Б.В. ГНЕДЕНКО

КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

ИЗДАНИЕ ШЕСТОЕ,
ПЕРЕРАБОТАННОЕ И ДОПОЛНЕННОЕ

Допущено Министерством высшего и среднего специального образования СССР в качестве учебника для студентов математических специальностей университетов

МОСКВА "НАУКА"
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1988

ББК 22.171
Г56
УДК 519.21 (075.8)

Гнеденко Б.В. **Курс теории вероятностей**: Учебник – Изд. 6-е, перераб. и доп. – М.: Наука. Гл. ред. физ.-мат. лит., 1988. – 448 с.
ISBN 5-02-013761-8

Дается систематическое изложение основ теории вероятностей, проиллюстрированное большим числом подробно рассмотренных примеров, в том числе и прикладного содержания. Серьезное внимание уделено рассмотрению вопросов методологического характера.

Настоящее издание значительно отличается по содержанию от 5-го (1969 г.): введены дополнительные параграфы математического и прикладного характера, добавлен большой очерк истории теории вероятностей, содержащий результаты исследований самого последнего времени.

Для студентов математических специальностей университетов и педагогических институтов.

Табл. 22. Ил. 19. Библиогр. 28 назв.

Рецензенты:
Кафедра математической статистики факультета ВМК МГУ
(заведующий кафедрой академик *Ю.В. Прохоров*).
Член-корреспондент АН СССР *Б.А. Севастьянов*.

$Г \dfrac{1702060000-050}{053(02)-88}$ 71-88

ISBN 5-02-013761-8

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ К ШЕСТОМУ ИЗДАНИЮ

Более трети века прошло со времени выхода в свет первого издания настоящей книги. С тех пор в нашей стране и за ее пределами вышли многочисленные учебники по теории вероятностей, заслуживающие самой высокой оценки. Отличительная черта подавляющего большинства этих книг — стремление дать возможно более строгое в теоретическом плане изложение теории и показать силу математической абстракции. Настоящая книга ставит перед собой совсем иную цель: восходя от интуитивных представлений и рассматривая большое число примеров, подойти хотя бы к некоторым исследованиям, активно развивающимся в наши дни.

Это издание значительно отличается от предшествующего: введен ряд параграфов, содержащих изложение некоторых новых результатов, вполне доступных читателям настоящей книги; вновь помещена небольшая глава, содержащая элементы математической статистики; приведено добавление, излагающее довольно подробно период возникновения и развития теории вероятностей. Этот очерк базируется на исследованиях последних лет автора и его учеников. Следует сказать, что многие вопросы истории теории вероятностей еще ожидают своих исследователей. В частности, в таком состоянии находится теория случайных процессов. Однако многое еще требует выяснения и в классической теории вероятностей.

Всем хорошо известно, что абстрактное изложение предмета дает возможность быстрее подвести читателя к современному состоянию науки, а также выиграть страницы, которые необходимы для изложения материала. Я считаю, что при первоначальном знакомстве с математическими дисциплинами, а особенно с теорией вероятностей, необходимо рассмотрение большого числа примеров, которые помогли бы развить своеобразную теоретико-вероятностную интуицию, способность увязывать абстрактные идеи и методы с практическими ситуациями. Это приобретение необходимо каждому математику, а особенно подавляющему большинству студентов-математиков, которым предстоит работать в научно-исследовательских институтах прикладного плана. К тому же в настоящее время с теорией вероятностей вынуждены знакомиться многие специалисты, поскольку в их повседневной работе теоретико-вероятностные концепции крайне необходимы. Им знакомиться с необходимым разделом науки по абстрактным книгам и трудно, и не нужно, поскольку такие книги не создадут так необходимого мостика между потребностями практики и математической теорией. Впрочем, для этой категории читателей, быть может, нужны совсем особые книги, написанные в специальном методическом и психологическом ключе.

Когда книга уже написана, видишь, как много в ней недостатков, как много мест следовало бы в ней переделать. Однако приходится смириться и отложить переделки до возможного переиздания. В связи с этим я прошу читателя направлять мне критические замечания и пожелания, к которым я отнесусь со всем необходимым вниманием.

Я счастлив поблагодарить Ю.В. Прохорова, Б.А. Севастьянова и Д.М. Чибисова за большое число замечаний, которые они мне сделали в результате знакомства с рукописью. К сожалению, я не имел возможности в полной мере использовать все их пожелания, постараюсь это сделать впоследствии.

ИЗ ПРЕДИСЛОВИЯ КО ВТОРОМУ ИЗДАНИЮ

Настоящее издание значительно отличается от первого. Я постарался возможно полнее учесть в нем замечания и пожелания, которые содержались в печатных рецензиях на первое издание книги, а также были сообщены мне устно и письменно. Пожалуй, наиболее существенным изменением является добавление задач для упражнений в первых девяти главах.

Значительные добавления сделаны в главе 10; они касаются главным образом расширения сведений по теории стационарных случайных процессов. Большим изменениям подверглась последняя глава, посвященная математической статистике. В этой главе имеются некоторые новые параграфы, но в то же время исключен частично материал, содержавшийся в первом издании. Пользуюсь случаем сердечно поблагодарить товарищей, высказавших откровенное мнение о недостатках книги и способствовавших своей критикой их исправлению. Особенно я благодарен Ю.В. Линнику за его постоянный интерес к настоящей книге, большое число замечаний к первому изданию и за дискуссию по рукописи второго издания.

ИЗ ПРЕДИСЛОВИЯ К ПЕРВОМУ ИЗДАНИЮ

Настоящий курс разбивается на две части – элементарную (главы 1–6) и специальную (главы 7–11). Последние пять глав могут служить базой для спецкурсов – теории суммирования случайных величин, теории стохастических процессов, элементов математической статистики.

Теория вероятностей рассматривается в книге исключительно как математическая дисциплина, поэтому получение конкретных естественнонаучных или технических результатов в ней не является самоцелью. Все примеры в тексте книги имеют целью только разъяснение общих положений теории и указание на связь этих положений с задачами естествознания. Конечно, одновременно эти примеры дают указания на возможные области приложения общетеоретических результатов, а также развивают умение применять эти результаты в конкретных задачах. Хорошо, если изучающий теорию вероятностей имеет перед глазами какие-нибудь явления материального мира для того, чтобы общая математическая схема наполнялась определенным смыслом. Такое направление изучения дает возможность читателю выработать своеобразную теоретико-вероятностную интуицию, которая позволяет предвидеть в общих чертах выводы раньше, чем применен аналитический аппарат. Заметим далее, что без систематического решения задач изучать теорию вероятностей нельзя, в особенности на первых порах.

Первые четыре параграфа главы 1 являются незначительной переработкой неопубликованных рукописей А.Н. Колмогорова.

Я счастлив поблагодарить здесь моих дорогих учителей А.Н. Колмогорова и А.Я. Хинчина, много помогавших мне своими советами и беседами, касавшимися узловых вопросов теории вероятностей.

ВВЕДЕНИЕ

Цель настоящей книги состоит в изложении основ *теории вероятностей* — математической науки, изучающей закономерности случайных явлений.

Возникновение теории вероятностей относится к середине XVII века и связано с именами Гюйгенса (1629–1695), Паскаля (1623–1662), Ферма (1601–1665) и Якоба Бернулли (1654–1705). В переписке Паскаля и Ферма, вызванной задачами, поставленными азартными игроками и не укладывающимися в рамки математики того времени, выкристаллизовывались постепенно такие важные понятия, как вероятность и математическое ожидание. При этом, конечно, нужно отдавать себе ясный отчет, что выдающиеся ученые, занимаясь задачами азартных игроков, предвидели и фундаментальную натурфилософскую роль науки, изучающей случайные явления. Они были убеждены в том, что на базе массовых случайных событий могут возникать четкие закономерности И только состояние естествознания привело к тому, что азартные игры еще долго продолжали оставаться тем почти единственным конкретным материалом, на базе которого создавались понятия и методы теории вероятностей. Это обстоятельство накладывало отпечаток и на формально-математический аппарат, посредством которого решались возникавшие в теории вероятностей задачи: он сводился исключительно к элементарно арифметическим и комбинаторным методам. Последующее развитие теории вероятностей, а также широкое привлечение ее результатов и методов исследования в естествознание, и в первую очередь в физику, показали, что классические понятия и классические методы не потеряли своего интереса и в настоящее время.

Серьезные требования со стороны естествознания и общественной практики (теория ошибок наблюдений, задачи теории стрельбы, проблемы статистики, в первую очередь статистики народонаселения) привели к необходимости дальнейшего развития теории вероятностей и привлечения более развитого аналитического аппарата. Особенно значительную роль в развитии аналитических методов теории вероятностей сыграли Муавр (1667–1754), Лаплас (1749–1827), Гаусс (1777–1855), Пуассон (1781–1840). С формально-аналитической стороны к этому же направлению примыкает работа творца неевклидовой геометрии Н.И. Лобачевского (1792–1856), посвященная теории ошибок при измерениях на сфере и выполненная с целью установления геометрической системы, господствующей во вселенной.

С половины XIX столетия и приблизительно до двадцатых годов нашего века развитие теории вероятностей связано в значительной мере с именами русских ученых – П.Л. Чебышева (1821–1894), А.А. Маркова (1856–1922), А.М. Ляпунова (1857–1918). Этот успех русской науки был подготовлен деятельностью В.Я. Буняковского (1804–1889), широко культивировавшего в России исследования по применению теории вероятностей к статистике, в особенности к страховому делу и демографии. Им был написан первый в России курс теории вероятностей, оказавший большое влияние на развитие интереса к этой области науки. Основное непреходящее значение работ Чебышева, Маркова и Ляпунова в области теории вероятностей состоит в том, что ими было введено в качестве объекта систематического изучения и широко использовано понятие случайной величины. С результатами Чебышева относительно закона больших чисел, с "цепями Маркова" и с предельной теоремой Ляпунова мы познакомимся в соответствующих разделах настоящей книги.

Современное развитие теории вероятностей характеризуется всеобщим подъемом интереса к ней, а также расширением круга ее практических приложений. В этой напряженной научной работе советская школа теории вероятностей продолжает занимать выдающееся положение. Среди представителей первого поколения советских ученых прежде всего должны быть названы имена С.Н. Бернштейна (1880–1968), А.Н. Колмогорова (1903–1987) и А.Я. Хинчина (1894–1959). В процессе изложения мы будем вынуждены самим существом дела вводить читателя в курс преобразовавших лицо теории вероятностей идей и результатов. Так, уже в первой главе будем говорить о фундаментальных работах С.Н. Бернштейна, Р. Мизеса (1883–1953) и А.Н. Колмогорова по основаниям теории вероятностей. В двадцатых годах нашего столетия А.Я. Хинчин, А.Н. Колмогоров, Е.Е. Слуцкий (1880–1948) и П. Леви (1886–1971) установили тесную связь между теорией вероятностей и метрической теорией функций. Эта связь оказалась весьма плодотворной. На этом пути удалось найти окончательное решение классических задач, поставленных еще П.Л. Чебышевым, а также значительно расширить содержание теории вероятностей. Полностью к советскому периоду относится создание А.Н. Колмогоровым и А.Я. Хинчиным в тридцатых годах основ теории стохастических (вероятностных, случайных) процессов, которая теперь стала основным направлением исследований в теории вероятностей. Указанная теория служит прекрасным образцом того органического синтеза математического и естественнонаучного мышления, когда математик, овладев физическим существом узловой проблемы естествознания, находит для нее адекватный математический язык. Нам важно заметить, что решение классических задач теории вероятностей оказалось тесно связанным с теорией стохастических процессов. Элементы этой важной главы теории вероятностей будут изложены нами в главе десятой.

За последние десятилетия неизмеримо выросла роль, которую играет теория вероятностей в современном естествознании. После того как молекулярные представления о строении вещества получили всеобщее признание, стало неизбежным широкое использование теории вероятностей и в физике и в химии. Заметим, что с точки зрения молекулярной физики каждое вещество состоит из огромного числа малых частиц, находящихся в непрерывном движении и в процессе этого движения воздействующих друг на друга. При этом о природе этих частиц, о существующем между ними взаимодействии, характере их движения и пр. известно мало. В основных чертах эти сведения исчерпываются тем, что частиц, из которых состоит вещество, очень много и что в однородном теле они близки по своим свойствам. Естественно, что при таких условиях обычные для физических теорий методы математических исследований становились бессильными. Так, например, аппарат дифференциальных уравнений не мог привести в указанной обстановке к серьезным результатам. Действительно, ни строение, ни законы взаимодействия между частицами вещества в достаточной мере не изучены, и при таких условиях применение аппарата дифференциальных уравнений должно носить элементы грубого произвола. Но даже если бы этой трудности не существовало, уже одно количество этих частиц представляет собой такую трудность в изучении их движения, которую преодолеть с помощью обычных уравнений механики нет возможности.

К тому же и методологически такой подход неудовлетворителен. Действительно, задача, которая здесь возникает, состоит не в изучении индивидуальных движений частиц, а в изучении тех закономерностей, которые возникают в совокупностях большого числа движущихся и взаимодействующих частиц. Закономерности же, возникающие вследствии участвующих в их возникновении ингредиентов, имеют свое собственное своеобразие и не сводятся к простому суммированию индивидуальных движений. Более того, эти закономерности в известных пределах оказываются не зависящими от индивидуальных особенностей участвующих в их порождении частиц. Конечно, для изучения этих новых закономерностей должны быть найдены и соответствующие новые математические методы исследования. Какие же требования должны быть в первую очередь предъявлены к этим методам? Понятно, что в первую очередь они должны учитывать то, что изучаемое явление носит массовый характер; таким образом, для этих методов наличие большого числа взаимодействующих частиц должно представлять не дополнительную трудность, а облегчать изучение возникающих закономерностей. Далее, недостаточность знаний о природе и строении частиц, а также о характере их взаимодействия также не должна ограничивать эффективности их применения. Этим требованиям лучше всего удовлетворяют методы теории вероятностей.

Чтобы сказанное не было понято ошибочно, мы еще раз подчеркнем следующее обстоятельство. Говоря, что аппарат теории вероятностей лучше приспособлен для изучения молекулярных явлений, мы ни в какой мере не хотим сказать, что философские предпосылки использования теории вероятностей в естествознании лежат в "недостаточности знаний". Основной принцип состоит в том, что при изучении "массовых" явлений возникают *своеобразные новые закономерности.* При изучении явлений, обусловленных действием большого числа молекул, учет свойств каждой молекулы не нужен. Действительно, при изучении явлений природы необходимо отвлекаться от учета несущественных подробностей. Рассмотрение же всех деталей, всех существующих связей, в том числе и несущественных для данного явления, приводит лишь к тому, что само явление затемняется и овладение им отодвигается ввиду такой искусственной усложненной обстановки.

Насколько удачно произведена схематизация явлений, насколько удачно выбран математический аппарат для его изучения, мы можем судить по согласию теории с опытом, с практикой. Развитие естествознания, в частности физики, показывает, что аппарат теории вероятностей оказался весьма хорошо приспособленным к изучению многочисленных явлений природы.

Указанная связь теории вероятностей с потребностями современной физики лучше всего поясняет те причины, в силу которых в последние десятилетия теория вероятностей превратилась в одну из наиболее быстро развивающихся областей математики. Новые теоретические результаты открывают новые возможности для естественнонаучного использования метода теории вероятностей. Всестороннее изучение явлений природы толкает теорию вероятностей на разыскание новых закономерностей, порождаемых случаем. Теория вероятностей не отмежевывается от запросов других наук, а идет в ногу с общим развитием естествознания. Понятно, что сказанное не означает, что теория вероятностей является лишь вспомогательным средством для решения тех или иных практических задач. Наоборот, следует подчеркнуть, что за последние три десятилетия теория вероятностей превратилась в стройную математическую дисциплину с собственными проблемами и методами доказательств. В то же время выяснилось, что наиболее существенные проблемы теории вероятностей служат делу решения различных задач естествознания.

Мы определили в самом начале теорию вероятностей как науку, изучающую случайные явления. Отложив выяснение смысла понятия "случайное явление (событие)" до первой главы, мы сейчас ограничимся несколькими замечаниями. Если в обыденных представлениях, в житейской практике считается, что случайные события представляют собой нечто крайне редкое, идущее вразрез установившемуся порядку вещей, закономерному развитию событий, то в теории вероятностей мы откажемся от этих представлений. Случайные события, как они понимаются в теории вероятностей, обладают рядом характерных особенностей; в частности, все они происходят в массовых явлениях. Под массовыми явлениями мы понимаем такие, которые имеют место в совокупностях большого числа равноправных или почти равноправных объектов и определяются именно этим массовым характером явлений и лишь в незначительной мере зависят от природы составляющих объектов.

Теория вероятностей, подобно другим разделам математики, развилась из потребностей практики; в абстрактной форме она отражает закономерности, присущие случайным событиям массового характера. Эти закономерности играют исключительно важную роль в физике и других областях естествознания, военном деле, разнообразнейших технических дисциплинах, экономике и т.д. В последнее время в связи с широким развитием предприятий, производящих массовую продукцию, результаты теории вероятностей используются не только для браковки уже изготовленной продукции, но, что важнее, для организации самого процесса производства (статистический контроль в производстве). Большое значение в этом круге идей имеет разработка статистических методов управления качеством продукции в процессе производства. Для всего инженерного дела серьезную роль приобрела теория надежности, широко использующая методы теории вероятностей. Здесь уместно заметить, что в свою очередь теория надежности выдвинула перед теорией вероятностей ряд новых теоретических вопросов. Связь теории вероятностей с практическими потребностями, как уже было отмечено, была основной причиной бурного развития ее в последние десятилетия. Многие ее разделы были развиты как раз в связи с ответами на запросы практиков. Здесь кстати вспомнить замечательные слова основателя нашей отечественной школы теории вероятностей П.Л. Чебышева: "Сближение теории с практикой дает самые благотворные результаты, и не одна только практика от этого выигрывает; сами науки развиваются под влиянием ее: она открывает им новые предметы для исследования или новые стороны в предметах давно известных... Если теория много выигрывает от новых приложений старой методы или от новых развитий ее, то она еще более приобретает открытием новых метод, и в этом случае наука находит себе верного руководителя в практике".

ГЛАВА 1

СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

§ 1. Интуитивные представления о случайных событиях

На протяжении длительного времени человечество изучало и использовало для своей деятельности лишь так называемые детерминистические закономерности. Подавляющая часть сведений, получаемых учащимися в школьных курсах физики, математики, химии, относится именно к этому кругу идей. Приведем примеры.

Если в основании пирамиды находится квадрат со стороной a и ее высота равна h, то объем пирамиды равен $\frac{1}{3}a^2h$.

Если тело свободно падает на земную поверхность, то путь, пройденный им за t секунд после начала падения, равен $s = gt^2/2$.

Если химически чистую воду при атмосферном давлении 760 мм рт. ст. нагреть до 100 °С, то вода начнет превращаться в пар.

При любых химических реакциях каких угодно веществ без обмена с окружающей средой общее количество вещества остается неизменным (закон сохранения вещества).

Число подобных примеров можно увеличивать неограниченно. Однако закономерности этого типа не в состоянии описать все разнообразие ситуаций, с которыми приходится сталкиваться в научной и практической деятельности. Для примера спросим себя: сколько дорожных происшествий произойдет завтра в Москве? Сколько вызовов от больных поступит в пункт скорой медицинской помощи? Сколько лет проживет только что родившийся младенец? Как много времени придется затратить на поиск неисправности в определенном сложном техническом устройстве? Все эти вопросы обладают одной особенностью – на них невозможно дать однозначного ответа, поскольку процессы, с которыми они связаны, по самому их существу лишены полной определенности. Действительно, дорожные происшествия зависят от огромного числа причин, которые невозможно предусмотреть: погода, состояние дорожного покрытия, освещенность, психологическое состояние водителей и пешеходов, взаимное расположение автомобилей на дороге и множество других. Аналогичные заключения мы можем высказать и относительно остальных предложенных нами вопросов.

Во всех подобных случаях мы говорим, что интересующее нас событие является случайным.

Поскольку случайные события врываются в нашу жизнь помимо нашего желания и постоянно окружают нас и, более того, поскольку все явления природы, согласно современным воззрениям, являются случайными, необходимо научиться их изучать и разработать для этой цели методы их изучения.

Сейчас уместно подчеркнуть, что со случайными явлениями мы вынуждены сталкиваться не время от времени, а постоянно и зачастую именно они определяют структуру того или иного интересующего нас процесса. Так, при организации работы телефонной станции необходимо учитывать, что моменты поступления вызовов от абонентов, так же как и длительности разговоров между абонентами, случайны. В грузовой морской порт суда дальнего плавания поступают не точно по расписанию, а в моменты времени, нередко существенно отличные от запланированных. Точно так же длительность погрузо-разгрузочных работ (обработки судна) коренным образом зависит не только от погрузочных средств причала, но и от устройства трюма, характера и количества прибывших грузов, их упаковки и многих других обстоятельств. Таким образом, как при организации работы телефонной станции, так и при организации работы грузового порта, мы должны считаться с, так сказать, двойной случайностью — случайностью поступления требовании (вызовов абонентов, прибытия судов) на обслуживание и случайностью длительности их обслуживания. Мы видим, таким образом, что случайные события играют большую роль как в научных исследованиях, так и в практической деятельности. Более того, нередко при проектировании ответственных сооружений (телефонных узлов, морских портов и т.д.) мы должны опираться на эти случайные явления. Это обстоятельство привело к тому, что за последние три столетия, и в особенности за последние десятилетия, случайные явления были подвергнуты систематическому исследованию. Краткий исторический очерк этого процесса приложен в конце настоящей книги.

Прежде чем переходить к изложению основных результатов теории вероятностей, мы должны формализовать те понятия, с которыми она имеет дело. Пока же понятие "случайного" явления имеет лишь чисто описательный, интуитивный и весьма расплывчатый облик. Мы увидим, что теория вероятностей занимается изучением не любых событий, которые в житейской практике называются случайными, а только тех из них, которые обладают определенными свойствами.

Прежде всего, она ограничивается изучением лишь тех событий, которые в принципе могут быть осуществлены неограниченное число раз, притом в неизменных условиях. Приведем примеры. Игральная кость может быть подброшена в одних и тех же условиях столько раз, сколько заблагорассудится нам, исследователям. Мы можем производить неограниченно большое число наблюдений за числом вызовов абонентов телефонной сети, поступающих за четверть часа дневного времени. Выходные дни и ночные часы мы при этом исключим из рассмотрения, поскольку могут возникнуть опасения, что условия формирования потока вызовов в эти дни и часы будут отличны от будней.

Далее, теория вероятностей занимается лишь теми событиями, которые обладают так называемой *статистической устойчивостью* или, иначе, устойчивостью частот. Это требование к случайным событиям следует рассмотреть более подробно.

Представим себе, что производится последовательность испытаний, в каждом из которых может появиться, а может и не появиться некоторое событие A. Эти испытания производятся в одинаковых условиях, и результаты одних испытаний не оказывают влияния на результаты других (как говорят, испытания независимы). Обозначим через μ число появлений события A в каких-то n заранее назначенных номерах испытаний, например в n последовательных испытаниях; тогда *частота*, т.е. отношение μ/n при больших n для статистически устойчивых событий A близка к постоянной и лишь слегка изменяется от одной серии в n испытаний к другой.

Проверка статистической устойчивости представляет собой довольно сложную задачу, и сейчас мы не станем ею заниматься. Позднее же к этому вопросу мы будем возвращаться неоднократно.

Теперь подчеркнем ту мысль, что теория вероятностей не занимается изучением уникальных событий, которые не допускают повторений. Так, не имеет смысла говорить о том, какова вероятность, что данный студент сдаст экзамен по теории вероятностей на ближайшей экзаменационной сессии, поскольку здесь речь идет о единичном событии, повторить которое в тех же самых условиях нет возможности. Мы можем об этом событии высказывать лишь некоторые субъективные суждения, основанные на нашем знакомстве со знаниями этого студента. Теория же вероятностей ставит перед собой задачу изучения объективных закономерностей, которые не зависят от субъективных суждений того или иного лица. И как бы ни были интересны вопросы, касающиеся единичных, неповторимых событий, теория вероятностей к ним не имеет отношения, если только относительно них нет возможности провести длительные независимые испытания в одинаковых условиях. Так, все следующие высказывания –

6 августа 1999 г. в Термезе произойдет землетрясение силой в 8 баллов;

к двухтысячному году будут найдены радикальные методы излечения всех форм рака;

в 1989 г. родится поэт, по таланту равный А.С. Пушкину – носят характер уникальности, и ответ на них будет получен (положительный или отрицательный) в свое время. Пока же эти события носят неопределенный характер. И хотя они относятся к категории "может произойти, а может и не произойти", к ним понятия и методы теории вероятностей не имеют отношения.

Теория вероятностей изучает лишь такие случайные события, в отношении которых имеет смысл не только утверждение об их случайности, но и возможна объективная оценка доли случаев их появления. Эта оценка может быть выражена предложением вида:

1. Вероятность того, что при осуществлении определенного комплекса условий $\mathfrak{S}$ произойдет событие A, равна p.*)

Закономерности такого рода называются *стохастическими* или *вероятностными*. С ними приходится иметь дело в самых разнообразных ситуациях, связанных с изучением как природных, так и общественных явлений, а также в самых разнообразных прикладных вопросах.

Для примера, пусть у нас имеется некоторое техническое устройство, скажем, электрическая лампочка. Нас интересует вопрос: проработает ли она безотказно 1500 часов? Заранее мы не можем ответить ни утвердительно, ни отрицательно. Для этого у нас нет данных. Но мы можем и должны ответить иначе, что доля тех лампочек из большого числа находящихся в одинаковых условиях эксплуатации, которые проработают по меньшей мере 1500 часов, равна p. Эта величина существенно зависит от того, на каком предприятии изготовлена лампочка и, безусловно, от того, в каких условиях ей приходится работать (исправность проводки и патрона, размеры колебаний напряжения и силы тока и пр.).

Конечно, каждая электрическая лампочка индивидуальна по своим качествам, но этих лампочек изготовляется очень много и испытать можно большое их число, притом изготовленных в одних и тех же производственных условиях. Мы встречаемся, таким образом, с повторением испытаний двух различных типов:

а) повторение испытаний для одного и того же объекта изучения (повторное извлечение шара из урны, содержащей несколько одинаковых шаров; извлечение наудачу карты из полной колоды и т.д.);

б) испытание многих сходных объектов. Именно по этому образцу на заводах производится испытание качества изготовленной продукции.

Формулировка детерминистических закономерностей, к которым все мы привыкли со школьных лет, звучит так:

2. При каждом осуществлении комплекса условий $\mathfrak{S}$ обязательно происходит событие A.

В следующем параграфе мы увидим, что вероятность случайного события измеряется числом, заключенным между 0 и 1. Единице соответству-

---

*) Мы пока не даем определения понятия вероятности и полагаемся на имеющиеся у читателей интуитивные представления.

ют те события, которые обязательно наступают при каждом осуществлении комплекса $\mathfrak{S}$. Такие события называются достоверными. Если же событие невозможно, то ему соответствует вероятность 0. Мы видим, что детерминистические закономерности можно рассматривать как частный случай стохастических, для которых вероятность p равна 0 или 1. Таким образом, стохастические закономерности являются более широкими, чем детерминистические, и позволяют точные, количественные методы применять и в тех случаях, когда о классическом детерминизме не может быть и речи.

