t}. \quad (2.2)$$

Правая часть представляет собой математическое ожидание ω по отношению к некоторому распределению вероятностей и потому не превосходит максимума $|\omega|$. При $a=0$ подынтегральное выражение слева не отрицательно и стремится к $u(x)$ при $t \rightarrow 0$. Из ограниченности интеграла вытекает интегрируемость u . Переходя в (2.2) к пределу при $t \rightarrow 0$, получаем

$$\int_{-\infty}^{+\infty} u(x) e^{iax} dx = \omega(a) \quad (2.3)$$

(в левой части применяется теорема об ограниченной сходимости, в правой части распределение, по которому интегрируется ω , сходится к распределению, сосредоточенному в точке a). Полагая $a=0$, видим, что u — плотность вероятности, которой соответствует характеристическая функция ω . ►

Условие интегрируемости, появляющееся в лемме, выглядит более ограничительным, чем оно в действительности является.

¹⁾ Как и всегда, это означает абсолютную интегрируемость.

В самом деле, по теореме непрерывности непрерывная функция ω будет характеристической в том и только том случае, когда $\omega(\xi)e^{-\varepsilon\xi^2}$ есть характеристическая функция при каждом $\varepsilon > 0$. Из леммы следует, что ограниченная и непрерывная функция с $\omega(0) = 1$ является характеристической тогда и только тогда, когда при всех x и $\varepsilon > 0$

$$\int_{-\infty}^{+\infty} e^{-i\xi x} \omega(\xi) e^{-\varepsilon\xi^2} d\xi \geq 0. \quad (2.4)$$

Этот критерий имеет совершенно общий характер, однако его трудно применять в частных случаях. К его недостаткам относятся и произвольный выбор «множителя сходимости» $e^{-\varepsilon\xi^2}$. Поэтому мы сформулируем критерий в новой форме.

Лемма 2. *Ограниченная непрерывная функция ω будет характеристической в том и только том случае, если $\omega(0) = 1$ и для любого распределения вероятностей A и всех x*

$$\int_{-\infty}^{+\infty} e^{-i\xi x} \omega(\xi) {}^0A\{d\xi\} \geq 0, \quad (2.5)$$

где ${}^0A = A \star A^-$ — распределение, полученное симметризацией.

Доказательство. а) *Необходимость.* Пусть α — характеристическая функция A . Тогда ${}^0A$ имеет характеристическую функцию $|\alpha|^2$, и необходимость (2.5) вытекает из равенства Парсеваля (1.3).

б) *Достаточность.* Как показывает (2.4), условие (2.5) является достаточным уже в том случае, когда в качестве A выбираются нормальные распределения с произвольными дисперсиями. ►

Мы видели, что (2.5) можно переписать в виде (1.6) с $B = A$. В частности, если A сосредоточено в конечном числе точек $t_1, t_2, \dots, t_n$ и приписывает им массы $p_1, p_2, \dots, p_n$, то (2.5) принимает форму

$$\sum_{j, k} \omega(t_j - t_k) e^{-ix(t_j - t_k)} p_j p_k \geq 0. \quad (2.6)$$

Если это неравенство верно при любом выборе t_j и p_j , то (2.5) выполняется для всех дискретных распределений с конечным числом атомов. Так как каждое распределение может быть представлено как предел подобных дискретных распределений, условие (2.6) оказывается необходимым и достаточным. Обозначим $z_j = p_j e^{-ixt_j}$, тогда оно принимает вид

$$\sum_{j, k} \omega(t_j - t_k) z_j \bar{z}_k \geq 0. \quad (2.7)$$

Для окончательной формулировки нашего критерия введем один часто используемый термин.

Определение. Комплекснозначная функция¹⁾ вещественного переменного t называется положительно определенной, если (2.7) выполняется при любом выборе вещественных чисел $t_1, \dots, t_n$ и комплексных чисел $z_1, \dots, z_n$.

Теорема. (Бохнер.) Непрерывная функция ω является характеристической функцией некоторого распределения вероятностей в том и только том случае, когда она положительно определена и $\omega(0) = 1$.

Доказательство. Мы показали раньше, что эти условия необходимы и что они достаточны, если ω ограничена. Доказательство заканчивается ссылкой на следующую далее лемму, которая утверждает, что все положительно определенные функции ограничены. ►

Лемма 3. Если ω положительно определена, то

$$\omega(0) \geq 0, \quad |\omega(t)| \leq \omega(0), \quad (2.8)$$

$$\omega(-t) = \overline{\omega(t)}. \quad (2.9)$$

Доказательство. Полагая в (2.7) $n=2$, $t_2=0$, $z_2=1$, $t_1=t$, $z_1=z$, получаем

$$\omega(0)[1 + |z|^2] + \omega(t)z + \omega(-t)\bar{z} \geq 0. \quad (2.10)$$

При $z=0$ видим, что $\omega(0) \geq 0$. Рассматривая (2.10) при положительных z , получаем (2.9), откуда следует, что $\omega \equiv 0$, если $\omega(0)=0$. Наконец, если $\omega(0) \neq 0$ и $z = -\overline{\omega(t)}/\omega(0)$, (2.10) сводится к $|\omega(t)|^2 \leq \omega^2(0)$.

§ 3. СТАЦИОНАРНЫЕ ПРОЦЕССЫ

Важные следствия имеет применение последней теоремы к случайным процессам со стационарными ковариациями. Под этим понимается семейство случайных величин $\{X_t\}$, определенных на некотором пространстве при всех $-\infty < t < \infty$ и таких, что

$$\text{Cov}(X_{s+t}, X_s) = \rho(t) \quad (3.1)$$

при любых s . До сих пор мы рассматривали только действительные случайные величины, но теперь мы введем комплексные случайные величины, что сделает обозначения более простыми и более симметричными. Комплексная случайная величина—это просто пара действительных случайных величин, записанная в форме $X = U + iV$. Мы не делаем никаких предположений относительно совместного распределения величин U и V . Случайная

¹⁾ Об «обобщенных» положительно определенных функциях (эквивалентных распределениям в смысле Л. Шварца) в произвольных пространствах см. Гельфанд и Виленкин (1964).

величина $\bar{X} = U - iV$ называется комплексно сопряженной с X , и произведение $X\bar{X}$ играет здесь ту же роль, что и X^2 для вещественных случайных величин. Отсюда некоторая асимметрия в определении дисперсии и ковариации.

Определение. Для комплексных случайных величин с

$$E(X) = E(Y) = 0$$

определим

$$\text{Cov}(X, Y) = E(X\bar{Y}). \quad (3.2)$$

Тогда $\text{Var}(X) = E(|X|^2) \geq 0$, а $\text{Cov}(Y, X)$ является числом, комплексно сопряженным с $\text{Cov}(X, Y)$.

Теорема. Пусть $\{X_t\}$ — такое семейство случайных величин, что функция

$$\rho(t) = E(X_{t+s}\bar{X}_s) \quad (3.3)$$

не зависит от s и непрерывна¹⁾. Тогда ρ положительно определена и потому

$$\rho(t) = \int_{-\infty}^{+\infty} e^{i\lambda t} R\{d\lambda\}, \quad (3.4)$$

где R — мера на вещественной прямой, приписывающая всей прямой массу $\rho(0)$.

Если случайные величины X_t действительны, то мера R симметрична и

$$\rho(t) = \int_{-\infty}^{+\infty} \cos \lambda t R\{d\lambda\}. \quad (3.5)$$

Доказательство. Выберем произвольно точки $t_1, \dots, t_n$ и комплексные числа $z_1, \dots, z_n$. Тогда

$$\begin{aligned} \sum \rho(t_j - t_k) z_j \bar{z}_k &= \sum E(X_{t_j} \bar{X}_{t_k}) z_j \bar{z}_k = \\ &= E(\sum X_{t_j} z_j \bar{X}_{t_k} \bar{z}_k) = E(|\sum X_{t_j} z_j|^2) \geq 0 \end{aligned} \quad (3.6)$$

и (3.4) вытекает из критерия предшествующего параграфа. Если ρ действительно, то соотношение (3.4) выполняется также и для «отраженной» меры, получаемой заменой x на $-x$. По теореме единственности R симметрична. ►

Меру R называют *спектральной мерой*²⁾ случайного процесса. Множество ее точек роста называют спектром $\{X_t\}$. В большинстве применений случайные величины центрированы так, что $E(X_t) = 0$. В этом случае $\rho(t) = \text{Cov}(X_{t+s}, X_s)$. По этой причине ρ

¹⁾ Непрерывность существенна: для взаимно независимых случайных величин X_t мы имеем $\rho(t) = 0$ для всех $t \neq 0$; эта функция не допускает представления (3.4) (см. задачу 2).

²⁾ В теории связи говорят также о «спектре мощностей».

называют обычно *ковариационной функцией* процесса. Центрирование X_t никак не влияет на те свойства процесса, которые мы будем изучать.

Примеры. а) Пусть $Z_1, \dots, Z_n$ — попарно некоррелированные случайные величины с нулевыми математическими ожиданиями и дисперсиями $\sigma_1^2, \dots, \sigma_n^2$. Положим

$$X_t = Z_1 e^{i\lambda_1 t} + \dots + Z_n e^{i\lambda_n t}, \quad (3.7)$$

где числа $\lambda_1, \dots, \lambda_n$ действительны. Тогда

$$\rho(t) = \sigma_1^2 e^{i\lambda_1 t} + \dots + \sigma_n^2 e^{i\lambda_n t}, \quad (3.8)$$

так что мера R сосредоточена в n точках $\lambda_1, \dots, \lambda_n$. Мы увидим позже, что общий стационарный процесс можно трактовать как предельный случай процессов, указанных в этом примере.

Если процесс (3.7) — действительный, его можно представить в виде

$$X_t = U_1 \cos \lambda_1 t + \dots + U_r \cos \lambda_r t + V_1 \sin \lambda_1 t + \dots + V_r \sin \lambda_r t, \quad (3.9)$$

где U_j и V_j — действительные попарно не коррелированные случайные величины и

$$E(U_j^2) = E(V_j^2) = \sigma_j^2.$$

Типичный пример см. в (7.23) гл. III. Соответствующие ковариации равны $\rho(t) = \sigma_1^2 \cos \lambda_1 t + \dots + \sigma_r^2 \cos \lambda_r t$.

б) *Марковские процессы.* Если случайные величины X_t нормальны и процесс марковский, то $\rho(t) = e^{-a|t|}$ [см. (8.14) гл. III]. Спектральная мера пропорциональна распределению Коши.

в) Пусть $X_t = Ze^{itY}$, где Y и Z — независимые действительные случайные величины и $E(Z) = 0$. Тогда

$$\rho(t) = E(Z\bar{Z}) E(e^{itY}).$$

Это равенство показывает, что спектральная мера R совпадает с распределением вероятностей Y , умноженным на $E(Z\bar{Z})$. ►

Теоретически не так важно, описывается ли процесс в терминах ковариационной функции $\rho(t)$ или эквивалентным образом в терминах спектральной меры R . Однако с практической точки зрения описание в терминах спектральной меры проще и предпочтительнее. В приложениях к теории связи спектральное описание имеет технические преимущества при проектировании приборов и измерениях, но мы не будем задерживаться на этом обстоятельстве. Более важным преимуществом с нашей точки зрения является то, что линейные операции на X_t (часто называемые «фильтрами») лучше описываются в терминах R , чем в терминах ρ .

Пример. г) *Линейные операции.* Начнем с простейшего при-

мера случайных величин Y_t , определяемых равенством

$$Y_t = \sum c_k X_{t-\tau_k}, \quad (3.10)$$

где c_k и τ_k — постоянные (τ_k действительны) и сумма содержит конечное число членов. Ковариационная функция для Y_t определяется двойной суммой

$$\rho_Y(t) = \sum c_j \bar{c}_k \rho(t - \tau_j + \tau_k). \quad (3.11)$$

Подставляя в (3.4), находим

$$\rho_Y(t) = \int_{-\infty}^{\infty} \left| \sum c_j e^{-i\tau_j \lambda} \right|^2 e^{it\lambda} R \{d\lambda\}.$$

Отсюда следует, что спектральная мера R_Y задается равенством

$$R_Y \{d\lambda\} = \left| \sum c_j e^{-i\tau_j \lambda} \right|^2 R \{d\lambda\}. \quad (3.12)$$

В отличие от (3.11) последнее соотношение допускает интуитивную интерпретацию: составляющая R , соответствующая «частоте» λ , умножается на «частотную характеристику фильтра» $f(\lambda)$, определяемую данным преобразованием (3.10).

Этот пример имеет значительно более широкую область применимости, чем это может показаться на первый взгляд. Причина этого в том, что интегралы и производные суть пределы сумм вида (3.10), и потому аналогичное замечание применимо и к ним. Пусть, например, X_t — сигнал, подаваемый на вход стандартной электрической цепи. Тогда выходной сигнал Y_t может быть представлен в виде некоторого интеграла, содержащего X_t , и спектральная мера R_Y опять может быть выражена через R и частотную характеристику. Последняя зависит от строения цепи. Поэтому наш результат может быть использован в двух направлениях: именно чтобы описать процесс на выходе и чтобы строить цепи, выходной сигнал которых обладает заданными свойствами. ►

Перейдем к теореме, обратной к доказанной выше. Мы покажем, что для любой меры R на числовой прямой найдется стационарный процесс $\{X_t\}$ со спектральной мерой R . Поскольку преобразование $X_t \rightarrow aX_t$ переводит R в $a^2 R$, мы можем, не ограничивая общности, считать R вероятностной мерой. Возьмем в качестве выборочного пространства λ -ось вместе с мерой R и обозначим X_t случайную величину $X_t(\lambda) = e^{it\lambda}$. Тогда

$$\rho(t) = E(X_{t+s} \bar{X}_s) = \int_{-\infty}^{+\infty} e^{it\lambda} R \{d\lambda\}, \quad (3.13)$$

и поэтому спектральная мера нашего процесса равна R . Мы построили, таким образом, модель стационарного процесса с заданной спектральной мерой. Приятной неожиданностью является тот факт, что в подобной модели возможно выбрать в качестве

выборочного пространства действительную прямую. Мы вернемся к этой модели в § 8.

Легко видоизменить модель так, чтобы получить случайные величины с нулевыми математическими ожиданиями. Пусть Y — случайная величина, не зависящая от всех X_t и принимающая значения ± 1 с вероятностью по $\frac{1}{2}$ каждое. Положим $X'_t = YX_t$. Тогда $E(X'_t) = 0$ и $E(X'_{t+s} \bar{X}'_s) = E(Y^2) E(X_{t+s} \bar{X}_s)$. Таким образом, $\{X'_t\}$ является стационарным процессом с нулевым математическим ожиданием и ковариационной функцией (3.13).

§ 4. РЯДЫ ФУРЬЕ

Арифметическое распределение, приписывающее точке n вероятность φ_n , имеет периодическую, с периодом 2π , характеристическую функцию

$$\varphi(\zeta) = \sum_{-\infty}^{+\infty} \varphi_n e^{in\zeta}. \quad (4.1)$$

Вероятности φ_n выражаются с помощью формулы обращения

$$\varphi_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\zeta} \varphi(\zeta) d\zeta, \quad (4.2)$$

которая легко выводится из (4.1) [см. гл. XV, (3.14)].

Возьмем теперь произвольную функцию φ с периодом 2π и определим числа φ_k с помощью (4.2). Задача состоит в том, чтобы узнать, является ли φ характеристической функцией, т. е. задают ли числа φ_n распределение вероятностей. Метод опирается на исследование семейства функций f_r , определенных при $0 < r < 1$ равенством

$$f_r(\zeta) = \sum_{-\infty}^{+\infty} \varphi_n r^{|n|} e^{in\zeta}. \quad (4.3)$$

Наши рассуждения, несмотря на их простоту, приведут к важным результатам относительно рядов Фурье и характеристических функций распределений, сосредоточенных на конечных интервалах.

В последующем удобнее рассматривать основной интервал $-\pi, \pi$ как *круг* (т. е. отождествить точки π и $-\pi$). Для интегрируемой φ число φ_k будет называться *k-м коэффициентом Фурье функции φ* . Ряд (4.1) называют «*формальным рядом Фурье*», соответствующим φ . Он не обязан сходиться, но так как последовательность $\{\varphi_n\}$ ограничена, то ряд (4.3) сходится к непрерывной (и, более того, дифференцируемой) функции f_r . При $r \rightarrow 1$ функции f_r могут стремиться к пределу φ , даже если ряд (4.1) расходится. В этом случае говорят, что ряд «*суммируем по Абелю*» к функции φ .

Примеры. а) Пусть $\varphi_n = 1$ для $n = 0, 1, 2, \dots$, но $\varphi_n = 0$ для

$n < 0$. Каждый член ряда в (4.1) равен по абсолютной величине единице, и поэтому ряд может сходиться, но не при любом значении ξ . С другой стороны, правая часть (4.3) есть сумма членов геометрической прогрессии, которая сходится к

$$f_r(\xi) = \frac{1}{1 - re^{i\xi}}. \quad (4.4)$$

Предел при $r \rightarrow 1$ существует во всех точках, кроме $\xi = 0$.

б) Важный частный случай (4.3) получается, когда $\varphi_n = \frac{1}{2\pi}$ при всех n :

$$p_r(t) = \frac{1}{2\pi} \sum_{-\infty}^{+\infty} r^{|n|} e^{int}. \quad (4.5)$$

Сумма членов с $n \geq 0$ указана в (4.4). По соображениям симметрии получаем

$$2\pi p_r(t) = \frac{1}{1 - re^{it}} + \frac{1}{1 - re^{-it}} - 1 \quad (4.6)$$

или

$$p_r(t) = \frac{1}{2\pi} \cdot \frac{1 - r^2}{1 + r^2 - 2r \cos t}. \quad (4.7)$$

Эта функция постоянно используется в теории гармонических функций [где $p_r(t - \xi)$ называют «ядром Пуассона»]. Для удобства ссылок мы сформулируем ее основное свойство в следующей лемме.

Лемма. При фиксированном $0 < r < 1$ функция p_r есть плотность некоторого распределения P_r на единичной окружности. При $r \rightarrow 1$ последнее сходится к распределению вероятностей, сосредоточенному в точке $t = 0$.

Доказательство. Ясно, что $p_r \geq 0$. Из (4.5) видно, что интеграл от p_r по $-\pi, \pi$ равен единице, так как при $n \neq 0$ интеграл от e^{int} равен нулю. При $\delta \leq t \leq \pi$ знаменатель в (4.7) больше некоторого положительного числа. При $r \rightarrow 1$ в каждом открытом интервале, не содержащем $t = 0$, $p_r(t)$ ограниченно сходится к нулю. Следовательно, P_r имеет предельное распределение, сосредоточенное в точке $t = 0$. ▶

Теорема 1. Непрерывная функция φ , имеющая период 2π , будет характеристической тогда и только тогда, когда ее коэффициенты Фурье (4.2) неотрицательны и $\varphi(0) = 1$. Если эти условия выполнены, ряд Фурье (4.1) равномерно сходится к φ .

[Иными словами, формальный ряд Фурье с неотрицательными коэффициентами φ_n из (4.2) сходится к непрерывной функции

в том и только том случае, когда $\sum \varphi_k < \infty$. Последнее условие влечет за собой (4.1).]