Каждый исследователь, имеющий дело с применениями теории вероятностей к физике, биологии, инженерному делу, организации производства или любой другой конкретной области знания, исходит в своей работе из убеждения, что вероятностные суждения выражают определенные объективные свойства реальных явлений. Поэтому утверждение, что при выполнении некоторого комплекса условий $\mathfrak{S}$ событие A имеет вероятность p, имеет серьезное познавательное значение. Оно указывает на наличие определенной, хотя и своеобразной, но от этого не менее объективной связи между комплексом условий $\mathfrak{S}$ и событием A. Даже только утверждение, что вероятность события A при осуществлении комплекса условий $\mathfrak{S}$ существует (хотя и неизвестна), является содержательным утверждением, нуждающимся в объективном обосновании и последующей проверке, если оно принято в качестве гипотезы. Философская задача состоит в выяснении природы этой связи. Ее решению посвящено большое число исследований, но задача до сих пор еще остается не решенной. Трудность этой проблемы привела к тому парадоксальному результату, что даже среди ученых, стоящих на материалистических философских позициях, встречается стремление не найти положительное ее решение, а снять ее, объявив вероятностные суждения имеющими отношение только к состоянию познающего субъекта (измеряющими степень его уверенности в наступлении события A и дающими основания для приписывания равных вероятностей исходов испытания при полном незнании). В последнее время такие субъективистские выводы нередко делаются по поводу событий, происходящих в условиях неопределенности, как теперь принято говорить.

§ 2. Поле событий. Классическое определение вероятности

Мы до сих пор не давали формализованного определения ни случайного события, ни его вероятности. Теперь приступим к этому, стремясь одновременно воспитать у читателей теоретико-вероятностную интуицию. Такая цель заставляет нас вводить определение вероятности постепенно, как бы повторяя исторический путь. Такой подход позволяет избежать формального восприятия, а отправляясь от простейших представлений, постепенно переходить к более сложным и общим.

Начнем с так называемого классического определения вероятности. При этом мы увидим, что оно в действительности является не определением, а скорее методом вычисления вероятностей во вполне определенных и сильно ограниченных условиях. Классическое определение исходит из предположения равновозможности как объективного свойства изучаемых явлений, основанного на их реальной симметрии. Понятие равновозможности (равновероятности) является первичным, не подлежащим формальному определению. Оно лишь поясняется рядом простых и доступных примеров.

При бросании на плоскость геометрически правильного куба, изготовленного из однородного материала, любая из шести граней (при подбрасывании наудачу) не имеет реальных преимуществ перед другими. Таким образом, если перенумеровать грани цифрами от 1 до 6, то при бросании куба могут произойти шесть равновероятных событий: выпадение граней 1, 2, 3, 4, 5, 6.

При подбрасывании однородного по плотности правильного двадцатигранника (икосаедра) выпадение каждой грани в силу симметрии одинаково возможно. Мы имеем случай, когда возможны двадцать равновероятных исходов. Представим себе теперь, что при каждом испытании единственно возможны *n несовместимых* и *равновозможных* исходов $E_1, E_2, \ldots, E_n$.

Слово "несовместимый" было введено в науку английским ученым Байесом (1702–1761) и означает, что если наступил какой-нибудь исход E_1, то ни один из остальных $n-1$ исходов в этом испытании наступить уже не мог. Так, если при бросании игральной кости выпала грань "3", то это означает, что при том же бросании не могла появиться грань "5". Каждый такой исход станем называть *элементарным событием.*

Наряду с элементарными событиями рассматриваются также случайные события. Часто представляет интерес наступление при испытании не какого-то элементарного события, а одного из нескольких определенных элементарных событий. Например, при бросании игральной кости нас может интересовать появление граней с числом очков, бóльших трех, т.е. появление какого-то из элементарных событий "4", "5", "6". Мы станем говорить в этом случае, что нас интересует случайное событие – выпадение числа очков, бóльших трех. Вообще, если нас интересует появление какого-то из определенных элементарных событий, например, одного из событий $E_{i_1}, E_{i_2}, \ldots, E_{i_m}$, то мы станем говорить, что нас интересует наступление случайного события A, состоящего в выпадении какого-то из m только что указанных элементарных событий.

Вероятностью случайного события A называется отношение числа несовместимых равновероятностных элементарных событий, составляющих A (т.е. числа m), к числу всех возможных элементарных событий (т.е. к числу n). Вероятность случайного события A обозначается символом $\mathsf{P}(A)$.

Согласно только что данному определению

$$\mathsf{P}(A) = \frac{m}{n}.$$

Например, при однократном бросании игральной кости полная группа попарно несовместимых и равновероятных событий состоит из событий

$$E_1, E_2, E_3, E_4, E_5, E_6,$$

которые состоят соответственно в выпадении 1, 2, 3, 4, 5, 6 очков. Событие

$$C = E_2 + E_4 + E_6,$$

состоящее в выпадении четного числа очков, подразделяется на три частных случая, входящих в состав несовместимых и равновероятных событий. Поэтому вероятность события C равна

$$\mathsf{P}(C) = 3/6 = 1/2.$$

Очевидно также, что в силу принятого определения

$$\mathsf{P}(E_i) = 1/6, \qquad 1 \leqslant i \leqslant 6,$$

$$\mathsf{P}(E_1 + E_2) = 2/6 = 1/3$$

и т.д.

Рассмотрим теперь бросание двух костей. Если кости правильны, то выпадение каждой из 36 возможных комбинаций числа очков на первой и второй кости можно считать равновероятными. Так, скажем, вероятность выпадения в сумме 12 очков равна 1/36. Выпадение в сумме 11 очков возможно двумя способами: на первой кости 5, а на другой 6, на первой кости 6, а на второй 5. Поэтому вероятность выпадения в сумме одиннадцати очков равна $2/36 = 1/18$. Читатель легко проверит, что вероятности выпадения той или иной суммы очков даются следующей таблицей:

Таблица 1

Число очков	2	3	4	5	6	7
Вероятность	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$

Число очков	8	9	10	11	12
Вероятность	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Подсчитаем теперь общее число возможных случайных событий, которое можно образовать из n элементарных. Очевидно, что можно образовать C_n^m событий, каждое из которых будет содержать по m каких то элементарных событий $(1 \leqslant m \leqslant n)$. При $m = n$ случайное событие всегда происходит, т.е. оно является достоверным. Всего, таким образом, образовано $\sum_{m=1}^{n} C_n^m = 2^n - 1$ событий. Добавим теперь ко всем построенным событиям еще одно, которому не соответствует ни одно элементарное событие, т.е. состоящее из пустого множества элементов. Очевидно, что оно никогда не может наступить (поскольку ему не соответствует ни одно элементарное событие). Это случайное событие носит название невозможного события. Таким образом, всех случайных событий в рассмотренном нами случае будет 2^n.

Ближайшие рассмотрения относятся не только к классическому определению вероятности, но и ко всем дальнейшим обобщениям. Будем считать фиксированным комплекс условий $\mathfrak{S}$ и станем рассматривать некоторую систему S событий $A, B, C \ldots$ *), каждое из которых должно при каждом осуществлении комплекса $\mathfrak{S}$ п р о и з о й т и или н е п р о и з о й т и **). Между событиями системы S могут существовать известные соотношения, с которыми мы постоянно будем иметь дело и которые поэтому прежде всего изучим.

1) Если при каждом осуществлении комплекса условий $\mathfrak{S}$, при котором происходит событие A, происходит и событие B, то мы будем говорить, что A *влечет за собой* ***) B, и обозначать это обстоятельство символом $\subset$:

$$A \subset B$$

или символом $\supset$:

$$B \supset A.$$

2) Если A влечет за собой B и в то же время B влечет за собой A, т.е. если при каждой реализации комплекса условий $\mathfrak{S}$ события A и B оба наступают или оба не наступают, то мы будем говорить, что события A и B *равносильны* и будем обозначать это обстоятельство символом $=$:

$$A = B.$$

3) Событие, состоящее в наступлении обоих событий A и B, будем называть *произведением* событий A и B и обозначать AB.

---

*) События в дальнейшем обозначаются латинскими прописными буквами $A, B, C, D, E, \ldots$

**) Вместо "произойти" говорят также "появиться", "иметь место" или "наступить"

***) Вместо "A влечет за собой B" говорят также "A является частным случаем B".

4) Событие, состоящее в наступлении хотя бы одного из событий A и B, будем называть *суммой* событий A и B и обозначать $A + B$.

5) Событие, состоящее в том, что событие A происходит, а событие B не происходит, будем называть *разностью* событий A и B и обозначать $A - B$.

6) Событие, состоящее в том, что событие A не происходит, называется *противоположным для* A и обозначается символом $\bar{A}$.

Пусть, например, комплекс условий $\mathfrak{S}$ состоит в том, что внутри квадрата, изображенного на рис. 1, выбирается наудачу точка. Обозначим через A событие "выбранная точка лежит внутри левого круга" и через B событие "выбранная точка лежит внутри правого круга". Тогда события A, $\bar{A}$, B, $\bar{B}$, $A + B$, AB, $A - B$, $B - A$, $\overline{A - B}$ состоят в попадании выбранной точки внутрь областей, заштрихованных на соответствующих фигурах рис. 1.

Рассмотрим другой пример. Допустим, что комплекс условий $\mathfrak{S}$ состоит в том, что на стол бросается (один раз) игральная кость. Обозначим через A

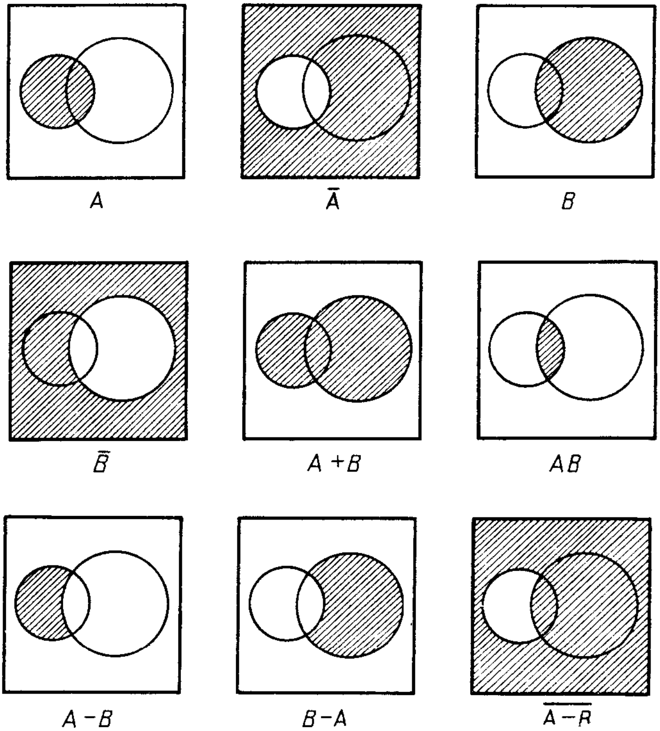

Рис. 1

выпадение на верхней грани кости *) шести очков, через B – выпадение трех очков, через C – выпадение какого-либо четного числа очков, через D – выпадение какого-либо числа очков, кратного трем. Тогда события A, B, C и D связаны следующими соотношениями:

$$A \subset C, \quad A \subset D, \quad B \subset D,$$

$$A + B = D, \quad CD = A.$$

Определение суммы и произведения двух событий обобщается на любое число событий:

$$A + B + \ldots + N$$

обозначает событие, заключающееся в наступлении хотя бы одного из событий $A, B, \ldots, N$, а

$$AB \ldots N$$

обозначает событие, заключающееся в наступлении всех событий $A, B, \ldots, N$.

7) Событие называется *достоверным*, если оно с необходимостью должно произойти (при каждой реализации комплекса условий $\mathfrak{S}$). Например, при бросании двух игральных костей достоверно, что сумма очков будет не меньше двух.

Событие называется *невозможным*, если оно заведомо не может произойти (ни при одной реализации комплекса условий $\mathfrak{S}$). Например, при бросании двух игральных костей невозможно появление суммы очков, равной тринадцати.

Очевидно, что все достоверные события равносильны между собой. Поэтому законно обозначать все достоверные события одной буквой. Мы будем употреблять для этого букву Ω. Все невозможные события тоже равносильны между собой. Мы будем обозначать любое невозможное событие знаком ϕ.

8) Два события A и $\bar{A}$ называются *противоположными*, если для них одновременно выполняются два соотношения:

$$A + \bar{A} = \Omega, \quad A\bar{A} = \phi.$$

Например, если при бросании одной игральной кости C обозначает выпадение четного числа очков, то

$$\Omega - C = \bar{C}$$

есть событие, состоящее в выпадении нечетного числа очков.

---

*) В Японии изготовляют теперь кости не только в виде кубов, но также в виде додекаэдров и икосаэдров.

8) Два события A и B называются *несовместимыми*, если их совместное появление невозможно, т.е. если

$$AB = \phi.$$

Если

$$A = B_1 + B_2 + \ldots + B_n$$

и события B_i попарно несовместимы, т.е.

$$B_i B_j = \phi \quad \text{при} \quad i \neq j,$$

то говорят, что событие A *подразделяется на частные случаи* $B_1, B_2, \ldots, B_n$. Например, при бросании игральной кости событие C, состоящее в выпадении четного числа очков, подразделяется на частные случаи E_2, E_4, E_6, состоящие соответственно в выпадении 2, 4 и 6 очков.

События $B_1, B_2, \ldots, B_n$ образуют *полную группу событий*, если хотя бы одно из них непременно должно произойти (при каждом осуществлении комплекса $\mathfrak{S}$), т.е. если

$$B_1 + B_2 + \ldots + B_n = \Omega.$$

Особенно существенны для нас в дальнейшем будут *полные группы попарно несовместимых событий*. Такова, например, при однократном бросании игральной кости система событий

$$E_1, E_2, E_3, E_4, E_5, E_6,$$

состоящая, соответственно, в появлении 1, 2, 3, 4, 5 и 6 очков.

9) В каждой задаче теории вероятностей приходится иметь дело с каким-либо определенным комплексом условий $\mathfrak{S}$ и с какой-либо определенной системой S событий, наступающих или не наступающих после каждой реализации комплекса условий $\mathfrak{S}$. Относительно этой системы целесообразно сделать следующие допущения:

а) *если системе S принадлежат события A и B, то ей принадлежат также события AB, $A + B$, $A - B$*;

б) *система S содержит достоверное и невозможное события.*

Система событий, удовлетворяющая этим допущениям, называется *полем событий.*

Заметим еще в заключение, что во всех рассмотрениях теории вероятностей равносильные между собой события могут заменять друг друга. Поэтому мы условимся в дальнейшем любые два равносильных события считать просто тождественными друг другу.

Рассмотрим какую-либо полную систему G попарно несовместимых равновозможных событий (назовем их *элементарными событиями*):

$$E_1, E_2, \ldots, E_n,$$

и рассмотрим систему событий S, состоящую из невозможного события ϕ, всех событий E_k системы G и всех событий A, которые могут быть подразделены на частные случаи, входящие в состав системы G.

Например, если система G состоит из трех элементарных событий E_1, E_2, и E_3, то в систему S входят события $\phi, E_1, E_2, E_3, E_1 + E_2, E_2 + E_3, E_1 + E_3$, $\Omega = E_1 + E_2 + E_3$.

Легко установить, что система S есть поле событий. В самом деле, очевидно, что сумма, разность и произведение событий из S входят в S; невозможное событие ϕ входит в S по определению, а достоверное событие Ω входит в S так как оно представляется в виде

$$\Omega = E_1 + E_2 + \ldots + E_n.$$

В соответствии с приведенным определением каждому событию A, принадлежащему к построенному сейчас полю событий S, приписывается вполне определенная вероятность

$$\mathsf{P}(A) = m/n,$$

где m есть число тех событий E_i исходной группы G, которые являются частными случаями события A. Таким образом, вероятность $\mathsf{P}(A)$ можно рассматривать как *функцию от события A, определенную на поле событий S*.

Функция эта обладает следующими свойствами:

1. *Для каждого события A поля S*

$$\mathsf{P}(A) \geqslant 0.$$

2. *Для достоверного события Ω*

$$\mathsf{P}(\Omega) = 1.$$

3. *Если событие A подразделяется на частные случаи B и C и все три события A, B и C принадлежат полю S, то*

$$\mathsf{P}(A) = \mathsf{P}(B) + \mathsf{P}(C).$$

Это свойство называется теоремой сложения вероятностей.

Свойство 1 очевидно, так как дробь m/n не может быть отрицательной. Свойство 2 не менее очевидно, так как достоверному событию Ω благоприятствуют все n возможных результатов испытания, и поэтому

$$\mathsf{P}(\Omega) = n/n = 1.$$

Докажем свойство 3. Пусть событию B благоприятствуют m', а событию C – m'' *событий* E_i системы G. Так как события B и C по допущению несовместимы, то события E_i, благоприятствующие одному из них, отличны от событий E_i, благоприятствующих другому. Всего, таким образом, имеется $m' + m''$ событий E_i, благоприятствующих появлению одного из событий B

или C, т.е. благоприятствующих событию $B + C = A$. Следовательно,

$$\mathsf{P}(A) = \frac{m' + m''}{n} = \frac{m'}{n} + \frac{m''}{n} = \mathsf{P}(B) + \mathsf{P}(C),$$

что и требовалось доказать.

Мы ограничимся здесь указанием еще нескольких свойств вероятности.

4. *Вероятность события $\bar{A}$, противоположного событию A, равна*

$$\mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A).$$

Действительно, так как

$$A + \bar{A} = \Omega$$

то, согласно уже доказанному свойству 2,

$$\mathsf{P}(A + \bar{A}) = 1,$$

а так как события A и $\bar{A}$ несовместимы, то по свойству 3

$$\mathsf{P}(A + \bar{A}) = \mathsf{P}(A) + \mathsf{P}(\bar{A}).$$

Два последних равенства доказывают наше предложение.

5. *Вероятность невозможного события равна нулю.*

В самом деле, события Ω и ϕ несовместимы, поэтому

$$\mathsf{P}(\Omega) + \mathsf{P}(\phi) = \mathsf{P}(\Omega),$$

откуда следует, что

$$\mathsf{P}(\phi) = 0.$$

6. *Если событие A влечет за собой событие B, то*

$$\mathsf{P}(A) \leqslant \mathsf{P}(B).$$

Действительно, событие B может быть представлено как сумма двух событий A и $\bar{A}B$. Отсюда в силу свойств 3 и 1 получаем:

$$\mathsf{P}(B) = \mathsf{P}(A + \bar{A}B) = \mathsf{P}(A) + \mathsf{P}(\bar{A}B) \geqslant \mathsf{P}(A).$$

7. *Вероятность любого события заключена между нулем и единицей.*

Из того, что для любого события A имеют место соотношения

$$\phi \subset A + \phi = A = A\Omega \subset \Omega,$$

следует в силу предыдущего свойства, что имеют место неравенства

$$0 = \mathsf{P}(\phi) \leqslant \mathsf{P}(A) \leqslant \mathsf{P}(\Omega) = 1.$$

§ 3. Примеры

Мы рассмотрим теперь несколько примеров вычисления вероятностей событий, пользуясь классическим определением вероятности. Приводимые нами примеры носят исключительно иллюстративный характер и не претендуют на то, чтобы сообщить читателю все основные методы расчета вероятностей.

П р и м е р 1. Бросаются три игральные кости. Что вероятнее: получить в сумме выпавших очков 11 или 12?

Эта задача, как это видно из исторического дополнения, была одной из первичных, на которой формировались понятия и методы теории вероятностей. Утверждают, что с ней связана и следующая легенда: однажны к Галилею (а кто говорит, что к Гюйгенсу) за консультацией обратился ландскнехт. Его интересовал именно предложенный нам вопрос. Ландскнехт оказался мыслящим человеком, склонным к теоретическому и экспериментальному мышлению. Он заявил Галилею, что согласно логике обе эти суммы должны появляться одинаково часто, но опыт учит другому, а именно, что сумма 11 появляется чаще, чем 12. В чем здесь дело?

Обоснование ландскнехта на первый взгляд звучит убедительно: числа 11 и 12 оба могут быть разложены на сумму трех положительных слагаемых лишь шестью различными способами, а именно,

$$11 = 1 + 5 + 5 = 1 + 4 + 6 = 2 + 3 + 6 = 2 + 4 + 5 =$$
$$= 3 + 3 + 5 = 3 + 4 + 4;$$
$$12 = 1 + 5 + 6 = 2 + 4 + 6 = 2 + 5 + 5 = 3 + 4 + 5 =$$
$$= 3 + 3 + 6 = 4 + 4 + 4;$$

отсюда по его мнению вытекает равновозможность обоих интересующих нас событий.

Однако Галилей возразил ему, сказав, что каждое из этих разложений следует снабдить еще определенным весом и пояснил свою мысль таким рассуждением. Назовем кости "первой", "второй" и "третьей"; тогда разложение $1 + 5 + 5$ на самом деле может произойти не одним, а т р е м я различными способами:

$$1 + 5 + 5 = 5 + 1 + 5 = 5 + 5 + 1,$$

т.е. единица может выпасть на первой, второй или третьей кости. Точно также разложение $1 + 4 + 6$ может произойти следующими ш е с т ь ю различными способами:

$$1 + 4 + 6 = 1 + 6 + 4 = 4 + 1 + 6 = 4 + 6 + 1 =$$
$$= 6 + 1 + 4 = 6 + 4 + 1.$$

Таким образом, 11 в сумме может появиться не шестью, а 27 различными равновозможными способами. Сумма же 12, оказывается, разлагается лишь 25 различными способами. Здесь все дело в том, что разложение $4 + 4 + 4$ осуществимо лишь одним способом.

Теперь заметим, что общее число всех возможных равновероятных выпадений трех игральных костей равно $6^3 = 216$. Это было правильно подсчитано еще в XII веке.

Обозначим через A и B случайные события, состоящие в выпадении соответственно 11 и 12 очков в сумме. Тогда согласно данному нами классическому определению вероятности,

$$\mathsf{P}(A) = 27/216 \quad \text{и} \quad \mathsf{P}(B) = 25/126.$$

В математической статистике показывается, что для уверенного разделения двух вероятностей, отличающихся менее чем на одну сотую, нужно произвести много тысяч испытаний.

П р и м е р 2. Из колоды карт (36 карт) наудачу вынимаются три карты. Найти вероятность того, что среди них окажется точно один туз.

Р е ш е н и е. Полная группа равновероятных и несовместимых событий в нашей задаче состоит из всевозможных комбинаций по три карты, их число равно C_{36}^3. Число благоприятствующих событий можно подсчитать следующим способом. Один туз мы можем выбрать C_4^1 различными способами, а две другие карты (не тузы) можно выбрать C_{32}^2 различными способами. Так как для каждого определенного туза две остальные карты могут быть выбраны C_{32}^2 способами, то всего благоприятствующих случаев будет $C_4^1 \cdot C_{32}^2$. Искомая вероятность, таким образом, равна

$$p = \frac{C_4^1 \cdot C_{32}^2}{C_{36}^3} = \frac{\dfrac{4}{1} \cdot \dfrac{32 \cdot 31}{1 \cdot 2}}{\dfrac{36 \cdot 35 \cdot 34}{1 \cdot 2 \cdot 3}} = \frac{31 \cdot 16}{35 \cdot 3 \cdot 17} = \frac{496}{1785} \approx 0{,}2778,$$

т.е. немного больше 0,25.

П р и м е р 3. Из колоды карт (36 карт) наудачу вынимаются три карты. Найти вероятность того, что среди них окажется хотя бы один туз.

Р е ш е н и е. Обозначим интересующее нас событие буквой A: оно может быть представлено в виде суммы трех следующих несовместимых событий: A_1 — появление одного туза, A_2 — появления двух тузов, A_3 — появления трех тузов.

Рассуждениями, аналогичными тем, которые мы привели при решении предыдущей задачи, легко установить, что число случаев, благоприятствующих

событию A_1 равно $C_4^1 \cdot C_{32}^2$,

" A_2 " $C_4^2 \cdot C_{32}^1$,

" A_3 " $C_4^3 \cdot C_{32}^0$.

Так как число всевозможных случаев равно C_{36}^3, то

$$\mathsf{P}(A_1)=\frac{C_4^1\cdot C_{32}^2}{C_{36}^3}=\frac{16\cdot 31}{3\cdot 35\cdot 17}\approx 0{,}2778,$$

$$\mathsf{P}(A_2)=\frac{C_4^2 C_{32}^1}{C_{36}^3}=\frac{3\cdot 16}{3\cdot 35\cdot 17}\approx 0{,}0269,$$

$$\mathsf{P}(A_3)=\frac{C_4^3\cdot C_{32}^0}{C_{36}^3}=\frac{1}{3\cdot 35\cdot 17}\approx 0{,}0006.$$

В силу теоремы сложения

$$\mathsf{P}(A)=\mathsf{P}(A_1)+\mathsf{P}(A_2)+\mathsf{P}(A_3)=\frac{109}{3\cdot 119}\approx 0{,}3053.$$

Этот пример можно решить и иным методом. Событие $\overline{A}$, противоположное A, состоит в том, что среди вынутых карт не окажется ни одного туза. Очевидно, что три нетуза можно вынуть из колоды карт C_{32}^3 различными способами и, следовательно,

$$\mathsf{P}(\overline{A})=\frac{C_{32}^3}{C_{36}^3}=\frac{32\cdot 31\cdot 30}{36\cdot 35\cdot 34}=\frac{31\cdot 8}{3\cdot 17\cdot 7}\approx 0{,}6947.$$

Иско[illegible]ая вероятность равна

$$\mathsf{P}(A)=1-\mathsf{P}(\overline{A})\approx 0{,}3053.$$

Примечание. В обоих примерах выражение "наудачу" означало, что всевозможные комбинации по три карты равновероятны.

Пример 4. Рассмотрим теперь пример, который широко используется при проверке качества принимаемой продукции. В математическом отношении он близок к примеру 2.

В урне находится N одинаковых по размеру и внешнему виду шаров; среди них M белых и $N-M$ черных. Наудачу вынимаются n шаров (без возвращения в урну). Чему равна вероятность того, что среди них окажется m белых?

Из условия задачи ясно, что мы предполагаем выполнение неравенств $m\leqslant n$ и $m\leqslant M$, а также $n-m\leqslant N-M$.

Общее число всех равновероятных случаев, как легко понять, равно C_N^n. Число благоприятствующих случаев подсчитаем следующим путем: различных способов извлечь m белых шаров имеется C_M^m, а $n-m$ черных — C_{N-M}^{n-m}. Таким образом общее число благоприятных равновозможных

Таблица 2

Номер испытаний	Число красных карт	Число благопр. случаев	Частота	Номер испытаний	Число красных карт	Число благопр. случаев	Частота
1	8	0	0,00	51	9	13	0,25
2	9	1	0,50	52	8	13	0,25
3	11	1	0,33	53	7	13	0,25
4	9	2	0,50	54	9	14	0,26
5	11	2	0,40	55	7	14	0,26
6,	8	2	0,33	56	9	15	0,27
7	11	2	0,29	57	9	16	0,28
8	9	3	0,37	58	11	16	0,28
9	8	3	0,33	59	8	16	0,27
10	7	3	0,30	60	8	16	0,27
11	12	3	0,27	61	8	16	0,26
12	10	3	0,25	62	10	16	0,26
13	9	4	0,31	63	12	16	0,25
14	13	4	0,29	64	9	17	0,27
15	12	4	0,27	65	11	17	0,26
16	8	4	0,25	66	12	17	0,26
17	11	4	0,23	67	11	17	0,26
18	10	4	0,22	68	8	17	0,25
19	8	4	0,21	69	10	17	0,25
20	11	4	0,20	70	8	17	0,25
21	12	4	0,19	71	7	17	0,24
22	10	4	0,18	72	9	18	0,25
23	10	4	0,17	73	10	18	0,25
24	9	5	0,21	74	8	18	0,24
25	9	6	0,24	75.	11	18	0,24

случаев равно $C_M^m \cdot C_{N-M}^{n-m}$. Отсюда искомая вероятность

$$p = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^M}.$$

Пример 5. Колоду карт, состоящую из 36 карт, наудачу разделяют на две равные части. Чему равна вероятность, что в обеих частях окажется *по равному числу* красных и черных карт.

Выражение "наудачу" означает, что всевозможные разделения колоды на две равные части равновероятны.

Решение. Нам нужно найти вероятность того, что среди наудачу вынутых из колоды 18 карт 9 будут красными и 9 – черными. Согласно при-

Номер испытаний	Число красных карт	Число благопр. случаев	Частота	Номер испытаний	Число красных карт	Число благопр. случаев	Частота
26	14	6	0,23	76	8	18	0,24
27	9	7	0,26	77	9	19	0,25
28	10	7	0,25	78	9	20	0,26
29	10	7	0,24	79	5	20	0,26
30	7	7	0,23	80	8	20	0,25
31	10	7	0,22	81	7	20	0,25
32	7	7	0,22	82	10	20	0,24
33	8	7	0,21	83	9	21	0,25
34	10	7	0,21	84	6	21	0,24
35	9	8	0,23	85	10	21	0,25
36	9	9	0,25	86	10	21	0,24
37	10	9	0,24	87	9	22	0,25
38	10	9	0,24	88	7	22	0,25
39	8	9	0,23	89	7	22	0,25
40	7	9	0,22	90	10	22	0,24
41	9	10	0,24	91	8	22	0,24
42	10	10	0,24	92	8	22	0,24
43	10	10	0,23	93	10	22	0,24
44	9	11	0,25	94	8	22	0,23
45	8	11	0,24	95	11	22	0,23
46	7	11	0,24	96	9	23	0,24
47	12	11	0,23	97	9	24	0,25
48	9	12	0,25	98	10	24	0,25
49	6	12	0,25	99	7	24	0,24
50	7	12	0,24	100	7	24	0,24

меру 4 искомая вероятность равна, следовательно,

$$p = \frac{C_{18}^{9} \cdot C_{18}^{9}}{C_{36}^{18}} = \frac{(18!)^4}{36!(9!)^4}.$$

Чтобы составить себе представление о величине этой вероятности и при этом не производить утомительных вычислений, мы воспользуемся формулой Стирлинга, согласно которой имеет место следующее асимптотическое равенство:

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}.$$

Таким образом, имеем:

$$18! \sim 18^{18} e^{-18} \sqrt{2\pi \cdot 18},$$

$$9! \sim 9^9 e^{-9} \sqrt{2\pi \cdot 9},$$

$$36! \sim 36^{36} \cdot e^{-36} \sqrt{2\pi \cdot 36},$$

и значит,

$$p \approx \frac{(\sqrt{2\pi \cdot 18} \cdot 18^{18} \cdot e^{-18})^4}{\sqrt{2\pi \cdot 36}\, 36^{36} \cdot e^{-36} (\sqrt{2\pi \cdot 9} \cdot 9^9 \cdot e^{-9})^4}$$

После несложных преобразований находим, что

$$p \approx \frac{2}{\sqrt{18\pi}} \approx \frac{4}{15} \approx 0{,}26.$$

Естественно возникает вопрос: какое отношение имеют найденные вероятности к реальным явлениям? Чтобы наглядно это ощутить, на одной из лекций был проведен такой эксперимент: студенты принесли несколько колод карт по 36 карт каждая и затем сто раз было произведено разделение колод наудачу на две равные части. В таблице 2 приведены результаты этого эксперимента. В первом столбце указан номер испытания, во втором – число появившихся в одной из полуколод красных карт, в третьем – число случаев деления красных и черных карт пополам среди уже произведенных испытаний и, наконец, в четвертом столбце даны значения частот.

Приводимый на рис. 2 график наглядно представляет изменение частоты μ/n в зависимости от числа испытаний. Вначале, когда число опытов невелико, ломаная линия порой значительно уклоняется от прямой $y = p \approx 0{,}26$.

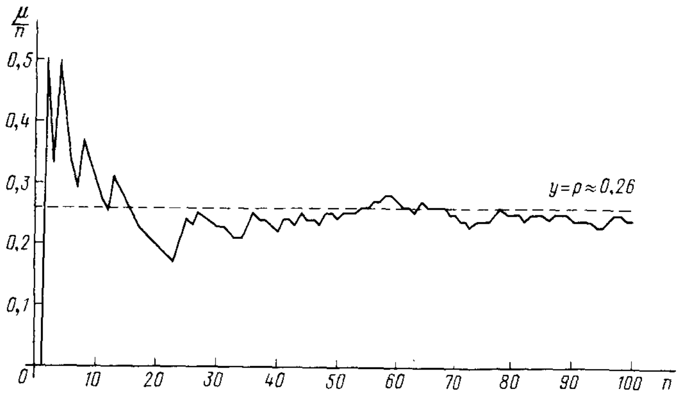

Рис. 2

Затем с увеличением числа опытов ломаная в общем все ближе и ближе подходит к этой прямой.

П р и м е р 6. Имеются n частиц, каждая из которых может находиться с одной и той же вероятностью $\frac{1}{N}$ в каждой из $N(N > n)$ ячеек. Найти вероятность того, что: 1) в определенных n ячейках окажется по одной частице, 2) в каких-то n ячейках окажется по одной частице.

Р е ш е н и е. Эта задача играет важную роль в современной статистической физике, и в зависимости от того, как образуется полная группа равновероятных событий, приходят к той или иной физической статистике: Больцмана, Бозе–Эйнштейна, Ферми–Дирака.

В статистике Больцмана равновероятными считаются любые мыслимые размещения, отличающиеся не только числом, но и индивидуальностью частиц: в каждой ячейке может помещаться любое число частиц от 0 до n.

Общее число возможных размещений мы подсчитаем следующим способом: каждая частица может находиться в каждой из N ячеек, следовательно, n частиц можно разметить по ячейкам N^n различными способами.

В первом вопросе число благоприятствующих случаев будет, очевидно, $n!$ и, значит, вероятность того, что в определенные n ячеек попадет по одной частице, равна

$$p_1 = \frac{n!}{N^n}.$$

Во втором вопросе число благоприятствующих случаев будет в C_N^n раз больше и, значит, вероятность того, что в какие-то n ячеек попадет по одной частице, равна

$$p_2 = \frac{C_N^n \cdot n!}{N^n} = \frac{N!}{N^n(N-n)!}.$$

В статистике Бозе–Эйнштнейна считаются тождественными случаи, когда частицы меняются местами между ячейками (важно только, сколько частиц попало в ячейку, но не индивидуальность попавших частиц), и полная группа равновероятных событий состоит из всевозможных размещений n частиц по N ячейкам, причем за одно размещение принимается целый класс больцмановских размещений, отличающихся не числами содержащихся в определенных ячейках частиц, а только самими частицами. Для наглядного представления о различии статистик Больцмана и Бозе–Эйнштейна рассмотрим частный пример: $N = 4$, $n = 2$. Всевозможные размещения в этом примере можно записать в виде следующей таблицы, в которой a и b – наименования частиц. В статистике Больцмана все 16 возможностей представляют собой различные равновероятные события, в статистике же

Бозе–Эйнштейна случаи 5 и 11, 6 и 12, 7 и 13, 8 и 14, 9 и 15, 10 и 16 попарно отождествляются и мы имеем группу из 10 равновероятных событий.

Таблица 3

Случаи	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Ячейки	*ab*				*a*	*a*	*a*				*b*	*b*	*b*			
		ab			*b*			*a*	*a*		*a*			*b*	*b*	
			ab			*b*		*b*		*a*		*a*		*a*		*b*
				ab			*b*		*b*	*b*			*a*		*a*	*a*

Вычислим теперь общее число равновероятных случаев в статистике Бозе–Эйнштейна. С этой целью заметим, что всевозможные размещения частиц по ячейкам мы можем получить следующим путем: расположим ячейки на прямой вплотную друг к другу, расположим далее рядом одну возле другой на той же прямой наши частицы. Рассмотрим теперь всевозможные перестановки частиц и перегородок между ячейками. Таким образом, как легко сообразить, будут учтены всевозможные заполнения ячеек, отличающихся как порядком расположения частиц в ячейках, так и порядком расположения перегородок.

Число этих перестановок равно $(N+n-1)!$. Среди этих перестановок имеются и тождественные: каждое распределение по ячейкам считается $(N-1)!$ раз, так как мы различали, какие перегородки были между ячейками, а кроме того, каждое распределение по ячейкам мы снова считали по $n!$ раз, так как мы учитывали не только число частиц в ячейке, но и то, какие это частицы и в каком порядке они расположены. Таким образом, каждое распределение по ячейкам мы считали $n!(N-1)!$ раз, отсюда число различных в смысле Бозе–Эйнштейна размещений частиц по ячейкам равно

$$\frac{(n+N-1)!}{n!(N-1)!}.$$

Таким образом, число равновероятных событий в полной системе событий нами найдено. Теперь мы легко можем ответить на вопросы нашей задачи. В статистике Бозе–Эйнштейна вероятности p_1 и p_2 равны

$$p_1 = \frac{1}{\dfrac{(n+N-1)!}{n!(N-1)!}} = \frac{n!(N-1)!}{(n+N-1)!},$$

$$p_2 = \frac{C_N^n}{\dfrac{(n+N-1)!}{n!(N-1)!}} = \frac{N!(N-1)!}{(N-n)!\,(N+n-1)!}.$$

Рассмотрим, наконец, статистику Ферми–Дирака. Согласно этой статистике в ячейке может находиться либо одна частица либо не находится ни одной: индивидуальность частиц уничтожается.

Общее число различных размещений частиц по ячейкам в статистике Ферми–Дирака подсчитать легко: первая частица может быть расположена N различными способами, вторая – только $N-1$, третья – $(N-2)$ и, наконец, n-я – $(N-n+1)$ различными способами. Таким образом, n частиц по N ячейкам могут быть расположены

$$N(N-1)\ldots(N-n+1) = \frac{N!}{(N-n).}$$

различными равновероятными способами.

Легко сообразить, что в статистике Ферми–Дирака искомые вероятности равны

$$p_1 = \frac{(N-n)!}{N!}.$$

$$p_2 = 1.$$

Рассмотренный пример показывает, насколько важно точно определять, какие события считаются в задаче равновероятными.

П р и м е р 7. У театральной кассы стоят в очереди $2n$ человек. Среди них n человек имеют монеты только рублевого достоинства, а остальные – только монеты по 50 копеек. Билет стоит 50 копеек. Каждый покупатель приобретает по одному билету. В начальный момент в кассе нет денег. Чему равна вероятность того, что ни один покупатель не будет ждать сдачу?

Эта задача является переформулировкой вопроса, который возник при изучении проблем управления качеством продукции в процессе производства.

Всевозможные расстановки покупателей равновероятны. Таким образом имеется C_{2n}^n всех возможных равновероятных случаев. Для разыскания числа благоприятствующих случаев прибегнем к следующему геометрическому приему. Рассмотрим плоскость xOy и допустим, что покупатели в порядке очередности располагаются в точках оси абсцисс с координатами $1, 2, \ldots, 2n$. В начале координат расположена касса. Припишем каждому лицу, имеющему рубли, значение $+1$, а имеющему полтинники – значение -1. Будем теперь суммировать последовательно эти значения слева направо и в каждой целочисленной точке отмечать в виде ординаты полученную сумму (рис. 3). Припишем началу координат ординату, равную 0. Соединим концы ординат ломаной и назовем ее траекторией. Ясно, что она на концах отрезка $(0, 2n)$ имеет ординаты, равные 0.

Каждой траектории поставлено в соответствие определенное расположение лиц с рублями и полтинниками. Интересующему нас событию благоприятствуют те и только те траектории, которые не поднимаются над осью абсцисс.

Вычислим теперь общее число траекторий, достигающих или пересекающих хотя бы раз прямую $y = 1$. Эти и только эти траектории благоприятствуют противоположному событию, когда хотя бы одному лицу

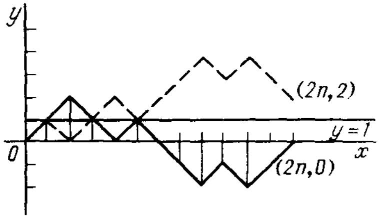

Рис. 3

придется ожидать сдачу. Для этой цели построим новую, фиктивную траекторию. До первого достижения прямой $y = 1$ она совпадает со старой, а от точки достижения этой прямой она является зеркальным отображением старой траектории относительно прямой $y = 1$ (на рис. 3 – пунктирная ломаная). Каждая новая траектория начинается в точке $(0, 0)$ и заканчивается в точке $(2n, 2)$. Отсюда вытекает, что единичных подъемов она имеет больше, чем спусков (именно: $n + 1$ подъем и $n - 1$ спуск). Отсюда следует, что всех новых траекторий будет C_{2n}^{n-1}. Значит число событий благоприятствующих событию нашей задачи, будет $C_{2n}^{n} - C_{2n}^{n-1}$ и тем самым искомая вероятность равна

$$p = 1 - \frac{C_{2n}^{n-1}}{C_{2n}^{n}} = 1 - \frac{n}{n+1} = \frac{1}{n+1}.$$

§ 4. Геометрические вероятности

Еще в самом начале развития теории вероятностей была замечена недостаточность "классического" определения вероятности, основанного на рассмотрении конечной группы равновероятных событий. Уже тогда частные примеры привели к некоторому видоизменению этого определения и построению понятия вероятности также для случаев, когда мыслимо бесконечное множество исходов. При этом по-прежнему основную роль играло понятие "равновероятности" некоторых событий.

Общая задача, которая ставилась и привела к распространению понятия вероятности, может быть сформулирована следующим способом.

Пусть имеется, например, на плоскости некоторая область G и в ней содержится другая область g с квадрируемой границей. В область G наудачу бросается точка и спрашивается, чему равна вероятность того, что точка попадет в область g. При этом выражению "точка бросается наудачу в область G" придается следующий смысл: брошенная точка может попасть в любую точку области G, вероятность попасть в какую-либо часть области G пропорциональна мере этой части (длине, площади и т.д.) и не зависит от ее расположения и формы.

Таким образом, по определению, вероятность попадания в область g при бросании наудачу точки в область G равна

$$p = \frac{\operatorname{mes} g}{\operatorname{mes} G}.$$

Рассмотрим несколько примеров.

Пример 1. Задача о встрече. Два лица A и B условились встретиться в определенном месте между 12 часами и часом. Пришедший первым ждет другого в течение 20 минут, после чего уходит. Чему равна вероятность встречи лиц A и B, если приход каждого из них в течение указанного часа может произойти наудачу и моменты прихода независимы *).

Решение. Обозначим моменты прихода лица A через x и лица B через y. Для того чтобы встреча произошла, необходимо и достаточно, чтобы

$$|x - y| \leqslant 20.$$

Станем изображать x и y как декартовы координаты на плоскости; в качестве единицы масштаба выберем минуту. Всевозможные исходы изобразятся точками квадрата со сторонами 60; благоприятствующие встрече – расположатся в заштрихованной области (рис. 4).

Искомая вероятность равна **) отношению площади заштрихованной фигуры к площади всего квадрата:

$$p = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

---

*) То есть момент прихода одного лица не влияет на момент прихода другого. Понятие независимости событий будет подробно рассмотрено в § 9.

**) В § 9 мы увидим, что в силу независимости моментов прихода лиц A и B вероятность того, что лицо A придет в промежуток от x до $x + h$, а лицо B – в промежуток от y до $y + s$, равна $\frac{h}{60} \cdot \frac{s}{60}$, т.е. пропорциональна площади прямоугольника со сторонами h и s.

Некоторые инженеры-исследователи применяли задачу о встрече к решению следующей проблеме организации производства. Рабочий обслуживает несколько однотипных станков, каждый из которых в случайные моменты времени может потребовать внимания рабочего. Может случиться, что в то время, когда рабочий занят у одного станка, потребовалось его вмешательство у других станков. Требуется найти вероятность этого события, т.е., иными словами, среднюю длительность ожидания станком рабочего (иначе говоря, простой станка). Заметим, однако, что схема задачи о

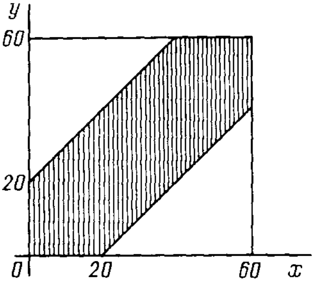

Рис. 4

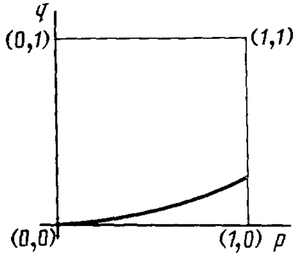

Рис. 5

встрече мало пригодна для решения этого производственного вопроса, так как никакого условленного времени, в течение которого станки обязательно требуют к себе внимания рабочего, не существует, а длительности операции рабочего у станка не постоянны. Помимо этой основной причины, нужно указать на сложность вычислений в задаче о встрече для случая большого числа лиц (станков). А эту задачу нередко нужно решать для большого числа станков (в текстильном производстве, например, некоторые ткачихи брали на обслуживание по несколько десятков станков).

Пример 2. Коэффициенты p и q квадратного уравнения

$$x^2 + px + q = 0$$

выбираются наудачу в промежутке $(0,\ 1)$. Спрашивается, чему равна вероятность того, что корни будут действительными числами?

Чтобы корни квадратного уравнения были действительными числами, необходимо и достаточно выполнение неравенства $p^2 \geqslant 4q$. В прямоугольных декартовых координатах (рис. 5) множество всех возможных пар чисел $(p,\ q)$ задается точками квадрата с вершинами $(0,\ 0)$, $(0,\ 1)$, $(1,\ 1)$, $(1,\ 0)$. Точки же, благоприятствующие нашему событию, лежат под параболой $q = \frac{1}{4}p^2$. Таким образом, согласно определению, искомая вероятность равна

$$\frac{\int\limits_0^1 \frac{p^2}{4}\,dp}{1} = \frac{1}{12}.$$

Задачи, подобные только что рассмотренной, нашли интересные применения в теории чисел и в ряде научных и технических применений.

Пример 3. Парадокс Бертрана. Теория геометрических вероятностей неоднократно подвергалась критике за произвольность определения вероятности событий. При этом авторы приходили к убеждению, что для бесконечного числа исходов нельзя дать объективного, на зависящего от способа расчета, определения вероятности. В качестве особенно яркого выразителя этого скептицизма можно привести французского математика прошлого века Жозефа Бертрана. В своем курсе теории вероятностей он привел ряд задач на геометрические вероятности, в которых результат зависел от метода решения. В качестве примера приведем одну из задач, рассмотренных Бертраном.

Наудачу берется хорда в круге. Чему равна вероятность того, что ее длина превосходит длину стороны вписанного равностороннего треугольника?

Решение 1. По соображениям симметрии можно заранее задать направление хорды. Проведем диаметр, перпендикулярный к этому направлению. Очевидно, что только хорды, пересекающие диаметр в промежутке от четверти до трех четвертей его длины, будут превосходить стороны правильного треугольника. Таким образом, искомая вероятность равна 1/2.

Решение 2. По соображениям симметрии можно заранее закрепить один из концов хорды на окружности. Касательная к окружности в этой точке и две стороны правильного треугольника с вершиной в этой точке образуют три угла по $60°$. Условию задачи благоприятствуют только хорды, попадающие в средний угол. Таким образом, при этом способе вычисления искомая вероятность оказывается равной 1/3.

Решение 3. Чтобы определить положение хорды, достаточно задать ее середину. Чтобы хорда удовлетворяла условию задачи, необходимо чтобы ее середина находилась внутри круга, концентрического данному, но половинного радиуса. Площадь этого круга равна одной четверти площади данного; таким образом, искомая вероятность равна 1/4.

Мы должны теперь выяснить, в чем причина неоднозначности решения нашей задачи. Лежит ли причина в принципиальной невозможности определить вероятность для случаев бесконечного числа возможных исходов или же причина лежит в том, что мы приняли в процессе решения какие-либо недопустимые предпосылки.

Дело, как легко усмотреть, заключается в том, что за решение одной и той же задачи, пользуясь тем, что в условии задачи не определено понятие проведения хорды наудачу, выдаются решения трех различных задач. В самом деле, в первом решении вдоль одного из диаметров заставляют катиться круглый цилиндрический стержень (рис. 6 *а*). Множество всех возможных мест остановки этого стержня есть множество точек отрезка AB длины, равной диаметру. Равновероятными считаются события, состоящие в том, что остановка произойдет в интервале длины h, где бы внутри диаметра ни был расположен этот отрезок. Во втором решении стержень, закрепленный на шарнире, расположенном в одной из точек окружности, заставляют совершать колебания размером не более 180° (рис. 6*б*). При этом предполагается, что остановка стержня внутри дуги окружности длины h зависит только от длины дуги, но не от ее положения. Таким образом, равновероятными событиями считаются остановки стержня в любых дугах окружности одинаковой длины. Несогласованность определений вероятности в первом и во втором решениях становится совершенно очевидной после такого простого расчета. Вероятность того, что стержень остановится в промежутке от A до x, согласно первому решению равна $\frac{x}{D}$. Вероятность того, что проекция точки пересечения стержня с окружностью во втором решении попадает в тот же интервал, как

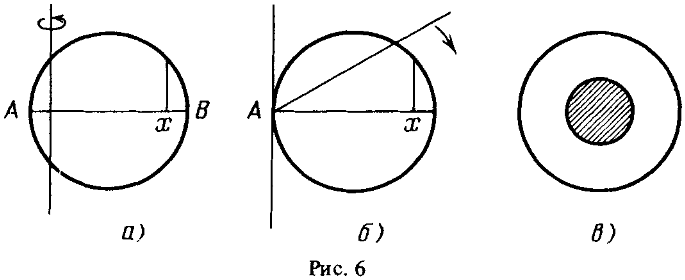

Рис. 6

показывают элементарно геометрические подсчеты, равна

$$1 - \frac{1}{\pi} \arccos \frac{2x - D}{D} \quad \text{при } x \geqslant D/2$$

и

$$\frac{1}{\pi} \arccos \frac{D - 2x}{D} \quad \text{при } x \leqslant D/2.$$

Наконец, в третьем решении мы бросаем наудачу точку внутрь круга и спрашиваем себя о вероятности попадания внутрь некоторого меньшего концентрического круга (рис. 6 *в*).

Различие постановок задач во всех трех случаях совершенно очевидно.

Пример 4. Задача Бюффона. Плоскость разграфлена параллельными прямыми, отстоящими друг от друга на расстоянии $2a$. На плоскость наудачу*) бросается игла длины $2l$ $(l < a)$. Найти вероятность того, что игла пересечет какую-нибудь прямую.

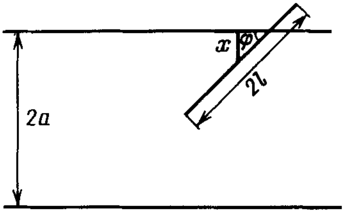

Рис. 7

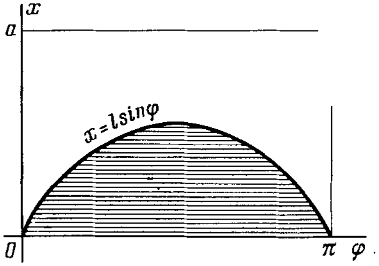

Рис. 8

Решение. Обозначим через x расстояние от центра до ближайшей параллели и через φ—угол, составленный иглой с этой параллелью. Величины x и φ полностью определяют положение иглы. Всевозможные положения иглы определяются точками прямоугольника со сторонами a и π. Из рис. 7 видно, что для пересечения иглы с параллелью необходимо и достаточно, чтобы

$$x \leqslant l \sin \varphi.$$

Искомая вероятность в силу сделанных предположений равна отношению площади заштрихованной на рис. 8 области к площади прямоугольника

$$p = \frac{1}{a\pi} \int_0^\pi l \sin \varphi \, d\varphi = \frac{2l}{a\pi}.$$

Заметим, что задача Бюффона является исходным пунктом для решения некоторых проблем теории стрельбы, учитывающих размеры снаряда.

Пример 5. На горизонтальную плоскость, разграфленную параллельными прямыми, отстоящими друг от друга на расстоянии $2a$, нау-

---

*) Под словом "наудачу" здесь подразумевается следующее: во-первых, центр иглы наудачу падает на отрезок длины $2a$, перпендикулярный к проведенным прямым, во-вторых, вероятность того, что угол φ, составленный иглой и проведенными прямыми, будет заключаться между φ_1 и $\varphi_1 + \Delta\varphi$ пропорциональна $\Delta\varphi$ и, в-третьих, величины x и φ независимы (см. § 7).

дачу*) брошен выпуклый контур, диаметр которого меньше $2a$. Найти вероятность того, что контур пересечет одну из параллельных прямых.

Решение. Положим сначала, что выпуклый контур является многоугольником с n сторонами. Пусть его стороны пронумерованы от номера 1 до номера n. Если многоугольник пересекается с какой-либо параллельной прямой, то это пересечение должно произойти по каким-либо двум сторонам. Обозначим через $p_{ij} = p_{ji}$ вероятность того, что пересечение произойдет по i-й и j-й сторонам. Очевидно, что событие A, состоящее в том, что брошенный многоугольник пересечет одну из параллельных прямых, может быть представлено в виде следующей суммы попарно несовместимых событий:

$$A = (A_{12} + A_{13} + \ldots + A_{1n}) +$$
$$+ (A_{23} + A_{24} + \ldots + A_{2n}) + \ldots + (A_{n-2,n-1} + A_{n-2,n}) + A_{n-1,n},$$

где через A_{ij} $(i < j,\ i = 1, 2, \ldots;\ j = 1, 2, \ldots)$ обозначено событие, состоящее в пересечении с параллельной прямой i-й и j-й сторон. По теореме сложения вероятностей

$$p = \mathsf{P}(A) = [\mathsf{P}(A_{12}) + \mathsf{P}(A_{13}) + \ldots + \mathsf{P}(A_{1n})] +$$
$$+ [\mathsf{P}(A_{23}) + \ldots + \mathsf{P}(A_{2n})] + \ldots + \mathsf{P}(A_{n-1,n}) =$$
$$= (p_{12} + p_{13} + \ldots + p_{1n}) + (p_{23} + p_{24} + \ldots + p_{2n}) + \ldots + p_{n-1,n}.$$

Пользуясь равенством $p_{ij} = p_{ji}$, мы может записать вероятность p иным способом:

$$p = \frac{1}{2}\,[(p_{12} + p_{13} + \ldots + p_{1n}) + (p_{21} + p_{23} + \ldots + p_{2n}) + \ldots$$
$$\ldots + (p_{n1} + p_{n2} + \ldots + p_{n,n-1})].$$

Но сумма $\sum\limits_{j=1}^{n}{}' p_{ij}$, где положено $p_{ii} = 0$, представляет собой не что иное, как вероятность пересечения i-й стороны многоугольника с одной из параллельных прямых. Если длину i-й стороны обозначить через $2l_i$, то из задачи Бюффона находим, что

$$\sum_{j=1}^{n}{}' p_{ij} = \frac{2l_i}{\pi a},$$

*) В этом примере "наудачу" значит, что мы берем какой-либо отрезок, жестко связанный с контуром, и бросаем его "наудачу" в смысле предыдущего примера. Нетрудно показать, что таким образом определенное понятие не зависит от выбора указанного отрезка.