Доказательство. В силу (4.2) и (4.5) функция f_r из (4.3) может быть представлена в виде

$$f_r(\xi) = \int_{-\pi}^{\pi} \varphi(t) \cdot p_r(\xi - t) dt. \quad (4.8)$$

Справа стоит свертка φ и распределения вероятностей P_r , поэтому при $r \rightarrow 1$

$$f_r(\xi) \rightarrow \varphi(\xi). \quad (4.9)$$

Далее, если m — верхняя грань $|\varphi|$, то в силу (4.8)

$$f_r(0) = \sum_{-\infty}^{+\infty} \varphi_n r^{|n|} \leq m. \quad (4.10)$$

Так как члены последнего ряда неотрицательны, то переходом к пределу при $r \rightarrow 1$ получаем неравенство $\sum \varphi_n \leq m$. Поэтому ряд $\sum \varphi_n e^{in\xi}$ сходится равномерно и из (4.3) с очевидностью вытекает, что $f_r(\xi)$ сходится к его сумме. Таким образом, (4.1) верно. Доказательство закончено. ►

Заметим, что (4.9) — прямое следствие свойств свертки и не зависит от предположения о неотрицательности коэффициентов φ_n . То есть нами попутно доказана

Теорема 2¹). Если φ непрерывна и имеет период 2π , то (4.9) выполняется равномерно относительно ξ .

(Обобщение на случай разрывных функций см. в следствии 2 и задачах 6—8).

Следствие 1. (Фейер.) Любая непрерывная периодическая функция φ может быть представлена как предел равномерно сходящейся последовательности тригонометрических многочленов.

Другими словами, для любого $\varepsilon > 0$ найдутся числа $a_{-N}, \dots, a_N$, такие, что

$$\left| \varphi(\xi) - \sum_{n=-N}^N a_n e^{in\xi} \right| < \varepsilon \quad (4.11)$$

¹) Теорема может быть сформулирована по-другому: ряд Фурье непрерывной периодической функции φ суммируем по Абелю к φ . Теорема (и метод доказательства) легко распространяется на другие методы суммирования.

Указанное явление было впервые замечено Л. Фейером (для суммирования по Чезаро, см. задачу 9) в то время, когда расходящиеся ряды все еще представлялись чем-то таинственным. Открытие Фейера было сенсационным, и по историческим причинам в учебниках до сих пор употребляют суммирование по Чезаро, хотя метод Абеля более удобен и делает доказательства единообразными.

при всех ζ .

Доказательство. Для произвольного N и $0 < r < 1$

$$\left| \varphi(\zeta) - \sum_{n=-N}^N \varphi_n r^{|n|} e^{in\zeta} \right| \leq |\varphi(\zeta) - f_r(\zeta)| + \sum_{|n| > N} |\varphi_n| \cdot r^{|n|}. \quad (4.12)$$

Мы вправе выбрать r настолько близким к 1, чтобы первый член справа был меньше $\varepsilon/2$ для всех ζ . После этого можно выбрать N столь большим, чтобы сумма последнего ряда в (4.12) была меньше $\varepsilon/2$. Тогда (4.11) выполнено с $a_n = \varphi_n r^{|n|}$. ►

Полноты ради укажем следующий результат, который фактически содержится в лемме 1 § 6.

Следствие 2. *Две интегрируемые периодические функции с одинаковыми коэффициентами Фурье могут отличаться друг от друга только на множестве меры нуль (т. е. неопределенные интегралы совпадают).*

Доказательство. Для интегрируемой периодической функции φ с коэффициентами Фурье φ_n положим

$$\Phi(x) = \int_{-\pi}^x [\varphi(t) - \varphi_0] dt. \quad (4.13)$$

Эта функция Φ непрерывна и периодична, и интегрирование по частям показывает, что при $n \neq 0$ ее n -й коэффициент Фурье равен $-i\varphi_n/n$.

Соотношения (4.9) и (4.3) вместе показывают, что непрерывная функция φ однозначно определяется своими коэффициентами Фурье φ_n . Поэтому коэффициенты φ_n при $n \neq 0$ определяют φ с точностью до аддитивной постоянной. Отсюда следует, что коэффициенты Фурье произвольной интегрируемой функции φ определяют интеграл Φ и, следовательно, φ однозначно определяется с точностью до значений на множестве меры нуль. ►

§ 5*). ФОРМУЛА СУММИРОВАНИЯ ПУАССОНА

В этом параграфе будем считать, что характеристическая функция φ такова, что $|\varphi|$ интегрируема на всей прямой. Из формулы обращения для интеграла Фурье (формула (3.5) из гл. XV) вытекает существование *непрерывной плотности* f . По теореме Римана—Лебега (лемма 3 гл. XV, 4) как φ , так и f обращаются в нуль на бесконечности. В случае когда φ стремится к нулю достаточно быстро, можно строить периодические функции способом, который заключается, грубо говоря, в «наматывании ζ -оси»

*) Здесь рассматриваются важные, но специальные вопросы. Содержание параграфа не используется в дальнейшем и не зависит от предыдущих параграфов, кроме того, что здесь используется теорема 1 из § 4.

на окружность длины 2λ . Новая функция может быть представлена в виде

$$\psi(\zeta) = \sum_{k=-\infty}^{+\infty} \varphi(\zeta + 2k\lambda) \quad (5.1)$$

(сумма бесконечного в обе стороны ряда $\sum_{k=-\infty}^{+\infty} a_k$ определяется здесь как предел $\lim_{N \rightarrow \infty} \sum_{k=-N}^N a_k$, если он существует). В случае сходимости ряда функция ψ является, очевидно, периодической с периодом 2λ . В своем простейшем вероятностном варианте формула суммирования Пуассона утверждает, что если ψ непрерывна, то $\psi(\zeta)/\psi(0)$ есть характеристическая функция арифметического распределения, приписывающего точкам $n\pi/\lambda$ массы, пропорциональные $f(n\pi/\lambda)$ ($n=0, \pm 1, \pm 2, \dots$). На первый взгляд этот результат может показаться только забавным, однако он включает в себя известную теорему теории передачи сообщений и много частных случаев, представляющих несомненный интерес.

Формула суммирования Пуассона¹⁾. Предположим, что характеристическая функция φ абсолютно интегрируема и, следовательно, соответствующая плотность вероятности f непрерывна. Тогда

$$\sum_{-\infty}^{\infty} \varphi(\zeta + 2k\lambda) = \frac{\pi}{\lambda} \sum_{-\infty}^{\infty} f(n\pi/\lambda) e^{in(\pi/\lambda)\zeta}, \quad (5.2)$$

если ряд в левой части сходится к непрерывной функции ψ .

При $\zeta=0$ отсюда следует, что

$$\sum_{-\infty}^{\infty} \varphi(2k\lambda) = (\pi/\lambda) \sum f(n\pi/\lambda) \quad (5.3)$$

есть положительное число, скажем A , и поэтому $\psi(\zeta)/A$ — характеристическая функция.

Доказательство. Достаточно показать, что правая часть (5.2) представляет собой формальный ряд Фурье периодической функции ψ , стоящей слева в (5.2), т. е. что

$$\frac{1}{2\lambda} \int_{-\lambda}^{\lambda} \psi(\zeta) e^{-in(\pi/\lambda)\zeta} d\zeta = \frac{\pi}{\lambda} f(n\pi/\lambda). \quad (5.4)$$

Действительно, коэффициенты Фурье (5.4) неотрицательны и функ-

¹⁾ Тождество (5.2) обычно устанавливается при различных дополнительных условиях. Наша простая формулировка, так же как и очень простое доказательство, стала возможна благодаря систематическому использованию положительности f . По поводу других вариантов теоремы и их доказательств см. задачи 12 и 13.

ния ψ по предположению непрерывна; таким образом, по теореме 1 предыдущего параграфа ряд Фурье сходится к ψ , и поэтому (5.2) справедливо.

Вклад k -го члена ряда (5.1) для ψ в левую часть (5.4) равен

$$\frac{1}{2\lambda} \int_{-\lambda}^{\lambda} \varphi(\xi + 2k\lambda) e^{-in(\pi/\lambda)\xi} d\xi = \frac{1}{2\lambda} \int_{(2k-1)\lambda}^{(2k+1)\lambda} \varphi(s) e^{-in(\pi/\lambda)s} ds, \quad (5.5)$$

и его абсолютная величина меньше чем

$$\int_{(2k-1)\lambda}^{(2k+1)\lambda} |\varphi(s)| ds. \quad (5.6)$$

Интервалы $(2k-1)\lambda < s \leq (2k+1)\lambda$ попарно не пересекаются и покрывают всю действительную прямую. Поэтому в сумме величины (5.6) дают интеграл от $|\varphi|$, который конечен. Суммируя (5.5) по $-N < K < N$ и устремляя $N \rightarrow \infty$, получаем в силу теоремы о мажорированной сходимости

$$\frac{1}{2\lambda} \int_{-\lambda}^{\lambda} \psi(\xi) e^{-in(\pi/\lambda)\xi} d\xi = \frac{1}{2\lambda} \int_{-\infty}^{\infty} \varphi(s) e^{-in(\pi/\lambda)s} ds. \quad (5.7)$$

По формуле обращения для интегралов Фурье правая часть (5.7) равна $(\pi/\lambda)f(n\pi/\lambda)$. Теорема доказана. ►

В наиболее интересном частном случае функция φ тождественно равна нулю при $|\xi| \geq a$, где $a < \lambda$. Ряд (5.1) сводится в этом случае к единственному слагаемому, и ψ представляет собой периодическое продолжение φ с периодом 2λ . Соотношение (5.2) выполняется. Левая часть (5.3) равна 1, откуда следует, что ψ является характеристической функцией некоторого распределения вероятностей. Таким образом, справедливо

Следствие. Если характеристическая функция равна нулю при $|\xi| \geq a$, то все ее периодические продолжения с периодом $2\lambda > 2a$ тоже являются характеристическими функциями. Это утверждение в действительности несколько сильнее¹⁾ приводимой ниже теоремы 2 (известной под названием «sampling theorem»), которая обычно присутствует в трактатах по теории передачи сообщений и приписывается то Найквисту, то Шеннону²⁾.

Теорема 2. («Sampling theorem».) Допустим, что плотность вероятности f имеет характеристическую функцию, равную нулю

¹⁾ Обычно дополнительно вводятся ненужные условия. Это вызвано тем, что доказательства опираются на общую теорию Фурье, которая пренебрегает в отличие от теории вероятностей свойством положительности f .

²⁾ А также В. А. Котельникову. — Прим. перев.

вне $-\overline{a}$, a . Тогда f однозначно определяется¹⁾ по значениям $\frac{\pi}{\lambda} f\left(n \frac{\pi}{\lambda}\right)$ при любом фиксированном $\lambda > a$ ($n=0, \pm 1, \dots$). Эти значения индуцируют распределение вероятностей с характеристической функцией, которая является периодическим продолжением φ с периодом 2λ .

Примеры. а) Рассмотрим плотность $f(x) = (1 - \cos x)/\pi x^2$ с характеристической функцией $\varphi(\xi) = 1 - |\xi|$ для $|\xi| \leq 1$ и $\varphi(\xi) = 0$ для $|\xi| > 1$. При $\lambda = 1$ и $\xi = 1$, учитывая, что $f(0) = 1/(2\pi)$, получаем из (5.3)

$$\frac{1}{2} + \frac{4}{\pi^2} \sum_{v=0}^{\infty} \frac{1}{(2v+1)^2} = 1. \quad (5.8)$$

Периодическое продолжение φ с периодом $2\lambda = 2$ изображено на рис. 2 в гл. XV, 2, а продолжение с периодом $\lambda > 2$ на рис. 3 там же.

б) Простой пример к (5.2) указывается в задаче 11. ►

Как часто бывает в подобных ситуациях, формулу (5.2) можно переписать в форме, которая *кажется* более общей. В самом деле, применяя (5.2) к плотности $f(x+s)$, мы получаем *другую запись формулы суммирования Пуассона*:

$$\sum_{-\infty}^{+\infty} \varphi(\xi + 2k\lambda) e^{-is(\xi + 2k\lambda)} = \frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} f\left(n \frac{\pi}{\lambda} + s\right) e^{in(\pi/\lambda)\xi}. \quad (5.9)$$

Примеры. в) Для нормальной плотности и $\xi = \lambda$ выводим из (5.9)

$$\sum_{-\infty}^{+\infty} e^{-\frac{1}{2}(2k+1)^2\lambda^2} \cos(2k+1)\lambda s = \frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} (-1)^k \Pi\left(\frac{(2k+1)\pi}{\lambda} + s\right). \quad (5.10)$$

¹⁾ Явное выражение для $f(x)$ может быть получено следующим путем. При $|\xi| \leq \lambda$ имеем $\varphi(\xi) = \psi(\xi)$, и, следовательно, по формуле обращения для интегралов Фурье

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi(\xi) e^{-ix\xi} d\xi.$$

Далее используется то, что ψ определяется правой частью (5.2), и, интегрируя, получаем окончательную формулу:

$$f(x) = \frac{1}{2\pi} \frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} \int_{-\lambda}^{\lambda} f\left(n \frac{\pi}{\lambda}\right) e^{in(\pi/\lambda)\xi - i\xi x} d\xi = \frac{\sin \lambda x}{\lambda} \sum_{-\infty}^{+\infty} f\left(n \frac{\pi}{\lambda}\right) \frac{(-1)^n}{x - n\pi/\lambda}.$$

Это выражение, имеющее много приложений, иногда называется «кардинальным рядом». [См. теорему 16 в книге: Whittaker J. M., Interpolatory function theory, Cambridge Tracts. 33, 1935. Многомерный аналог см.: Petersen D. P., Middleton D., Information and Control, 5 (1962), 279—323.]

Это — знаменитая формула теории зэта-функций, которая была доказана более элементарным методом в гл. X, 5. В самом деле, производя дифференцирование по x в выражениях (5.8) и (5.9) гл. X, мы видим, что равенство этих выражений равносильно (5.10) с $\lambda = (\pi/a)\sqrt{t}$ и $s = x/\sqrt{t}$.

г) Для плотности $f(x) = \pi^{-1}(1+x^2)^{-1}$ с характеристической функцией $\varphi(\xi) = e^{-|\xi|}$ мы получаем из (5.2), полагая $\xi = 0$,

$$\frac{e^\lambda + e^{-\lambda}}{e^\lambda - e^{-\lambda}} = \sum_{n=-\infty}^{+\infty} \frac{\lambda}{\lambda^2 + n^2\pi^2}. \quad (5.11)$$

Это не что иное, как разложение гиперболического котангенса на простейшие дроби.

д) Плотности на окружности длины 2π можно получить, «наматывая» действительную ось на окружность, как это объяснено в гл. II, 8. При этом плотности f на прямой ставится в соответствие плотность на окружности, задаваемая рядом $\sum f(2\pi n + s)$. Из (5.9) при $\xi = 0$ мы получаем новое представление этой плотности в терминах первоначальной характеристической функции. В частном случае $f = \pi$ мы получаем аналог *нормальной плотности* на единичной окружности в виде

$$\frac{1}{\sqrt{2\pi t}} \sum_{-\infty}^{+\infty} \exp\left(-\frac{1}{2t}(s + 2n\pi)^2\right) = \frac{1}{2\pi} \sum_{-\infty}^{+\infty} e^{-\frac{1}{2}n^2t} \cos ns. \quad (5.12)$$

Представление справа ясно показывает, что свертка двух нормальных плотностей с параметрами t_1 и t_2 есть нормальная плотность с параметром $t_1 + t_2$. ►

§ 6. ПОЛОЖИТЕЛЬНО ОПРЕДЕЛЕННЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

Этот параграф посвящен распределениям вероятностей на конечном интервале; для определенности будем считать, что длина интервала равна 2π . Как и в § 4, мы отождествим две концевые точки интервала и будем интерпретировать интервал как круг единичного радиуса. Таким образом мы рассматриваем F как *распределение вероятностей на окружности* единичного радиуса и определяем *коэффициенты Фурье* распределения F равенством

$$\varphi_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} F\{dt\}, \quad k=0, \pm 1, \dots \quad (6.1)$$

Отметим, что $\overline{\varphi_k} = \varphi_{-k}$. Покажем, что распределение F однозначно восстанавливается по коэффициентам φ_k . Отвлекаясь от тривиального изменения масштаба, мы видим, что наше утверждение эквивалентно следующему: *распределение, сосредоточенное на $-\lambda, \lambda$,*

однозначно определяется по значениям $\varphi(n\pi/\lambda)$, принимаемым характеристической функцией в точках с абсциссами, кратными π/λ . Это утверждение двойственно теореме 2 предыдущего параграфа, в соответствии с которой характеристическая функция, равная нулю вне $[-\lambda, \lambda]$, однозначно определяется по значениям $f(n\pi/\lambda)$ соответствующей плотности.

Теорема 1. *Распределение F на окружности однозначно определяется своими коэффициентами Фурье φ_k .*

Доказательство. Как и в (4.3), положим при $0 \leq r < 1$

$$f_r(\zeta) = \sum_{-\infty}^{+\infty} \varphi_n \cdot r^{|n|} \cdot e^{in\zeta}. \quad (6.2)$$

Вычисления, подобные тем, которые привели к (4.8), показывают, что теперь

$$f_r(\zeta) = \int_{-\pi}^{\pi} p_r(\zeta - t) F \{dt\}, \quad (6.3)$$

где p_r — ядро Пуассона, определенное в (4.7). Мы знаем, что p_r является плотностью вероятности, и мы обозначим через f_r соответствующее распределение. Таким образом, f_r есть плотность свертки $P_r \star F$, которая сходится к F при $r \rightarrow 1$. Поэтому F восстанавливается по f_r . ►

Лемма 1. *Пусть $\{\varphi_n\}$ — произвольная ограниченная последовательность комплексных чисел. Для того чтобы на окружности существовала мера F с коэффициентами Фурье φ_n , необходимо и достаточно, чтобы при каждом $r < 1$ функция f_r , определяемая формулой (6.2), была неотрицательной.*

Доказательство. Необходимость, очевидно, следует из (6.3) и строгой положительности p_r . Умножая теперь (6.2) на $e^{-ik\zeta}$ и интегрируя, получаем

$$\varphi_k \cdot r^{|k|} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_r(\zeta) e^{-ik\zeta} d\zeta. \quad (6.4)$$

Полагая $k=0$, видим, что $\varphi_0 > 0$. Не ограничивая общности, допустим, что $\varphi_0 = 1/(2\pi)$. Тогда f_r есть плотность некоторого распределения F_r на окружности и (6.4) показывает, что $\varphi_k r^{|k|}$ есть k -й коэффициент Фурье F_r . По теореме о выборе найдется такая последовательность значений $r \rightarrow 1$, что F_r сходятся к некоторому распределению F . Из (6.4) вытекает, очевидно, что числа φ_k удовлетворяют (6.1). Доказательство закончено. ►

Отметим, что эта лемма сильнее, чем следствие 2 из § 4. Поступая так же, как и в § 2, мы установим аналог теоремы Бохнера (принадлежащий Герглотцу).

Определение. Последовательность $\{\varphi_k\}$ называется положительно определенной, если для любого конечного набора комплексных чисел $z_1, \dots, z_n$

$$\sum_{j, k} \varphi_{j-k} z_j \bar{z}_k \geq 0. \quad (6.5)$$

Лемма 2. Если последовательность $\{\varphi_n\}$ положительно определена, то $\varphi_0 \geq 0$ и $|\varphi_n| \leq \varphi_0$.

Доказательство. Применимо доказательство леммы 3 из § 2 (см. также задачу 14).

Теорема 2. Последовательность $\{\varphi_n\}$ представляет собой последовательность коэффициентов Фурье некоторой меры F на окружности в том и только том случае, когда она положительно определена.