и, следовательно,

$$p = \frac{\sum\limits_{i=1}^{n} 2l_i}{2\pi a}.$$

Если обозначить через $2s$ периметр многоугольника, то мы найдем, что

$$p = \frac{s}{\pi a}.$$

Мы видим, таким образом, что вероятность p не зависит ни от числа сторон, ни от величины сторон многоугольника. Отсюда мы заключаем, что найденная формула верна и для любого выпуклого контура, так как мы всегда можем рассматривать этот последний как предел выпуклых многоугольников с безгранично возрастающим числом сторон.

§ 5. О статистической оценке неизвестной вероятности

Классическое определение вероятности при переходе от простейших примеров к рассмотрению сложных задач, в особенности же задач естественнонаучного или технического характера, наталкивается на непреодолимые трудности принципиального порядка. Прежде всего, в большинстве случаев возникает вопрос о возможности нахождения разумного способа выделения "равновозможных случаев". Так, например, из соображений симметрии, на которых основаны наши суждения о равновероятности событий, вывести вероятность распада атома радиоактивного вещества за определенный промежуток времени, или же определить вероятность того, что родившийся ребенок окажется мальчиком, представляется невозможным. В этих случаях еще на заре возникновения теории вероятностей был замечен иной способ приближенной оценки неизвестной вероятности случайного события.

Длительные наблюдения над появлением или непоявлением события A при большом числе независимых испытаний, производимых при одном и том же комплексе условий $\mathfrak{S}$, в ряде случаев показывают, что число появлений события A подчиняется устойчивым закономерностям. А именно, если мы через μ обозначим число появлений события A при n независимых испытаниях, то оказывается, что отношение μ/n (частота события A) при достаточно больших n для большинства таких серий наблюдений сохраняет почти постоянную величину. причем большие отклонения наблюдаются тем реже, чем многочисленнее произведенные испытания. Более того, оказывается, что для тех случаев, к которым применимо классическое определение вероятности, это колебание частоты происходит около вероятности события p. Мы увидим впоследствии, что этот эмпирический факт имеет глубокие основания в теореме Бернулли. То, что при большом числе испытаний частота события остается почти постоянной, дает нам возможность расширить круг явлений, для которых мы будем говорить об их вероятности.

Представим себе, что относительно события A принципиально возможно проведение неограниченной последовательности независимых друг от друга испытаний в неизменных условиях $\mathfrak{S}$. Если в результате достаточно многочисленных наблюдений замечено, что частота события A ведет себя достаточно правильно и почти всегда колеблется около некоторой, вообще говоря, неизвестной, постоянной, то мы скажем, что это событие имеет вероятность.

За численное значение этой вероятности может быть приближенно при большом числе n независимых испытаний, производящихся в неизменных условиях $\mathfrak{S}$, принята частота события A.

Однако испытания позволяют нам делать заключения и иного характера. Представим себе, что некоторые соображения дают нам основание считать, что вероятность некоторого события A равна p. Пусть, далее, при проведении нескольких серий независимых испытаний оказалось, что частоты в подавляющем числе серий значительно отклоняются от величины p. Это обстоятельство дает нам основание высказать сомнение относительно правильности наших априорных суждений и предпринять более детальное исследование тех предпосылок, из которых мы исходили

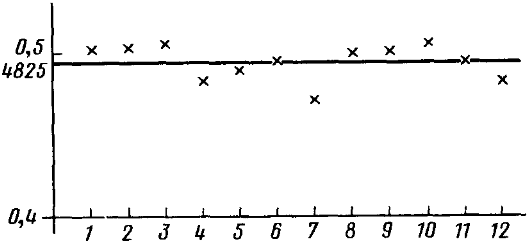

Рис. 9

в своих априорных выводах. Так, например, в отношении некоторой игральной кости мы делаем предположения о ее геометрической правильности и однородности материала, из которого она изготовлена. Из этих предварительных предпосылок мы вправе сделать вывод, что при бросании кости вероятность выпадения некоторой грани, например, грани с номером 5, должна быть равна 1/6. Если неоднократные серии достаточно многочисленных испытаний (бросаний) в нашем примере систематически показывают, что частота появления этой грани значительно отличается

Таблица 4

Месяц	1	2	3	4	5	6	7
Всех	7280	6957	7883	7884	7892	7609	7585
Мальчиков	3743	3550	4017	4173	4117	3944	3964
Девочек	3537	3407	3866	3711	3775	3665	3621
Частота рождения девочек	0,486	0,489	0,490	0,471	0,478	0,482	0,462

Месяц	8	9	10	11	12	За год
Всех	7393	7203	6903	6552	7132	88 273
Мальчиков	3797	3712	3512	3392	3761	45 682
Девочек	3596	3491	3391	3160	3371	42 591
Частота рождения девочек	0,484	0,485	0,491	0,482	0,473	0,4825

от 1/6, то мы усомнимся не в существовании определенной вероятности выпадения этой грани, а в наших предпосылках о правильности кости или в правильности организации процесса наших испытаний (бросаний).

В качестве иллюстрации почти постоянства частот при больших числах испытаний рассмотрим распределение новорожденных по полу по месяцам. Данные заимствованы из книги Г. Крамера "Математические методы статистики" и представляют собой официальные данные шведской статистики за 1935 г.

На рис. 9 показано уклонение частоты рождений девочек по месяцам от частоты рождений девочек за год.

Заметим, что в случае статистического определения снова имеют место такие свойства вероятности:

1) вероятность достоверного события равна единице;

2) вероятность невозможного события равна нулю;

3) если случайное событие C является суммой конечного числа несовместимых событий $A_1, A_2, \ldots, A_n$, имеющих вероятность, то его вероятность существует и равна сумме вероятностей слагаемых

$$\mathsf{P}(C) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n).$$

В заключении мы должны остановиться на весьма распространенной, в особенности среди естествоиспытателей, концепции вероятности, данной Р. Мизесом. Согласно Мизесу раз частота по мере увеличения числа опытов все меньше и меньше уклоняется от вероятности p, то в пределе должно быть

$$p = \lim_{n\to\infty} \frac{\mu}{n}.$$

Это равенство Мизес предлагает считать определением понятия вероятности. По его мнению, любое априорное определение обречено на неудачу и лишь данное им эмпирическое определение способно обеспечить интересы естествознания, математики и философии, причем раз классическое определение имеет лишь весьма ограниченное применение, а статистическое определение применимо ко всем имеющим научный интерес случаям, то классическое определение через равновозможность, основанную на симметрии, Мизес предлагает вовсе отбросить. Более того, Мизес считает вообще ненужным выяснение структуры явлений, для которых вероятность является объективной числовой характеристикой, для него достаточно наличия эмпирической устойчивости частоты.

Мы не будем останавливаться на деталях теории Мизеса, в частности, на тех ограничениях и условиях, которые он дополнительно накладывает на последовательность испытаний. За подробностями теории мы отошлем читателя к его книге "Вероятность и статистика". В то же время его положения не были безоговорочно приняты наукой. Критические замечания в развернутом виде изложены в статье А.Я. Хинчина*).

В концепции Мизеса вероятность теряет свой характер объективной числовой характеристики некоторых реальных явлений. Действительно, до производства бесконечного числа испытаний нельзя даже говорить про вероятность того или иного события, а поскольку этого нельзя осуществить, то и вообще мы лишены возможности в каких-либо условиях использовать теорию вероятности. Следует заметить при этом, что, требуя от частот сходимости к вероятносте, Мизес выставляет такое требование, какого не предъявляют ни в одной области естествознания. Ведь в самом деле. никто из нас не откажется от понятия температуры только потому, что мы не можем произвести бесконечного числа измерений и не можем проверить, будут ли результаты этих измерений стремиться к пределу, если бы мы их все же стали производить. Или не станем же мы говорить, что какой-либо предмет не имеет размеров только потому, что последовательность наших измерений не стремится к пределу. Более того, следуя за Мизесом, мы вообще не можем говорить о температуре тела или о существовании размеров предмета до тех пор, пока не появится мыслящий субъект, не начнет производить измерения и не убедится в том, что их результаты стремятся к пределу.

---

*) Частотная теория Р. Мизеса и современные идеи теории вероятностей//Вопросы философии. – 1961. – Вып. 1. – С. 91–102; Вып. 2. – С. 77–89.

В историческом очерке, помещенном в конце книги, отмечено, что как классическое определение вероятности, так и статистическое были впервые четко сформулированы Я. Бернулли

§ 6. Аксиоматическое построение теории вероятностей

До недавнего времени теория вероятностей представляла собой еще не сложившуюся математическую науку, в которой основные понятия были недостаточно четко определены. Эта нечеткость приводила нередко к парадоксальным выводам (вспомним парадоксы Бертрана). Естественно, что приложения теории вероятностей к изучению явлений природы были слабо обоснованы и встречали порой резкую и обоснованную критику. Нужно сказать, что эти обстоятельства мало смущали естествоиспытателей и их наивный теоретиковероятностный подход в различных областях науки приводил к крупным успехам. Развитие естествознания в начале текущего столетия предъявило к теории вероятностей повышенные требования. Возникла необходимость в систематическом изучении основных понятий теории вероятностей и выяснении тех условий, при которых возможно использование ее результатов. Вот почему особенно важное значение приобрело формально-логическое обоснование теории вероятностей, ее аксиоматическое построение. При этом в основу теории вероятностей как математической науки должны быть положены некоторые предпосылки, являющиеся обобщением многовекового человеческого опыта. Дальнейшее же ее развитие должно строиться посредством дедукции из этих основных положений без обращения к наглядным представлениям, к выводам "согласно здравому смыслу". Иными словами, теория вероятностей должна строиться из аксиом так же, как любая сформировавшаяся математическая наука — геометрия, теоретическая механика, абстрактная теория групп и т.д.

Впервые такая точка зрения была высказана и развита в 1917 г. советским математиком С.Н. Бернштейном. При этом С.Н. Бернштейн исходил из качественного сравнения случайных событий по их большей или меньшей вероятности.

Имеется иной подход, предложенный А.Н. Колмогоровым. Этот подход тесно связывает теорию вероятностей с современной метрической теорией функций, а также теорией множеств. Настоящая книга следует пути, предложенному Колмогоровым.

Мы увидим, что аксиоматическое построение основ теории вероятностей отправляется от основных свойств вероятности, подмеченных на примерах классического и статистического определений. Аксиоматическое определение вероятности, таким образом, как частные случаи включает в себя и классическое и статистическое определения и преодолевает недостаточность каждого из них. На этой базе удалось построить логически совершенное здание современной теории вероятностей и в то же время удовлетворить повышенные требования к ней современного естествознания.

Отправным пунктом аксиоматики Колмогорова является множество Ω, элементы которого называются *элементарными событиями*. Наряду с Ω рассматривается множество $\mathfrak{F}$ подмножеств элементарных событий. Множество $\mathfrak{F}$ называется *алгеброй множеств*, если выполнены следующие требования:

1) $\Omega \in \mathfrak{F}$, $\phi \in \mathfrak{F}$ (ϕ – пустое множество);

2) из того, что $A \in \mathfrak{F}$ следует, что так же $\bar{A} \in \mathfrak{F}$;

3) из того, что $A \in \mathfrak{F}$ и $B \in \mathfrak{F}$ следует, что

$A \cup B \in \mathfrak{F}$, $A \cap B \in \mathfrak{F}$

Если дополнительно к перечисленным выполняется еще следующее требование

4) из того, что $A_n \in \mathfrak{F}$ (при $n = 1, 2, \ldots$) вытекает, что

$\bigcup\limits_n A_n \in \mathfrak{F}$, $\bigcap\limits_n A_n \in \mathfrak{F}$

то множество $\mathfrak{F}$ называется *σ-алгеброй*. Элементы $\mathfrak{F}$ назывются *случайными событиями*.

Под операциями над случайными событиями понимаются операции над соответствующими множествами. В результате можно составить словарь переводов с языка теории множеств на язык теории вероятностей, приводимый нами в табл. 5

Таблица 5

Обозначения	Термины	
	теории множеств	теории вероятностей
Ω	Множество, пространство	Пространство элементарных событий, достоверное событие
ω	Элемент множества	Элементарное событие
A, B	Подмножество A, B	Случайное событие A, B
$A + B = A \cup B$	Объединение (сумма) множеств A и B	Сумма случайных событий A и B
$AB = A \cap B$	Пересечение множеств A и B	Произведение событий A и B
$\bar{A}$	Дополнение множества A	Событие, противоположное для A
$A \backslash B$	Разность множеств A и B	Разность событий A и B
ϕ	Пустое множество	Невозможное событие
$AB = A \cap B = \phi$	Множества A и B не пересекаются (не имеют общих элементов)	События A и B несовместимы
$A = B$	Множества A и B равны	События A и B равносильны
$A \subset B$	A есть подмножество B	Событие A влечет событие B

Теперь мы можем перейти к формулировке аксиом, определяющих вероятность.

Аксиома 1. *Каждому случайному событию A поставлено в соответствие неотрицательное число $\mathrm{P}(A)$, называемое его вероятностью.*

Аксиома 2. $\mathrm{P}(\Omega) = 1$.

Аксиома 3 (аксиома сложения). *Если события $A_1, A_2, \ldots$ $\ldots, A_n$ попарно несовместимы, то*

$$\mathrm{P}(A_1 + A_2 + \ldots + A_n) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \ldots + \mathrm{P}(A_n).$$

Для классического определения вероятности свойства, выраженные аксиомами 2 и 3, не нужно было постулировать, так как эти свойства вероятности были нами доказаны.

Из сформулированных аксиом мы выведем несколько важных элементарных следствий.

Прежде всего, из очевидного равенства

$$\Omega = \phi + \Omega$$

и аксиомы 3 мы заключаем, что

$$\mathrm{P}(\Omega) = \mathrm{P}(\phi) + \mathrm{P}(\Omega).$$

Таким образом,

1. Вероятность невозможного события равна нулю.
2. Для любого события A

$$\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A).\,^{*)}$$

3. Каково бы ни было случайное событие A,

$$0 \leqslant \mathrm{P}(A) \leqslant 1.$$

4. Если событие A влечет за собой событие B, то

$$\mathrm{P}(A) \leqslant \mathrm{P}(B).$$

5. Пусть A и B – два произвольных события. Поскольку в суммах $A + B = A + (B - AB)$ и $B = AB + (B - AB)$ слагаемые являются несовместимыми событиями, то в соответствии с аксиомой 3

$$\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B - AB); \quad \mathrm{P}(B) = \mathrm{P}(AB) + \mathrm{P}(B - AB).$$

Отсюда вытекает теорема сложения для произвольных событий A и B

$$\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(AB).$$

В силу неотрицательности $\mathrm{P}(AB)$ отсюда заключаем, что

$$\mathrm{P}(A + B) \leqslant \mathrm{P}(A) + \mathrm{P}(B).$$

---

*) Формулировка этого предложения имеется в трактате Я. Бернулли.

По индукции теперь выводим, что если $A_1, A_2, \ldots, A_n$ – произвольные события, то имеет место неравенство

$$\mathsf{P}\{A_1 + A_2 + \ldots + A_n\} \leqslant \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n).$$

Система аксиом Колмогорова *непротиворечива,* так как существуют реальные объекты, которые всем этим аксиомам удовлетворяют. Например, если за Ω принять произвольное конечное множество с конечным числом элементов $\Omega = \{a_1, a_2, \ldots, a_n\}$, за $\mathfrak{F}$ – совокупность всех подмножеств $\{a_{i_1}, a_{i_2}, \ldots, a_{i_s}\}$, $0 \leqslant i_1 \leqslant i_2 \leqslant \ldots \leqslant i_s \leqslant n$, $0 \leqslant s \leqslant n$, то положив $\mathsf{P}(a_1) = p_1$, $\mathsf{P}(a_2) = p_2, \ldots$ $\mathsf{P}(a_n) = p_n$, где $p_1, p_2, \ldots, p_n$ – произвольные неотрицательные числа, удовлетворяющие равенству $p_1 + p_2 + \ldots + p_n = 1$, а $\mathsf{P}(a_{i_1}, a_{i_2}, \ldots, a_{i_s}) = p_{i_1} + \ldots + p_{i_s}$, мы удовлетворим всем аксиомам Колмогорова.

Система аксиом Колмогорова *неполна:* даже для одного и того же множества Ω вероятности в множестве $\mathfrak{F}$ мы можем выбирать различными способами.

Так, в рассмотренном нами примере с игральной костью мы можем положить или

$$\mathsf{P}(E_1) = \mathsf{P}(E_2) = \ldots = \mathsf{P}(E_6) = 1/6 \tag{1}$$

или

$$\mathsf{P}(E_1) = \mathsf{P}(E_2) = \mathsf{P}(E_3) = 1/4,$$
$$\mathsf{P}(E_4) = \mathsf{P}(E_5) = \mathsf{P}(E_6) = 1/12 \tag{2}$$

и т.д.

Неполнота системы аксиом теории вероятностей не является свидетельством их неудачного выбора или недостаточной работы мысли при их создании, а вызвана существом дела: в различных задачах могут встретиться явления, при изучении которых требуется рассматривать одинаковые множества случайных событий, но с различными вероятностями. Например, могут встретиться игральные кости, из которых одна правильная (точный куб с одинаковой плотностью в каждой точке), другая неправильная. В первом случае система вероятностей будет задана системой равенств (1), а во втором, скажем, системой (2).

Дальнейшее развитие теории нуждается в дополнительном предположении, которое носит название *расширенной аксиомы сложения.* Необходимость введения новой аксиомы объясняется тем, что в теории вероятностей постоянно приходится рассматривать события, подразделяющиеся на бесконечное число частных случаев.

Р а с ш и р е н н а я   а к с и о м а   с л о ж е н и я. *Если событие А равносильно наступлению хотя бы одного из попарно несовместимых событий*

$A_1, A_2, \ldots, A_n, \ldots$, *то*

$$P(A) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots$$

Заметим, что расширенная аксиома сложения может быть заменена равносильной ей *аксиомой непрерывности.*

Аксиома непрерывности. *Если последовательность событий $B_1, B_2, \ldots, B_n, \ldots$ такова, что каждое последующее влечет за собой предыдущее и произведение всех событий B_n есть невозможное событие, то*

$$P(B_n) \to 0 \quad \text{при } n \to \infty.$$

Докажем эквивалентность только что сформулированных предложений.

1. Из расширенной аксиомы сложения следует аксиома непрерывности. Действительно, пусть события $B_1, B_2, \ldots, B_n, \ldots$ таковы, что

$$B_1 \supset B_2 \supset \ldots \supset B_n \supset \ldots$$

и при любом $n \geqslant 1$

$$\prod_{k \geqslant n} B_k = \phi. \tag{3}$$

Очевидно, что

$$B_n = \sum_{k=n}^{\infty} B_k \bar{B}_{k+1} + \prod_{k=n}^{\infty} B_k.$$

Так как события, стоящие в этой сумме, попарно несовместимы, то согласно расширенной аксиоме сложения

$$P(B_n) = \sum_{k=n}^{\infty} P(B_k \bar{B}_{k+1}) + P\left(\prod_{k=n}^{\infty} B_k\right).$$

Но в силу условия (3)

$$P\left(\prod_{k=n}^{\infty} B_k\right) = 0,$$

поэтому

$$P(B_n) = \sum_{k=n}^{\infty} P(B_k \bar{B}_{k+1}),$$

т.е. $P(B_n)$ есть остаток сходящегося ряда

$$\sum_{k=1}^{\infty} P(B_k \bar{B}_{k+1}) = P(B_1).$$

Поэтому $P(B_n) \to 0$ при $n \to \infty$.

2. Из аксиомы непрерывности следует расширенная аксиома сложения. Пусть события $A_1, A_2, \ldots, A_n, \ldots$ попарно несовместимы и

$$A = A_1 + A_2 + \ldots + A_n + \ldots$$

Положим

$$B_n = \sum_{k=n}^{\infty} A_k.$$

Ясно, что $B_{n+1} \subset B_n$. Если событие B_n наступило, то наступило какое-нибудь из событий $A_i (i \geqslant n)$ и, значит, в силу попарной несовместимости событий A_k, события $A_{i+1}, A_{i+2}, \ldots$ уже не наступили. Таким образом события $B_{i+1}, B_{i+2} \ldots$ невозможны, и, следовательно, невозможно событие $\prod_{k=n}^{\infty} B_k$. По аксиоме непрерывности $\mathsf{P}(B_n) \to 0$ при $n \to \infty$. Так как

$$A = A_1 + A_2 + \ldots + A_n + B_{n+1},$$

то по обычной аксиоме сложения

$$\mathsf{P}(A) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n) + \mathsf{P}(B_{n+1}) =$$

$$= \lim_{n \to \infty} \sum_{k=1}^{n} \mathsf{P}(A_k) = \sum_{k=1}^{\infty} \mathsf{P}(A_k).$$

Мы видим из сказанного, что аксиоматика Колмогорова позволяет строить теорию вероятностей как часть теории меры, а вероятность рассматривать как неотрицательную нормированную аддитивную функцию множества.

Вероятностным пространством принято называть тройку символов $(\Omega, \mathfrak{F}, \mathsf{P})$, где Ω – множество элементарных событий, $\mathfrak{F}$ – σ-алгебра подмножеств Ω, называемых случайными событиями и $\mathsf{P}(A)$ – вероятность, определенная на σ-алгебра $\mathfrak{F}$

§ 7. Условная вероятность и простейшие основные формулы

Мы уже говорили, что в основе определения вероятности события лежит некоторая совокупность условий $\mathfrak{S}$. Если никаких ограничений, кроме условий $\mathfrak{S}$, при вычислении вероятности $\mathsf{P}(A)$ не налагается, то такие вероятности называются *безусловными.*

Однако в ряде случаев приходится рассматривать вероятности событий при дополнительном условии, что произошло некоторое событие B. Такие вероятности мы будем называть *условными* и обозначать символом $\mathsf{P}(A|B)$; это означает вероятность события A при условии, что событие B произошло. Строго говоря, безусловные вероятности также являются условными, так как исходным моментом построенной теории было предположение о существовании некоторого неизменного комплекса условий $\mathfrak{S}$

Пример 1. Брошены две игральные кости. Чему равна вероятность того, что сумма выпавших на них очков равна 8 (событие A), если известно, что эта сумма есть четное число (событие B)?

Таблица 6

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Все возможные случаи, которые могут представиться при бросании двух костей, мы запишем в табл. 6, каждая клетка которой содержит запись возможного события: на первом месте в скобках указывается число очков, выпавших на первой кости, на втором месте – число очков, выпавших на второй кости.

Общее число возможных случаев – 36, благоприятствующих событию A – 5. Таким образом, безусловная вероятность $\mathsf{P}(A) = 5/36$. Если событие B произошло, то осуществилась одна из 18 (а не 36) возможностей и, следовательно, условная вероятность равна $\mathsf{P}(A|B) = 5/18$.

Пример 2. Из колоды карт последовательно вынуты две карты. Найти а) безусловную вероятность того, что вторая карта окажется тузом (неизвестно, какая карта была вынута вначале) и б) условную вероятность, что вторая карта будет тузом, если первоначально был вынут туз.

Обозначим через A событие, состоящее в появлении туза на втором месте, а через B – событие, состоящее в появлении туза на первом месте. Ясно, что имеет место равенство

$$A = AB + A\bar{B}.$$

В силу несовместимости событий AB и $A\bar{B}$ имеем:

$$\mathsf{P}(A) = \mathsf{P}(AB) + \mathsf{P}(A\bar{B}).$$

При вынимании двух карт из колоды в 36 карт могут произойти $36 \cdot 35$ (учитывая порядок!) случаев. Из них благоприятствующих событию AB – $4 \cdot 3$ случаев, а событию $A\bar{B}$ – $32 \cdot 4$ случая. Таким образом,

$$\mathsf{P}(A) = \frac{4 \cdot 3}{36 \cdot 35} + \frac{32 \cdot 4}{36 \cdot 35} = \frac{1}{9} .$$

Если первая карта есть туз, то в колоде осталось 35 карт и среди них только три туза. Следовательно,

$$\mathsf{P}(A \mid B) = 3/35.$$

Общее решение задачи нахождения условной вероятности для классического определения вероятности не представляет труда. В самом деле, пусть из n единственно возможных, несовместимых и равновероятных событий $A_1, A_2, \ldots, A_n$

событию A благоприятствует m событий,

" B " k "

" AB " r "

(понятно, что $r \leqslant k, r \leqslant m$). Если событие B произошло, то это означает, что наступило одно из событий A_j, благоприятствующих B. При этом условии событию A благоприятствуют r и только r событий A_j, благоприятствующих AB. Таким образом,

$$\mathsf{P}(A \mid B) = \frac{r}{k} = \frac{r/n}{k/n} = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}. \tag{1}$$

Понятие независимости событий играет значительную роль в теории вероятностей и в ее приложениях. В частности, большая часть результатов, изложенных в настоящей книге, получена в предположении независимости тех или иных рассматриваемых событий.

В практических вопросах для определения независимости данных событий редко обращаются к проверке выполнения для них равенств (3) и (4). Обычно для этого пользуются интуитивными соображениями, основанными на опыте.

Так, например, ясно, что выпадение герба на одной монете не изменяет вероятности появления герба (решки) на другой монете, если только эти монеты во время бросания не связаны между собой (например, жестко не скреплены). Точно так же рождение мальчика у одной матери не изменяет вероятности появления мальчика (девочки) у другой матери. Это события независимые.

Для независимых событий теорема умножения принимает особенно простой вид, а именно, *если события A и B независимы*,

то

$$\mathsf{P}(AB) = \mathsf{P}(A) \cdot \mathsf{P}(B).$$

Точно так же, если $\mathsf{P}(A) > 0$, то

$$\mathsf{P}(B \mid A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)}. \qquad (1')$$

Понятно, что если B (соответственно, A) есть невозможное событие, то равенство (1) (соответственно (1$'$)) теряет смысл.

Заметим, что рассуждения, проведенные нами в примерах 1 и 2, не являются доказательствами, а представляют только мотивировки определений, данных равенствами (1) и (1$'$).

При $\mathsf{P}(A)\mathsf{P}(B) > 0$ каждое из равенств (1), (1$'$) эквивалентно так называемой т е о р е м е у м н о ж е н и я, согласно которой

$$\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B \mid A) = \mathsf{P}(B)\,\mathsf{P}(A \mid B), \qquad (2)$$

т.е. *вероятность произведения двух событий равна произведению вероятности одного из этих событий на условную вероятность другого при условии, что первое произошло.*

Теорема умножения применима и в том случае, когда одно из событий A или B есть невозможное событие, так как в этом случае вместе с $\mathsf{P}(A) = 0$ имеют место равенства $\mathsf{P}(A \mid B) = 0$ и $\mathsf{P}(AB) = 0$.

Говорят, что событие A *независимо* от события B, если имеет место равенство

$$\mathsf{P}(A \mid B) = \mathsf{P}(A). \qquad (3)$$

т.е. если наступление события B не изменяет вероятности события A*)

Если событие A независимо от B, то в силу (2) имеет место равенство

$$\mathsf{P}(A)\,\mathsf{P}(B \mid A) = \mathsf{P}(B)\,\mathsf{P}(A).$$

Отсюда находим, что

$$\mathsf{P}(B \mid A) = \mathsf{P}(B), \qquad (4)$$

т.е. что событие B также независимо от A. Таким образом, при сделанном предположении свойство независимости событий взаимно.

Если независимость событий A и B определить посредством равенства

$$\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B),$$

то это определение верно всегда, в том числе и тогда, когда $\mathsf{P}(A) = 0$ или $\mathsf{P}(B) = 0$.

---

*) Понятия условной вероятности и независимости, а также формулировка теоремы умножения даны А.Муавром в 1718 г.

Мы обобщим теперь понятие независимости двух событий на совокупность нескольких событий.

События $B_1, B_2, \ldots, B_s$ называются *независимыми в совокупности*, если для любого события B_p из их числа и произвольных $B_{i_1}, B_{i_2}, \ldots$, $\ldots, B_{i_r} (i_n \neq p)$ из их же числа события B_p и $B_{i_1} B_{i_2} \ldots B_{i_r}$ взаимно независимы.

В силу предыдущего это определение эквивалентно следующему: при любых $1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant s$ и $r (1 \leqslant r \leqslant s)$

$$\mathbf{P}(B_{i_1} B_{i_2} \ldots B_{i_r}) = \mathbf{P}(B_{i_1}) \mathbf{P}(B_{i_2}) \ldots \mathbf{P}(B_{i_r}).$$

Заметим, что для независимости в совокупности нескольких событий недостаточно их попарной независимости. В этом можно убедиться на следующем простом примере. Представим себе, что грани тетраэдра окрашены: 1-я в – красный цвет (A), 2-я – в зеленый (B), 3-я – в синий (C) и 4-я – во все эти три цвета (ABC). Легко видеть, что вероятность грани, на которую упадет тетраэдр при бросании, в своей окраске иметь красный цвет равна 1/2: граней четыре и две из них имеют в окраске красный цвет. Таким образом,

$$\mathbf{P}(A) = 1/2.$$

Точно так же можно подсчитать, что

$$\mathbf{P}(B) = \mathbf{P}(C) = \mathbf{P}(A \mid B) = \mathbf{P}(B \mid C) = \mathbf{P}(C \mid A) =$$

$$= \mathbf{P}(B \mid A) = \mathbf{P}(C \mid B) = \mathbf{P}(A \mid C) = 1/2,$$

события A, B, C, таким образом, попарно независимы.

Однако если нам известно, что осуществились события B и C, то заведомо осуществилось и событие A, т.е.

$$\mathbf{P}(A \mid BC) = 1.$$

Таким образом, события A, B, C в совокупности зависимы.

Формула ($1'$), которая в случае классического определения была нами выведена из определения условной вероятности, в случае аксиоматического определения вероятности будет взята нами в качестве определения. Таким образом, в общем случае при $\mathbf{P}(A) > 0$ по определению

$$\mathbf{P}(B \mid A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)}.$$

(В случае $\mathbf{P}(A) = 0$ условная вероятность $\mathbf{P}(B \mid A)$ остается неопределенной.) Это позволяет нам перенести автоматически на общее понятие вероятности все определения и результаты настоящего параграфа.

Предположим теперь, что событие B может осуществиться с одним и только с одним из n несовместимых событий $A_1, A_2, \ldots, A_n$. Иными словами, положим, что

$$B = \sum_{i=1}^{n} BA_i, \tag{5}$$

где события BA_i и BA_j с разными индексами i и j несовместимы. По теореме сложения вероятностей имеем:

$$\mathsf{P}(B) = \sum_{i=1}^{n} \mathsf{P}(BA_i).$$

Использовав теорему умножения, находим, что

$$\mathsf{P}(B) = \sum_{i=1}^{n} \mathsf{P}(A_i)\,\mathsf{P}(B \mid A_i).$$

Это равенство носит название *формулы полной вероятности**) и играет основную роль во всей дальнейшей теории.

В качестве иллюстрации рассмотрим два примера.

Пример 1. Имеется пять урн:

2 урны состава A_1 – по два белых шара и одному черному,

1 урна состава A_2 – по 10 черных шаров,

2 урны состава A_3 – по 3 белых шара и одному черному.

Наудачу выбирается урна и из нее наудачу вынимается шар.

Чему равна вероятность, что вынутый шар белый (событие B)?

Решение. Так как вынутый шар может быть только из урны 1-го, 2-го или 3-го состава, то

$$B = A_1B + A_2B + A_3B.$$

По формуле полной вероятности находим, что

$$\mathsf{P}(B) = \mathsf{P}(A_1)\,\mathsf{P}(B \mid A_1) + \mathsf{P}(A_2)\,\mathsf{P}(B \mid A_2) + \mathsf{P}(A_3)\,\mathsf{P}(B \mid A_3).$$

Но ясно, что

$$\mathsf{P}(A_1) = 2/5, \quad \mathsf{P}(A_2) = 1/5, \quad \mathsf{P}(A_3) = 2/5,$$

$$\mathsf{P}(B \mid A_1) = 2/3, \quad \mathsf{P}(B \mid A_2) = 0, \quad \mathsf{P}(B \mid A_3) = 3/4.$$

Таким образом,

$$\mathsf{P}(B) = 2/5 \cdot 2/3 + 1/5 \cdot 0 + 2/5 \cdot 3/4 = 17/30.$$

Пример 2. Известно, что вероятность поступления k вызовов на телефонную станцию за промежуток времени t равна $P_t(k)$ $(k = 0, 1, 2, \ldots)$.

---

*) Формула полной вероятности широко использовалась математиками начала XVIII века, но впервые была сформулирована как основное предложение теории вероятностей П. Лапласом лишь в конце XVIII века.

Считая, что появления какого-либо числа вызовов за два соседних промежутка времени являются событиями независимыми найти вероятность поступления s вызовов за промежуток времени длительности $2t$.

Решение. Обозначим через $A^k_{a,\ a+\tau}$ событие, состоящее в поступлении k вызовов за время от a до $a+\tau$. Очевидно, что мы имеем следующее равенство:

$$A^s_{0,\ 2t} = A^0_{0,t}A^s_{t,2t} + \ldots + A^s_{0,t}A^0_{t,2t},$$

которое означает, что событие $A^s_{0,2t}$ можно рассматривать как сумму $s+1$ несовместимых событий, состоящих в том, что за первый промежуток времени длительности t поступает i вызовов, а за следующий промежуток той же продолжительности поступает $s-i$ вызовов $(i = 0, 1, 2,..., s)$. По теореме сложения вероятностей

$$\mathsf{P}(A^s_{0,2t}) = \sum_{i=0}^{s} \mathsf{P}(A^i_{0,t}A^{s-i}_{t,2t}).$$

По теореме умножения вероятностей для независимых событий

$$\mathsf{P}(A^i_{0,t}A^{s-i}_{t,2}) = \mathsf{P}(A^i_{0,t})\,\mathsf{P}(A^{s-i}_{t,2t}).$$

Таким образом, если положить

$$P_{2t}(s) = \mathsf{P}(A^s_{0.2t}),$$

то

$$P_{2t}(s) = \sum_{i=0}^{s} P_t(i)\cdot P_t(s-i). \tag{6}$$

Впоследствии мы увидим, что при некоторых весьма общих условиях $(k = 0, 1, 2\ \ldots)$

$$P_t(k) = \frac{(at)^k}{k!}\,e^{-at}, \tag{7}$$

где a – некоторая константа.

Из формулы (6) мы находим, что

$$P_{2t}(s) = \sum_{i=0}^{s}\frac{(at)^s e^{-2at}}{i!(s-i)!} = (at)^s\,e^{-2at}\sum_{i=0}^{s}\frac{1}{i!(s-i)!}.$$

Но

$$\sum_{i=0}^{s}\frac{1}{i!(s-i)!} = \frac{1}{s!}\sum_{i=0}^{s}\frac{s!}{i!(s-i)!} = \frac{1}{s!}\,(1+1)^s = \frac{2^s}{s!}.$$

Поэтому

$$P_{2t}(s) = \frac{(2at)^s e^{-2at}}{s!} \qquad (s = 0, 1, 2, \ldots).$$

Таким образом, если для промежутка времени длительности t имеет место формула (7), то для промежутков времени, в два раза больших, и, как легко убедиться, для любых кратных t промежутков времени характер формулы для вероятности сохраняется.

Мы в состоянии теперь вывести важные формулы Байеса или, как иногда говорят, вероятности гипотез. Пусть попрежнему имеет место равенство (5). Требуется найти вероятность события A_i, если известно, что B произошло. Слогласно теореме умножения имеем:

$$\mathsf{P}(A_iB) = \mathsf{P}(B)\,\mathsf{P}(A_i|B) = \mathsf{P}(A_i)\,\mathsf{P}(B|A_i).$$

Отсюда

$$\mathsf{P}(A_i\,|\,B) = \frac{\mathsf{P}(A_i)\,\mathsf{P}(B\,|\,A_i)}{\mathsf{P}(B)},$$

используя формулу полной вероятности, находим, что

$$\mathsf{P}(A_i\,|\,B) = \frac{\mathsf{P}(A_i)\,\mathsf{P}(B\,|\,A_i)}{\sum\limits_{j=1}^{n} \mathsf{P}(A_j)\,\mathsf{P}(B\,|\,A_j)}.$$

Полученные нами формулы носят название *формул Байеса**). Общая схема применения этих формул к решению практических задач такова. Пусть событие B может протекать в различных условиях, относительно характера которых может быть сделано n гипотез: $A_1, A_2, \ldots, A_n$. По тем или иным причинам нам известны вероятности $\mathsf{P}(A_i)$ этих гипотез до испытания. Известно также, что гипотеза A_i сообщает событию B вероятность $\mathsf{P}(B|A_i)$. Произведен опыт, в котором событие B наступило. Это должно вызвать переоценку вероятностей гипотез A_i – формулы Байеса количественно решают этот вопрос.

В артиллерийской практике производится так называемая пристрелка, имеющая своей целью уточнить наши знания относительно условий стрельбы (например, правильность прицела). В теории пристрелки широко используется формула Байеса. Мы ограничимся приведением чисто схематического примера исключительно ради иллюстрации характера задач, решаемых этой формулой.

---

*) Т. Байес приведенных формул не выводил, он ограничился записью формулы (1) настоящего параграфа. Приведенные формулы были выписаны лишь П. Лапласом в конце XVIII века.

Пример 1. Имеются пять урн следующего состава:
2 урны (состава A_1) по 2 белых и 3 черных шара,
2 урны (состава A_2) по 1 белому и 4 черных шара,
1 урна (состав A_3) – 4 белых и 1 черный шар.

Из одной наудачу выбранной урны взят шар. Он оказался белым (событие B). Чему равна после опыта вероятность (апостериорная вероятность) того, что шар вынут из урны третьего состава?

Решение. Согласно предположению

$$\mathsf{P}(A_1) = 2/5, \quad \mathsf{P}(A_2) = 2/5, \quad \mathsf{P}(A_3) = 1/5;$$

$$\mathsf{P}(B \mid A_1) = 2/5, \quad \mathsf{P}(B \mid A_2) = 1/5, \quad \mathsf{P}(B \mid A_3) = 4/5.$$

Согласно формуле Байеса имеем:

$$\mathsf{P}(A_3 \mid B) = \frac{\mathsf{P}(A_3)\,\mathsf{P}(B \mid A_3)}{\mathsf{P}(A_1)\,\mathsf{P}(B \mid A_1) + \mathsf{P}(A_2)\,\mathsf{P}(B \mid A_2) + \mathsf{P}(A_3)\,\mathsf{P}(B \mid A_3)} =$$

$$= \frac{\frac{1}{5}\cdot\frac{4}{5}}{\frac{2}{5}\cdot\frac{2}{5} + \frac{1}{5}\cdot\frac{2}{5} + \frac{1}{5}\cdot\frac{4}{5}} = \frac{4}{10} = \frac{2}{5}.$$

Точно так же находим:

$$\mathsf{P}(A_1 \mid B) = 2/5, \quad \mathsf{P}(A_2 \mid B) = 1/5.$$

§ 8. Примеры.

Мы приведем несколько более сложных примеров на использование изложенной теории.

Пример 1*). Два игрока A и B продолжают некоторую игру до полного разорения одного из них. Капитал первого равняется a руб., капитал второго – b руб. Вероятность выигрыша каждой партии для игрока A равна p, а для игрока B равна q; $p + q = 1$ (ничьи отсутствуют). В каждой партии выигрыш одного игрока (и, значит, проигрыш другого) равняется

---

*) Мы сохраняем для этой задачи "о разорении игрока" ее классическую формулировку, но возможны и иные формулировки, например: материальная частица находится на прямой в точке O и каждую секунду подвергается случайному толчку, в результате которого передвигается на 1 *см* вправо с вероятностью p или на 1 *см* влево с вероятностью $q = 1 - p$. Чему равна вероятность того, что материальная частица окажется правее точки с координатой b ($b > 0$), прежде чем она попадет в положение, расположенное левее точки с координатой a ($a < 0$, a и b – целые числа)?

Задача о разорении игрока была предложена и впервые изучена Х. Гюйгенсом. Мы предполагаем, что вероятность события "разорение игрока" существует.

1 рублю. Найти вероятность разорения каждого из игроков (результаты отдельных партий предполагаются независимыми).

Р е ш е н и е. Обозначим через p_n вероятность разорения игрока A, когда он имеет n руб. Очевидно, что искомая вероятность есть p_a и что

$$p_{a+b} = 0, \quad p_0 = 1, \tag{1}$$

поскольку в первом случае игрок A уже сосредоточил в своих руках весь капитал, а во втором он уже ничего не имеет.

Если игрок A имел n руб. перед некоторой партией, то его разорение может осуществиться двумя различными способами: или он очередную партию выиграет, а всю игру проиграет, или он проиграет и партию и игру. По формуле полной вероятности поэтому

$$p_n = p \cdot p_{n+1} + q \cdot p_{n-1}.$$

Относительно p_n мы получили уравнение в конечных разностях; легко видеть, что его можно записать в следующем виде:

$$q(p_n - p_{n-1}) = p(p_{n+1} - p_n). \tag{2}$$

Рассмотрим сначала решение этого уравнения при $p = q = 1/2$. При этом допущении

$$p_{n+1} - p_n = p_n - p_{n-1} = \ldots = p_1 - p_0 = c,$$

где c — постоянная. Отсюда находим, что

$$p_n = p_0 + nc.$$

Поскольку $p_0 = 1$ и $p_{a+b} = 0$, то

$$p_n = 1 - \frac{n}{a+b}.$$

Таким образом, вероятность разорения игрока A равняется

$$p_a = 1 - \frac{a}{a+b} = \frac{b}{a+b}.$$

Подобным же путем найдем, что в случае $p = 1/2$ вероятность разорения игрока B равна

$$q_b = \frac{a}{a+b}.$$

В общем случае при $p \neq q$ из (2) находим, что

$$q^n \prod_{k=1}^{n} (p_k - p_{k-1}) = p^n \prod_{k=1}^{n} (p_{k+1} - p_k).$$

После сокращений и учета соотношений (1) находим, что

$$p_{n+1} - p_n = (q/p)^n (p_1 - 1).$$

Рассмотрим разность $p_{a+b} - p_n$; очевидно, что

$$p_{a+b} - p_n - \sum_{k=n}^{a+b-1} (p_{k+1} - p_k) = \sum_{k=n}^{a+b-1} (q/p)^k (p_1 - 1) =$$

$$= (p_1 - 1) \frac{(q/p)^n - (q/p)^{a+b}}{1 - q/p}.$$

Поскольку $p_{a+b} = 0$, то

$$p_n = (1 - p_1) \frac{(q/p)^n - (q/p)^{a+b}}{1 - q/p},$$

а так как $p_0 = 1$, то

$$1 = (1 - p_1) \frac{(q/p)^0 - (q/p)^{a+b}}{1 - q/p}.$$

Исключив из двух последних равенств величину p_1, находим, что

$$p_n = \frac{(q/p)^{a+b} - (q/p)^n}{(q/p)^{a+b} - 1}.$$

Отсюда вероятность разорения игрока A

$$p_a = \frac{q^{a+b} - q^a q^b}{q^{a+b} - p^{a+b}} = \frac{1 - (p/q)^b}{1 - (p/q)^{a+b}}.$$

Подобным же путем находим, что вероятность разорения игрока B при $p \neq q$ равна

$$q_b = \frac{1 - (q/p)^a}{1 - (q/p)^{a+b}}.$$

Из этих формул мы можем сделать следующие выводы: если капитал одного из игроков, например B, несравненно больше капитала игрока A, так что практически b можно считать бесконечно большим по сравнению с a, а игроки одинаково искусны, то разорение B практически невозможно. Вывод будет совсем иной, если A играет лучше, чем B, и, значит, $p > q$. Считая $b \sim \infty$, находим, что

$$q_b \sim 1 - (q/p)^a$$

и

$$p_a \sim (q/p)^a.$$

Отсюда мы делаем тот вывод, что умелый игрок даже с малым капиталом может иметь меньше шансов на разорение, чем игрок с большим капиталом, но менее умелый.

К задаче о разорении игрока сводится решение некоторых задач физики и техники.

Пример 2. Найти вероятность того, что станок, работающий в момент t_0, не остановится до момента $t_0 + t$, если известно, что 1) эта вероятность зависит только от величины промежутка времени $(t_0, t_0 + t)$, 2) вероятность того, что станок остановится за промежуток времени Δt пропорциональна Δt с точностью до бесконечно малых высших порядков*) относительно Δt.

Решение. Обозначим вероятность через $p(t)$. Вероятность того, что станок остановится за промежуток времени Δt равна

$$1 - p(\Delta t) = a\Delta t + o(\Delta t),$$

где a – некоторая постоянная.

Определим вероятность того, что станок, работавший в момент t_0, не остановится до момента $t_0 + t + \Delta t$. Для осуществления этого события необходимо, чтобы станок не остановился за периоды времени длины t и Δt; в силу теоремы умножения, таким образом,

$$p(t + \Delta t) = p(t) \cdot p(\Delta t) = p(t)(1 - a\Delta t - o(\Delta t)).$$

Отсюда

$$\frac{p(t + \Delta t) - p(t)}{\Delta t} = -ap(t) - o(1). \tag{3}$$

Перейдем теперь к пределу, положив $\Delta t \to 0$; из того, что существует предел правой части равенства (3), вытекает, что существует также предел левой части. В результате находим, что

$$\frac{dp(t)}{dt} = -ap(t).$$

Решение этого уравнения есть функция

$$p(t) = Ce^{-at},$$

где C – постоянная. Эта постоянная находится из того очевидного условия, что $p(0) = 1$. Таким образом,

$$p(t) = e^{-at}.$$

---

*) В дальнейшем для записи того факта, что некоторая величина α бесконечно мала сравнительно с величиной β, мы будем пользоваться записью $\alpha = o(\beta)$. Если же отношение α/β ограничено по абсолютной величине, то мы будем писать $\alpha = O(\beta)$.

Первое условие задачи налагает на режим работы станка большие ограничения, однако существуют производства, где оно выполняется с большой степенью точности. В качестве примера можно привести работу автоматического ткацкого станка. Заметим, что к рассмотренной задаче сводится много других вопросов, например, вопрос о распределении вероятностей длины свободного пути молекулы в кинетической теории газов.

Пример 3. При составлении таблиц смертности часто исходят из таких допущений: 1) вероятность того, что некоторое лицо умрет в возрасте от t до $t+\Delta t$ равна

$$p(t, t+\Delta t) = a(t)\Delta t + o(\Delta t),$$

где $a(t)$ — неотрицательная, непрерывная функция.

2) считается, что смерть данного лица (или его выживание) за рассматриваемый промежуток (t_1, t_2) возраста не зависит от того, что было до момента t_1, 3) вероятность смерти в момент рождения равна нулю.

Исходя из высказанных предположений, найти вероятность смерти лица A до того, как оно достигнет возраста t.

Решение. Обозначим через $\pi(t)$ вероятность того, что лицо A доживет до возраста t, и вычислим $\pi(t+\Delta t)$. Очевидно, что из допущений, принятых в задаче, вытекает равенство

$$\pi(t+\Delta t) = \pi(t)\pi(t+\Delta t; t),$$

где $\pi(t+\Delta t;\ t)$ обозначает вероятность дожить до возраста $t+\Delta t$, если лицо A уже дожило до возраста t. В соответствии с первым и вторым допущениями

$$\pi(t+\Delta t; t) = 1 - p(t, t+\Delta t) = 1 - a(t)\Delta t - o(\Delta t);$$

поэтому

$$\pi(t+\Delta t) = \pi(t)[1 - a(t)\Delta t - o(\Delta t)].$$

Отсюда находим, что $\pi(t)$ удовлетворяет следующему дифференциальному уравнению:

$$\frac{d\pi(t)}{dt} = -a(t)\pi(t).$$

Решением этого уравнения при учете третьего условия задачи будет функция

$$\pi(t) = e^{-\int_0^t a(z)dz}$$

Вероятность умереть прежде, чем будет достигнут возраст t, таким образом, равна

$$1-\pi(t)=1-e^{-\int\limits_0^t a(z)dz}$$

При составлении таблиц смертности для взрослого населения нередко пользуются формулой Макегама, согласно которой

$$a(t)=\alpha+\beta e^{\gamma t},$$

постоянные α, β, γ – положительны*). При выводе этой формулы исходили из допущения, что взрослый человек может умереть от причин, не зависящих от возраста, и причин, зависящих от возраста, причем вероятность смерти растет с увеличением возраста в геометрической прогрессии. При таком дополнительном предположении

$$\pi(t)=e^{-\alpha t-\frac{\beta}{\gamma}(e^{\gamma t}-1)}.$$

Пример 4. В современной ядерной физике для измерения интенсивности источника частиц используются счетчики Гейгера. Частица, попавшая в счетчик, вызывает в нем разряд, длящийся время τ, в протяжение которого счетчик не регистрирует частицы, попадающие в счетчик. Найти вероятность того, что счетчик сосчитает все частицы, попавшие в него за время t, если выполняются следующие условия: 1) вероятность того, что за промежуток времени t в счетчик попадут k частиц, не зависит от того, сколько частиц попало в счетчик до начала этого промежутка; 2) вероятность того, что за промежуток времени от t_0 до t_0+t в счетчик попадет k частиц, задается формулой **)

$$p_k(t_0,\, t_0+t)=\frac{(at)^k e^{-at}}{k!},$$

где a – положительная постоянная; 3) τ – постоянная величина.

Решение. Обозначим через $A(t)$ – событие, состоящее в том, что все попавшие за время t в счетчик частицы были сосчитаны; через $B_k(t)$ – событие, состоящее в том, что за время t в счетчик попало k частиц.

---

*) Их значение определяется условиями, в которых находится группа лиц, подлежащих изучению, и прежде всего социальными условиями.

**) Позднее мы выясним, почему в этом примере и в примере 2 предыдущего параграфа мы считали, что

$$p_k=\frac{(at)^k e^{-at}}{k!}.$$

В силу первого условия задачи при $t \geqslant \tau$

$$\mathsf{P}\{A(t+\Delta t)\} = \mathsf{P}\{A(t)\}\mathsf{P}\{B_0(\Delta t)\} +$$
$$+ \mathsf{P}\{A(t-\tau)\}\mathsf{P}\{B_0(\tau)\}\mathsf{P}\{B_1(\Delta t)\} + o(\Delta t),$$

а при $0 \leqslant t \leqslant \tau$

$$\mathsf{P}\{A(t+\Delta t)\} = \mathsf{P}\{A(t)\}\mathsf{P}\{B_0(\Delta t)\} + \mathsf{P}\{B_0(t)\} + \mathsf{P}\{B_1(\Delta t)\} + o(\Delta t).$$

Обозначим для краткости записи $\pi(t) = \mathsf{P}\{A(t)\}$; тогда на основании второго и третьего условий задачи при $0 \leqslant t \leqslant \tau$

$$\pi(t+\Delta t) = \pi(t)e^{-a\Delta t} + e^{-a\Delta t}a\Delta t e^{-at} + o(\Delta t)$$

и при $t \geqslant \tau$

$$\pi(t+\Delta t) = \pi(t)e^{-a\Delta t} + \pi(t-\tau)e^{-a\Delta t}a\Delta t e^{-a\tau} + o(\Delta t).$$

Путем перехода к пределу при $\Delta t \to 0$ находим, что при $0 \leqslant t \leqslant \tau$ имеет место равенство

$$\frac{d\pi(t)}{dt} = -a\pi(t) + ae^{-at}, \tag{4}$$

а при $t \geqslant \tau$ – равенство

$$\frac{d\pi(t)}{dt} = -a[\pi(t) - \pi(t-\tau)e^{-a\tau}]. \tag{5}$$

Из уравнений (4) находим, что при $0 \leqslant t \leqslant \tau$

$$\pi(t) = e^{-at}(c + at).$$

Из условия

$$\pi(0) = 1$$

определяем постоянную c. Окончательно при $0 \leqslant t \leqslant \tau$

$$\pi(t) = e^{-at}(1 + at). \tag{6}$$

При $\tau \leqslant t \leqslant 2\tau$ вероятность $\pi(t)$ определяется из уравнения

$$\frac{d\pi(t)}{dt} = -a[\pi(t) - \pi(t-\tau)e^{-a\tau}] =$$
$$= -a[\pi(t) - e^{-a(t-\tau)}(1 + a(t-\tau))e^{-a\tau}] =$$
$$= -a[\pi(t) - e^{-at}(1 + a(t-\tau))].$$

Решение этого уравнения дает нам:

$$\pi(t) = e^{-at}\left(c_1 + at + \frac{a^2(t-\tau)^2}{2!}\right).$$

Постоянное c_1 может быть найдено из того, что согласно (6)

$$\pi(\tau) = e^{-a\tau}(1 + a\tau).$$

Таким образом, $c_1 = 1$ и для $\tau \leqslant t \leqslant 2\tau$

$$\pi(t) = e^{-at}\left[1 + at + \frac{a^2(t-\tau)^2}{2!}\right].$$

Методом полной индукции можно доказать, что для $(n-1)\tau \leqslant t \leqslant n\tau$ имеет место равенство

$$\pi(t) = e^{-at}\sum_{k=0}^{n}\frac{a^k[t-(k-1)\tau]^k}{k!}.$$

Упражнения

A, B, C – случайные события.

1. Каков смысл равенств

а) $ABC = A$;

б) $A + B + C = A$?

2. Упростить выражения

а) $(A + B)(B + C)$;

б) $(A + B)(A + \bar{B})$;

в) $(A + B)(A + \bar{B})(\bar{A} + B)$.

3. Доказать равенства

а) $\overline{\bar{A}\bar{B}} = A + B$;

б) $\overline{\bar{A} + \bar{B}} = AB$;

в) $\overline{A_1 + A_2 + \ldots + A_n} = \bar{A}_1\bar{A}_2\ldots\bar{A}_n$;

г) $\overline{A_1A_2\ldots A_n} = \bar{A}_1 + \bar{A}_2 + \ldots + \bar{A}_n$.

4. Четырехтомное сочинение расположено на полке в случайном порядке. Чему равна вероятность того, что тома стоят в должном порядке справа налево или слева направо?

5. Числа 1, 2, 3, 4, 5 написаны на пяти карточках. Наудачу последовательно вынимаются три карточки, и вынутые таким образом цифры ставятся слева направо. Чему равна вероятность того, что полученное таким образом трехзначное число окажется четным?

6. В партии, состоящей из N изделий, имеются M бракованных. Неудачу выбираются n изделий из этой партии $(n < N)$. Чему равна вероятность того, что среди них окажутся m бракованных $(m \leqslant M)$?