Доказательство. а) Простые вычисления показывают, что для φ_k , задаваемых (6.1), левая часть (6.5) равна интегралу по мере F от $\frac{1}{2\pi} |\sum e^{-ijt} z_j|^2$. Поэтому условие (6.5) необходимо.

б) Для доказательства достаточности положим $z_k = r^k e^{ikt}$ при $k \geq 0$ и $z_k = 0$ при $k < 0$. При такой последовательности левая часть (6.5) принимает вид

$$\begin{aligned} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \varphi_{j-k} r^{j+k} e^{i(j-k)t} &= \sum_{n=-\infty}^{+\infty} \varphi_n e^{int} \sum_{k=0}^{\infty} r^{|n|+2k} = \\ &= (1-r^2)^{-1} \sum_{n=-\infty}^{\infty} \varphi_n r^{|n|} e^{int}, \end{aligned} \quad (6.6)$$

и по определению (6.2) последняя сумма равна $f_r(t)$. Хотя неравенство (6.5) постулировалось только для конечных наборов $\{z_n\}$, простой предельный переход показывает, что оно справедливо также и для нашей бесконечной последовательности. Таким образом мы доказали, что $f_r(t) \geq 0$. В силу леммы 1 отсюда следует, что φ_k действительно являются коэффициентами Фурье некоторой меры, сосредоточенной на единичной окружности. ►

Используя этот критерий, получаем аналог теоремы § 3.

Теорема 3. Пусть $\{X_n\}$ — последовательность случайных величин, заданных на некотором вероятностном пространстве, и пусть

$$\rho_n = E(X_{n+v} \bar{X}_v) \quad (6.7)$$

не зависит от v . Тогда существует и притом только одна мера R на окружности $\overline{[-\pi, \pi]}$, такая, что числа ρ_n суть ее коэффициенты Фурье.

Доказательство. Очевидно,

$$\sum \rho_{j-k} z_j \bar{z}_k = \sum E(X_j z_j \bar{X}_k \bar{z}_k) = E(|\sum X_j z_j|^2), \quad (6.8)$$

т. е. последовательность $\{\rho_n\}$ положительно определена. ►

Обратное утверждение также верно: какова бы ни была мера на окружности, найдется такая последовательность $\{X_n\}$, что ρ_n из (6.7) будут ее коэффициентами Фурье. Это можно было бы увидеть из построения в конце § 3, но мы еще вернемся к этому вопросу в § 8.

Примеры. а) Пусть X_n — действительные независимые одинаково распределенные случайные величины с $E(X_n) = \mu$ и $\text{Var}(X_n) = \sigma^2$. Тогда $\rho_0 = \sigma^2 + \mu^2$ и $\rho_k = \mu^2$ для всех $k \neq 0$. Спектральная мера представляет собой сумму вырожденной в нуле меры, имеющей массу μ^2 , и равномерного распределения с плотностью $\sigma^2/(2\pi)$.

б) Из построения § 4 видно, что плотности, определенной при фиксированных $r < 1$ и θ как $p_r(t - \theta)$, соответствуют коэффициенты Фурье $\rho_n = r^{|n|} e^{in\theta}$.

в) *Марковские процессы.* В гл. III, 8, было показано, что нормальная стационарная марковская последовательность имеет ковариации вида $\rho_n = r^{|n|}$ с $0 \leq r \leq 1$. Сходные рассуждения показывают, что ковариации произвольной комплексной стационарной марковской последовательности имеют вид $r^{|n|} e^{in\theta}$. При $r < 1$ спектральная мера имеет плотность $p_r(t - \theta)$, а при $r = 1$ она сосредоточена в точке θ . ►

§ 7. L^2 -ТЕОРИЯ

Для целей теории вероятностей мы ввели характеристические функции, определив их как некоторые преобразования мер. Однако в равной степени естественны и другие подходы к гармоническому анализу. В частности, возможно ввести преобразования Фурье для функций (а не для мер!). Формула обращения интегралов Фурье делает правдоподобным предположение о том, что на этом пути можно достичь большей симметрии. Оказывается, что простота и красота теории становятся наибольшими, когда рассматриваются лишь функции с интегрируемым квадратом. Мы разберем сейчас этот случай. Он интересен и сам по себе, и по тем обширным применениям, которые он находит в теории случайных процессов.

Определим *норму* $\|u\| \geq 0$ комплекснозначной функции u действительного аргумента x равенством

$$\|u\|^2 = \int_{-\infty}^{+\infty} |u(x)|^2 dx. \quad (7.1)$$

Две функции, отличающиеся только на множестве меры нуль,

мы будем рассматривать как идентичные (другими словами, мы будем фактически иметь дело с классами эквивалентности, но присоединимся к обычному и безвредному неправильному словопотреблению). При этом соглашении $\|u\|=0$ тогда и только тогда, когда $u=0$. Класс всех функций с конечной нормой обозначим L^2 . Расстояние между u и v из L^2 определяется как $\|u-v\|$. При таком определении L^2 становится метрическим пространством, и последовательность функций u_n из L^2 сходится в этой метрике тогда и только тогда, когда $\|u_n-u\|\rightarrow 0$. Эта сходимость¹⁾ будет обозначаться символами $u=\text{l. i. m. } u_n$ или $u_n\rightarrow u$.

Эту сходимость называют «сходимостью в среднем квадратичном». $\{u_n\}$ называют последовательностью Коши, если и только если

$$\|u_n-u_m\|\rightarrow 0 \quad \text{при } n, m\rightarrow\infty.$$

Отметим без доказательства, что метрическое пространство L^2 полно в том смысле, что каждая последовательность Коши $\{u_n\}$ имеет единственный предел $u\in L^2$.

Примеры. а) Функция u из L^2 интегрируема на любом конечном интервале, так как $|u(x)| < |u(x)|^2 + 1$ при всех значениях x . Это утверждение *перестает* быть верным в случае бесконечных интервалов. Действительно, функция $(1+|x|)^{-1}$ принадлежит L^2 , но не является интегрируемой.

б) Любая *ограниченная* интегрируемая функция принадлежит L^2 , поскольку неравенство $|u|\leq M$ влечет неравенство $|u|^2\leq M|u|$. Это неверно в случае неограниченных функций: функция $x^{-1/2}$ интегрируема на отрезке $\overline{0, 1}$, но не интегрируема с квадратом. ►

Скалярное (внутреннее) произведение (u, v) двух функций u и v из L^2 определяется равенством

$$(u, v) = \int_{-\infty}^{+\infty} u(x) \overline{v(x)} dx. \quad (7.2)$$

Оно существует для любых двух функций из L^2 , так как по неравенству Шварца

$$\int_{-\infty}^{+\infty} |uv| dx \leq \|u\| \cdot \|v\|. \quad (7.3)$$

¹⁾ Поточечная сходимость u_n к пределу v не влечет $u_n \xrightarrow{\text{l. m.}} v$ [см. пример 2, д) гл. IV]. Однако если известно, что $u=\text{l. i. m. } u_n$ существует, то $u=v$. В самом деле, по лемме Фату

$$\int_{-\infty}^{+\infty} |u(x)-v(x)|^2 dx \leq \liminf \int_{-\infty}^{+\infty} |u(x)-u_n(x)|^2 dx = 0.$$

В частности, $(u, u) = \|u\|^2$. При таком определении скалярного произведения L^2 становится *гильбертовым пространством*. Аналогия между скалярным произведением (7.2) и ковариацией двух случайных величин с нулевыми математическими ожиданиями очевидна и позже будет нами использована.

После этих предварительных замечаний перейдем к нашей основной задаче — определению преобразований вида

$$\hat{u}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u(x) e^{i\xi x} dx. \quad (7.4)$$

В случае если u — плотность вероятности, $\hat{u}$ лишь множителем $\sqrt{2\pi}$ отличается от характеристической функции. Чтобы избежать недоразумений, назовем $\hat{u}$ *преобразованием Планшереля функции u* . Определение (7.4) применимо только к интегрируемым функциям, однако мы расширим область определения u до всего пространства L^2 . Следующие примеры должны облегчить понимание процедуры и природы обобщенного преобразования.

Примеры. в) Поскольку любая функция u из L^2 интегрируема на любом конечном интервале, можно определить урезанные преобразования

$$\hat{u}^{(n)}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-n}^n u(x) e^{i\xi x} dx.$$

Заметим, что $\hat{u}^{(n)}$ есть истинное преобразование Планшереля функции $u^{(n)}$, которая определяется как $u^{(n)}(x) = u(x)$ при $|x| < n$ и $u^{(n)}(x) = 0$ при всех остальных x . При $n \rightarrow \infty$ последовательность $\hat{u}^{(n)}(\xi)$ не обязана сходиться ни при каком фиксированном ξ , однако мы покажем, что $\{\hat{u}^{(n)}\}$ является последовательностью Коши, и, следовательно, существует функция $\hat{u}$ из L^2 , такая, что $\hat{u} = \text{l.i.m. } \hat{u}^{(n)}$. Эту функцию $\hat{u}$ следует определить как преобразование Планшереля функции u , если даже интеграл в (7.4) не сходится. Конкретный вид урезания не играет роли: та же самая функция $\hat{u}$ может быть получена для любой другой последовательности интегрируемых функций $u^{(n)}$, сходящихся в среднем квадратичном к u .

г) Если u — плотность равномерного распределения на $[-1, 1]$, то ее преобразование Планшереля — функция $\hat{u} = \sin x / (x \sqrt{2\pi})$ не интегрируема. Однако $\hat{u}$ принадлежит L^2 , и мы увидим, что ее преобразование Планшереля совпадает с исходной плотностью u . Таким образом, мы получаем обобщение формулы обращения для интегралов Фурье (3.5) из гл. XV, применимой к плотностям, чьи характеристические функции не интегрируемы. ►

Перейдем к определению общего преобразования Планшереля. Для любой интегрируемой функции u ее преобразование $\hat{u}$ определяется формулой (7.4). Такая функция $\hat{u}$ непрерывна (в силу теоремы о мажорируемой сходимости), и величина $|\hat{u}|$ ограничена умноженным на $(2\pi)^{-1/2}$ интегралом от $|u|$. В общем случае $\hat{u}$ не интегрируема. Для краткости договоримся называть функцию u «хорошей», если она ограничена, непрерывна и интегрируема вместе с $\hat{u}$. Тогда $\hat{u}$ тоже «хорошая» функция, и обе, как u , так и $\hat{u}$, принадлежат L^2 [см. пример б)].

Сначала мы покажем, что для «хороших» функций справедлива формула обращения

$$u(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{u}(\xi) e^{-i\xi t} d\xi. \quad (7.5)$$

(Мы повторим рассуждение, приведенное для формулы обращения в гл. XV, 3). Умножая (7.4) на $\frac{1}{\sqrt{2\pi}} e^{-i\xi t - \frac{1}{2}\varepsilon^2 \xi^2}$ и интегрируя, получим

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{u}(\xi) e^{-i\xi t - \frac{1}{2}\varepsilon^2 \xi^2} d\xi = \int_{-\infty}^{+\infty} u(x) \Pi\left(\frac{t-x}{\varepsilon}\right) \frac{dx}{\varepsilon}, \quad (7.6)$$

где Π обозначает стандартную нормальную плотность. При $\varepsilon \rightarrow 0$ интеграл слева стремится к интегралу в (7.5), в то время как свертка справа стремится к $u(t)$. Таким образом, (7.5) для «хороших» функций выполняется.

Пусть теперь v — другая «хорошая» функция. Умножим (7.5) на функцию $\bar{v}$, комплексно сопряженную с v , и произведем интегрирование по множеству $-\infty < t < \infty$. Левая часть будет равна скалярному произведению (u, v) , а правая, после перемены порядка интегрирования, сведется к $(\hat{u}, \hat{v})$. Таким образом, «хорошие» функции удовлетворяют соотношению

$$(\hat{u}, \hat{v}) = (u, v), \quad (7.7)$$

которое будем называть равенством Парсеваля в L^2 . Для $v = u$ оно сводится к

$$\|\hat{u}\| = \|u\|. \quad (7.8)$$

Отсюда следует, что расстояние между преобразованиями $\hat{u}$ и $\hat{v}$ равно расстоянию между u и v , т. е. можно сказать, что преобразование Планшереля на множестве «хороших» функций изометрично.

Далее, покажем, что соотношения (7.7) и (7.8) остаются справедливыми для произвольных интегрируемых функций u и v ,

принадлежащих L^2 . Преобразования не обязаны быть интегрируемыми, однако из (7.8) вытекает, что они принадлежат L^2 (сравните примеры а) и г)).

Убедимся сначала в том, что интегрируемая функция ω , у которой две первые производные интегрируемы, обязательно является «хорошей». В самом деле, из леммы 4 гл. XV, 4, получаем, что $|\hat{\omega}(\xi)| = o(\xi^{-2})$ при $\xi \rightarrow \pm\infty$ и, следовательно, $\hat{\omega}$ интегрируема.

Предположим теперь, что u ограничена и интегрируема (и, следовательно, принадлежит L^2). По теореме об аппроксимации в среднем из гл. IV, 2, существует такая последовательность «хороших» функций u_n , что

$$\int_{-\infty}^{+\infty} |u(x) - u_n(x)| dx \rightarrow 0. \quad (7.9)$$

Если $|u| < M$, то все u_n можно выбрать так, чтобы $|u_n| < M$. Тогда u_n сходится в среднем квадратичном к u , так как $\|u - u_n\|^2$ не превосходит значения интеграла в (7.9), умноженного на $2M$. Таким образом, из равенства (7.8) для «хороших» функций вытекает, что $\{\hat{u}_n\}$ есть последовательность Коши. С другой стороны, при каждом фиксированном ξ $\hat{u}_n(\xi) \rightarrow \hat{u}(\xi)$, поскольку $|\hat{u}(\xi) - \hat{u}_n(\xi)|$ не превосходит значения интеграла в (7.9). Но, как было отмечено в последней сноске, поточечная сходимость элементов последовательности Коши влечет за собой сходимость в среднем квадратичном, и, следовательно,

$$\hat{u} = \text{l. i. m. } \hat{u}_n. \quad (7.10)$$

Применяя (7.7) к паре u_n и v и устремляя $n \rightarrow \infty$, видим, что (7.7) остается верным в том случае, когда функция u ограничена и интегрируема, а v — «хорошая» функция. Еще один переход к пределу показывает, что (7.7) остается справедливым для любой пары ограниченных интегрируемых функций.

Остается показать, что (7.7) справедливо также для неограниченных функций u и v при условии, что они интегрируемы и принадлежат L^2 . Для доказательства мы повторим предыдущие рассуждения с тем единственным изменением, что аппроксимирующие функции u_n определяются теперь с помощью следующего урезания: $u_n(x) = u(x)$ для таких x , что $|u(x)| < n$, и $u_n(x) = 0$ i. m.

для всех остальных x . Тогда (7.10) выполняется и $u_n \rightarrow u$. Дальнейшее доказательство проводится без изменений.

Теперь мы готовы сделать заключительный шаг, а именно распространить определение преобразования Планшереля на все пространство L^2 . Как показано в примере в), каждая функция u из L^2 является пределом последовательности Коши интегрируе-

мых функций u_n из L^2 . Мы только что показали, что преобразования $\hat{u}_n$, определенные по формуле (7.4), образуют последовательность Коши, и мы теперь определим $\hat{u}$ как предел этой последовательности. Так как две последовательности Коши могут быть соединены в одну, предел $\hat{u}$ не зависит от выбора аппроксимирующей последовательности $\{u_n\}$. Кроме того, если u окажется интегрируемой, мы можем положить $u_n = u$ при всех n , и, таким образом, мы видим, что новое определение согласуется с (7.4), если u интегрируема.

Подведем итоги.

Преобразование Планшереля $\hat{u}$ определено для каждой функции u из L^2 ; для интегрируемых u оно задается формулой (7.4), а в общем случае — с помощью следующего правила:

$$\text{если } u = 1. \text{ i. m. } u_n, \text{ то } \hat{u} = 1. \text{ i. m. } \hat{u}_n. \quad (7.11)$$

Равенство Парсеваля (7.7) и свойство изометрии (7.8) справедливы в общем случае. Отображение $u \rightarrow \hat{u}$ является взаимно однозначным; преобразованием $\hat{u}$ является $u(-x)$.

Последнее утверждение является вариантом формулы обращения Фурье (7.5), который применим в случае, когда u или $\hat{u}$ не интегрируемы, так что интегралы (7.4) и (7.5) не определены в обычном смысле. Эта полная симметрия между исходными функциями и их преобразованиями является основным преимуществом анализа Фурье в гильбертовых пространствах.

Изложенная теория находит широкое применение в теории прогнозирования случайных процессов. В качестве примера применения в теории вероятностей приведем критерий, обычно приписываемый А. Я. Хинчину, хотя он встречается еще в классической работе Н. Винера. В силу исторических традиций этот результат до сих пор связывается с их именами, хотя по существу он является частным случаем равенства Парсеваля и не требует отдельного доказательства.

Критерий Винера — Хинчина. Для того чтобы функция f была характеристической функцией плотности вероятностей f , необходимо и достаточно, чтобы существовала функция u , такая, что $\|u\|^2 = 1$ и

$$f(\lambda) = \int_{-\infty}^{+\infty} u(x) \overline{u(x+\lambda)} dx. \quad (7.12)$$

В этом случае $f = |\hat{u}|^2$.

Доказательство. При фиксированном λ положим $v(x) = u(x+\lambda)$.

Тогда $\hat{v}(\zeta) = \hat{u}(\zeta) e^{-i\lambda\zeta}$. Равенство Парсеваля (7.7) принимает вид

$$\int_{-\infty}^{+\infty} |\hat{u}(x)|^2 e^{i\lambda x} dx = \int_{-\infty}^{+\infty} u(x) \overline{u(x+\lambda)} dx, \quad (7.13)$$

и так как $\|\hat{u}\|^2 = 1$, то левая часть представляет собой характеристическую функцию некоторой плотности вероятности. Обратно, для заданной плотности вероятности f можно выбрать u так, что $f = |\hat{u}|^2$, и тогда (7.12) выполняется. Выбор u возможен не единственным способом. (Решение одной из задач теории прогнозирования случайных процессов связано с возможностью выбора функции u , равной нулю на полупрямой.) ▶

Изложенная здесь L^2 -теория интегралов Фурье переносится на ряды Фурье. В этом случае функции определены на окружности, а преобразования Фурье (или Планшереля) соответствуют теперь последовательности коэффициентов Фурье. За исключением этого формального различия, обе теории по существу одинаковы. Поэтому достаточно краткого изложения основных фактов.

Пространством L^2 в данном случае будет гильбертово пространство $L^2(-\pi, \pi)$ функций, интегрируемых с квадратом на окружности. Норма и скалярное произведение определяются формулами

$$\|u\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |u(x)|^2 dx, \quad (u, v) = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) \overline{v(x)} dx, \quad (7.14)$$

где подразумевается, что интегралы берутся по всей окружности (точки $-\pi$ и π отождествляются). Роль «хороших» функций играют конечные тригонометрические полиномы вида

$$u(x) = \sum u_n e^{inx}. \quad (7.15)$$

Их коэффициенты Фурье равны, очевидно,

$$u_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) e^{-inx} dx. \quad (7.16)$$

Каждой «хорошей» функции соответствует конечная последовательность $\{u_n\}$ ее коэффициентов, и, обратно, каждой конечной последовательности комплексных чисел соответствует «хорошая» функция. Соотношения (7.16) и (7.15) определяют преобразование Фурье $\hat{u} = \{u_n\}$ и обратное преобразование. Умножение и интегрирование показывают, что для двух «хороших» функций

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) \overline{v(x)} dx = \sum u_k \bar{v}_k. \quad (7.17)$$

Рассмотрим теперь гильбертово пространство $\mathfrak{H}$ бесконечных последовательностей $\hat{u} = \{u_n\}$, $\hat{v} = \{v_n\}$ и т. д. Определим норму и скалярное произведение равенствами

$$\|\hat{u}\| = \sum |u_n|^2, \quad (\hat{u}, \hat{v}) = \sum u_n \bar{v}_n. \quad (7.18)$$

Это пространство обладает свойствами, сходными со свойствами L^2 . В частности, конечные последовательности образуют всюду плотное множество. Можно показать, что существует взаимно однозначное соответствие между последовательностями $\hat{u} = \{u_n\}$ из $\mathfrak{H}$ и функциями u из $L^2(-\pi, \pi)$. Каждой последовательности $\{u_n\}$, для которой $\sum |u_n|^2 < \infty$, соответствует функция с интегрируемым квадратом и коэффициентами Фурье u_n . Верно и обратное утверждение. Отображение $u \leftrightarrow \{u_n\}$ снова изометрично. Снова имеет место равенство Парсеваля $(u, v) = (\hat{u}, \hat{v})$. Ряд Фурье не обязан сходиться, но его частные суммы $\sum_{-n}^n u_k e^{ikx}$ образуют последовательность непрерывных функций, сходящуюся к u в метрике L^2 . Это же верно и по отношению к другим непрерывным аппроксимациям: так, $\sum u_k r^{|k|} e^{ikx}$ сходится к u при $r \rightarrow 1$.

Рассмотрим, как и выше, специальный случай равенства Парсеваля

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) \bar{v}(x) e^{-inx} dx = \sum_k u_{k+n} \bar{v}_k. \quad (7.19)$$

При $u=v$ получаем, что последовательность $\{\varphi_n\}$ является последовательностью коэффициентов Фурье некоторой плотности вероятности на $-\pi, \pi$ в том и только том случае, когда она допускает представление

$$\varphi_n = \sum u_{k+n} \bar{u}_k, \quad \text{где } \sum |u_k|^2 = 1. \quad (7.20)$$

Ковариация такого вида встречалась в (7.4) гл. III (см. также задачу 17).

§ 8. СЛУЧАЙНЫЕ ПРОЦЕССЫ И СТОХАСТИЧЕСКИЕ ИНТЕГРАЛЫ

Ради упрощения обозначений мы рассмотрим в этом параграфе лишь последовательности $\{X_n\}$ случайных величин. Однако, и это станет очевидным из дальнейшего, все изложение пригодно и для семейств, зависящих от непрерывного параметра. Отличие лишь в том, что спектральная мера не будет сосредоточена на конечном промежутке, и суммы должны быть заменены интегралами.

Пусть тогда $\{X_n\}$ обозначает бесконечную в обе стороны пос-

ледовательность случайных величин, заданных в некотором вероятностном пространстве $\mathfrak{E}$ и имеющих конечные вторые моменты. Эта последовательность предполагается стационарной «в широком смысле»¹⁾, т. е.

$$E(X_{n+v}\overline{X}_v) = \rho_n$$

не зависит от v . По теореме 3 из § 6 существует единственная мера R на окружности $\overline{[-\pi, \pi]}$, такая, что

$$\rho_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} R\{dx\}, \quad n=0, \pm 1, \dots \quad (8.1)$$

Мы рассмотрим теперь более детально отмеченную еще в § 3 идею о том, что окружность вместе со спектральной мерой R могут быть использованы для построения *конкретного представления случайного процесса* $\{X_n\}$ (представления, пригодного во всяком случае для изучения свойств, зависящих только от вторых моментов). Мы будем применять терминологию гильбертовых пространств и рассмотрим два гильбертовых пространства.

а) *Пространство L_R^2* . Построим пространство функций, определенных на окружности, повторяя (с заменой прямой на окружность $\overline{[-\pi, \pi]}$ и меры Лебега на меру R) рассуждения, касающиеся L^2 (см. § 7). Норма и скалярное произведение (комплекснозначных) функций на окружности определяются при этом равенствами

$$\|u\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |u(x)|^2 R\{dx\}, \quad (u, v) = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) \overline{v(x)} R\{dx\}. \quad (8.2)$$

Примем основное соглашение: считать две функции равными, если они отличаются друг от друга на множестве R -меры нуль. Последствия этого соглашения серьезны. Так, если R сосредоточено в двух точках 0 и 1, то «функция» в нашем понимании вполне определяется своими значениями в этих точках. Например, функция πx равна нулю. Однако даже в этих крайних случаях использование обычных формул для непрерывных функций и ссылки на их графики не принесут никакого вреда. Точно так же осмысленно (и упрощает язык) применение термина «ступенчатая функция».

Гильбертово пространство L_R^2 состоит из всех функций, заданных на окружности и имеющих конечную норму. Если $\|u_n - u\| \rightarrow 0$, то мы назовем последовательность $\{u_n\}$ сходящейся к u в нашей метрике (или в среднем квадратичном по отношению

¹⁾ У автора — «in restricted sense». — Прим. перев.

к мере R). Гильбертово пространство L_R^2 есть полное метрическое пространство. Непрерывные функции образуют в нем плотное множество (определения см. в § 7).

б) *Гильбертово пространство $\mathfrak{H}$, натянутое на $\{X_n\}$* . Обозначим символом $\mathfrak{H}_0$ семейство случайных величин с конечными вторыми моментами, заданных на произвольном, но фиксированном выборочном пространстве $\mathfrak{S}$. По неравенству Шварца $E(U\bar{V})$ существует для любой пары таких случайных величин. Естественно обобщить (8.2), рассматривая вместо круга выборочное пространство $\mathfrak{S}$ и вместо меры R основное распределение вероятностей в $\mathfrak{S}$. Мы опять условимся отождествлять две случайные величины, если они отличаются друг от друга на множестве меры нуль. Определим скалярное произведение как $E(U\bar{V})$ и норму U как положительный корень из $E(U\bar{U})$. Тогда $\mathfrak{H}_0$ превращается в гильбертово пространство. Это — полное метрическое пространство, в котором последовательность принадлежащих ему случайных величин U_n называют сходящейся к U , если $E(|U_n - U|^2) \rightarrow 0$.

Изучая последовательность $\{X_n\}$, обычно интересуются только случайными величинами, которые являются функциями X_k , а во многих случаях ограничиваются лишь линейными функциями, т. е. рассматривают конечные линейные комбинации $\sum a_k X_k$ и их пределы. Получаемые таким образом случайные величины образуют подпространство $\mathfrak{H}$ пространства $\mathfrak{H}_0$, называемое *гильбертовым пространством, натянутым на величины X_k* . В нем скалярные произведения, нормы и сходимость определены, как только что описано, и $\mathfrak{H}$ является полным метрическим пространством.

В настоящем контексте математические ожидания $E(X_n)$ не играют никакой роли, но, так как термин «ковариация» звучит лучше, чем «скалярное произведение», мы примем обычное предположение, что $E(X_n) = 0$. Единственная цель этого предположения — добиться совпадения ρ_n и ковариации. Никакое центрирование не нужно, если согласиться назвать $E(X\bar{Y})$ ковариацией X и Y . Мы подходим теперь к решающему моменту. Мы покажем, что для наших целей возможно взять пространство L_R^2 (простое для интуитивного восприятия) в качестве конкретной модели $\mathfrak{H}$. В самом деле, по определению ковариация $\rho_{j-k} = \text{Cov}(X_j, X_k)$ любой пары X равна скалярному произведению функций e^{ijx} и e^{ikx} в L_R^2 . Отсюда следует, что ковариация конечных двух линейных комбинаций $U = \sum a_j X_j$ и $V = \sum b_k X_k$ равна скалярному произведению соответствующих линейных комбинаций $u = \sum a_j e^{ijx}$ и $v = \sum b_k e^{ikx}$. Определения сходимости в рассматриваемых пространствах позволяют распространить отображение на все случайные величины. Мы получаем, таким образом, важный результат: отображение $X_k \leftrightarrow e^{ikx}$ порождает взаимно однозначное соответствие между случайными величинами из $\mathfrak{H}$ и функциями из L_R^2 ; это соответствие сохраняет скалярные произведения и нормы

(и, следовательно, сходимость). Иными словами, эти пространства изометричны ¹⁾. Теперь мы в состоянии изучать $\mathfrak{H}$ и $\{X_n\}$, рассматривая лишь конкретное пространство L^2_R . Эта процедура, помогая интуиции, в то же время имеет и теоретические преимущества. Так как функции на окружности—объект хорошо знакомый, то сравнительно просто построить последовательность $\{u^{(n)}\}$ функций из L^2_R с заданными структурными свойствами. Функциям u^n соответствуют некоторые случайные величины Z_n , определенные на исходном выборочном пространстве $\mathfrak{S}$. Если известны коэффициенты Фурье $u^{(n)}$, то Z_n возможно выразить явно в виде предела конечных линейных комбинаций величин X_k . Если совместные распределения X_k нормальны, то и распределение Z_n будет нормальным.

На практике указанная процедура часто производится в обратном порядке. Если данный процесс $\{X_k\}$ —сложный, то наша цель—выразить его в терминах величин Z_n , образующих более простой случайный процесс. Практически это достигается переходом от исходного пространства к пространству L^2_R . Несколько примеров пояснят это лучше, чем теоретические рассуждения.

Примеры. а) *Представление $\{X_n\}$ в терминах независимых случайных величин.* Как и всюду в этом параграфе, $\{X_n\}$ обозначает случайный процесс с ковариациями ρ_n и спектральной мерой R , определенными формулой (8.1). При нашем отображении случайная величина X_n соответствует функции e^{inx} на окружности. Покажем теперь, что для определенных функций γ случайные величины, соответствующие $e^{inx}/\gamma(x)$, не коррелированы. Рассмотрим только случай, когда спектральная мера R имеет плотность r . Для простоты допустим, что r строго положительна и непрерывна ²⁾. Выберем функцию γ так, что

$$|\gamma(x)|^2 = r(x). \quad (8.3)$$

¹⁾ Читателям, знакомым с теорией гильбертовых пространств, полезно иметь в виду связь изложенного со стандартной *спектральной теоремой для унитарных операторов*. Линейный оператор, отображающий $\mathfrak{H}$ на себя таким образом, что $X_n \rightarrow X_{n+1}$, называют *оператором сдвига*, и L^2_R служит моделью, в которой действием этого оператора сдвига соответствует умножение на e^{ix} . Обратно, пусть дан произвольный унитарный оператор T на гильбертовом пространстве $\mathfrak{H}_0$ и произвольный элемент $X_0 \in \mathfrak{H}_0$. Тогда последовательность элементов $X_n = T^n X_0$ можно рассматривать как стационарную последовательность и T —как оператор сдвига в подпространстве $\mathfrak{H}$, натянутом на эту последовательность. Если X_0 можно выбрать так, что $\mathfrak{H} = \mathfrak{H}_0$, мы получаем стандартную спектральную теорему для T (с тем отличием, что для «разложения единицы» взято конкретное представление, основанное на выборе X_0). Если $\mathfrak{H} \subset \mathfrak{H}_0$, то $\mathfrak{H}_0$ можно представить как прямую сумму двух инвариантных подпространств и можно повторить рассуждения. Таким путем возможно прийти к общей теории спектральных разложений, включая теорию кратности спектра.

²⁾ Эти ограничения используются лишь затем, чтобы избежать скучных объяснений того, как понимать $r(x)/\gamma(x)$ при $r(x) = \gamma(x) = 0$ и как понимать сходимость рядов,

Ряд Фурье γ сходится в L^2 -норме, как это было объяснено в § 7. Обозначая коэффициенты Фурье γ через γ_k , имеем $\sum |\gamma_k|^2 < \infty$ и по равенству Парсеваля (7.20)

$$\rho_n = \sum_{k=-\infty}^{+\infty} \gamma_{k+n} \overline{\gamma_k}. \quad (8.4)$$

Рассмотрим теперь бесконечную в обе стороны последовательность $u^{(n)}$, определенную соотношением

$$u^{(n)}(x) = \frac{e^{in x}}{\gamma(x)}. \quad (8.5)$$

Подставляя $u^{(n)}$ в (8.2), видим, что

$$\|u^{(n)}\| = 1, \quad (u^{(n)}, u^{(m)}) = 0 \quad (8.6)$$

при $m \neq n$. Отсюда вытекает, что соответствующие функциям $u^{(n)}$ случайные величины Z_n не коррелированы и имеют единичные дисперсии. В частности, если X_k нормально распределены, то Z_k взаимно независимы.

Интересно отметить, что пространство, натянутое на случайные величины X_k , содержит стационарную последовательность $\{Z_n\}$ из некоррелированных величин. Явное выражение Z_n в терминах X_k может быть получено из разложения Фурье функции $u^{(n)}$. Однако предпочтительнее идти в обратном направлении — выразить X_k в терминах Z_n , так как структура последовательности $\{Z_n\}$ проще, чем структура $\{X_k\}$. Мы имеем

$$\sum_{n=-N}^N \gamma_n u^{(n+k)}(x) = \frac{e^{ikx}}{\gamma(x)} \sum_{n=-N}^N \gamma_n e^{inx}. \quad (8.7)$$

Сумма справа представляет собой отрезок ряда Фурье для γ и сходится (в метрике гильбертова пространства) к γ . Отсюда следует, что величина (8.7) стремится к e^{ikx} . При нашем отображении $u^{(n+k)}$ соответствует Z_{n+k} , так что ряд $\sum \gamma_n Z_{n+k}$ сходится, и мы можем написать

$$X_k = \sum_{n=-\infty}^{+\infty} \gamma_n Z_{n+k}. \quad (8.8)$$

Мы получили, таким образом, явное выражение X_k в форме «скользящих средних» стационарной последовательности некоррелированных величин Z_n .

Очевидно, что представление (8.8) не единственно. Возникает естественный вопрос, возможно ли выразить X_k только через Z_k , Z_{k-1} , Z_{k-2} , ... (через «прошлое» процесса $\{Z_n\}$), или, что то же самое, возможно ли функцию γ в (8.3) выбрать так, что $\gamma_n = 0$ при $n \geq 1$. Эта проблема имеет фундаментальный характер для теории прогнозирования случайных процессов, но она лежит вне

рамок этой книги. Типичные примеры см. в (7.5) гл. III и задаче 18.

б) *Соответствующий процесс с некоррелированными приращениями.* Для каждого t , $-\pi < t \leq \pi$, положим

$$y_t(x) = \begin{cases} 1 & \text{при } x \leq t, \\ 0 & \text{при } x > t \end{cases} \quad (8.9)$$

и обозначим через Y_t соответствующую случайную величину в §. Очевидно, что ковариация приращений $Y_t - Y_s$ за непересекающиеся интервалы равна 0; более того, $\text{Var}(Y_t) = R\{-\pi, t\}$. Таким образом, Y_t — процесс с *некоррелированными приращениями и дисперсией, определяемой по R* . Если X_t нормально распределены, то приращения процесса Y_t будут независимы.

Итак, для каждой стационарной последовательности $\{X_k\}$ существует связанный с ней указанным образом процесс с некоррелированными приращениями. Явное выражение для Y_t в терминах X_k можно получить стандартным путем — разлагая функцию y_t из (8.9) в ряд Фурье. Однако снова предпочтительнее идти в обратном направлении. Это будет сделано в следующем примере.

в) *Стохастические интегралы.* Представление какой-либо случайной величины U в терминах X_k определяется (как мы видели) разложением Фурье функции, соответствующей U . По сравнению с этим представление в терминах Y_t может показаться «слишком простым». Оно связано с графиком функции, соответствующей рассматриваемой случайной величине. Для простоты мы предположим эту функцию непрерывной.

Рассмотрим сначала «ступенчатую функцию» w , т. е. функцию вида

$$w = a_1 y_{t_1} + a_2 (y_{t_2} - y_{t_1}) + \dots + a_n (y_{t_n} - y_{t_{n-1}}), \quad (8.10)$$

где a_j — постоянные и $-\pi < t_1 < t_2 < \dots < t_{n-1} < \pi$. Соответствующая случайная величина W получается заменой в (8.10) y_{t_j} на Y_{t_j} . Теперь каждая непрерывная функция w может быть равномерно приближена ступенчатыми функциями $w^{(n)}$ вида (8.10). Равномерная сходимость $w^{(n)}$ к w влечет за собой сходимость в L_R^2 -норме, а следовательно, и сходимость соответствующих случайных величин $W^{(n)}$ к W . Это дает нам рецепт отыскания образа W произвольной непрерывной функции w с помощью простого предельного перехода: надо аппроксимировать w ступенчатыми функциями типа (8.10) и заменить y_{t_j} на Y_{t_j} . Напомним, что (8.10) определяет функцию, а не число; точно так же как и w , предел $w^{(n)}$, является функцией, а не числом. Но (8.10) *выглядит* как риманова сумма, и формально наша процедура напоминает определение интеграла Римана. Поэтому общепринятым стало

обозначение

$$W = \int_{-\pi}^{\pi} w(t) dY_t, \quad (8.11)$$

указывающее на описанный выше предельный переход. Случайная величина (8.11) называется *стохастическим интегралом* от непрерывной функции w . Название это до некоторой степени произвольно, а обозначение — не более как краткое выражение предельного перехода, который был строго определен выше. По определению функция e^{int} соответствует случайной величине X_n , и потому мы можем написать

$$X_n = \int_{-\pi}^{\pi} e^{int} dY_t. \quad (8.12)$$

Это есть *спектральное представление произвольной стационарной последовательности $\{X_n\}$ в терминах соответствующего процесса с некоррелированными приращениями*.

Обозначение стохастического интеграла, пожалуй, более наглядно, чем логично, но мы не будем касаться этого вопроса. Наша цель состояла в том, чтобы показать, как это полезное понятие и важное представление (8.12) могут быть получены средствами анализа Фурье. Этим иллюстрируются возможности канонического отображения, использованного в этом разделе (и введенного впервые Крамером). ►

Изложенная теория использует только вторые моменты $\{X_n\}$ и практически применима лишь в случаях, где эти моменты в самом деле значимы. Таков случай, когда процесс является нормальным, так как нормальные распределения полностью определяются своими ковариациями. В других применениях можно верить, что процесс «не слишком отклоняется от нормального» (подобно тому как регрессионный анализ основывался на вере в универсальную приложимость методов, развитых для нормального закона). К сожалению, один лишь факт существования прекрасной теории никоим образом не оправдывает этой веры. Так, в примере 3, в) выборочные функции процесса строго периодичны. Будущее отдельной реализации вполне определяется ее заданием на полном периоде. Однако методы прогнозирования, опирающиеся на L^2 -теорию, не принимают во внимание этот факт и отождествляют все процессы с одной и той же спектральной мерой. Наблюдатель, получающий последовательность $\dots, 1, -1, 1, -1, \dots$, может однозначно предсказать следующий член ряда, но L^2 -методы заставляют его предсказывать появление нуля. Эти методы, таким образом, не являются универсально приложимыми. Однако они дают идеальное средство изучения нормальных процессов.