7. Технический контроль проверяет изделия в партии, состоящей из m изделий первого сорта и n изделий второго сорта. Проверка первых b изделий, выбранных из партии наудачу, показала, что все они второго сорта $(b < m)$. Чему равна вероятность того, что среди следующих двух наудачу выбранных из среды непроверенных изделий по меньшей мере одно окажется также второго сорта?

8. Пользуясь теоретико-вероятностными соображениями, показать тождество

$$1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \ldots + \frac{(A-a)\ldots 2\cdot 1}{(A-1)\ldots(a+1)a} = \frac{A}{a}.$$

Указание. Из урны, содержащей A шаров и среди них a белых, наудачу вынимаются шары без возвращения. Найти вероятность того, что рано или поздно натолкнутся на белый шар.

9. Из ящика, содержащего m белых и n черных шаров $(m > n)$, вынимают наудачу один шар за другим. Чему равна вероятность того, что наступит момент, когда число вынутых черных шаров будет равно числу вынутых белых?

10. Некто написал n адресатам письма, в каждый конверт вложил по одному письму и затем наудачу написал на каждом конверте один из n адресов. Чему равна вероятность того, что хотя бы одно письмо попало по назначению?

11. В урне имеется n билетов с номерами от 1 до n. Билеты вынимаются наудачу по одному (без возвращения). Чему равна вероятность того, что хотя бы при одном вынимании номер вынутого билета совпадает с номером произведенного испытания?

12. Из урны, содержащей n белых и n черных шаров, наудачу вынимается четное число шаров (все различные способы вынуть четное число шаров, независимо от их числа, считаются равновероятными). Найти вероятность того, что среди вынутых шаров окажется одинаковое число черных и белых.

13. Задача кавалера де Мере. Что вероятнее: при бросании четырех игральных костей хотя бы на одной получить 1 или при 24 бросаниях двух костей хотя бы раз получить две единицы?

14. На отрезок $(0, a)$ наудачу брошены три точки. Найти вероятность того, что из отрезков, равных расстояниям от точки 0 до точек падения, можно составить треугольник.

15. Стержень длины l разломан в двух наудачу выбранных точках. Чему равна вероятность того, что из полученных отрезков можно составить треугольник?

16. На отрезок AB длины a наудачу бросается точка. На отрезок BC длины b также наудачу бросается точка. Чему равна вероятность того, что из отрезков: 1) от точки A до первой брошенной точки, 2) между двумя брошенными точками, 3) от второй брошенной точки до точки C можно составить треугольник?

17. В сфере радиуса R случайно и независимо друг от друга разбросано N точек.

а) Чему равна вероятность того, что расстояние от центра до ближайшей точки будет не менее r?

б) К чему стремится вероятность, найденная в вопросе а), если $R \to \infty$ и $\dfrac{N}{R^3} \to \dfrac{4}{3}\pi\lambda$?

Примечание. Задача заимствована из звездной астрономии: в окрестности Солнца $\lambda \approx 0{,}0063$, если R измерено в парсеках.

18. События $A_1, A_2, \ldots, A_n$ независимы; $\mathsf{P}(A_k) = p_k$. Найти вероятность

а) появления хотя бы одного из этих событий,

б) непоявления всех этих событий,

в) появления точно одного (безразлично какого) события.

19. Доказать, что если события A и B несовместимы, $\mathsf{P}(A) > 0$ и $\mathsf{P}(B) > 0$, то события A и B зависимы.

20. Пусть $A_1, A_2, \ldots, A_n$ – случайные события. Доказать формулу

$$\mathsf{P}\left\{\sum_{k=1}^{n} A_k\right\} = \sum_{i=1}^{n} \mathsf{P}(A_j) - \sum_{1 \leqslant i < j \leqslant n} \mathsf{P}(A_i A_j) + \sum_{1 \leqslant i < j < k \leqslant n} \mathsf{P}(A_i A_j A_k) \pm \ldots \pm \mathsf{P}(A_1 A_2 \ldots A_n).$$

Посредством этой формулы решить задачи 10 и 11.

21. Вероятность того, что молекула, испытавшая в момент $t = 0$ столкновение с другой молекулой и не имевшая других столкновений до момента t, испытывает столкновение в промежуток времени между t и $t + \Delta t$, равна $\lambda \Delta t + o(\Delta t)$. Найти вероятность того, что время свободного пробега (т.е. время между двумя соседними столкновениями) будет больше t.

22. Считая, что при размножении бактерий делением (на две бактерии) вероятность бактерии разделиться за промежуток времени Δt равна $a\Delta t + o(\Delta t)$ и не зависит от числа предшествующих делений, а также от числа имеющихся бактерий, найти вероятность того, что если в момент 0 была 1 бактерия, то в момент t окажется i бактерий.

ГЛАВА 2

ПОСЛЕДОВАТЕЛЬНОСТЬ НЕЗАВИСИМЫХ ИСПЫТАНИЙ

§ 9. Вводные замечания

Проведение различного рода испытаний и экспериментов является непременным условием развития науки и прогресса прикладных областей деятельности. Прежде чем внести в регистр новый сорт пшеницы, необходимо произвести многочисленные испытания, которые должны дать убедительные данные о тех или иных преимуществах нового сорта по сравнению с прежними: повышенная урожайность, устойчивость к погодным условиям, более короткие сроки вегетации, устойчивость к заболеваниям и пр. Перед тем как ввести в массовое производство новый тип телевизора (вычислительной машины, станка, самолета и т.д.) производятся представительные испытания на его безотказность в работе, простоту наладки, приспособленность к ремонту, долговечность. Новые методы преподавания и измененное содержание обучения также требуют длительных и представительных наблюдений и экспериментов, которые могли бы продемонстрировать их преимущества. То же самое можно сказать и о проблемах медицины, экономики, организации производства, социальных исследований. Все новое, прежде чем стать достоянием практики, должно быть предварительно тщательно проверено и подтверждено испытаниями, экспериментами и наблюдениями.

Приходится сталкиваться и с другой ситуацией, когда производятся систематические наблюдения за явлениями, происходящими независимо от исследователя. Так, для примера, метеорологи производят наблюдения за числом облачных дней, температурой воздуха в определенные часы суток, его влажностью и пр. Точно также организатор производства наблюдает за производительностью труда при различных формах его организации.

В научной и практической деятельности постоянно приходится проводить многократно повторяющиеся испытания в сходных условиях. Как правило, при этом результаты предшествующих испытаний никак не сказываются на последующих. Очень важен простейший тип таких испытаний, когда в каждом из испытаний некоторое событие A может появиться с одной и той же вероятностью p и эта вероятность остается одной и той же, независимо от результатов предшествующих или последующих испытаний. Этот тип испытаний был впервые исследован знаменитым швейцарским ученым Якобом Бернулли (1654–1705) в произведении "Ars conjectandi" (Искусство предположений), изданном после смерти автора в 1713 г., и потому получил наименование схемы Бернулли. Подробное исследование таких последовательностей испытаний заслуживает внимания как в силу исключительного их значения в теории вероятностей и в приложениях, так и в силу выявившейся в процессе развития теории вероятностей возможности обобщения тех закономерностей, которые впервые были открыты при изучении схемы последовательных независимых испытаний. Многие факты, подмеченные на схеме Бернулли, впоследствии служили путеводной нитью при изучении более сложных схем. Сделанное замечание относится как к прошлому, так и современному развитию теории вероятностей. Мы убедимся в этом на примерах закона больших чисел и теоремы Муавра – Лапласа.

Рассмотрим теперь следующий вопрос: что следует понимать в схеме Бернулли под элементарным событием? Очевидно, что это последовательность наступлений и ненаступлений интересующего нас события A в последовательных испытаниях. Отнесем наступлению события A единицу, а ненаступлению – нуль. Тогда элементарным событием для n испытаний будет последовательность из n нулей и единиц. Например, если $n = 3$, то все возможные элементарные события записываются следующими тройками названных нами символов: (0,0,0), (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,1,0), (1,0,1), (1,1,1). Смысл каждой из написанных троек чисел 0 и 1 ясен. Первая из перечисленных троек означает, что во всех трех испытаниях событие A не наступило. Вторая тройка означает, что в первых двух испытаниях событие A не наступило, а в третьем – произошло. Легко понять, что множество всех элементарных событий при n испытаниях состоит из 2^n элементов.

Теперь мы должны ввести вероятностную меру на множестве элементарных событий. Это делается однозначно. Действительно, вероятность наступления события A в испытании с номером k равна p, а его ненаступления – $q = 1 - p$. Наступление или ненаступление события A в испытаниях с разными номерами для схемы Бернулли независимы. Значит, в силу теоремы умножения вероятностей, вероятность того, что событие A наступит в m определенных испытаниях (например, в испытаниях с номерами $s_1, s_2, \ldots$ $\ldots, s_m$), а при остальных $n - m$ не наступит, равна $p^m q^{n-m}$. Эта вероятность не зависит от того, как расположены номера $s_1, s_2, \ldots, s_m$.

Простейшая задача, относящаяся к схеме Бернулли, состоит в определении вероятности $P_n(m)$ того, что в n испытаниях событие A произойдет m раз $(0 \leqslant m \leqslant n)$.

Мы только что нашли, что вероятность того, что событие A наступит в испытаниях с определенными m номерами, а в остальных не наступит равна $p^m q^{n-m}$. По теореме сложения искомая вероятность равна сумме только что вычисленных вероятностей для всех различных способов размещения m появлений события A и $n - m$ непоявлений среди n испытаний. Число таких способов известно из теории сочетаний, оно равно $C_n^m = \dfrac{n!}{m!(n-m)!}$ и, следовательно,

$$P_n(m) = C_n^m p^m q^{n-m} \qquad (m = 0, 1, 2, \ldots, n). \tag{1}$$

Полученная формула носит наименование *формулы Бернулли.* Легко заметить, что вероятность $P_n(m)$ равна коэффициенту при x^m в разложении бинома $(q + px)^n$ по степеням x. В силу этого свойства совокупность вероятностей $P_n(m)$ называют *биномиальным законом распределения* вероятностей.

Лишь немного изменив проведенные рассуждения легко обобщить полученный результат. А именно, если в каждом испытании может произойти одно и только одно из k событий $A_1, A_2, \ldots, A_k$, испытания независимы и в каждом из них событие A_k происходит с вероятностью p_k, то вероятность того, что в n независимых испытаниях появятся m_1 событий A_1, m_2 событий $A_2, \ldots, m_k$ событий A_k, равна

$$P_n(m_1, m_2, \ldots, m_k) = \frac{n!}{m_1!m_2!\ldots m_k!}\, p_1^{m_1} p_2^{m_2} \ldots p_k^{m_k}. \tag{1'}$$

Легко также убедиться в том, что эта вероятность является коэффициентом при $x_1^{m_1} x_2^{m_2} \ldots x_k^{m_k}$ в разложении полинома $(p_1x_1 + p_2x_2 + \ldots + p_kx_k)^n$ по степеням $x_1, x_2, \ldots, x_k$. Естественно, что вероятности (1′) называются *полиномиальным распределением.* Полиномиальные распределения находят применения в естествознании, экономических задачах, инженерном деле.

Так как все возможные несовместимые между собой исходы испытаний состоят в появлении события A 0 раз, 1 раз, 2 раза, ..., n раз, то ясно, что

$$\sum_{m=0}^{n} P_n(m) = 1.$$

Это соотношение может быть выведено и без учета теоретико-вероятностных соображений, поскольку по формуле бинома Ньютона

$$\sum_{m=0}^{n} P_n(m) = (p + q)^n = 1^n = 1.$$

Имея в виду постановку общих задач, относящихся к схеме независимых испытаний, рассмотрим теперь числовые примеры. Встречающиеся в них расчеты мы не станем доводить до окончательного числового результата, поскольку эти подсчеты лучше оставить до того момента, когда будут подготовлены удобные и достаточно точные методы для их осуществления.

Пример 1. Вероятность изделию некоторого производства оказаться бракованным равна 0,005. Чему равна вероятность того, что из 10 000 наудачу взятых изделий бракованных изделий окажется а) равно 40, б) не более 70?

В нашем примере $n = 10\,000$, $p = 0{,}005$, поэтому по формуле (1) находим:

$$\text{а) } P_{10000}(40) = C_{10000}^{40}(0{,}995)^{9960}(0{,}005)^{40}.$$

Вероятность того, что число бракованных изделий окажется не большим 70, равна сумме вероятностей числу бракованных изделий оказаться равным 0, 1, 2, . . . , 70. Таким образом

$$\text{б) } \mathsf{P}\{\mu \leqslant 70\} = \sum_{m=0}^{70} P_n(m) = \sum_{m=0}^{70} C_{10000}^{m} 0{,}995^{10000-m} 0{,}005^{m}.$$

Пример 2. Имеются два сосуда A и B, каждый объемом в 1 дм^3. В каждом из них содержится по $2{,}7 \cdot 10^{22}$ молекул газа. Эти сосуды приведены в соприкосновение так, что между ними происходит свободный обмен молекулами, но нет общения с внешней средой. Чему равна вероятность того, что по истечению некоторого времени в одном из сосудов число молекул будет отличаться от числа молекул в другом по меньшей мере на одну десятимиллиардную часть?

Для каждой молекулы вероятность оказаться в определенном сосуде равна половине. Таким образом, как бы производится $5{,}4 \cdot 10^{22}$ испытаний, для каждого из которых вероятность попасть в сосуд A равна 0,5. Пусть μ — число молекул, попавших в сосуд A, и, следовательно $5{,}4 \cdot 10^{22} - \mu$ есть число молекул, попавших в сосуд B. Нам нужно определить вероятность того, что

$$|\mu - (5{,}4 \cdot 10^{22} - \mu)| \geqslant \frac{5{,}4 \cdot 10^{22}}{10^{10}} = 5{,}4 \cdot 10^{12}$$

Иначе говоря, нужно найти вероятность

$$p = \mathsf{P}\{|\mu - 2{,}7 \cdot 10^{22}| \geqslant 2{,}7 \cdot 10^{12}\}.$$

Согласно теореме сложения $p = \Sigma P_n(m)$, где сумма распространена на те значения m, для которых $|m - 2{,}7 \cdot 10^{22}| \geqslant 2{,}7 \cdot 10^{12}$.

Рассмотренные примеры показывают, что при решении реальных задач постоянно возникают задачи, требующие приближенного вычисления сумм $\sum_{m=s}^{t} P_n(m)$ для заданных и t при достаточно больших n. Точно также необходимы приближенные формулы для вычисления вероятностей $P_n(m)$ при больших значениях m и n или же при малых m, но больших n. Эти задачи будут решены нами в ближайших параграфах. Сейчас же мы обратимся к установлению некоторых элементарных фактов, относящихся к изучению поведения $P_n(m)$ как функции m. Для $0 \leqslant m < n$, как легко подсчитать,

$$\frac{P_n(m+1)}{P_n(m)} = \frac{n-m}{m+1} \cdot \frac{p}{q}.$$

Отсюда следует, что

$$P_n(m+1) > P_n(m),$$

если $(n-m)p > (m+1)q$, т.е. если $np - q > m$;

$$P_n(m+1) = P_n(m),$$

если $m = np - q$ и, наконец,

$$P_n(m+1) < P_n(m)$$

если $m > np - q$.

Мы видим, что вероятность $P_n(m)$ с увеличением m сначала возрастает, затем достигает максимума и при дальнейшем росте m убывает. При этом если $np - q$ является целым числом, то максимальное значение вероятность $P_n(m)$ принимает для двух значений m, именно для $m_0 = np - q$ и $m_0' = np - q + 1 = np + p$. Если же $np - q$ не является целым числом, то максимального значения вероятность $P_n(m)$ достигает при $m = m_0$, равном наименьшему целому числу, большему m_0. Число $\bar{m}_0$ называют *вероятнейшим значением* μ. Мы видели, что если $np - q$ есть целое число, то имеет два вероятнейших значения: m_0 и $m_0' = m_0 + 1$.

Таблица 7

m	$P_n(m)$	m	$P_n(m)$	m	$P_n(m)$	m	$P_n(m)$
<5	0,0000	11	0,0287	18	0,1080	25	0,0059
5	0,0001	12	0,0470	19	0,0910	26	0,0028
6	0,0004	13	0,0679	20	0,0704	27	0,0012
7	0,0012	14	0,0879	21	0,0503	28	0,0005
8	0,0033	15	0,1077	22	0,0332	29	0,0002
9	0,0077	16	0,1178	23	0,0202	30	0,0001
10	0,0157	17	0,1178	24	0,0113	>30	0,0000

Отметим, что если $np - q < 0$, то

$$P_n(0) > P_n(1) > \ldots > P_n(n),$$

а если $np - q = 0$, то

$$P_n(0) = P_n(1) > P_n(2) > \ldots > P_n(n).$$

В дальнейшем мы увидим, что при больших значениях n все вероятности $P_n(m)$ становятся близкими к нулю, но только для m, близких к вероятнейшему значению m_0, вероятности $P_n(m)$ сколько-нибудь заметно отличаются от нуля. Этот факт впоследствии будет доказан нами, а сейчас проиллюстрируем сказанное числовым примером.

Пример 3. Пусть $n = 50$, $p = 1/3$.

Вероятнейших значений два: $m_0 = np - q = 16$ и $m_0 + 1 = 17$. Значения вероятностей $P_n(m)$ с точностью до четвертого десятичного знака представлены в табл. 7.

§ 10. Локальная предельная теорема

При рассмотрении числовых примеров предыдущего параграфа было замечено, что при больших значениях n и m вычисление вероятностей $P_n(m)$ превращается в технически сложную задачу. Это обстоятельство было отмечено в ряде работ математиков начала XVIII века, посвященных демографическим проблемам. Возникла необходимость в асимптотических формулах как для $P_n(m)$, так и для $\sum\limits_{m=a}^{b} P_n(m)$. Эта задача была исчерпывающе решена Абрахамом де Муавром (1667–1754), французским математиком, всю жизнь прожившим в Англии. Позднее неоднократно две замечательные его теоремы, к формулировке и доказательству которых мы и перейдем, находили применения и широкие обобщения. Первая из них получила наименование локальной предельной теоремы.

Локальная теорема Муавра. *Если вероятность наступления некоторого события в n независимых испытаниях постоянна и равна p $(0 < p < 1)$, то вероятность $P_n(m)$ того, что в этих испытаниях событие A наступит ровно m раз, удовлетворяет при $n \to \infty$ соотношению*

$$\sqrt{npq}\, P_n(m) : \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \to 1 \tag{1}$$

равномерно для всех m, для которых

$$x = x_{mn} = \frac{m - np}{\sqrt{npq}} \tag{2}$$

находится в каком-либо конечном интервале.

Доказательство. Приводимое нами доказательство опирается на известную из курса математического анализа формулу Стирлинга (одновременно открытую и Муавром)

$$s! = \sqrt{2\pi s}\, s^s e^{-s} e^{\theta_s}$$

в которой остаточный показатель θ_s удовлетворяет неравенству

$$|\theta_s| \leqslant \frac{1}{12s}. \tag{2'}$$

Заметим, что равенство (2) может быть записано в виде

$$m = np + x\sqrt{npq}.$$

Отсюда следует, что

$$n - m = nq - x\sqrt{npq}.$$

Последние равенства показывают, что если x остается в каком-либо ограниченном отрезке, то числа m и $n - m$ возрастают до бесконечности вместе с n. После этого замечания мы можем использовать формулу Стирлинга. Ее применение дает нам

$$P_n(m) = \frac{n!}{m!(n-m)!} p^m q^{n-m} =$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{m(n-m)}}\, e^{-\theta} \left(\frac{n}{m} p\right)^m \left(\frac{n}{n-m} q\right)^{n-m},$$

где

$$\theta = \theta_n - \theta_m - \theta_{n-m} < \frac{1}{12}\left(\frac{1}{n} + \frac{1}{m} + \frac{1}{n-m}\right).$$

Отсюда мы видим, что, каков бы ни был отрезок $a \leqslant x \leqslant b$, величина θ равномерно относительно x в этом отрезке стремится к нулю при $n \to \infty$. Следовательно, множитель $e^{-\theta}$ при том же условии равномерно стремится к единице.

Рассмотрим теперь величину

$$\ln A_n = \ln \left(\frac{n}{m} p\right)^m \left(\frac{n}{n-m} q\right)^{n-m} =$$

$$= -(np + x\sqrt{npq}) \ln\left(1 + x\sqrt{\frac{q}{np}}\right) - (nq - x\sqrt{npq}) \ln\left(1 - x\sqrt{\frac{p}{np}}\right).$$

В условиях теоремы величины $\sqrt{\frac{q}{np}}$ и $\sqrt{\frac{p}{nq}}$ при достаточно больших n могут быть сделаны как угодно малыми, поэтому можно воспользоваться разложением логарифма в степенной ряд. Ограничившись двумя первыми членами, находим

$$\ln\left(1 + x\sqrt{\frac{q}{np}}\right) = x\sqrt{\frac{q}{np}} - \frac{1}{2}\,\frac{qx^2}{np} + O\left(\frac{1}{n^{3/2}}\right),$$

$$\ln\left(1 - x\sqrt{\frac{p}{nq}}\right) = -x\sqrt{\frac{p}{nq}} - \frac{1}{2}\,\frac{px^2}{nq} + O\left(\frac{1}{n^{3/2}}\right).$$

Несложные подсчеты показывают, что $\ln A_n = -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{n}}\right)$ и равномерно относительно n в любом конечном отрезке x

$$A_n : e^{-\frac{x^2}{2}} \to 1 \qquad (n \to \infty).$$

Далее, $\sqrt{npq} \cdot \sqrt{\frac{n}{m(n-m)}} \to 1$ равномерно в каждом конечном отрезке x.

Приведенные подсчеты доказывают теорему.

По существу теми же подсчетами можно доказать аналогичную локальную теорему и для полиномиального распределения.

Локальная теорема. *Если вероятности $p_1, p_2, \ldots, p_k$ появления соответственно событий $A_1^{(s)}, A_2^{(s)}, \ldots, A_k^{(s)}$ в s-м испытании не зависят от номера испытания и отличны от 0 и от 1 $(0 < p_i < 1,\ i = 1, 2, \ldots, k)$, то вероятность $P_n(m_1, m_2, \ldots, m_k)$ того, что при n независимых испытаниях события $A_i^{(s)}$ $(i = 1, 2, \ldots, k)$ появятся m_i раз $(m_1 + m_2 + \ldots + m_k = n)$, удовлетворяет соотношению*

$$\sqrt{n^{k-1}}\, P_n(m_1, m_2, \ldots, m_k) : \frac{e^{-\frac{1}{2}\sum_{i=1} q_i x_i^2}}{(2\pi)^{\frac{k-1}{2}}\sqrt{p_1 p_2 \ldots p_k}} \to 1$$

$(n \to \infty)$

равномерно для всех m_i $(i = 1, 2, \ldots, k)$ для которых $x_i = \frac{m_i - np_i}{\sqrt{np_i q_i}}$ находятся в произвольных конечных интервалах $a_i \leqslant x_i \leqslant b_i$.

Из равенства $\sum\limits_{i=1}^{k} m_i = n$ вытекает соотношение $\sum\limits_{i=1}^{k} x_i \sqrt{np_i q_i} = 0$, которое позволяет одно из x_i выразить через остальные. Заметим вдобавок, что $\sum\limits_{i=1}^{k} p_i = 1$. При $k = 2$ из этой теоремы, как частный случай, получается теорема Муавра.

П р и м е р 1. В примере 2 предыдущего параграфа нам нужно было определить $P_n(m)$ при $n = 10\,000$, $m = 40$, $p = 0{,}005$. По только что доказанной теореме имеем:

$$P_n(m) \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}\left(\frac{m-np}{\sqrt{npq}}\right)^2}$$

Для нашего примера

$$\sqrt{npq} = \sqrt{10\,000 \cdot 0{,}005 \cdot 0{,}995} = \sqrt{49{,}75} \sim 7{,}05,$$

$$\frac{m-np}{\sqrt{npq}} \sim -1{,}42.$$

Следовательно,

$$P_n(m) \sim \frac{1}{7{,}05\sqrt{2\pi}} e^{-\frac{1{,}42^2}{2}}$$

Функция

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

табулирована; краткая таблица значений этой функции приведена в конце книги. По этой таблице находим, что

$$P_n(m) \sim \frac{0{,}1456}{7{,}05} = 0{,}00206.$$

Точные подсчеты без использования теоремы Муавра–Лапласа дают нам $P_n(m) \sim 0{,}00197$.

Для иллюстрации характера приближений, даваемых теоремой Муавра–Лапласа, а также для геометрического пояснения проделанных при ее

Таблица 8

$n = 4$

m	0	1	2	3	4
$P_n(m)$	0,4096	0,4096	0,1536	0,0256	0,0016
x	−1,00	0,25	1,50	2,75	4,00
$\sqrt{npq}\,P_n(m)$	0,3277	0,3277	0,1229	0,0205	0,0013
$\varphi(x)$	0,2420	0,3867	0,1295	0,0091	0,0001

Таблица 9

$n = 25$

m	x	$P_n(m)$	$\sqrt{npq}\,P_n(m)$	$\varphi(x)$	m	x	$P_n(m)$	$\sqrt{npq}\,P_n(m)$	$\varphi(x)$
0	−2,5	0,0037	0,0075	0,0175	8	1,5	0,0623	0,1247	0,1295
1	−2,0	0,0236	0,0472	0,0540	9	2,0	0,0294	0,0589	0,0540
2	−1,5	0,0708	0,1417	0,1295	10	2,5	0,0118	0,0236	0,0175
3	−1,0	0,1358	0,2715	0,2420	11	3,0	0,0040	0,0080	0,0044
4	−0,5	0,1867	0,3734	0,3521	12	3,5	0,0012	0,0023	0,0009
5	0,0	0,1960	0,3920	0,3989	13	4,0	0,0003	0,0006	0,0001
6	0,5	0,1633	0,3267	0,3521	14	4,5	0,0000	0,0000	0,0000
7	1,0	0,1108	0,2217	0,2420	>14	>4,5	0,0000	0,0000	0,0000

Таблица 10

$n = 100$

m	x	$P_n(m)$	$\sqrt{npq}\,P_n(m)$	$\varphi(x)$	m	x	$P_n(m)$	$\sqrt{npq}\,P_n(m)$	$\varphi(x)$
8	−3,00	0,0006	0,0023	0,0044	21	0,25	0,0946	0,3783	0,3867
9	−2,75	0,0015	0,0059	0,0091	22	0,50	0,0849	0,3396	0,3521
10	−2,50	0,0034	0,0134	0,0175	23	0,75	0,0720	0,2879	0,3011
11	−2,25	0,0069	0,0275	0,0317	24	1,00	0,0577	0,2309	0,2420
12	−2,00	0,0127	0,0510	0,0540	25	1,25	0,0439	0,1755	0,1826
13	−1,75	0,0216	0,0863	0,0862	26	1,50	0,0316	0,1266	0,1295
14	−1,50	0,0335	0,1341	0,1295	27	1,75	0,0217	0,0867	0,0862
15	−1,25	0,0481	0,1923	0,1826	28	2,00	0,0141	0,0565	0,0540
16	−1,00	0,0638	0,2553	0,2420	29	2,25	0,0088	0,0351	0,0317
17	−0,75	0,0788	0,3154	0,3011	30	2,50	0,0052	0,0208	0,0175
18	−0,50	0,0909	0,3636	0,3521	31	2,75	0,0029	0,0117	0,0091
19	−0,25	0,0981	0,3923	0,3867	32	3,00	0,0016	0,0063	0,0044
20	−0,00	0,0993	0,3972	0,3989					

Таблица 11

$n = 400$

m	x	$P_n(m)$	$\sqrt{npq}\,P_n(m)$	$\varphi(x)$	m	x	$P_n(m)$	$\sqrt{npq}\,P_n(m)$	$\varphi(x)$
56	−3,000	0,0004	0,0034	0,0044	81	0,125	0,0492	0,3956	0,3957
57	−2,875	0,0006	0,0051	0,0064	82	0,250	0,0478	0,3828	0,3867
58	−2,750	0,0009	0,0076	0,0091	83	0,375	0,0458	0,3666	0,3719
59	−2,625	0,0014	0,0104	0,0127	84	0,500	0,0432	0,3459	0,3521
60	−2,500	0,0019	0,0156	0,0175	85	0,625	0,0402	0,3215	0,3282
61	−2,375	0,0027	0,0218	0,0238	86	0,750	0,0368	0,2944	0,3011
62	−2,250	0,0037	0,0298	0,0317	87	0,875	0,0332	0,2656	0,2721
63	−2,125	0,0050	0,0399	0,0417	88	1,000	0,0295	0,2362	0,2420
64	−2,000	0,0066	0,0525	0,0540	89	1,125	0,0259	0,2070	0,2119
65	−1,875	0,0089	0,0679	0,0684	90	1,250	0,0223	0,1788	0,1826
66	−1,750	0,0108	0,0862	0,0862	91	1,375	0,0190	0,1523	0,1550
67	−1,625	0,0134	0,1075	0,1065	92	1,500	0,0160	0,1279	0,1295
68	−1,500	0,0164	0,1316	0,1295	93	1,625	0,0132	0,1059	0,1065
69	−1,375	0,0198	0,1583	0,1550	94	1,750	0,0108	0,0865	0,0862
70	−1,250	0,0234	0,1871	0,1827	95	1,875	0,0087	0,0696	0,0684
71	−1,125	0,0271	0,2175	0,2119	96	2,000	0,0069	0,0553	0,0540
72	−1,000	0,0310	0,2483	0,2420	97	2,125	0,0054	0,0433	0,0417
73	−0,875	0,0349	0,2789	0,2721	98	2,250	0,0042	0,0335	0,0317
74	−0,750	0,0385	0,3081	0,3011	99	2,375	0,0032	0,0255	0,0238
75	−0,625	0,0419	0,3317	0,3282	100	2,500	0,0024	0,0192	0,0175
76	−0,500	0,0447	0,3580	0,3521	101	2,625	0,0018	0,0142	0,0127
77	−0,375	0,0471	0,3766	0,3719	102	2,750	0,0013	0,0105	0,0091
78	−0,250	0,0487	0,3919	0,3867	103	2,875	0,0009	0,0075	0,0064
79	−0,125	0,0497	0,3973	0,3957	104	3,000	0,0008	0,0054	0,0044
80	−0,000	0,0498	0,3985	0,3989					

доказательстве аналитических преобразований, мы рассмотрим численный пример.