§ 9. ЗАДАЧИ

1. Любопытные примеры характеристических функций. Пусть $\tau_h(x) = 1 - |x|/h$ для $|x| \leq h$ и $\tau_h(x) = 0$ для $|x| \geq h$. Положим

$$\alpha(x) = \sum_{n=-\infty}^{\infty} a_n \tau_h(x-n). \quad (9.1)$$

Когда a_n вещественны и $h=1$, график α представляет собой ломаную линию с вершинами (n, a_n) . При $h < \frac{1}{2}$ график α состоит из отрезков оси x и боковых сторон равнобедренных треугольников с вершинами (n, a_n) .

Используя критерий леммы 1 из § 2, покажите, что $\alpha(\zeta)/\alpha(0)$ является характеристической функцией, если a_n действительные, $a_{-n} = a_n$ и $|a_1| + |a_2| + \dots \leq \frac{1}{2} a_0$.

2. Обобщение. Последнее утверждение остается верным, если τ_h заменить произвольной четной интегрируемой характеристической функцией. (Используя характеристические функции, изображенные на рис. 1 гл. XV, 2, можно построить характеристические функции, графики которых представляют собой весьма причудливые ломаные линии.)

3. Продолжение. Результат предыдущей задачи является специальным случаем следующего утверждения: если τ — четная интегрируемая характеристическая функция, если λ_n — действительные, а a_n — комплексные постоянные и $\sum |a_n| < \infty$, то

$$\alpha(\zeta) = \sum a_k \tau(\zeta - \lambda_k) \quad (9.2)$$

есть характеристическая функция в том и только том случае, когда $\alpha(0) = 1$ и $\sum a_k e^{-i\lambda_k \zeta} \geq 0$ для всех ζ . (На самом деле достаточно, чтобы последний ряд был суммируем по Абелю к положительной функции.)

4. Ковариационная функция ρ , определенная в (3.3), непрерывна всюду тогда и только тогда, когда она непрерывна в нуле. Последнее верно в том и только том случае, когда $E(X_t - X_0)^2 \rightarrow 0$ при $t \rightarrow 0$.

5. Разностные отношения и производные. Пусть $\{X_t\}$ — стационарный процесс с $\rho(t) = E(X_{t+s} \bar{X}_s)$ и спектральной мерой R . Для $h > 0$ определим новый процесс как $X_t^{(h)} = (X_{t+h} - X_t)/h$.

а) Покажите, что спектральная мера $R^{(h)}$ нового процесса определяется соотношением $R^{(h)}\{dx\} = 2h^{-2} [1 - \cos hx] R\{dx\}$. Ковариации $\rho^{(h)}(t)$ стремятся к пределу при $h \rightarrow 0$ в том и только том случае, если существует непрерывная вторая производная $\rho''(t)$, т. е. если мера $x^2 R\{dx\}$ конечна.

б) В последнем случае $E(|X_t^{(\varepsilon)} - X_t^{(\delta)}|^2) \rightarrow 0$ при $\varepsilon \rightarrow 0$ и $\delta \rightarrow 0$.

Замечание. В терминологии гильбертовых пространств из § 7 это означает, что при фиксированном t и $\varepsilon_n \rightarrow 0$ последовательность $\{X_t^{(\varepsilon_n)}\}$ есть последовательность Коши и поэтому производная $X'_t = \text{i. t. m. } X_t^{(h)}$ существует.

6. См. теорему 2 из § 4. Если φ непрерывна всюду, кроме нуля, где она имеет разрыв первого рода, то $f_r(\zeta) \rightarrow \varphi(\zeta)$ равномерно вне некоторой окрестности нуля и $f_r(0) \rightarrow \frac{1}{2} [\varphi(0+) - \varphi(0-)]$.

7. Продолжение. Если φ может быть представлена как разность двух монотонных функций, то $f_r(\zeta) \rightarrow \frac{1}{2} [\varphi(\zeta+) - \varphi(\zeta-)]$ во всех точках.

8. Ограниченная периодическая функция φ с неотрицательными коэффициентами Фурье φ_n необходимо непрерывна и $\sum \varphi_n < \infty$. Пример $\varphi_n = \frac{1}{n}$ показывает, что утверждение неверно, если предполагается только интегрируемость φ . Указание. Используйте основную идею доказательства теоремы 1 из § 4.

9. Суммируемость по Чезаро. Заменим (4.3) на

$$f_r(\zeta) = \sum \varphi_n a_n e^{in\zeta},$$

где $a_n = 1 - |n| (2N+1)^{-1}$ при $|n| \leq 2N$ и $a_n = 0$ при $|n| > 2N$. Покажите, что при замене $p_r(t)$ на

$$q_N(t) = \frac{1}{2N+1} \frac{\sin^2\left(N + \frac{1}{2}\right)t}{\sin^2 \frac{1}{2}t}$$

сохраняется теория § 4 ($q_N(t)$ снова плотность вероятности).

10. Продолжение. Докажите более общее утверждение, что эта теория сохраняется, если в качестве a_n взять коэффициенты Фурье симметризованной плотности вероятности, заданной на окружности.

11. Покажите, используя формулу суммирования Пуассона (5.2), что

$$\sum_{-\infty}^{\infty} \left\{ n \left(\frac{y-x+2k\lambda}{\sqrt{t}} \right) + n \left(\frac{y+x+2k\lambda}{\sqrt{t}} \right) \right\} = \\ = \frac{1}{\lambda} \sum_{-\infty}^{\infty} \exp \left(-\frac{1}{2} t n^2 \frac{\pi^2}{\lambda^2} \right) \cos \frac{n\pi}{\lambda} x \cdot \cos \frac{n\pi}{\lambda} y,$$

где n обозначает стандартную нормальную плотность. [Это есть решение задачи об отражающем экране, см. гл. X, § 5, д.)]

12. В формуле суммирования Пуассона (5.2) условие о том, что $\sum \varphi(\zeta + 2k\lambda)$ сходится к непрерывной функции ψ , можно заменить на условие $\sum f(n\pi/\lambda) < \infty$. Указание: ψ — при любых обстоятельствах интегрируемая функция. Используйте следствие 2 § 4.

13. Другой вывод формулы суммирования Пуассона. Пусть φ — характеристическая функция произвольного распределения F . Если p_r обозначает ядро Пуассона из (4.5), то (4.7) показывает (без дальнейших вычислений), что при $0 < r < 1$

$$\frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \varphi(\zeta+n) r^{|n|} e^{in\lambda} = \int_{-\infty}^{\infty} e^{i\zeta x} p_r(x+\lambda) F\{dx\}. \quad (9.3)$$

Отсюда следует, что левая часть есть характеристическая функция. Полагая $r \rightarrow 1$, выведите отсюда, что если F имеет плотность f , то

$$\frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \varphi(\zeta+n) e^{in\lambda} = \sum_{k=-\infty}^{+\infty} e^{i\zeta(-\lambda+2k\pi)} f(-\lambda+2k\pi) \quad (9.4)$$

в предположении $\sum f(-\lambda+2k\pi) < \infty$. Покажите, что (9.4) равносильно варианту (5.9) как формуле суммирования.

Замечание. Результат может быть переформулирован следующим образом: левая часть (9.4) суммируема по Абелю к правой части, если последняя непрерывна.

14. Последовательность $\{\varphi_n\}$ положительно определена тогда и только тогда, когда последовательность $\{\varphi_n r^{|n|}\}$ положительно определена при любом r , $0 < r < 1$. Для этого необходимо и достаточно условие $\sum \varphi_n r^{|n|} e^{in\lambda} \geq 0$ для всех λ и $0 < r < 1$.

15. Выведите (без вычислений) из задач 1 и 14 следующую теорему (отмеченную Л. Шеппом). Пусть $\{\varphi_n\}$ — положительно определенная последова-

тельность. Обозначим через α кусочно-линейную функцию с вершинами (n, φ_n) . Тогда α положительно определена.

16. Пусть φ — характеристическая функция. Обозначим через α кусочно-линейную функцию с вершинами $(n, \varphi(n))$. Тогда α — характеристическая функция (это лишь иная формулировка задачи 15 и частный случай задачи 1 с $h=1$). Используйте задачу 1 и опишите другие любопытные характеристические функции, которые можно получить из φ .

17. При $r < 1$ ковариации $\rho_n = r^{|n|} e^{in\theta}$ марковских последовательностей удовлетворяют (7.20) с $u_k = \sqrt{1-r^2} r^k e^{ik\theta}$ при $k \geq 0$ и $u_k = 0$ при $k < 0$. Найдите другие представления.

18. Продолжение. Пусть $\{X_n\}$ — марковская последовательность с ковариациями $\rho_n = r^{|n|} e^{in\theta}$. Если $r < 1$, то $X_n = \sqrt{1-r^2} \sum_{k=0}^{\infty} r^k e^{ik\theta} Z_{n-k}$, где величины Z_k не коррелированы. При $r=1$ имеем $X_n = e^{in\theta} Z_0$.

Глава I

1. (i) $\frac{\alpha}{3} \frac{2}{\sqrt[3]{x^2}} e^{-\alpha x^{1/3}}$ (ii) $\frac{\alpha}{2} e^{-\alpha(x-3)/2}$ при $x > 3$
 (iii) $\frac{\alpha}{2} e^{-\alpha|x|}$ при всех x
 (iv) $\alpha e^{-\alpha x}$ при $x > 0$
 (v) $\alpha \left(1 + \frac{1}{4} \frac{1}{\sqrt[3]{x^2}}\right) e^{-\alpha x - \alpha x^{1/3}}$
 (vi) $\alpha e^{-\alpha x} + \frac{\alpha}{3} \frac{1}{\sqrt[3]{x^2}} e^{-\alpha x^{1/3}} - \alpha \left(1 + \frac{1}{3} \frac{1}{\sqrt[3]{x^2}}\right) e^{-\alpha x - \alpha x^{1/3}}.$

2. (i) $\frac{1}{6} \frac{1}{\sqrt[3]{x^2}}$ при $|x| < 1$

(ii) $\frac{1}{4}$ при $1 < t < 5$

(iii) $\frac{1}{2} \left(1 - \frac{|x|}{2}\right)$ при $|x| < 2$

(iv) $1 - \frac{x}{2}$ при $0 < x < 2$

(v) $\frac{1}{4} - \frac{1}{3} x^{1/3} + \frac{1}{12} x^{-2/3}$ при $|x| < 1$

(vi) $\frac{1}{4} + \frac{1}{3} x^{1/3} + \frac{1}{12} x^{-2/3}$ при $|x| < 1.$

3. (i) $h^{-1}(1 - e^{-\alpha x})$ при $0 < x < h$ и $h^{-1}(e^{\alpha h} - 1) e^{-\alpha x}$ при $x > h$

(ii) $h^{-1}(1 - e^{-\alpha(x+h)})$ при $-h < x < 0$

и $h^{-1}(1 - e^{-\alpha h}) e^{-\alpha x}$ при $x > 0.$

4. (i) $\frac{1}{3} h$, если $h \leq 1$, и $\frac{2}{3} \frac{1}{\sqrt[3]{h}}$, если $h \geq 1$

(ii) $\sqrt[3]{\alpha \pi} e^{\frac{1}{4} \alpha} (1 - \Re(\sqrt[3]{\alpha/2})).$

5. (i) $1 - x^{-1}$ при $x > 1$;

(ii) $x^2(x+1)^{-2}$ при $x > 0.$

7. $P\{Z \leq x\} = 1 - e^{-\alpha x}$ при $x < t$ и 1 при $x > t.$

10. б) Колонна вместе с возглавляющим и замыкающим автомобилями образует выборку, в которой наименьший элемент на последнем месте, наименьший из оставшихся на втором.

16. $p = \sum_{k=0}^{m-1} \binom{n+k-1}{k} 2^{-n-k}$, При $m=1$, $n=2$ получаем $p = \frac{1}{4}.$

18. $nt^{n-1} - (n-1)t^n.$

19. (i) $2 \int_0^1 dx \int_x^1 (1-z) dz = \frac{1}{3}$
 (ii) Плотность равна $2t - t^2$ при $0 < t \leq 1$
 и $(2-t)^2$ при $1 \leq t < 2$
 (iii) Плотность равна $2t^2$ при $0 < t < 1$ и снова $(2-t)^2$ при $1 \leq t < 2$.
20. $2 \int_0^1 x(1-x) dx = \frac{1}{3}$. Две из шести перестановок приводят к пересечению.
21. $X_{11}: 4 \log \frac{1}{4x}$ при $x < \frac{1}{4}$; X_{12} и $X_{21}: 4 \log 2$ при $x < \frac{1}{4}$; $4 \log \frac{1}{2x}$ при $\frac{1}{4} < x < \frac{1}{2}$;
 $X_{22}: 4 \log 4x$ при $\frac{1}{4} < x < \frac{1}{2}$; $4 \log \frac{1}{x}$ при $\frac{1}{2} < x < 1$. Математические ожидания равны $\frac{1}{16}, \frac{3}{16}, \frac{9}{16}$.
27. Функции распределения $\frac{2}{\pi} \arcsin \frac{1}{2}x$ и $\frac{1}{4}x^2$; плотности $\frac{2}{\pi} \frac{1}{\sqrt{4-x^2}}$ и $\frac{1}{2}x$ при $0 < x < 2$.
28. $2\pi^{-1} \arcsin \frac{1}{2}x$.
30. а) $\log \frac{1}{x}$,
 б) $\frac{2}{\pi} \log \frac{1 + \sqrt{1-x^2}}{x}$, где $0 < x < 1$.
31. $\frac{4}{\pi} \int_{0 < \cos \theta < x} \sin^2 \theta d\theta = \frac{4}{\pi} \int_0^x \sqrt{1-y^2} dy = 2\pi^{-1} [\arcsin x + x\sqrt{1-x^2}]$, где $0 < x < 1$.
32. $F(t) = \frac{2}{\pi} \int_0^{\pi/2} V\left(\frac{t}{\cos \theta}\right) (1 - \cos 2\theta) d\theta$.
35. Подстановка $z = F(y)$ преобразует интеграл к соответствующему интегралу для равномерного распределения. Заметим, что $F(m+x) \approx \frac{1}{2} + f(m)x$ для малых x .

Глава II

4. $g * g(x) = \frac{1}{4} e^{-|x|} (1 + |x|)$
 $g^{3*}(x) = \frac{1}{16} e^{-|x|} (3 + 3|x| + x^2)$
 $g^{4*}(x) = \frac{1}{32} e^{-|x|} \left(5 + 5|x| + 2x^2 + \frac{1}{3}|x|^3\right)$.
10. а) $\lambda \mu (e^{-\lambda t} - e^{-\mu t}) (\mu - \lambda)$ как свертки двух показательных плотностей.
 б) Используя а), получаем $\lambda e^{-\lambda t}$.
12. Плотность равна $1 - \frac{1}{2}t^2$ при $0 < t < 1$ и $\frac{1}{2}(2-t)^2$ при всех $1 < t < 2$.
 Математическое ожидание равно $\frac{7}{12}$.

$$13. \binom{j+k+\gamma-1}{j} \binom{m+n+\mu-j-k-1}{m-j} / \binom{m+n+\mu+\gamma-1}{m}.$$

Глава III

7. а) e^{-x} и $1 - e^{-x} - e^{-y} + e^{-x-y-ax}$ при $x > 0, y > 0$.

$$б) E(Y|X) = \frac{1+a+ax}{(1+ax)^2},$$

$$\text{Var}(Y|X) = \frac{1}{(1+ax)^2} + \frac{2a}{(1+ax)^3} - \frac{a^2}{(1+ax)^4}.$$

8. Если f имеет математическое ожидание μ и дисперсию σ^2 , то $E(X) = E(Y) = \frac{1}{2}\mu$, $\text{Var}(X) = \text{Var}(Y) = \frac{1}{3}\sigma^2 + \frac{1}{12}\mu^2$, $\text{Cov}(X, Y) = \frac{1}{6}\sigma^2 - \frac{1}{12}\mu^2$.

9. Плотность равна $2x_2$ в единичном квадрате. Для выборки объема n она равна $(n-1)! x_2 x_3^2, \dots, x_{n-1}^{n-2}$ в $(n-1)$ -мерном единичном кубе.

10. $\frac{1}{3}e^{-(x+y)}$ при $y > x > 0$ и $\frac{1}{3}e^{-y+2x}$ при $y > x, x < 0$. Если $y < x$, поменяйте x и y местами.

$$11. а) 4 \int_{2x}^{\frac{1}{2}} f(s) f\left(\frac{x}{s}\right) \frac{ds}{s}, \quad 0 < x < \frac{1}{4}.$$

б) $8 \int f(s) f\left(\frac{x}{s}\right) f\left(\frac{y}{1-s}\right) \frac{ds}{s(1-s)}$, где $0 < x < \frac{1}{4} < y < 1$ и область интегрирования удовлетворяет условиям: $2x < s < \frac{1}{2}$ и $1-2y < s < 1-y$.

12. Двумерная нормальная плотность с дисперсиями m, n и ковариацией $\sqrt{m/n}$. Условная плотность имеет математическое ожидание $\frac{m}{n}t$ и дисперсию $m \frac{n-m}{n}$.

13. Сумма $X_1^2 + \dots + X_n^2$ имеет гамма-плотность $f_{\frac{1}{2}, \frac{n}{2}}$ [см. гл. II, (2.2)].

Поэтому из (3.1)

$$u_t = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n-m}{2}\right)} \left(\frac{x}{t}\right)^{\frac{1}{2}m-1} \left(1-\frac{x}{t}\right)^{\frac{1}{2}(n-m)-1} \frac{1}{t}.$$

При $m=2, n=4$ мы получим пример 3, а).

15. а) $4xy$, когда $x+y < 1, x > 0, y > 0$

$4x(2-x-y)$, когда $x+y > 1, 0 < x, y < 1$

$4y(2-x-y)$, когда $y > 1, x+y < 2, x > 0$

$4y(2-x-y)$, когда $x > 1, x+y < 2, y > 0$.

б) $2(1-x-y)^2$ при $0 < x, y < 1, x+y < 1$

$2(1-x)^2$ при $x > 0, y < 0, x+y > 0$

$2(1+y)^2$ при $x > 0, y < 0, x+y < 0$.

При $x < 0$ по симметрии.

$$16. 2 \frac{r}{\pi^2} \left(\arccos \frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{r^2}{4}} \right).$$

17. $\int_0^{\infty} f(\rho) \rho d\rho \int_0^{2\pi} g(\sqrt{r^2 + \rho^2 - 2r\rho \cos \theta}) d\theta$.
20. а) $X_n = \left(U \cos \frac{1}{2} \pi n + V \sin \frac{1}{2} \pi n \right) \lambda$
 б) $\lambda U + \mu V (-1)^n$
 в) $\lambda \left(U \cos \frac{1}{2} \pi n + V \sin \frac{1}{2} \pi n \right) + \mu W$.
21. а) $\text{Var}(Y_{n+1}) - \text{Var}(Y_n) = \text{Var}(C_n) - 2 \text{Cov}(Y_n, C_n) + 1$, откуда
 б) $\alpha^2 - 2\alpha\sigma\rho + 1 = 0$
 в) $\sigma = \frac{1}{2} \left(\alpha + \frac{1}{\alpha} \right)$, $Y_n = \sum_{k=0}^{n-1} q^k X_{n-1-k} + q^n Y_0 + (bp-a)(1-q^n)/p$,
 где $q = 1 - p$.
22. $\sigma^2 \geq \frac{1}{4} \left(\alpha + \frac{1}{\alpha} \right)^2 + N$.

Глава VI

11. Не обязательно. Необходимое условие: $n[1 - F(en)] \rightarrow 0$.
12. При $x > 0$ плотности равны 1 и $\frac{1}{2}(1 - e^{-2x})$.
13. $qU(x) = 1 - qe^{-pct}$. Число моментов восстановления всегда распределено геометрически.
19. $Z = z + F \star Z$, где $z(t) = 1 - e^{-ct}$ при $t \leq \xi$, $z(t) = z(\xi)$ при $t \geq \xi$ и $F(t) = e^{-c\xi} - e^{-ct}$ при $t > \xi$.
20. $V = A + B \star V$, где $A\{dx\} = [1 - G(x)]F\{dx\}$ и $B\{dx\} = G(x)F\{dx\}$.
23. Плотность арксинуса $g(y) = \frac{1}{\pi} \frac{1}{\sqrt{y(1-y)}}$.

Глава VII

6. а) $\binom{n}{k} p^k (1-p)^{n-k}$. Распределение F сосредоточено в p .
 б) $\frac{1}{n+1}$, плотность $f(x) = 1$.
 в) $\frac{2(k+1)}{(n+1)(n+2)}$, плотность $2x$.

А. Общие курсы

- Krickeberg K. Probability theory. (Translated from the German 1963.) Addison — Wesley, Reading, Mass, 1965, 230 pp.
 Loève M. Probability theory. 3rd ed. Van Nostrand, Princeton, N. J., 1963, 685 pp.
 Имеется перевод: Лозв М. Теория вероятностей,— Перевод с англ.— М.: ИЛ, 1962.
 Neveu J. Mathematical Foundations of the Calculus of Probability. (Translated from the French 1964.) Holden Day, San Francisco, Calif, 233 pp.
 Имеется перевод: Неве Ж. Математические основы теории вероятностей,— Перев. с фр.— М.: Мир, 1969.

Б. Специальные темы

- Bochner S. Harmonic Analysis and the Theory of Probability. Univ. of California Press, 1955, 176 pp.
 Grenander U. Probabilities on Algebraic Structures, John Wiley, New York, 1963, 218 pp.
 Имеется перевод: Гренандер У. Вероятности на алгебраических структурах.— Перев. с англ.— М.: Мир, 1965, 275 с.
 Lukacs E. Characteristic Functions, Griffin, London, 1960, 216 pp.
 Имеется перевод: Лукач Е. Характеристические функции.— М.: Наука, 1964.
 Lukacs E., Laha R. G. Applications of Characteristic Functions. Griffin, London, 1964, 202 pp.

В. Случайные процессы (теория)

- Chung K. L. Markov Chains with Stationary Transition Probabilities, 2nd ed., Springer, Berlin, 1967, 301 pp.
 Имеется перевод первого издания: Чжун Кай-Лай. Однородные цепи Маркова.— Перев. с англ.— М.: Мир, 1964, 425 с.
 Дынкин Е. Б. Марковские процессы.— М.: Физматгиз, 1963, 859 с.
 Ito K., McKean H. P. Jr., Diffusion Processes and their Sample Paths. Springer, Berlin, 1965, 321 pp.
 Имеется перевод: Ито К., Маккин Х. Диффузионные процессы и их траектории.— Перев. с англ.— М.: Мир, 1968.
 Kemperman J. H. B. The Passage Problem for a Stationary Markov Chain. University of Chicago Press, 1961, 127 pp.
 Levy P. Processus Stochastiques et Mouvement Brownien. 2nd ed. Gauthier-Villars, Paris, 1965, 438 pp.
 Имеется перевод: Леви П. Стохастические процессы и броуновское движение.— Перев. с фр.— М.: Наука, 1972.
 Spitzer F. Principles of Random Walk. Van Nostrand, Princeton, 1964, 406 pp.
 Имеется перевод: Спитцер Ф. Принципы случайного блуждания.— М.: Мир, 1969.

- Скорород А. В. Случайные процессы с независимыми приращениями.— М.: Наука, 1964, 278 с.
- Яглом А. М. Введение в теорию стационарных случайных функций. Успехи математических наук, т. 7, № 5, 1952, 3—168.

Г. Случайные процессы (приложения и примеры)

- Barucha-Reid A. T. Elements of the Theory of Stochastic Processes and their Applications. McGraw-Hill, New York, 1960, 468 pp.
Имеется перевод: Баруча-Рид А. Т. Элементы теории марковских процессов и их приложения. Перев. с англ.— М.: Наука, 1969, 511 с.
- Beneš V. E. General Stochastic Processes in the theory of Queues. Addison-Wesley, Reading, Mass., 1963, 88 pp.
- Grenander U., Rosenblatt M. Statistical Analysis of Stationary Time Series. John Wiley, New York, 1957, 300 pp.
- Хинчин А. Я. Математические методы теории массового обслуживания.— М.: Физматгиз, 1963, 120 с.
- Prabhu N. U. Stochastic Processes. Macmillan, New York, 1965, 233 pp.
- Prabhu N. U. Queues and Inventories. John Wiley, New York, 1965, 275 pp.
Имеется перевод: Прабху Н. Методы теории массового обслуживания и управления запасами. Изучение основных случайных процессов. Перев. с англ.— М.: Машиностроение, 1969, 356 с.
- Riordan J. Stochastic Service Systems. John Wiley, New York, 1962, 139 pp.
Имеется перевод: Риордан Дж. Вероятностные системы обслуживания. Перев. с англ.— М.: Связь, 1966, 184 с.
- Wax N. (Editor). Selected Papers on Noise and Stochastic Processes. Dover, New York, 1954, 337 pp.

Д. Книги, имеющие исторический интерес

- Cramer H. Random Variables and Probability Distributions, 2nd ed. (The first appeared in 1937), Cambridge Tracts, 1962, 179 pp.
Имеется перевод с 1-го изд.: Крамер Г. Случайные величины и распределение вероятностей.— М.: ИЛ, 1947.
- Doob J. L. Stochastic Processes. John Wiley, New York, 1953, 654 pp.
Имеется перевод: Дуб Дж. Вероятностные процессы.— Перев. с англ.— М.: ИЛ, 1965, 605 с.
- Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин.— М.: Гостехиздат, 1949.
- Колмогоров А. Н. Основные понятия теории вероятностей, 2-е изд.— М.: Наука, 1974.
- Levy P. Calcul des Probabilités. Gauthier Villars, Paris, 1925, 350 pp.
- Levy P. Theorie de l'Addition des Variables aléatoires. Gauthier Villars, Paris, 1937 and 1954, 384 pp.

Е. Полугруппы и общий анализ

- Hille E., Phillips R. S. Functional Analysis and semigroups (Revised Edition). Amer. Math. Soc., 1957, 808 pp.
Имеется перевод: Хилле Э., Филлипс Р. Функциональный анализ и полугруппы.— Перев. с англ.— М.: ИЛ, 1962, 829 с.
- Karlin S. Studden W. Tchebycheff Systems: with Applications in Analysis and Statistics. Interscience, New York, 1966, 586 pp.
Имеется перевод: Карлин С., Стадден В. Чебышевские системы и их применение в анализе и статистике.— Перев. с англ.— М.: Наука, 1976, 586 с.
- Yosida K. Functional Analysis. Springer, Berlin, 1965, 458 pp.
Имеется перевод: Йосида К. Функциональный анализ.— Перев. с англ.— М.: Мир, 1967.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Абелевы теоремы 472, 501
Абеля интегральное уравнение 48
Абсолютная непрерывность 165, 167
Абсолютно беспристрастная последовательность 246
— монотонные функции 260, 495
Абсолютные вероятности 243—245
Аварии 73, 213
Адамара теорема факторизации 588
Аддитивные функции множеств 128, 129
Алгебра множеств 134, 135, 139, 142
— — порожденная 192
Ансамбли кругов и шаров 22
— случайные точек в пространстве 22
Арифметическое распределение 164, 561
— — конечное 681 *См. также* Решетчатое распределение
Арксинуса распределение 66
— — в случайном блуждании 471
— — и предельные теоремы 529
Арцела — Асколи теорема 310
Асимптотическая пренебрежимость 208, 653
Асимптотически несмещенная оценка 256
— плотное множество 174
Асимптотические оценки 426, 463
— разложения и центральная предельная теорема 595
— — для распределений 604
— —, различно распределенные слагаемые 612
— свойства правильно меняющихся функций 320
Астрономия, применения к 48, 202, 203—204, 242, 252
Атомические меры 163
Атомы (мер) 163
— при свертке 174

Банаха пространство 296, 397, 546
См. также Гильбертово пространство
Башелье процесс *См.* Броуновское движение
Безгранично делимые полугруппы 514
— — распределения 206, 334, 506
— — — в $\mathcal{R}^r$ 665
Безобидность абсолютная 246
Бернулли испытания 168
Берри — Эссена теорема 608
Бернштейна полиномы 258
— — в $\mathcal{R}^2$ 283
Бесконечная дифференцируемость 295, 336
Бесконечные свертки 305, 361, 664, 669
Бесселева плотность 564
Бесселевы функции 74, 75, 586
— —, безграничная делимость 207, 508
— — в стохастических процессах 75—77, 368
— —, *Лапласа* преобразование 493, 540
— —, распределения, связанные с ними 76, 77, 176, 196, 197
— —, характеристические функции 563
Бета-интеграл 62
— плотность 65
— распределение 529
Ближайшие соседи 23
Борелевская алгебра 135, 142,
— —, измеримые функции 139
Борелевское множество 135
— —, аппроксимация 137, 148
— —, соглашения об обозначениях 151
Бореля — Кантелли лемма 126, 362
Бохнера интеграл 511
Бросание монеты 246, 248, 471
— — и случайный выбор 49
Броуновское движение 120, 212, 378, 534
— — в $\mathcal{R}^r$ 205, 390
— —, два поглощающих экрана 388, 537
— —, один поглощающий экран 387, 536
— —, непрерывность траекторий 212, 348, 379

- —, первое прохождение 204, 387, 535
- —, подчиненность 394
- Бэровские функции 127, 131, 136, 137, 140, 151, 154, 348, 398
- Бюффона задача об игле 78
- Вальда* тождество 453, 481, 482, 673, 675
- Вейбула* распределение 68
- Вейерштрасса* теорема об аппроксимации 259
- Вероятностные меры и пространства 138, 155, 159
 - распределения 151, 155
- Ветвящиеся процессы 282, 497, 531
- Взаимные пары функций 564
- Видимость в x -направлении 23
- Винера — Башелье* процесс 120, 212
- Винера* процесс *См.* Броуновское движение
- Винера — Хинчина* критерий 716
- Винера — Хопфа* интегральное уравнение 455, 459, 479
 - метод 465
 - факторизация 440, 452, 482, 676, 691
- Вложенные процессы восстановления 224, 431
- Вогнутые функции 182
- Возвратные (устойчивые) процессы *См.* Невозвратный или обрывающийся процесс восстановления
- Возвращение в нуль 478
- Восстановления моменты 216, 417, 422
 - процессы 25, 216, 253
 - вложенные 224, 431
 - двустепенные 431
 - нелинейные 439
 - обрывающиеся 424
 - применения 427
 - суперпозиция 420
 - теорема 236, 408, 411, 432, 439
 - доказательство 412
 - уравнение 217, 407
- Вполне монотонные функции 260, 495, 507, 522
 - — — абстрактные 511
- Вращения 97, 104, 122
- Времена ожидания 31
- Время возвращения 216
 - максимальное 222, 437
- Выборка, медиана, экстремумы 32
 - среднее значение, дисперсия 106
- Выпуклые функции 181
 - и мартингалы 251
- Вырожденное распределение 102, 107
- Вырожденные процессы 110
- Гамеля* уравнение 341, 348
- Гамма-плотность 24, 62
 - процесс, направляемый пуассоновским процессом 396
 - —, направляющий пуассоновский процесс 395
 - распределение 24, 62, 63, 207
 - —, безграничная делимость 207, 363, 509, 635
 - как предел порядковых статистик 38
 - —, подчинение 396
 - —, приближение 256
 - —, рандомизированное 75
 - функция 62
- Гармонические функции 282
- Гейгера* счетчики 222, 223, 422, 438, 526
- Гельдера* неравенство 184
- Гильбертово пространство 312, 713, 719
- Гиперболические функции и плотности 563, 590, 635, 665, 708
- Гнеденко — Королюка* теорема 54, 59
- Гравитационные поля 203, 252, 253
- Граничные условия 383, 536
- Грина* функция 380, 535, 556
- Группировка данных 17
- Гьюгенса* принцип 67
- Двойные орбиты 48
- Двойственность 445, 682
- Двустороннее преобразование *Лапласа* 489
- Двусторонняя показательная плотность 65, 563
 - — —, факторизация 681
 - — —, характеристическая функция 563
- Дисперсия 18, 60, 162
 - инфинитезимальная 381
 - условная 91
- Диффузионные процессы в генетике 382
 - — — многомерном случае 390
 - — и процессы размножения и гибели 557
 - — с упругой силой 381
- Длина случайной цепи 242
- Длительность диффузии 388—389
 - периода занятости 533, 542
 - — нечувствительности 223
- процесса восстановления 219, 253, 425
 - — —, оценки 426
 - — размножения 551
- Дробная доля 175, 308
- Дробовой эффект 209, 667

Естественный масштаб в диффузии 379

Жордана разложения 164, 169

Задача о разорении 213, 233, 371, 527

— — — в обобщенном пуассоновском процессе 213, 231, 527

— — —, оценки 427, 464

Задержки в движении 222, 429, 438, 533, 556

Закон больших чисел 255, 271, 328, 401, 575

— — — для мартингалов 280

— — — для одинаково распределенных случайных чисел 271

— — — для серий 361, 668

— — — для стационарных последовательностей 284

— — —, обратная теорема 279

— — — слабый 273

— — — усиленный 276, 279

— — — Хинчина 272

— нуля или единицы 148, 149

Заражение 74

Значащие цифры 80

Иенсена неравенство 181, 251

Измеримое пространство 135

Измеримость 136

Изометрия 721

Индикатор множества 125

Индуктированное разбиение 36

Интеграл Дирихле 572

Интегрирование по частям 178

Интервал непрерывности 286

Инфинитезимальная скорость и дисперсия 381

Инфинитезимальные операторы 513

Исчезающие на бесконечности функции 286

«Исчезновение», случайные блуждания 478

Канонические меры 627

Кантора диагональный метод 307

— тип распределения 50, 167, 665

— — — (свертки) 173

Квазустойчивое распределение 202

Коварияция 86

—, матрица 99

— процесса 108, 698, 724

Колмогорова дифференциальные уравнения 370, 543

— неравенство 280, 284

— — — для мартингалов 280

— обратное уравнение 372, 544

— прямое уравнение 369, 373, 544

— Смирнова теорема 54, 389

— теорема о трех рядах 362

— Чепмена уравнение 77, 119, 380, 384, 392—394, 399, 400, 635

— — —, дискретное время 119

— — — и полугруппы 399

— — —, непрерывное время 367, 546

Компактность в схеме серий 353

Конкордантные (согласованные) функции 247, 283

Координатные случайные величины 16, 86

Корреляция 86

Коши полугруппы 346

— распределение 66, 81, 203, 564

— — — в броуновском движении 205, 395

— — — в $\mathcal{R}^r$ 89, 92, 121, 587, 666

— — — двумерное 587

— — — симметричное в $\mathcal{R}^2$ 89, 666

— — — — — $\mathcal{R}^3$ 89

— — —, случайное блуждание 227, 692

— — —, устойчивость 203, 622

Крамера — Леви теорема 587

— оценка для вероятности разорения 214, 427, 456, 465

Критерий значимости 95

Кронекера лемма 277

— символ 543

— ядро 241, 543

Круг, равномерное распределение на нем 308, 314

Круговая симметрия 585—586

Кэмпбелла теорема 209, 329, 667

Лапласа второй закон распределения: 65

— преобразование 270, 484, 524

— — —, примеры 492

— — —, производные и моменты 490

— — —, случайные блуждания 687

— — —, элементарные свойства 489

— Стильбеса преобразование 487, 528, 556

Лебега интеграл 128

— мера 48, 150

Никодима теорема 167

— пополнение 150

— Стильбеса интеграл 128, 132, 156

— теорема о разложении 169

Леви каноническое представление 632

— Крамера теорема 587

— метрика 327

— Парето распределение 206

— примеры 252, 364, 636

Лежандра формула удвоения 81

- Лестничные величины 225, 421, 441, 467
 — — — возрастающие слабо 444
 — — — строго 443
 — — — в процессе ожидания 232
 — — убывающие 445
 — высоты 225, 421, 443, 450
 — индексы 465
 — моменты, распределения 443, 465, 466
 — точки 441, 443
Линдберга условие для диффузии 380
 — условия 301, 328, 580
 Линейные операции над стохастическими процессами 699
 — приращения в скачкообразных процессах 369
 Линейный функционал 143
 Логарифмы комплексных чисел 622
 Логистическое распределение 68
 Локально компактные пространства 144, 147, 286
Ляпунова условие 328
- Мажорированная сходимость 133
Максвелла распределение 46, 63, 97
 Максимальная оценка 463
 — характеристическая функция 686, 691
 — частная сумма 226, 233, 365, 455, 461, 473
 Максимальное время возвращения 222, 437
 Максимальный член 195, 202, 318, 329, 448, 455
 — — влияние на сумму 522 *см. также*
 Порядковые статистики. Рекордные значения
 Маргинальные (частные) распределения 85, 159, 185
 — — заданные 195
 — — нормальные 121
 Марковские процессы с дискретным временем 115, 121, 240, 254
 — — — — и мартингалы 282
 — — — —, спектральная функция 711, 721
 — — — —, эргодические теоремы 311
 — — — — непрерывным временем 118, 366, 697
 — — — — в процессах восстановления 419
 — — — — счетных пространствах 543
 — — — — и полугруппы 396, 511
- — — — — эргодические теоремы 429, 553
 Марковское свойство 21 *См. также*
 Строго марковское свойство
 Мартингалы 245, 280
 —, неравенства 280, 284
 Масштабные параметры 60, 163
 Математические ожидания 17, 140, 156
 — — условные 191
 Медиана 32, 163
 Медленно меняющиеся функции *см.*
 Правильно меняющиеся функции
 Мера скорости (speed measure) 380
 Метрики 327 *См. также* Банахово пространство, Гильбертово пространство,
 Минимальное решение, *Винера — Хопфа* уравнение 454
 — —, дифференциальное уравнение *Колмогорова* 545
 — —, диффузия 386
 — —, полумарковские процессы 557
 — —, скачкообразные процессы 375
Миттаг-Леффлера функции 510
 Моменты 18, 162, 178, 639
 — восстановления 216, 417
 — в процессах восстановления 425
 — и производные 490
 —, неравенство 183
 —, проблема однозначности 264, 271, 576
 — — — в $\mathcal{R}^r$ 592
 —, производящие функции 489
 —, сходимость 290, 309
 —, *Хаусдорфа* проблема 260, 284
 Монотонной сходимости принцип 132
 — — свойство 398
 Монотонные последовательности 126
 — функции 316—318
 — — изменяющиеся доминируемо 317, 330
- Направляющий процесс 394
 Невозвратное распределение 688
 Невозвратные случайные блуждания 236
 — —, критерий 239, 448, 468, 687, 692
 — —, уравнение восстановления и теорема для них 235, 432, 433, 482
 Невозвратный или обрывающийся процесс восстановления 219, 424
 Невозможность систем игры 250, 251
 Независимость случайных величин 18, 87, 89, 145, 159
 — — —, комплексные величины 559