Пусть вероятность p равна 0,2. В табл. 8–11 собраны значения m, $x = \dfrac{m - np}{\sqrt{npq}}$, вероятностей $P_n(m)$, величин $\sqrt{npq}\,P_n(m)$, а также функции $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ с точностью до четвертого десятичного знака соответственно для числа испытаний n = 4,25, 100, 400. На рис. 10*а* ординаты изображают значения вероятностей $P_n(m)$ при различных целочисленных значениях абсциссы m. По рисунку видно, что с увеличением n величины $P_n(m)$ равномерно убывают. Для того чтобы на рисунке точки $[m, P_n(m)]$

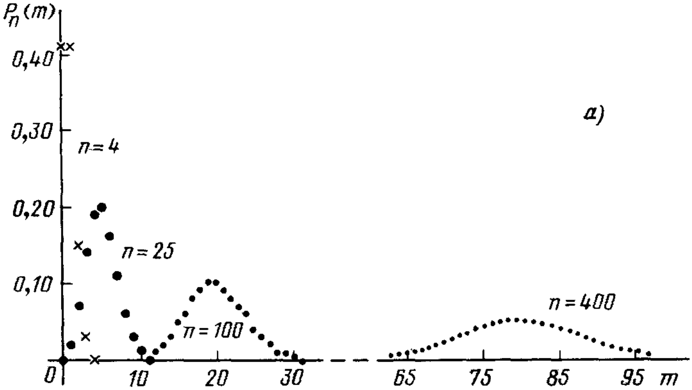

Рис. 10

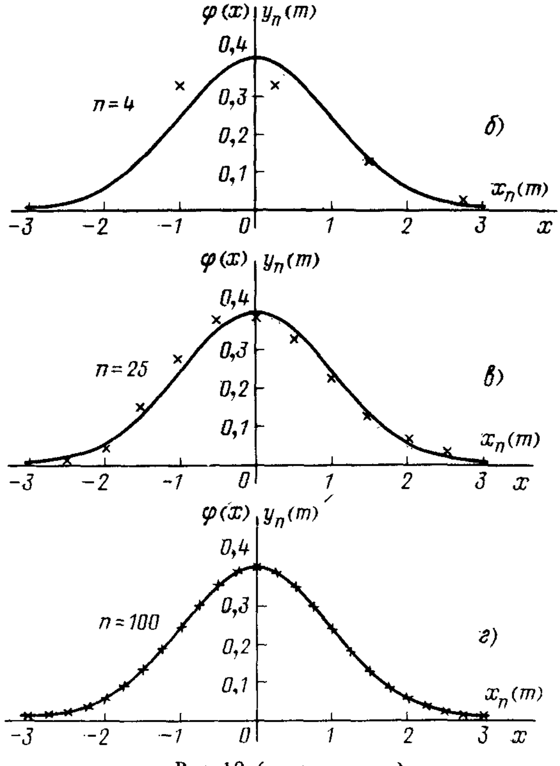

Рис. 10 (продолжение)

уже для рассматриваемых значений n не слились с осью абсцисс, мы выберем резко различные масштабы по осям координат.

Рассмотрение вместо абсцисс m и ординат $P_n(m)$ абсцисс $x_n = \dfrac{m - np}{\sqrt{npq}}$ и ординат $y_n(m) = \sqrt{npq}\, P_n(m)$ означает 1) перенос начала координат в точку $(np, 0)$, находящуюся вблизи от максимальной ординаты $P_n(m)$, 2) увеличение единицы масштаба по оси абсцисс в $\sqrt{npq}$ раз (иными словами, сжатие графика по оси абсцисс в $\sqrt{npq}$ раз), 3) уменьшение единицы масштаба по оси ординат в $\sqrt{npq}$ раз (иными словами, растяжение графика по оси ординат в $\sqrt{npq}$ раз).

На рис. 10*б, в, г* изображены кривая $y = \varphi(x)$ и преобразованные только что описанным способом точки $[m, P_n(m)]$, т.е. точки $[x_n, y_n(m)]$. Мы видим, что уже при $n = 25$ точки $[x_n, y_n(m)]$ сливаются на графике с соответствующими точками кривой $y = \varphi(x)$. Это совпадение становится еще лучше при больших чем 25 значениях n.

Чтобы получить наглядное представление о том, в какой мере можно пользоваться асимптотической формулой Муавра–Лапласа при конечных n*), т.е. заменять биномиальный закон при вычислении вероятностей $P_n(m)$ функцией $y = \varphi(x)$, приведем пример. Для простоты рассмотрим

Таблица 12

n	P_n	Q_n	P_n-Q_n	P_n/Q_n
25	0,09742	0,09679	0,00063	1,0065
100	0,04847	0,04839	0,00008	1,0030
400	0,024207	0,024194	0,000013	1,0004
1156	0,014236	0,014234	0,000002	1,0001

случай $p = q = 1/2$ и возьмем лишь те n, для которых возможно значение $x_{nm} = 1$; такими могут быть, например, $n = 25, 100, 400, 1156$. Именно для них $x_{nm} = 1$ при $m = 15, 55, 210, 595$.

Положим для краткости $P_n(m) = P_n$ и $\dfrac{1}{\sqrt{2\pi npq}} e^{-x_{nm}^2/2} = Q_n$ при $p = q = 1/2$ и $x_{nm} = 1$.

Согласно локальной теореме Муавра–Лапласа отношение P_n/Q_n должно стремиться к единице, когда $n \to \infty$. Вычисление при названных выше значениях n дает

---

*) Очень точные оценки остаточного члена даны в работе С.Н. Бернштейна "Возврат к вопросу о точности предельной формулы Лапласа" (Изв. АН СССР. – 1943. – Т. 7.)

§ 11. Интегральная предельная теорема

Только что выведенную локальную предельную теорему мы используем для вывода другого предельного соотношения теории вероятностей – интегральной предельной теоремы. Изложение начнем с простейшего частного случая этой теоремы – интегральной теоремы Муавра–Лапласа.

Интегральная теорема Муавра–Лапласа. *Если* μ *есть число наступлений события в* n *независимых испытаниях, в каждом из которых вероятность этого события равна* p, *причем* $0 < p < 1$, *то равномерно относительно* a *и* b $(-\infty \leqslant a \leqslant b \leqslant +\infty)$ *при* $n \to \infty$ *имеет место соотношение*

$$\mathsf{P}\left\{a \leqslant \frac{\mu - np}{\sqrt{npq}} < b\right\} \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}}\, dz.$$

Доказательство. Введем для краткости письма обозначение

$$P_n(a, b) = \mathsf{P}\left\{a \leqslant \frac{\mu - np}{\sqrt{npq}} < b\right\}.$$

Эта вероятность, очевидно, равна сумме $\Sigma P_n(m)$, распространенной на те значения m, для которых $a \leqslant x_m < b$, где попрежнему обозначено $x_m = \dfrac{m - np}{\sqrt{npq}}$.

Определим теперь функцию $y = \Pi_n(x)$ следующим образом:

$$y = \Pi_n(x) = \begin{cases} 0 & \text{для } x < x_0 = -\dfrac{np}{\sqrt{npq}} \text{ и} \\ 0 & \text{для } x \geqslant x_n + \dfrac{1}{\sqrt{npq}} = \dfrac{1 + nq}{\sqrt{npq}}, \\ \sqrt{npq}\, P_n(m) & \text{для } x_m \leqslant x < x_{m+1} \ (m = 0, 1, \ldots, n). \end{cases}$$

Очевидно, что вероятность $P_n(m)$ равна площади, ограниченной кривой $y = \Pi_n(x)$, осью OX и ординатами в точках $x = x_m$ и $x = x_{m+1}$, т.е.

$$P_n(m) = \sqrt{npq}\, P_n(m)\,(x_{m+1} - x_m) = \int_{x_m}^{x_{m+1}} \Pi_n(x)\, dx.$$

Отсюда следует, что искомая вероятность $P_n(a, b)$ равна площади, заключенной между кривой $y = \Pi_n(x)$, осью OX и ординатами в точках $x_{\underline{m}}$ и $x_{\overline{m}}$, где $\overline{m}$ и $\underline{m}$ определяются посредством неравенств

$$a \leqslant x_{\underline{m}} \leqslant a + \frac{1}{\sqrt{npq}}, \quad b \leqslant x_{\overline{m}} < b + \frac{1}{\sqrt{npq}}.$$

Таким образом,

$$P_n(a, b) = \int\limits_{x_{\underline{m}}}^{x_{\overline{m}}} \Pi_n(x)\,dx =$$

$$= \int\limits_a^b \Pi_n(x)\,dx + \int\limits_b^{x_{\overline{m}}} \Pi_n(x)\,dx - \int\limits_a^{x_{\underline{m}}} \Pi_n(x)\,dx.$$

Так как максимальное значение вероятности $P_n(m)$ приходится на значение $m_0 = [(n+1)p]$, то максимальное значение $\Pi_n(x)$ приходится на интервал

$$0 \leqslant \frac{m_0 - np}{\sqrt{npq}} \leqslant x < \frac{m_0 + 1 - np}{\sqrt{npq}} \leqslant \frac{2}{\sqrt{npq}}.$$

В этом интервале действует локальная теорема Муавра–Лапласа, и мы можем поэтому заключить, что при всех достаточно больших значениях n

$$\max \Pi_n(x) < 2\,\frac{1}{\sqrt{2\pi}} \max e^{-x^2/2} = \sqrt{2/\pi}.$$

Отсюда мы прежде всего выводим, что

$$|\rho_n| = \left| \int\limits_b^{x_{\overline{m}}} \Pi_n(x)\,dx - \int\limits_a^{x_{\underline{m}}} \Pi_n(x)\,dx \right| \leqslant \int\limits_b^{x_{\overline{m}}} \max \Pi_n(x)\,dx +$$

$$+ \int\limits_a^{x_{\underline{m}}} \max \Pi_n(x)\,dx < \sqrt{\frac{2}{\pi}}\,(-b + x_{\overline{m}} + x_{\underline{m}} - a) \leqslant 2\sqrt{\frac{2}{\pi npq}}$$

и что, следовательно

$$\lim_{n\to\infty} \rho_n = 0.$$

Таким образом, $P_n(a, b)$ только на величину бесконечно малую отличается от $\int\limits_a^b \Pi_n(x)\,dx$.

Мы предположим сначала, что a и b – конечные числа. При этом предположении согласно локальной теореме при $a \leqslant x_m < b$

$$\Pi_n(x_m) = \frac{1}{\sqrt{2\pi}} e^{-x_m^2/2}[1 + \alpha_n(x_m)],$$

где $\alpha_n(x_m) \to 0$ при $n \to \infty$ равномерно относительно x_m. Очевидно, что и при промежуточных зиачениях аргумента

$$\Pi_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} [1 + \alpha_n(x)],$$

причем $\lim\limits_{n\to\infty} \lim\limits_{a \leqslant x < b} \alpha_n(x) = 0$. Действительно, при любом m в интервале $x_m \leqslant x < x_{m+1}$ имеем:

$$\Pi_n(x) = \Pi_n(x_m) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} [1 + \alpha_n(x)],$$

где

$$\alpha_n(x) = e^{\frac{x^2 - x_m^2}{2}} [\alpha_n(x_m) + 1] - 1.$$

Так как

$$\frac{x^2 - x_m^2}{2} \leqslant |x| \cdot |x - x_m| < \frac{\max(|a|, |b|)}{\sqrt{npq}},$$

то ясно, что

$$\lim_{n\to\infty} \max_{a \leqslant x < b} \alpha(x) = 0.$$

Собрав вместе найденные оценки, получаем, что

$$P_n(a, b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx + R_n,$$

где

$$R_n = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \alpha_n(x)\, dx + \rho_n.$$

Так как

$$|R_n| \leqslant \max_{a \leqslant x < b} |\alpha_n(x)| \cdot \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx + \rho_n,$$

то из сказанного ясно, что

$$\lim_{n\to\infty} R_n = 0.$$

В сделанном нами по ходу доказательства частном предположении теорема доказана. Нам остается освободиться от этого ограничения.

С этой целью прежде всего заметим, что

$$\frac{1}{\sqrt{2\pi}} \int e^{-z^2/2}\, dz = 1.$$

Поэтому для любого $\epsilon > 0$ можно выбрать столь большое A, что

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-A}^{A} e^{-z^2/2}\, dz > 1 - \frac{\epsilon}{4},$$

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{A} e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \int\limits_{A}^{\infty} e^{-z^2/2}\, dz < \frac{\epsilon}{8}.$$

Выберем далее, в соответствии с доказанным, столь большое n, что при $-A \leqslant a \leqslant b < A$ будет:

$$\left| P_n(a, b) - \frac{1}{\sqrt{2\pi}} \int\limits_{a}^{b} e^{-z^2/2}\, dz \right| < \frac{\epsilon}{4}.$$

Тогда очевидно, что

$$P_n(-A, A) > 1 - \epsilon/2,$$

$$P(-\infty, -A) + P(A, +\infty) = 1 - P(-A, A) < \epsilon/2.$$

Теперь докажем, что при любых a и b $(-\infty \leqslant a \leqslant b \leqslant +\infty)$ будет:

$$\left| P_n(a, b) - \frac{1}{\sqrt{2\pi}} \int\limits_{a}^{b} e^{-z^2/2}\, dz \right| < \epsilon,$$

чем, очевидно, и закончится доказательство теоремы Лапласа.

Для этого надо разобрать отдельно различные случаи расположения на прямой точек a и b относительно интервала $(-A, A)$. Разберем, например, случай $a \leqslant -A$, $b \geqslant A$ (остальные предоставим читателю).

В этом случае

$$\frac{1}{\sqrt{2\pi}} \int\limits_{a}^{b} e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \left(\int\limits_{a}^{-A} + \int\limits_{-A}^{A} + \int\limits_{A}^{b} e^{-z^2/2}\, dz \right),$$

$$P_n(a, b) = P_n(a, -A) + P_n(-A, A) + P_n(A, b).$$

Поэтому

$$\left|P_n(a,b)-\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-z^2/2}\,dz\right| \leqslant \left|P_n(a,-A)-\frac{1}{\sqrt{2\pi}}\int\limits_a^{-A} e^{-z^2/2}\,dz\right| +$$

$$+\left|P_n(-A,A)-\frac{1}{\sqrt{2\pi}}\int\limits_{-A}^{A} e^{-z^2/2}\,dz\right| + \left|P_n(A,b)-\right.$$

$$\left.-\frac{1}{\sqrt{2\pi}}\int\limits_A^b e^{-z^2/2}\,dz\right| \leqslant P_n(-\infty,-A)+\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{-A} e^{-z^2/2}\,dz +$$

$$+\left|P_n(-A,A)-\frac{1}{\sqrt{2\pi}}\int\limits_{-A}^{A} e^{-z^2/2}\,dz\right| + P_n(A,+\infty) +$$

$$+\frac{1}{\sqrt{2\pi}}\int\limits_A^{\infty} e^{-z^2/2}\,dz < \frac{\epsilon}{2}+\frac{\epsilon}{4}+\frac{\epsilon}{8}+\frac{\epsilon}{8}=\epsilon.$$

Мы перейдем теперь к формулировке интегральной предельной теоремы в общем случае схемы последовательности независимых испытаний. Пусть попрежнему μ_i $(i = 1, 2, \ldots, k)$ означает число появлений событий $A_i^{(s)}$ $(s = 1, 2, \ldots, n)$ в n последовательных испытаниях. В зависимости от случая μ_i могут принимать лишь значения $0, 1, 2, \ldots, n$, причем так как в каждом испытании возможны k исходов и эти исходы несовместимы, то должно иметь место равенство

$$\mu_1 + \mu_2 + \ldots + \mu_k = n. \tag{1}$$

Станем теперь на величины $\mu_1, \mu_2, \ldots, \mu_k$ смотреть как на прямоугольные координаты точки в k-мерном евклидовом пространстве.

При этом результаты n испытаний изобразятся точкой с целочисленными координатами, не меньшими нуля и не большими n; будем в дальнейшем называть такие точки ц е л о ч и с л е н н ы м и. Равенство (1) показывает, что результаты испытаний изобразятся не произвольными целочисленными точками в гиперкубе $0 \leqslant \mu_i \leqslant n$ $(i = 1, 2, \ldots, k)$, а лишь теми из них, которые находятся в гиперплоскости (1). На рис. 11 изображено положение возможных результатов испытаний в гиперплоскости (1) для случая $n = 3$, $k = 3$.

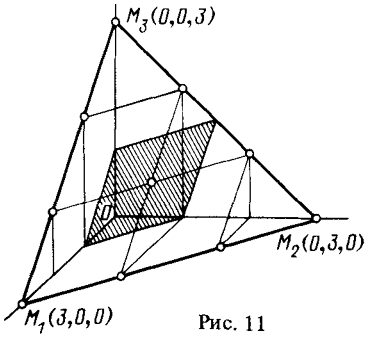

Рис. 11

Произведем преобразования координаты по формулам

$$x_i = \frac{\mu_i - np_i}{\sqrt{np_i q_i}} \qquad (i = 1, 2, \ldots, k;\ q_i = 1 - p_i).$$

Уравнение гиперплоскости (1) в новых координатах перепишется в следующем виде:

$$\sum_{i=1}^{k} x_i \sqrt{np_i q_i} = 0. \tag{2}$$

Точки гиперплоскости (2), в которые перешли целочисленные точки гиперплоскости (1), условимся также называть "целочисленными".

Обозначим через $P_n(G)$ вероятность того, что в результате n испытаний числа μ_i $(i = 1, 2, \ldots, k)$ появлений каждого из возможных исходов окажутся такими, что точка с координатами $x_i = \dfrac{\mu_i - np_i}{\sqrt{np_i q_i}}$ попадет внутрь области G.

Тогда имеет место следующая

Теорема. *Если в схеме последовательности независимых испытаний в каждом из испытаний возможны k исходов, причем вероятность каждого из исходов не зависит от номера испытания и отлична от* 0 *и от* 1, *то какова бы ни была область G гиперплоскости* (2), *для которой* $(k-1)$-*мерный объем ее границы равен нулю, равномерно относительно G при* $n \to \infty$, *имеет место соотношение*

$$P_n\{G\} \to \sqrt{\frac{q_1 q_2 \ldots q_k}{(2\pi)^{k-1} \sum\limits_{i=1}^{k} p_i q_i}} \int_G e^{-\frac{1}{2} \sum q_i x_i^2} dv,$$

где dv означает элемент объема области G и интеграл распространяется на область G.

Доказательство и по идее и по осуществлению является почти полной копией рассуждений, проведенных при доказательстве интегральной теоремы Муавра–Лапласа.

Замечание. Только что сформулированной теореме мы придали форму, в которой все переменные $x_1, x_2, \ldots, x_n$ играют одинаковую роль.

В интегральной теореме Муавра–Лапласа мы, однако, предпочли проводить рассуждения, нарушив однородность переменных x_1 и x_2, только с переменным $x = x_1$. Геометрически это означало, что мы рассматриваем не сами результаты испытаний (целочисленные точки на прямой $x_1 + x_2 = 0$), а их проекции на ось OX. Подобным же образом мы можем, нарушив однородность и в общем случае, рассмотреть интегрирование не по области G, а по ее проекции G' на какую-либо координатную гиперплоскость, скажем, на плоскость $x_k = 0$. Элемент объема dv' в гиперплоскости $x_k = 0$ связан с элементом объема dv гиперплоскости (2) соотношением

$$dv' = dv \cos\varphi,$$

где φ – угол между указанными гиперплоскостями. Легко подсчитать, что

$$\cos\varphi = \frac{\sqrt{p_k q_k}}{\sqrt{\Sigma p_i q_i}}.$$

В координатной гиперплоскости элемент объема $dv' = dx_1 dx_2 \ldots dx_{k-1}$, поэтому имеет место равенство

$$\sqrt{\frac{q_1 q_2 \ldots q_k}{(2\pi)^{k-1} \Sigma p_i q_i}} \int_G e^{-\frac{1}{2} \Sigma q_i x_i^2} dv =$$

$$= \sqrt{\frac{q_1 q_2 \ldots q_{k-1}}{(2\pi)^{k-1} p_k}} \int_{G'} e^{-\frac{1}{2} \sum\limits_1^k q_i x_i^2} dx_1 \ldots dx_{k-1}.$$

В подинтегральной функции мы должны произвести замену x_k на его выражение через $x_1, x_2, \ldots, x_{k-1}$:

$$x_k = -\frac{1}{\sqrt{p_k q_k}} \sum_{i=1}^{k-1} \sqrt{p_i q_i}\, x_i.$$

В результате этой замены мы имеем:

$$\sum_{i=1}^{k} q_i x_i^2 = \sum_{i=1}^{k-1} q_i \left(1 + \frac{p_i}{p_k}\right)^2 x_i^2 + 2 \sum_{1 \leqslant i < j \leqslant k-1} x_i x_j \frac{\sqrt{p_i q_i p_j q_j}}{p_k} =$$

$$= Q(x_1, x_2, \ldots, x_{k-1}).$$

Интегральная предельная теорема может быть, таким образом, сформулирована иначе, а именно:

В условиях интегральной предельной теоремы при $n \to \infty$

$$P(G) \to \sqrt{\frac{q_1 q_2 \ldots q_{k-1}}{(2\pi)^{k-1} p_k}} \int_{G'} e^{-\frac{1}{2} Q(x_1, x_2, \ldots, x_{k-1})} dx_1 dx_2 \ldots dx_{k-1}. \tag{14}$$

Понятно, что интегральная теорема Муавра–Лапласа является частным случаем только что доказанной теоремы: она легко может быть получена из формулы (14).

Для этого достаточно заметить, что в схеме Бернулли $k = 2$, $p = p_1$, $q = p_2 = 1 - p$.

При $k = 3$ формула (14) принимает следующий вид:

$$P(G) \to \sqrt{\frac{q_1 q_2}{(2\pi)^2 p_3}} \int_{G'} e^{-\frac{1}{2} Q(x_1, x_2)} \, dx_1 \, dx_2,$$

где

$$p_3 = 1 - p_1 - p_2,$$

$$Q(x_1, x_2) = q_1 \left(1 + \frac{p_1}{p_3}\right) x_1^2 + q_2 \left(1 + \frac{p_2}{p_3}\right) x_2^2 + 2 \, \frac{\sqrt{p_1 q_1 p_2 q_2}}{p_3} x_1 x_2 =$$

$$= \frac{q_1 q_2}{p_3} \left(x_1^2 + x_2^2 + 2 \sqrt{\frac{p_1 p_2}{q_1 q_2}} x_1 x_2\right).$$

Простой подсчет показывает, что

$$p_3 = 1 - p_1 - p_2 = q_1 q_2 - p_1 p_2,$$

поэтому

$$Q(x_1, x_2) = \frac{1}{1 - \dfrac{p_1 p_2}{q_1 q_2}} \left(x_1^2 + x_2^2 + 2 \sqrt{\frac{p_1 p_2}{q_1 q_2}} \, x_1 x_2\right).$$

§ 12. Применения интегральной теоремы Муавра–Лапласа

В качестве первого приложения интегральной теоремы Муавра–Лапласа мы оценим вероятность неравенства

$$\left| \frac{\mu}{n} - p \right| < \epsilon,$$

где $\epsilon > 0$ – постоянное.

Имеем

$$\mathbf{P}\left\{ \left| \frac{\mu}{n} - p \right| < \epsilon \right\} = \mathbf{P}\left\{ -\epsilon \sqrt{\frac{n}{pq}} < \frac{\mu - np}{\sqrt{npq}} < \epsilon \sqrt{\frac{n}{pq}} \right\}$$

и, значит, в силу интегральной теоремы Муавра–Лапласа

$$\lim_{n \to \infty} \mathbf{P}\left\{\left|\frac{\mu}{n} - p\right| < \epsilon\right\} = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{z^2}{2}} dz = 1.$$

Итак, каково бы ни было постоянное $\epsilon > 0$, вероятность неравенства $\left|\frac{\mu}{n} - p\right| < \epsilon$ стремится к единице.

Обнаруженный нами факт был впервые найден Я. Бернулли; он носит название *закона больших чисел* или *теоремы Бернулли.* Теорема Бернулли и ее многочисленные обобщения являются одними из важнейших теорем теории вероятностей. Через них именно теория соприкасается с практикой, именно в них заложен фундамент успехов применения теории вероятностей к различным проблемам естествознания и техники. Об это будет подробнее сказано в главе, посвященной закону больших чисел; там же мы дадим доказательство теоремы Бернулли более простым методом, отличным, как от только что изложенного, так и от употребленного Я. Бернулли.

Мы рассмотрим теперь типичные задачи, приводящие к теореме Муавра–Лапласа.