- — —, критерий для нее 161
- Независимые приращения 116, 210, 335, 347, 362, 723
- —, подчиненность 394
- —, разрывы 334, 362
- Неймана тождество 77
- Нелинейное восстановление 439
- Непосредственная интегрируемость по Риману 410
- Непрерывность полугрупп 399 *См. также* Фиксированные разрывы
- Несмещенная оценка 256
- Несобственные вероятностные меры 138, 155
- (дефектные) распределения 153, 155, 241
- — — в процессах восстановления 218
- Неудачи, устойчивость 29—31
- Номер первого максимума 471
- Норма 296, 397, 711, 719
- , топология, порождаемая 327
- Нормальное распределение 61, 82, 106, 203, 563, 634
- — в $\mathcal{R}$ 102, 585
- — вырожденное 107
- — двумерное 88, 91, 121
- — маргинальное 103, 105, 121, 195
- — марковское 115
- —, область притяжения 358, 643
- —, характеристические свойства 82, 96, 103, 587
- Нормальные полугруппы 342, 351, 363
- Нормальный стохастический процесс 107, 718
- Носитель меры 60
- Нулевая схема серий 208, 654
- Нулевое множество 149, 166, 711, 719
- Область притяжения 201
- — и устойчивые распределения 643
- —, критерий для нее 356, 365, 643
- — нормальная 651
- — частичная 365, 366
- Обобщенный пуассоновский процесс 211, 348, 368, 395
- — — и полугруппы 340, 342
- — —, разорение в нем 213, 233, 527
- Образование колонн 55, 556
- Обратное уравнение 380, 382, 390, 535
- — для полугрупп 400, 403
- — — полумарковских процессов 537
- — Колмогорова 372
- —, минимальное решение 375, 545
- —, скачкообразный процесс 372, 544
- Обрывающийся или невозвратный процесс восстановления 219, 424
- Обслуживание больных 214
- Обслуживающие устройства *См. Очереди*
- Одновершинные (унимодальные) распределения 186, 591
- —, свертки 197
- Ожидания процесс 229, 244
- Окружности и плотности на них 708
- — равномерное распределение на них 308, 314
- — распределение вероятностей на них 314
- — распределения на них 43, 78, 702
- , теорема о покрытии 43
- Оператор перехода 397
- сдвига 269, 295, 721
- Операторы, связанные с распределениями 295
- Операционное время 212, 392
- Орнштейна — Уленбека процесс 120, 381—382
- Остаточные события 148
- Остающееся время ожидания 220, 419
- —, предельная теорема 419, 437, 530
- Осциллирующие случайные блуждания 240, 447
- Отношение правдоподобия 247
- Отражающий экран 386, 390, 522
- Оценка 57
- Очереди 70, 71, 82, 231
- и парадоксы 25
- — предельные теоремы 431
- параллельные 31, 32, 56 *См. также* Периоды занятости
- , совместное распределение для прошедшего и остающегося времени ожидания 437
- Ошибки округления 36, 79
- Парадокс контроля 219, 421
- Парадоксы 25, 37, 219, 421
- времени ожидания 24, 421
- Парето распределение 66, 206
- Парсеваля равенство 521, 568, 688, 693, 694, 714, 716, 718
- — как критерий Хинчина 710
- Первое возвращение 444, 479, 555
- Первые прохождения, диффузия 204, 386, 535

- —, марковские цепи 553
- —, процесс размножения и гибели 77, 540, 554
- Периодограмма 95
- Периоды занятости 222, 234, 533, 542
- Петербургская игра 274
- Пирсона система распределений 63, 66
- Планиреля преобразование 713
- тождество 571
- Плотность 15, 65, 84, 164
- , обозначения и определения 60
- Поглощающие экраны 386, 522, 536, 537
- Поглощение (физическое) 39, 45
- Подчиненные процессы 392, 402, 508, 515, 642, 668
- Пойа критерий 566, 570
- распределение 74
- урновая схема 74, 247, 267
- Показательная формула для полугрупп 268, 400
- Показательное распределение 13, 21, 54—59, 92
- — двумерное 121
- — двустороннее 65, 176
- — и равномерное распределение 59
- — как предел 38, 59, 420 *См. также* Гамма-распределение
- Положительно определенные матрицы 100
- — последовательности 708
- — функции 695
- Положительные случайные величины 640
- Полугруппы 269
- и безгранично делимые процессы 332
- — марковские процессы 396
- *Коши* 346
- , *Лапласа* преобразование 511
- нормальные 342
- обобщенные пуассоновские 342
- , показательная формула 400
- , производящие операторы 403
- разрывные 348
- с конечными дисперсиями 341
- со сверткой 335
- устойчивые 349
- Полумарковский процесс 543, 557
- Полярные координаты 86
- Попадания в случайных процессах 224, 238
- Пополнение множеств 150
- Популяция, рост 383, 385
- , случайное рассеяние 301
- Порядковые статистики 32, 34, 37, 121
- —, предельные теоремы 38, 58
- —, применение к статистическому оцениванию 57
- Потенциал 547
- Потеря энергии 39, 368 *См. также* Столкновение частиц
- Потерянные вызовы 223, 556
- Правильно меняющиеся функции 315, 320
- Предельные теоремы 38, 285, 389
- — и очереди 431
- — — распределение арксинуса 529
- Принцип инвариантности 389
- отражения 205, 387, 536, 537
- Притяжение *См.* Область притяжения
- Проекция случайных векторов 44, 46, 58
- Произведение мер и пространств 145
- Производные и преобразование *Лапласа* 490
- Производящие операторы 336, 403, 512, 534
- функции, интерпретация с помощью «исчезновения» 478
- —, характеристика 259
- Произвольная остановка 250
- Пропуски (пробелы) большие 221, 428, 526
- малые 254
- Прохождение света через сферу 45
- Процесс авторегрессии 109, 117
- восстановления 25
- — с запаздыванием 219, 417
- Прошедшее время ожидания 220, 437
- —, предельные теоремы 421, 530
- Прямое уравнение 383, 544
- — и диффузия 383, 535
- — — полугруппы 400
- — — процесс скачков 370, 373
- — *Колмогорова* 369, 373
- —, минимальное решение 375, 377, 545
- Прямоугольная плотность 35, 65
- Псевдопуассоновский процесс 367
- —, показательная формула 400
- —, полугруппа 401, 517
- — с линейными приращениями 369
- Пуанкаре* задача о рулетке 79
- Пуассона* распределение 25, 73, 634
- — обобщенное 622, 624, 625
- — симметризованное 176, 196, 635
- формула суммирования 80, 389, 704, 726
- ядро 702, 726
- Пуассоновские ансамбли 22, 23, 28
- См. также* Гравитационные поля

- распределения, аппроксимация с их помощью 328
- Пуассоновский процесс 24, 28
- — как предел в процессах восстановления 420
- —, направляемый гамма-процессом 395
- —, направляющий гамма-процесс 396
- —, пробелы 428
- —, *supremum* 214

Равенство Парсеваля 521

- Равномерное распределение 35, 65, 562, 563
- — на сфере 87
- Равностепенная непрерывность 291, 310
- Радо-Никодима* производная 165
- — теорема 167
- Разбиения шахматные 158
- Различные распределения 360
- Разложения 639
- Размножения и гибели процессы 538, 557
- — — — и диффузия 557
- — — —, периоды занятости в них 542 *См. также* Рандомизированные случайные блуждания
- процесс 56, 306, 548
- — двусторонний 551
- Разности (обозначения) 257
- Разрывы 363
- фиксированные (скачки) 370, 374
- Райкова* теорема 639
- Рандомизация 69
- и подчиненные процессы 392
- , полугруппы 269, 401
- , симметрично зависимые случайные величины 265
- Рандомизированные случайные блуждания 75, 538, 634
- Расслоение 73
- Расстояние между распределениями 327
- Расхождение *См.* Эмпирические распределения
- Регрессия 90, 109
- Регулярное стохастическое ядро 312
- Резольвентное уравнение 509, 519, 546
- Резольвенты 484, 509, 512, 546
- и полная монотонность 519
- Рекордные значения 29, 55 *См. также*
- Порядковые статистики
- Решающая функция 250
- Решетчатые распределения 164, 174

- —, характеристические функции 561, 572
- —, центральная предельная теорема 579, 605
- Римана—Лебега* теорема 575, 603, 704
- Рисса* теорема о представлении 144, 160, 289
- Рулетка 36, 79
- Руше* теорема 460

Самовосстанавливающиеся совокупности 219

- Свертка ядер 241
- Свертки 169, 312, 319
- бесконечные 305, 361, 664
- и преобразование Лапласа 489
- — теоремы о покрытии 41
- на окружности 81, 170, 314
- определение 19
- плотностей 19, 61, 89
- сингулярных распределений 173
- формула 15, 16
- Свободный пробег 23
- Сглаживание 108, 170, 601
- Сдвиг, оператор 269, 295, 721
- , полугруппы 337, 404
- принцип 488
- Сдвиги случайные 47
- Сепарабельность 309
- Сервостохастический процесс (механизм) 123
- Серия 5 *См. также* Рекордные значения
- Сжатие 397, 516
- Сильная непрерывность в нуле 399
- сходимость 296, 399
- Симметризация 175
- , неравенства 176
- Симметрично зависимые величины 265, 477
- — —, центральная предельная теорема 328
- Симметричные распределения 202, 589
- события 149
- Сингулярные (вырожденные) распределения 51, 129, 167
- — —, свертки 173, 665
- Скалярное произведение 712
- Скачкообразный процесс 371
- — с бесконечным числом скачков 376, 544
- Скользящее среднее 108, 722
- Скрытая периодичность 95
- Слабая сходимость 288
- Случайная величина 16, 86, 139, 155, 558

- —, независимость, последовательности 147—149
- Случайные блуждания биномиальные 680
- — в $\mathcal{R}^1$ 224, 235, 440, 670
- — — и эмпирическое распределение 53
- — — — простые (*Бернулли*) 250, 363, 445, 447, 479, 493
- — — — сопряженные 458
- — — $\mathcal{R}^r$ 47
- — — —, центральная предельная теорема 300
- —, порождаемые распределением 441
- направления 44, 58, 169, 173, 243, 585
- последовательности 147
- разбиения 36, 39, 57, 93, 122 *См. также* Теоремы о покрытии
- суммы 70, 188, 197, 564
- —, характеристическая функция 564
- —, центральная предельная теорема 304, 593
- цепи 242
- —, длина 22
- Случайный вектор 46, 47, 129 *См. также* Случайные направления
- выбор 14, 35, 87
- — и бросание монеты 49 *См. также* Теоремы о покрытии, Случайные разбиения
- Смеси 69, 92, 187, 194
- , преобразование *Лапласа* 493, 564
- Снедекора* плотность 64
- Снос в диффузии 381
- — случайных блужданиях 448, 683
- Собственное распределение 155
- —, сходимость 286, 326
- Совместное распределение 90, 477
- Совпадения 254
- Сопряженные случайные блуждания 458
- Спарре-Андерсена* преобразование 475
- Спектр мощностей 698
- Спектральная мера 698, 700, 725
- —, представление для унитарных операторов 721
- Статистика 32
- Стационарная марковская последовательность 118
- мера и вероятность 244, 313
- Стационарное приращение 211, 335
- Стационарные процессы 107, 122, 697
- — закон больших чисел 284 *См. также* Стационарный режим, Эргодическая теорема
- Стационарный режим (steady state) 27, 243, 371
- — в процессах восстановления 420, 437
- Стирлинга* формула 63
- Столкновение частиц 242, 368, 370
- Стохастическая ограниченность 293, 353
- Стохастические интегралы 724
- и случайные процессы 718
- Стохастическое разностное уравнение 110—111
- ядро 188, 240, 311, 396, 398
- Строгая непрерывность полугрупп 511
- Строго марковское свойство 34
- устойчивые распределения 199
- Стьюдента* плотность 64
- Субмартингал 251, 280
- Субстохастическое ядро 241
- Суммируемость по *Абелю* 701, 703, 726
- — *Чезаро* 703, 726
- Суперпозиция процессов восстановления 420
- Схемы серий 208, 253, 352, 653, 668
- Сходимость в среднем квадратичном 712
- мажорированная 133
- мер 285, 307—310, 326—327
- моментов 290, 309
- операторов 296, 327, 399
- , основные понятия и обозначения 286
- плотностей 290
- по вероятности 292
- сильная 296, 399
- слабая 288
- Счетно-аддитивные функции (σ -аддитивные) 130, 142, 154
- Счетные пространства состояний 377
- Счетчики 422 *См. также* Гейгера счетчики, Очереди
- Такача* интегральное уравнение 532
- Тауберовы теоремы 411, 498
- —, применение 472, 504, 526, 529, 685
- Тейлора* ряд (разложение) 573, 576, 577, 598, 600
- формула 356
- — обобщенная 267
- Телефонные вызовы 73, 210
- — потерянные 223, 556 *См. также* Периоды занятости
- Теорема непрерывности 484, 569
- — для плотностей 571

- —, *Лапласа* преобразования 486
- —, характеристические функции 570, 571
- об аппроксимации в среднем 133
- о продолжении 142, 147
- — равномерном распределении дробных долей 308
- — среднем значении 131
- — трех рядах 362
- Теоремы о выборе 307
- — покрытия 96, 253, 527
- — — и свертки 41
- Теория надежности 68
- страхования 214 *См. также* Задача о разорении
- Тип распределения 60, 162
- —, сходимость 291
- Точка роста 163, 174
- Точки первого вхождения (достижения) в процессах восстановления 220, 421, 480 *См. также* Лестничные величины
- — — — случайных блужданиях 480, 670
- Точные выражения 227
- Транспортные задачи 55, 222, 429, 438, 533
- Треугольная плотность 41, 65, 562, 563
- Треугольника неравенство 295
- Тэта-функции 389, 708

- Универсальные законы *Дёблина* 663
- Унимодальность по *Хинчину* 187, 591
- Унитарные операторы в гильбертовом пространстве 721
- Упругая сила 381
- Уравнение восстановления 217, 407, — — для всей прямой 236, 482
- —, процесс 25, 216, 417
- Уравнения восстановления, теоремы 406, 411, 432, 439
- Условная вероятность 186
- Условные распределения и математические ожидания 90, 185, 189
- Устойчивое распределение 198, 504, 643, 645
- — в $\mathcal{R}^2$ 666
- — — узком смысле (строго устойчивое) 199
- — — ширококом смысле 199
- — — экономике 205
- —, область притяжения 201, 643
- —, подчинение 395, 508
- —, положительное 504
- —, порядка $1/2$ 68, 82, 203, 363, 492
- —, произведение 206
- —, симметричное 202, 669

- —, характеристический показатель 199
- Устойчивые плотности 651

- Фату* лемма 132, 712
- — для граничных значений 283
- Фиксированные разрывы (скачки) 370, 374
- Фильтр 108, 699
- Фишера* Z-статистика 64
- Фоккера — Планка* уравнение 338, 368, 369, 370, 373
- Формула обращения 167, 257, 568, 714
- — и проблема моментов 264
- —, характеристические функции 570, 572, 587
- удвоения 81
- Фреге* максимальное распределение 195
- Фубини* теорема 133, 171
- Функции интервалов 128, 153
- простые 156
- Функционал линейный 143
- Фурье* интегралы, обращение 570, 572, 713
- коэффициенты 702, 708, 717, 721
- преобразование 559, 597 *См. также* *Планишерея* преобразование
- ряд 701, 717
- *Стилтьеса* преобразование 559, 679

- Характеристические функции 558
- — безгранично делимые 621, 624, 628, 631
- — и их логарифмы и корни 621
- — многомерных распределений 584
- — нормальных распределений 585
- — периодические 566, 572, 701, 703, 704
- —, производные 633
- —, разложение *Тейлора* 576 *См. также* *Пуассона* формула суммирования
- Характеристический показатель распределения R 199
- Хи-квадрат, плотность 63
- Хилле — Йосида* теорема 516, 534
- Хинчина* закон больших чисел 272, 491—492
- каноническое представление 633
- критерий 716
- — унимодальности 187, 591
- *Поллачека* формула 463, 528, 691
- Хольцмарка* распределение 202, 203, 252, 253

- Центральная предельная теорема 297—305, 328, 333, 593, 595
 — — —, асимптотические разложения 595 и т. д.
 — — —; бесконечная дисперсия 299
 — — —, большие отклонения 615
 — — — для одинаково распределённых слагаемых 576
 — — — — плотностей 597
 — — —, приложения 245, 592—594
 — — —, процессы восстановления 422
 Центрирование 60, 163, 655
 — безгранично делимых распределений 625
 Цепи, прочность 22
 — случайные 242
 Цифры, распределение 49, 80
- Частичное притяжение 365, 636, 661
 Частицы, изменения направления движения 368
 —, расщепление 39, 57, 122 *Ср.* Столкновение частиц
 Частные (маргинальные) распределения 85, 159, 185
 — — —, заданные 195
 — — — нормальные 121
 Чебышева неравенство 179, 354, 402
 — —, обобщение 272
 — — — на мартингалы 284
 — Эрмита полином 609
 Чепмена — Колмогорова уравнение, дискретное время 119
- — —, непрерывное время 367, 546
 — — —, полугруппы 399
 Чистого размножения процессы (birth processes) 56, 306, 548
 — — —, двусторонние 551
- Шаг распределения 164
 Шварца неравенство 180, 559, 590, 720
 Шлёмилха формула 77, 635
 Шум 328 *См. также* Дробовой эффект
- Эджворта разложение 599, 607
 Эквивалентные функции 149, 712, 719
 Экраны *см.* Поглощающие экраны, Граничные условия
 Эмпирические распределения 51
 Эндоморфизм 397
 Эргодические стохастические ядра 311
 — пределы 551
 — теоремы 311, 551
 — *См. также* Стационарный режим
 Эрланга распределение 63
 Эрмита полиномы 597, 607
- Ядро стохастическое 188, 240, 311, 396, 398
 Якоби тэта-функция 389, 708
 — σ -аддитивность 130
 — алгебра 134
 — конечная мера 138

ИМЕННОЙ УКАЗАТЕЛЬ

Амбарцумян В. А. 242, 370, 371
Анникян (Неппекин Р. L.) 124

Байес (Bayes T.) 72
Бакстер (Baxter G.) 456, 478, 640, 677
Баруча-Рид (Barucha-Reid A. T.) 733
Бенеш (Benes V. E.) 429, 733
Бенфорд (Benford F.) 80
Бергстрём (Bergström H.) 651
Бернштейн С. Н. 98, 495
Берри (Berry A. C.) 595, 601, 608
Бизли (Bizley M. T. L.) 474
Бикел (Bickel P. J.) 439
Биллингсли (Billingsley P.) 54, 304, 390
Блекуэлл (Blackwell D.) 408, 432
Блам (Blum J. R.) 328
Боль (Bohl) 308
Боровков А. А. 228
Боттс (Botts T. A.) 124
Бохнер (Bochner S.) 366, 394, 511, 695, 697, 709, 732
Брело (Brelot M.) 283
Будро (Boudreau P. E.) 231
Бурбаки (Bourbaki N.) 124
Бюльман (Bühlmann H.) 266

Вальд (Wald A.) 265, 449, 459, 675
Вебер (Weber H.) 493
Вейль (Weyl H.) 308
Виленкин Н. Я. 697
Винер (Wiener N.) 212, 411, 455, 716
Винтнер (Wintner A.) 197
Вольд (Vold H.) 585
Вольфовиц (Wolfowitz J.) 235, 314, 432

Гальтон (Galton F.) 91
Гейтлер (Heitler W.) 368
Гельфанд И. М. 697
Герглотц (Herglotz G.) 709
Гилберт (Gilbert E.) 254
Гири (Geary R. C.) 106
Гнеденко Б. В. 53, 54, 318, 595, 605, 651, 663, 733
Гренандер (Grenander U.) 96, 732, 733

Гринвуд (Greenwood M.) 74
Гриффиш (Griffin J. S.) 231
Гуд (Good D. I.) 532
Гумбел (Gumbel E. J.) 195

Дарлинг (Darling D. A.) 523
Дворецкий (Dvoretzky A.) 314
Дени (Deny C. R.) 433
Дерман (Derman C.) 554
Дёблин (Doblin W.) 202, 663
Джозеф (Joseph A. W.) 474
Донскер (Donsker M.) 54, 390
Дуб (Doob J. L.) 124, 193, 246, 283, 733
Дынкин Е. Б. 366, 380, 530, 732

Зигмунд (Zygmund A.) 633
Золотарев В. М. 608

Иахав (Yahav J. A.) 439
Ибрагимов И. А. 197
Иосида (Yosida K.) 366, 733
Ито (Ito K.) 380, 732

Кантор (Cantor G.) 307
Карамата (Karamata J.) 202, 285, 315, 320, 502
Карлеман (Carleman T.) 264, 576
Карлин (Karlin S.) 265, 432, 733
Кац (Kac M.) 98, 231, 389
Кац (Katz M. L.) 611
Кельвин, лорд (Kelvin) 387
Кемперман (Kemperman J. H. B.) 732
Кендалл (Kendall D. G.) 228, 269, 532
Кингман (Kingman J. F. C.) 217
Кифер (Kiefer J.) 235
Колмогоров А. Н. 54, 147, 148, 210, 370, 379, 595, 605, 733
Королюк В. С. 53, 54, 59
Котельников В. А. 706
Коши (Cauchy A.) 202, 570
Крамер (Cramer H.) 214, 428, 456, 585,

587, 595, 607, 613, 614, 619, 724, 733
 Крикеберг (Krickeberg K.) 124, 732
 Ламперти (Lamperti A.) 222
 Ландау Л. Д. 368
 Ландау (Landau E.) 389, 502
 Лаха (Laha R. G.) 732
 Леви (Levy P.) 202, 204, 210, 212, 246,
 252, 301, 314, 348, 358, 363, 557,
 576, 587, 632, 636, 640, 644, 659,
 663, 732, 733
 Ле Кам (Le Cam L.) 328
 Линдберг (Lindeberg J. W.) 576
 Линдли (Lindley D. V.) 228, 440
 Литл (Little J. D. C.) 533
 Литлвуд (Littlewood J. E.) 183, 502
 Лозв (Loève M.) 124, 125, 266, 304, 366,
 732
 Лукач (Lukacs E.) 106, 732
 Лундберг (Lundberg F.) 214

Маккин (McKean H. P. Jr.) 380, 732
 Макшейн (McShane E. J.) 124
 Мандельброт (Mandelbrot B.) 206, 329
 Марков А. А. 265
 Маршалл (Marshall A. W.) 284
 Мидлтон (Middleton D.) 707
 Миллер (Miller H. D.) 675
 Миллс (Mills H. D.) 123
 Мюнц (Müntz H. Ch.) 284

Найквист (Nyquist N.) 706
 Невё (Neveu J.) 124, 732
 фон Нейман (Neumann J.) 59
 Нейман (Neuman J.) 213
 Нелсон (Nelson E.) 121, 394
 Ньюэлл (Newell G. F.) 55

Орей (Orey S.) 432
 Осипов Л. В. 611

Пайк (Pyke R.) 214, 441, 529, 557
 Партасарати (Parthasarathy K. R.)
 124
 Петерсен (Petersen D. P.) 707
 Петров В. В. 611, 619
 Пинкхэм (Pinkham R. S.) 81
 Пирсон (Pearson K.) 63
 Питмэн (Pittman E. J.) 633
 Пойа (Polya G.) 183, 202, 213
 Полард (Pollard H.) 408, 432
 Полак (Pollak H. O.) 254
 Полачек (Pollaczek F.) 233
 Порт (Port S. C.) 320
 Прабху (Prabhu N. U.) 733
 Прохоров Ю. В. 54, 390
 Пуанкаре (Poincaré H.) 79

Райт (Wright S.) 383
 Райт (Wright E. M.) 308
 Рвачева Е. Л. 54
 Реньи (Renyi A.) 390
 Риордан (Riordan J.) 733
 Рисс (Riesz M.) 269
 Рихтер (Richter W.) 619
 Роббинс (Robbins H. E.) 121, 304, 408
 Розенблат (Rosenblatt M.) 96, 328,
 733
 Ройден (Royden H. L.) 265
 Рэй (Ray D.) 379
 Рэлей, лорд (Rayleigh; Стрэтт Дж. У.)
 47, 586

Серпинский (Sierpinski W.) 308
 Скелам (Skellam J. G.) 301
 Скитович В. П. 98
 Скороход А. В. 54, 733
 Смирнов Н. В. 54, 58
 Смит (Smith W. L.) 411, 432, 557
 Спарре-Андерсен (Sparre-Andersen E.)
 440, 466, 472—475
 Спитцер (Spitzer F.) 233, 441, 456, 472,
 677, 685, 691, 732
 Стадден (Studden W.) 265, 733
 Стейн (Stein C.) 673
 Стейтель (Steutel F. W.) 509
 Стивенс (Stevens W. L.) 96
 Стильтъес (Stieltjes T. J.) 265
 Сэвидж (Savage L. J.) 148, 266

Тамаркин Я. Д. 265, 284
 Таннер (Tanner J. C.) 222
 Тейчер (Teicher H.) 328
 Титчмарш (Titchmarsh E. C.) 628
 Тортра (Tortrat A.) 124, 665
 Троттер (Trotter H. F.) 301

Угахери (Ugaheri T.) 254
 Уилкс (Wilks S. S.) 55
 Уиттекер (Whittaker J. M.) 707
 Уоллес (Wallace D. L.) 608
 Уолш (Walsh J. W.) 67
 Уэкс (Wax N.) 47, 733

Фейер (Fejer L.) 703
 Феллер (Feller W.) 69, 78, 210, 228,
 301, 320, 330, 370, 377, 379, 383,
 408, 432, 486, 543, 557, 612, 619,
 651
 Филлипс (Phillips R. S.) 269, 336, 511,
 733
 де Финетти (Finetti B.) 58, 210, 267
 Фишер (Fisher R. A.) 96, 318, 383
 Фукс (Fuchs W. H. J.) 687, 689

- Халмош (Halmos P. R.) 124, 250
Хант (Hunt G.) 247
Харди (Hardy G. H.) 183, 308, 502
Харрис (Harris T. E.) 282
Хаусдорф (Hausdorff F.) 262
Хейди (Heyde C. C.) 264
Хелли (Helly E.) 307
Хилле (Hille E.) 268, 269, 336, 460, 511, 588, 733
Хинчин А. Я. 162, 202, 210, 459, 591, 632, 659, 663, 716, 733
Хоби (Hobby C.) 441
Хопф (Hopf E.) 455
Хьюитт (Hewitt E.) 148, 266
- Ч**
Чандрасекар (Chandrasekhar S.) 47
Чебышев П. Л. 179
Чернов (Chernoff H.) 328
Чжоу (Chow Y. S.) 408
Чжун (Chung K. L.) 126, 197, 269, 402, 432, 543, 687, 689, 732
- Ш**
Шапиро (Schapiro J. M.) 640
фон Шеллинг (Schelling H.) 244
Шеннон (Shannon C.) 706
Шепп (Shepp L.) 82, 187, 692, 726
Шмидт (Schmidt R.) 502
Шоке (Choquet G.) 433, 665
Шохат (Shohat J. A.) 265, 284
- Э**
Эйнштейн (Einstein A. Jr.) 213, 379
Эрдёш (Erdős P.) 389, 408
Эссеев (Esséen G.) 605, 608, 609, 611
Эссер (Esscher F.) 619
- Ю**
Юл (Yule G. U.) 74
- Я**
Яглом А. М. 733
Яноши (Janossy L.) 368

ОГЛАВЛЕНИЕ

Из предисловия к русскому изданию 1967 г.	5
От переводчика	6
Предисловие к первому изданию	7
Предисловие ко второму изданию	10
Обозначения	12
Глава I. Показательные и равномерные плотности	13
§ 1. Введение	13
§ 2. Плотности. Свертки	15
§ 3. Показательная плотность	21
§ 4. Парадоксы, связанные с временем ожидания. Пуассоновский процесс	24
§ 5. Устойчивость неудач	29
§ 6. Времена ожидания и порядковые статистики	31
§ 7. Равномерное распределение	35
§ 8. Случайные разбиения	39
§ 9. Свертки и теоремы о покрытии	41
§ 10. Случайные направления	44
§ 11. Использование меры Лебега	48
§ 12. Эмпирические распределения	51
§ 13. Задачи	54
Глава II. Специальные плотности. Рандомизация	60
§ 1. Обозначения и определения	60
§ 2. Гамма-распределения	62
§ 3. Распределения математической статистики, связанные с гамма-распределением	63
§ 4. Некоторые распространенные плотности	65
§ 5. Рандомизация и смеси	69
§ 6. Дискретные распределения	72
§ 7. Бесселевы функции и случайные блуждания	74
§ 8. Распределения на окружности	78
§ 9. Задачи	81
Глава III. Многомерные плотности. Нормальные плотности и процессы	84
§ 1. Плотности	84
§ 2. Условные распределения	90
§ 3. Возвращение к показательному и равномерному распределениям	92
§ 4. Характеризация нормального распределения	96
§ 5. Матричные обозначения. Ковариационная матрица	99
§ 6. Нормальные плотности и распределения	102
§ 7. Стационарные нормальные процессы	107
§ 8. Марковские нормальные плотности	115
§ 9. Задачи	120

Глава IV. Вероятностные меры и пространства	124
§ 1. Бэровские функции	125
§ 2. Функции интервалов и интегралы в $\mathcal{R}^r$	128
§ 3. σ -алгебры. Измеримость	134
§ 4. Вероятностные пространства. Случайные величины	138
§ 5. Теорема о продолжении	142
§ 6. Произведения пространств. Последовательности независимых случайных величин	145
§ 7. Нулевые множества. Пополнение	149
Глава V. Вероятностные распределения в $\mathcal{R}^r$	151
§ 1. Распределения и математические ожидания	152
§ 2. Предварительные сведения	162
§ 3. Плотности	164
§ 4. Свертки	164
§ 5. Симметризация	175
§ 6. Интегрирование по частям. Существование моментов	178
§ 7. Неравенство Чебышева	179
§ 8. Дальнейшие неравенства. Выпуклые функции	180
§ 9. Простые условные распределения. Смеси	185
§ 10. Условные распределения	189
§ 11. Условные математические ожидания	191
§ 12. Задачи	195
Глава VI. Некоторые важные распределения и процессы	198
§ 1. Устойчивые распределения в $\mathcal{R}^1$	198
§ 2. Примеры	203
§ 3. Безгранично делимые распределения в $\mathcal{R}^1$	206
§ 4. Процессы с независимыми приращениями	210
§ 5. Обобщенные пуассоновские процессы и задачи о разорении	213
§ 6. Процессы восстановления	215
§ 7. Примеры и задачи	219
§ 8. Случайные блуждания	224
§ 9. Процессы массового обслуживания	228
§ 10. Возвратные и невозвратные случайные блуждания	235
§ 11. Общие марковские цепи	240
§ 12. Мартингалы	245
§ 13. Задачи	252
Глава VII. Законы больших чисел. Применения в анализе	255
§ 1. Основная лемма. Обозначения	255
§ 2. Полиномы Бернштейна. Абсолютно монотонные функции	258
§ 3. Проблема моментов	260
§ 4. Применение к симметрично зависимым случайным величинам	265
§ 5. Обобщенная формула Тейлора и полугруппы	267
§ 6. Формулы обращения для преобразования Лапласа	270
§ 7. Законы больших чисел для одинаково распределенных случайных величин	271
§ 8. Усиленный закон больших чисел	275
§ 9. Обобщение для мартингалов	280
§ 10. Задачи	283

Глава VIII. Основные предельные теоремы	285
§ 1. Сходимость мер	285
§ 2. Специальные свойства	291
§ 3. Распределения как операторы	293
§ 4. Центральная предельная теорема	297
§ 5. Бесконечные свертки	305
§ 6. Теоремы о выборе	307
§ 7. Эргодические теоремы для цепей Маркова	311
§ 8. Правильно меняющиеся функции	315
§ 9. Асимптотические свойства правильно меняющихся функций	320
§ 10. Задачи	326
Глава IX. Безгранично делимые распределения и полугруппы	332
§ 1. Общее знакомство с темой	332
§ 2. Полугруппы со сверткой	335
§ 3. Подготовительные леммы	339
§ 4. Случай конечных дисперсий	341
§ 5. Основные теоремы	343
§ 6. Пример: устойчивые полугруппы	349
§ 7. Схемы серий с одинаковыми распределениями	352
§ 8. Области притяжения	356
§ 9. Различные распределения. Теорема о трех рядах	360
§ 10. Задачи	363
Глава X. Марковские процессы и полугруппы	366
§ 1. Псевдопуассоновский тип	367
§ 2. Вариант: линейные приращения	369
§ 3. Скачкообразные процессы	371
§ 4. Диффузионные процессы в $\mathcal{R}^1$	378
§ 5. Прямое уравнение. Граничные условия	383
§ 6. Диффузия в многомерном случае	390
§ 7. Подчиненные процессы	392
§ 8. Марковские процессы и полугруппы	396
§ 9. «Показательная формула» в теории полугрупп	400
§ 10. Производящие операторы. Обратное уравнение	403
Глава XI. Теория восстановления	406
§ 1. Теорема восстановления	406
§ 2. Доказательство теоремы восстановления	412
§ 3. Уточнения	415
§ 4. Устойчивые (возвратные) процессы восстановления	417
§ 5. Число N_t моментов восстановления	422
§ 6. Обрывающиеся (невозвратные) процессы	424
§ 7. Различные применения	427
§ 8. Существование пределов в случайных процессах	429
§ 9. Теория восстановления на всей прямой	431
§ 10. Задачи	437
Глава XII. Случайные блуждания в $\mathcal{R}^1$	440
§ 1. Основные понятия и обозначения	441
§ 2. Двойственность. Типы случайных блужданий	445
§ 3. Распределение лестничных высот. Факторизация Винера — Хопфа	450
§ 4. Примеры	457
§ 5. Применения	461
§ 6. Одна комбинаторная лемма	465
§ 7. Распределение лестничных моментов	468

§ 8. Закон арксинуса	470
§ 9. Различные дополнения	477
§ 10. Задачи	479

Глава XIII. Преобразование Лапласа. Тауберовы теоремы. Резольвенты 484

§ 1. Определения. Теорема непрерывности	484
§ 2. Элементарные свойства	489
§ 3. Примеры	492
§ 4. Вполне монотонные функции. Формулы обращения	495
§ 5. Тауберовы теоремы	498
§ 6. Устойчивые распределения	504
§ 7. Безгранично делимые распределения	506
§ 8. Многомерный случай	509
§ 9. Преобразования Лапласа для полугрупп	511
§ 10. Теорема Хилле—Иосиды	516
§ 11. Задачи	521

Глава XIV. Применение преобразования Лапласа 524

§ 1. Уравнение восстановления: теория	524
§ 2. Уравнение типа уравнения восстановления: примеры	526
§ 3. Предельные теоремы, включающие распределения арксинуса	529
§ 4. Периоды занятости и соответствующие ветвящиеся процессы	531
§ 5. Диффузионные процессы	534
§ 6. Процессы размножения и гибели. Случайные блуждания	538
§ 7. Дифференциальные уравнения Колмогорова	543
§ 8. Пример: чистый процесс размножения	548
§ 9. Вычисление эргодических пределов и времен первого прохождения	551
§ 10. Задачи	555

Глава XV. Характеристические функции 558

§ 1. Определение. Основные свойства	558
§ 2. Специальные плотности. Смеси	562
§ 3. Единственность. Формулы обращения	568
§ 4. Свойства регулярности	573
§ 5. Центральная предельная теорема для одинаково распределенных слагаемых	576
§ 6. Условие Линдеберга	580
§ 7. Характеристические функции многомерных распределений	584
§ 8. Две характеристики нормального распределения	587
§ 9. Задачи	590

Глава XVI. Асимптотические разложения, связанные с центральной предельной теоремой 595

§ 1. Обозначения	596
§ 2. Асимптотические разложения для плотностей	597
§ 3. Сглаживание	601
§ 4. Асимптотические разложения для распределений	604
§ 5. Теорема Берри—Эссеена	608
§ 6. Асимптотические разложения в случае различно распределенных слагаемых	612
§ 7. Большие отклонения	615

Глава XVII. Безгранично делимые распределения	621
§ 1. Безгранично делимые распределения	621
§ 2. Канонические формы. Основная предельная теорема	625
§ 3. Примеры и специальные свойства	634
§ 4. Специальные свойства	638
§ 5. Устойчивые распределения и их области притяжения	643
§ 6. Устойчивые плотности	651
§ 7. Схема серий	653
§ 8. Класс L	659
§ 9. Частичное притяжение. «Универсальные законы»	661
§ 10. Бесконечные свертки	664
§ 11. Многомерный случай	665
§ 12. Задачи	666
Глава XVIII. Применение методов Фурье к случайным блужданиям	670
§ 1. Основное тождество	670
§ 2. Конечные интервалы. Вальдовская аппроксимация	673
§ 3. Факторизация Винера—Хопфа	676
§ 4. Выводы и применения	682
§ 5. Две более основательные теоремы	685
§ 6. Критерии возвратности	687
§ 7. Задачи	690
Глава XIX. Гармонический анализ	693
§ 1. Равенство Парсеваля	693
§ 2. Положительно определенные функции	695
§ 3. Стационарные процессы	697
§ 4. Ряды Фурье	701
§ 5. Формула суммирования Пуассона	704
§ 6. Положительно определенные последовательности	708
§ 7. L^2 -теория	711
§ 8. Случайные процессы и стохастические интегралы	718
§ 9. Задачи	725
Ответы на задачи	728
Литература	732
Предметный указатель	734
Именной указатель	744

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».

Вильям Феллер

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

В 2-х томах

Том 2

Старший науч. редактор А. А. Брядинская
Младший науч. редактор Т. А. Денисова
Художественный редактор В. И. Шаповалов
Художник Е. И. Волков
Технический редактор Н. И. Борисова
Корректор А. Я. Шехтер

ИБ № 3129

Сдано в набор 26.10.83. Подписано к печати 14.05.84. Формат 60×90¹/₄. Бумага типографская № 2. Гарнитура литературная. Печать высокая. Объем 23,50 бум. л. Усл. печ. л. 47,00. Усл. кр.-отт. 47,00. Уч.-изд. л. 46,53. Изд. № 1/1767. Тираж 40 000 экз. Зак. 249. Цена 3 р. 50 к.

ИЗДАТЕЛЬСТВО «МИР»

129820, Москва, И-110, ГСП 1-й Рижский пер., 2

Набрано и сматрицировано в ордена Октябрьской Революции и ордена Трудового Красного Знамени Первой Образцовой типографии имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР по делам издательства, полиграфии и книжной торговли. Москва, М-54, Валуевская, 28.

Отпечатано в Ленинградской типографии № 2 головном предприятии ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгения Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.