Производится n независимых испытаний, при каждом из которых вероятность наступления события A равна p:

I. Спрашивается, чему равна вероятность того, что частота наступления события A отклонится от вероятности p не более чем на α? Эта вероятность равна

$$\mathbf{P}\left\{\left|\frac{\mu}{n} - p\right| \leqslant \alpha\right\} = \mathbf{P}\left\{-\alpha\sqrt{\frac{n}{pq}} \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant \alpha\sqrt{\frac{n}{pq}}\right\} \sim$$

$$\sim \frac{1}{\sqrt{2\pi}} \int_{-\alpha\sqrt{\frac{n}{pq}}}^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}} dx.$$

II. Какое наименьшее число испытаний нужно произвести для того, чтобы с вероятностью, не меньшей β, частота отклонялась от вероятности не больше чем на α. Нам нужно определить n из неравенства

$$\mathbf{P}\left\{\left|\frac{\mu}{n} - p\right| \leqslant \alpha\right\} \geqslant \beta.$$

Вероятность, фигурирующую в левой части неравенства, мы заменяем приближенно по теореме Муавра – Лапласа интегралом. Для определения n в результате получается неравенство

$$\frac{2}{\sqrt{2\pi}} \int\limits_0^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}}\, dx \geqslant \beta.$$

III. При данной вероятности β и числе испытаний n требуется определить границу возможных изменений $\left|\frac{\mu}{n} - p\right|$. Иными словами, зная β и n, нужно найти α, для которого

$$\mathbf{P}\left\{\left|\frac{\mu}{n} - p\right| < \alpha\right\} = \beta.$$

Применение интегральной теоремы Лапласа дает нам для определения α уравнение

$$\frac{2}{\sqrt{2\pi}} \int\limits_0^{\alpha\sqrt{\frac{n}{pq}}} e^{-x^2/2}\, dx = \beta.$$

Численное решение всех рассмотренных нами задач требует умения вычислять значения интеграла

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-z^2/2}\, dz \tag{1}$$

при любых значениях x и решать обратную задачу – по величине интеграла $\Phi(x)$ вычислять соответствующее значение аргумента x. Для этих расчетов требуются специальные таблицы, так как интеграл (1) при $0 < x < \infty$ в конечном виде через элементарные функции не выражается. Такие таблицы составлены и имеются в конце настоящей книги.

Рис. 12 дает наглядное представление о функции $\Phi(x)$. При помощи таблицы значений функции $\Phi(x)$ можно вычислять по формуле $J(a, b) = \Phi(b) - \Phi(a)$ также значение интеграла

$$J(a, b) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-z^2/2}\, dz.$$

Таблица функции $\Phi(x)$ составлена только для положительных x; для отрицательных x функция $\Phi(x)$ определяется из равенства

$$\Phi(-x) = -\Phi(x).$$

Мы теперь в состоянии довести до конца решение примера 2 § 9.

Пример 1. В примере 2 § 9 нам нужно было найти вероятность

$$p = \sum \mathsf{P}\{\mu = m\},$$

где сумма распространена на те значения m, для которых

$$|m - 2{,}7 \cdot 10^{22}| \geqslant 2{,}7 \cdot 10^{12},$$

при условии, что общее число испытаний $n = 5{,}4 \cdot 10^{22}$ и $p = 1/2$. Так как

$$p = \mathsf{P}\left\{\frac{|\mu - np|}{\sqrt{npq}} \geqslant \frac{2{,}7 \cdot 10^{12}}{\sqrt{5{,}4 \cdot 10^{22} \cdot \frac{1}{4}}}\right\} \approx \mathsf{P}\left\{\frac{|\mu - np|}{\sqrt{npq}} \geqslant 2{,}33 \cdot 10\right\},$$

в силу теоремы Лапласа

$$p \sim \frac{2}{\sqrt{2\pi}} \int_{2{,}33 \cdot 10}^{\infty} e^{-x^2/2}\, dx.$$

Так как

$$\int_z^{\infty} e^{-x^2/2}\, dx < \frac{1}{z} \int_z^{\infty} x e^{-x^2/2}\, dx = \frac{1}{z} e^{-z^2/2},$$

то

$$p < \frac{1}{\sqrt{2\pi} \cdot 10^5} e^{-2{,}7 \cdot 100} < 10^{-100}$$

О том, как мала эта вероятность, можно судить по следующему сравнению. Предположим, что шар радиуса 6000 км наполнен белым песком, в который попала одна черная песчинка. Размер песчинки равен 1 мм^3.

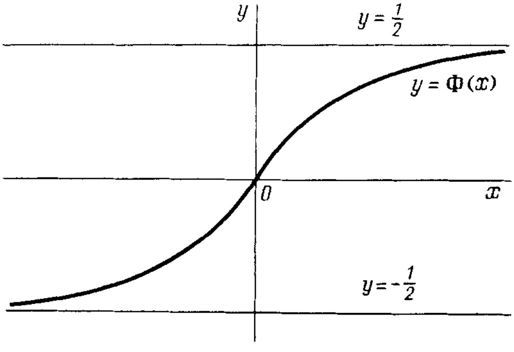

Рис. 12

Наудачу из всей этой массы песчинок берется одна; чему равна вероятность того, что она будет черного цвета?

Легко подсчитать, что объем шара радиуса 6000 км немногим меньше 10^{30} мм^3 и, следовательно, вероятность извлечь черную песчинку немногим больше 10^{-30}.

Пример 2. В примере 1 § 9 нам нужно было найти вероятность того, что число бракованных изделий окажется не больше семидесяти, если вероятность для каждого изделия быть бракованным равна $p = 0{,}005$ и число изделий равно 10 000. По только что доказанной теореме эта вероятность равна

$$\mathsf{P}\{\mu \leqslant 70\} = \mathsf{P}\left\{-\frac{50}{\sqrt{49{,}75}} \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant \frac{20}{\sqrt{49{,}75}}\right\} =$$

$$= \mathsf{P}\left\{-7{,}09 \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant 2{,}84\right\} \sim \frac{1}{\sqrt{2\pi}} \int_{-7{,}09}^{2{,}84} e^{-\frac{z^2}{2}}\, dz =$$

$$= \Phi(2{,}84) - \Phi(-7{,}09) = \Phi(2{,}84) + \Phi(7{,}09) = 0{,}9975.$$

Значения функции $\Phi(x)$ при $x = 7{,}09$ в таблицах нет, мы заменили его половиной, совершив при этом ошибку, меньшую 10^{-10}.

Естественно, что в примерах настоящего и предыдущего параграфов, равно как и в любых других задачах, относящихся к определению вероятностей $P_n(m)$ при каких-либо конечных значениях m и n по асимптотическим формулам Муавра–Лапласа требуется оценка совершаемой при такой замене ошибки. В течение очень долгого времени теоремы Муавра–Лапласа применялись к решению подобного рода задач без сколько-нибудь удовлетворительной оценки остаточного члена. Создалась чисто эмпирическая уверенность, что при n порядка нескольких сотен или еще большем, а также при p, не слишком близких к 0 или 1, употребление теорем Муавра–Лапласа приводит к удовлетворительным результатам. В настоящее время существуют достаточно хорошие оценки погрешностей, совершаемых при пользовании асимптотической формулой Муавра–Лапласа*).

Мы остановимся еще на обобщении теоремы Бернулли на случай общей схемы последовательности независимых испытаний. Пусть в каждом испытании возможны k исходов, вероятность каждого из них соответственно равна $p_1, p_2, \ldots, p_k$ и $\mu_1, \mu_2, \ldots, \mu_k$ — числа появлений каждого исхода

---

*) См., например, цитированную на стр. 84 работу С.Н. Бернштейна.

в последовательности n независимых испытаний. Определим вероятность одновременного осуществления неравенств

$$\left|\frac{\mu_1}{n} - p_1\right| < \epsilon_1, \quad \left|\frac{\mu_2}{n} - p_2\right| < \epsilon_2, \ldots, \left|\frac{\mu_k}{n} - p_k\right| < \epsilon_k, \tag{2}$$

т.е. неравенств

$$|x_1| < \epsilon_1 \sqrt{\frac{n}{p_1 q_1}}, \quad |x_2| < \epsilon_2 \sqrt{\frac{n}{p_2 q_2}}, \ldots, |x_k| < \epsilon_k \sqrt{\frac{n}{p_k q_k}}.$$

Последнее из этих неравенств, собственно, является следствием предыдущих, так как согласно (2) § 13 первые $k-1$ из неравенств (2) дают оценку

$$|x_k| = \left| -\sum_{i=1}^{k-1} \sqrt{\frac{p_i q_i}{p_k q_k}}\, x_i \right| \leqslant \sum_{i=1}^{k-1} \sqrt{\frac{p_i q_i}{p_k q_k}}\, \epsilon_i. \tag{3}$$

Согласно (16) § 13 вероятность первых $k-1$ неравенств (2) [а следовательно, также неравенства (3)] имеет своим пределом при $n \to \infty$ интеграл

$$\sqrt{\frac{q_1 q_2 \ldots q_{k-1}}{(2\pi)^{k-1} p_k}} \int \ldots \int e^{-\frac{1}{2} Q(x_1, \ldots, x_k)}\, dx_1\, dx_2 \ldots dx_{k-1} = 1.$$

§ 13. Теорема Пуассона

Мы видели при доказательстве локальной теоремы Муавра–Лапласа, что асимптотическое представление вероятности $P_n(m)$ посредством функции $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ действует тем хуже, чем больше вероятность p отличается от половины, т.е. чем меньшие значения p или q приходится рассматривать, и это представление отказывается служить при $p = 0$, $q = 1$, а также при $p = 1$, $q = 0$. Однако значительный круг задач связан с необходимостью вычисления вероятностей $P_n(m)$ именно при малых значениях p*). Для того чтобы в этом случае теорема Муавра–Лапласа дала результат с незначительной ошибкой, необходимо, чтобы число n испытаний было очень велико. Возникает, таким образом, задача разыскания асимптотической формулы, специально приспособленной для случая малых p. Такая формула была найдена Пуассоном.

---

*) Или также при малых значениях q, но очевидно, что задачи разыскания асимптотических формул для $P_n(m)$ при малых значениях p или q сводятся одна к другой.

Рассмотрим последовательность серий

$$
\begin{array}{l}
E_{11},\\
E_{21},\ E_{22},\\
E_{31},\ E_{32},\ E_{33}\\
\dots\dots\dots\dots\dots\\
E_{n1},\ E_{n2},\ E_{n3}, \dots,\ E_{nn},\\
\dots\dots\dots\dots\dots
\end{array}
$$

в которой события одной серии взаимно независимы между собой и имеют каждое вероятность p_n, зависящую только от номера серии. Через μ_n обозначается число фактически появившихся событий n-й серии.

Теорема Пуассона. *Если $p_n \to 0$ при $n \to \infty$, то*

$$
\mathsf{P}(\mu_n = m) - \frac{a_n^m}{m!} e^{-a_n} \to 0, \tag{1}
$$

где

$$
a_n = np_n.
$$

Доказательство. Очевидно, что

$$
P_n(m) = \mathsf{P}\{\mu_n = m\} = C_n^m p_n^m (1 - p_n)^{n-m} =
$$

$$
= \frac{n!}{m!(n-m)!}\left(\frac{a_n}{n}\right)^m \left(1 - \frac{a_n}{n}\right)^{n-m} =
$$

$$
= \frac{a_n^m}{m!}\left(1 - \frac{a_n}{n}\right)^n \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{m-1}{n}\right)}{\left(1 - \frac{a_n}{n}\right)^m}. \tag{2}
$$

Пусть m фиксировано. Выберем произвольно $\epsilon > 0$. Тогда можно выбрать $A = A(\epsilon)$ столь большим, чтобы при $a \geqslant A$ было

$$
\frac{a^m}{m!} e^{-\frac{1}{2}a} \leqslant \frac{\epsilon}{2}.
$$

Рассмотрим сначала те номера n, для которых $a_n \geqslant A$. Для этих n, по неравенству $1 - x < e^{-x}$, $0 \leqslant x \leqslant 1$:

$$P_n(m) \leqslant \frac{a_n^m}{m!} e^{-\frac{n-m}{n} a_n} \leqslant \frac{\epsilon}{2} \quad \text{при} \quad n \geqslant 2m$$

$$\frac{a_n^m}{m!} e^{-a_n} < \frac{\epsilon}{2}.$$

Поэтому для указанных n

$$\left| P_n(m) - \frac{a_n^m}{m!} e^{-a_n} \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Рассмотрим теперь те номера n, для которых $a_n \leqslant A$. Так как

$$\lim_{n\to\infty} \left\{ \left(1 - \frac{a_n}{n}\right)^n - e^{-a_n} \right\} = 0 \text{ при } a_n \leqslant A \text{ и при постоянном } m$$

$$\lim_{n\to\infty} \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right)}{\left(1 - \frac{a}{n}\right)^m} = 1,$$

то в силу формулы (2) при $n \geqslant n_0(\epsilon)$

$$\left| P_n(m) - \frac{a_n^m}{m!} e^{-a_n} \right| < \epsilon,$$

что и требовалось доказать.

Заметим, что теорема Пуассона имеет место и в том случае, когда вероятность события A в каждом испытании равна нулю. В этом случае $a_n = 0$.

Обозначим

$$P(m) = \frac{a^m}{m!} e^{-a}.$$

Полученное распределение вероятностей носит название *закона Пуассона*.

Легко подсчитать, что величины $P(m)$ удовлетворяют равенству $\sum_m P(m) = 1$. Изучим поведение $P(m)$ как функции m. С этой целью рассмотрим отношение

$$\frac{P(m)}{P(m-1)} = \frac{a}{m}.$$

Мы видим, что если $m > a$, то $P(m) < P(m-1)$, если же $m < a$, то $P(m) > P(m-1)$, если, наконец, $m = a$, то $P(m) = P(m-1)$. Отсюда мы выводим, что величина $P(m)$ возрастает при увеличении m от 0 до $m_0 = [a]$, и при дальнейшем увеличении m убывает. Если a – целое число, то $P(m)$ имеет два максимальных значения: при $m_0 = a$ и при $m_0' = a - 1$. Приведем примеры.

Пример 1. Вероятность попадания в цель при каждом выстреле равна 0,001. Найти вероятность попадания в цель двумя и более пулями, если число выстрелов равно 5000*).

Считая каждый выстрел за испытание и попадание в цель – за событие, мы можем для вычисления вероятности $\mathsf{P}(\mu_n \geqslant 2)$ воспользоваться теоремой Пуассона. В рассматриваемом примере

$$a_n = np = 0{,}001 \cdot 5000 = 5.$$

Искомая вероятность равна

$$\mathsf{P}\{\mu_n \geqslant 2\} = \sum_{m=2}^{\infty} P_n(m) = 1 - P_n(0) - P_n(1).$$

По теореме Пуассона

$$P_n(0) \approx e^{-5}, \quad P_n(1) \approx 5e^{-5}.$$

Поэтому

$$\mathsf{P}\{\mu_n \geqslant 2\} \approx 1 - 6e^{-5} \approx 0{,}9596.$$

Максимальное значение вероятность $P_n(m)$ принимает при $m = 4$ и $m = 5$. Эти вероятности равны с точностью до четвертого десятичного знака

$$P(4) = P(5) \approx 0{,}1751.$$

Вычисления по точной формуле дают с точностью до четвертого знака $P_{5000}(0) = 0{,}0071$, $P_{5000}(1) = 0{,}0354$ и, следовательно,

$$\mathsf{P}\{\mu_n \geqslant 2\} = 0{,}9575.$$

---

*) В Великой Отечественной войне реальное осуществление условий нашей задачи имело место при обстреливании самолета из пехотного оружия. Пулей самолет может быть подбит лишь при попадании в немногие уязвимые места – мотор, летчик, бензобаки и пр. Вероятность попадания в эти уязвимые места отдельным выстрелом весьма мала, но, как правило, по самолету вело огонь целое подразделение, и общее количество выстрелов, выпущенных по самолету, было значительным. В результате вероятность попадания хотя бы одной или двумя пулями имела заметную величину. Это обстоятельство было подмечено и чисто практически.

Ошибка от использования асимптотической формулы меньше 0,25% вычисляемой величины.

Пример 2. На прядильной фабрике работница обслуживает по нескольку сотен веретен, каждое из которых прядет свой моток пряжи. При вращении веретена пряжа из-за неравномерности натяжения, неровноты и других причин в моменты времени, зависящие от случая, рвется. Для производства важно знать, как часто могут происходить обрывы при тех или иных условиях работы (сорт пряжи, скорость веретен и т.д.).

Считая, что работница обслуживает 800 веретен и вероятность обрыва пряжи на каждом из веретен в течение некоторого промежутка времени τ равна 0,005, найти наиболее вероятное число обрывов и вероятность того, что в течение промежутка времени τ произойдет не более 10 обрывов.

Так как

$$a_n = np = 0{,}005 \cdot 800 = 4,$$

то наиболее вероятных чисел обрывов за промежуток времени τ будет два: 3 и 4. Их вероятности

$$P_{800}(3) = P_{800}(4) = C_{800}^{4} \cdot 0{,}005^{4} \cdot 0{,}995^{796}.$$

По формуле Пуассона имеем:

$$P_{800}(3) = P_{800}(4) \sim \frac{4^3}{3!}\, e^{-4} = \frac{32}{3} \cdot e^{-4} = 0{,}1954.$$

Точное значение $P_{800}(3) = P_{800}(4) = 0{,}1945$. Вероятность того, что число обрывов за промежуток времени τ будет не более 10 равна

$$\mathsf{P}\{\mu_n \leqslant 10\} = \sum_{m=0}^{10} P_{800}(m) = 1 - \sum_{m=11}^{\infty} P_{800}(m).$$

В силу теоремы Пуассона

$$P_{800}(m) \sim \frac{4^m}{m!}\, e^{-4} \quad (m = 0, 1, 2, \ldots),$$

поэтому

$$\mathsf{P}\{\mu_n \leqslant 10\} = 1 - \sum_{m=11}^{\infty} \frac{4^m}{m!} e^{-4}.$$

Но

$$\sum_{m=11}^{\infty} \frac{4^m}{m!} e^{-4} > \left(\frac{4^{11}}{11!} + \frac{4^{12}}{12!} + \frac{4^{13}}{13!} \right) e^{-4} =$$

$$= \frac{4^{12} \cdot 14}{11!39}\, e^{-4} = 0{,}00276.$$

С другой стороны,

$$\sum_{m=11}^{\infty} \frac{4^m}{m!} e^{-4} < \frac{4^{11}}{11!} e^{-4} + \frac{4^{12}}{12!} e^{-4} +$$

$$+ e^{-4} \frac{4^{13}}{13!} \left[1 + \frac{4}{14} + \left(\frac{4}{14}\right)^2 + \ldots \right] = \frac{4^{12} \cdot 24}{11!35} e^{-4} = 0{,}00284.$$

Таким образом,

$$0{,}99716 \leqslant \mathsf{P}\{\mu_n \leqslant 10\} \leqslant 0{,}99724.$$

Подобно тому, как и при использовании теоремы Муавра–Лапласа, возникает вопрос об оценке совершаемой ошибки при замене точной формулы для вычисления $P_n(m)$ на асимптотическую формулу Пуассона*).

Из равенства

$$P_n(0) = \left(1 - \frac{a_n}{n}\right)^n = e^{n \ln\left(1 - \frac{a_n}{n}\right)} =$$

$$= \exp\left\{-n \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{a_n}{n}\right)^k \right\} = e^{-a_n}(1 - R_n),$$

где

$$R_n = 1 - \exp\left\{-n \sum_{k=2}^{\infty} \frac{1}{k}\left(\frac{a_n}{n}\right)^k \right\},$$

мы легко можем найти эту оценку для случая $m = 0$. В самом деле, так как при любом положительном x $0 < 1 - e^{-x} < x$, то каковы бы ни были a_n и n

$$0 < R_n < n \sum_{k=2}^{\infty} \frac{1}{k}\left(\frac{a_n}{n}\right)^k.$$

Так как

$$\sum_{k=2}^{\infty} \frac{1}{k}\left(\frac{a_n}{n}\right)^k \leqslant \frac{a_n^2}{2n^2} + \frac{1}{3} \sum_{k=3}^{\infty}\left(\frac{a_n}{n}\right)^k =$$

$$= \frac{a_n^2}{2n^2} + \frac{a_n^3}{3n^3\left(1 - \frac{a}{n}\right)} = \frac{a_n^2}{6n^2} \cdot \frac{3n - a_n}{n - a_n} < \frac{a_n^2}{2n(n - a_n)},$$

---

*) Эта задача подробно исследована в статье Ю.В. Прохорова "Асимптотическое поведение биномиального распределения" (УМН. – 1953. – Т. 8. – С. 135–142.)

то

$$0 < R_n < \frac{a_n^2}{2(n - a_n)}\,.$$

Из того, что R_n неотрицательно, мы заключаем, что при замене $P_n(0)$ на e^{-a_n} мы несколько увеличиваем вероятность $P_n(0)$.

§ 14. Иллюстрация схемы независимых испытаний

В качестве иллюстрации использования предыдущих результатов для целей естествознания мы рассмотрим весьма схематически проблему случайных блужданий частицы на прямой линии. Эта задача может рассматриваться как прообраз реальных физических задач теории диффузии, броуновского движения и пр.

Представим себе, что в определенные моменты времени частица, находящаяся в начальный момент в положении $x = 0$, испытывает случайные толчки, в результате которых она получает смещение вправо или влево на единицу масштаба. Таким образом, в каждый из этих моментов частица с вероятностью $1/2$ смещается на единицу вправо или с такой же вероятностью — на единицу влево. В результате n толчков частица переместится на расстояние μ. Ясно, что в этой задаче мы имеем дело со схемой Бернулли в чистом виде. Отсюда следует, что при каждых n и m мы можем вычислить вероятность того, что $\mu = m$; а именно

$$\mathsf{P}\{\mu = m\} = \begin{cases} C_n^{\frac{m+n}{2}}\left(\dfrac{1}{2}\right)^n, & \text{если } -n \leqslant m \leqslant n, \\ 0 & , \text{если } |m| > n. \end{cases}$$

При больших значениях n, как это следует из локальной теоремы Муавра—Лапласа,

$$\mathsf{P}\{\mu = m\} \approx \frac{1}{\sqrt{2\pi n}}\, e^{-\frac{m^2}{2n}}. \tag{1}$$

На полученную формулу мы сможем смотреть следующим образом. Пусть в начальный момент имелось большое число частиц, имеющих координату $x = 0$. Все эти частицы независимо друг от друга начинают перемещаться по прямой под влиянием случайных толчков. Тогда после n толчков доля частиц, переместившихся на расстояние m, дается формулой (1).

Понятно, что мы рассматриваем *идеализированные* условия движения частиц и реальные молекулы движутся при гораздо более сложных условиях, однако полученный результат дает правильную *качественную* картину явления.

В физике приходится рассматривать более сложные примеры случайных блужданий. Мы ограничимся столь же схематическим рассмотрением влияния 1) отражающей стенки, 2) поглощающей частицы стенки.

Представим себе, что на расстоянии s единиц вправо от точки $x = 0$ имеется отражающая стенка; так что частица, попавшая в какой-либо момент времени на эту стенку, при следующем толчке с вероятностью единица выбивается в том же направлении, откуда она пришла.

Для наглядности станем изображать положение частицы на плоскости (x, t). Путь частицы изобразится при этом в виде ломаной линии. При каждом толчке частица передвигается на единицу "вверх" и на единицу вправо или влево (каждый раз, когда $x < s$, с вероятностью половина). Если же $x = s$, то при очередном толчке частица сдвигается на единицу влево.

Для подсчета вероятности $\mathsf{P}\{\mu = m\}$ поступим следующим образом: мысленно откинем стенку и разрешим частице двигаться свободно, как если бы не было стенки. На рис. 13 показаны такие идеализированные пути, приводящие в точки A и A', симметрично расположенные относительно стенки. Ясно, что для того, чтобы реальная частица, двигаясь с отражениями, достигла точки A, необходимо и достаточно, чтобы частица, двигающаяся в идеализированной обстановке (без отражающей стенки), достигла либо точки A, либо точки A'. Но вероятность попасть в точку A в идеализированной обстановке, очевидно, равна

$$\mathsf{P}\{\mu = m\} = \frac{n!}{\left(\frac{m+n}{2}\right)!\left(\frac{n-m}{2}\right)!}\left(\frac{1}{2}\right)^n .$$

Точно так же вероятность попасть в точку A' равна (абсцисса точки A' равна $2s - m$)

$$\mathsf{P}\{\mu = 2s - m\} = \frac{n!}{\left(s + \frac{n-m}{2}\right)!\left(\frac{n+m}{2} - s\right)!}\left(\frac{1}{2}\right)^n .$$

Искомая вероятность, следовательно, равна

$$P_n(m; s) = \mathsf{P}\{\mu = m\} + \mathsf{P}\{\mu = 2s - m\} .$$

Если воспользоваться локальной предельной теоремой Муавра–Лапласа, то находим, что

$$P_n(m, s) \sim \frac{1}{\sqrt{2\pi n}}\left\{ e^{-\frac{m^2}{2n}} + e^{-\frac{(2s-m)^2}{2n}} \right\}$$

Это известная формула из теории броуновского движения. Она приобретает более симметричный вид, если начало координат поместить в точке $x = s$ и, следовательно, перейти к новой координате z по формуле $z = x - s$. В результате этой замены получим, что

$$P_n(z = k) = P_n\{k + s, s\} = \frac{1}{\sqrt{2\pi n}} \left\{ e^{-\frac{(k+s)^2}{2n}} + e^{-\frac{(k-s)^2}{2n}} \right\}$$

Мы перейдем теперь к рассмотрению третьей схематической задачи, когда

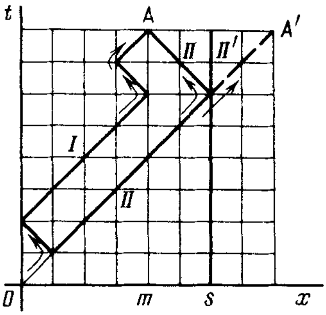

Рис. 13

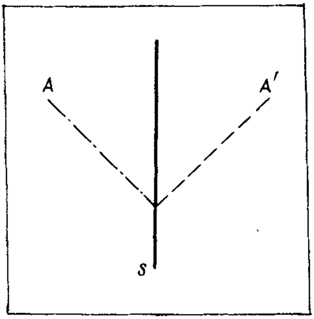

Рис. 14

на пути частицы поставлена в точке $x = s$ поглощающая перегородка. Частица, попавшая на перегородку, в дальнейшем движении участия уже не принимает. Очевидно, что в этом примере вероятность попасть в точку $x = m$ $(m < s)$ после n толчков будет меньше, чем $P_n(m)$ (т.е. меньше вероятности попадания в эту точку без поглощающей стенки); обозначим искомую вероятность символом $\overline{P}_n(m;\ s)$.

Для подсчета вероятности $\overline{P}_n(m;\ s)$ снова мысленно уберем поглощающую стенку и предоставим тем самым частице свободно двигаться по прямой. Частица, попавшая в некоторый момент времени в положение $x = s$, оказывается в последующие моменты времени справа и слева от прямой $x = s$ с одной и той же вероятностью. Точно так же после попадания на прямую $x = s$ частица с одной и той же вероятностью может попасть как в точку $A(m, n)$, так и в точку $A'(2s - m, n)$. Но в точку A' частица может попасть, только попав предварительно в положение $x = s$, поэтому для всякого пути, ведущего в точку A', имеется путь, симметричный относительно прямой $x = s$ и ведущий в точку A; точно так же для всякого запрещенного в действительном движении пути, приводящем в точку A, существует симметричный относительно прямой $x = s$ путь, приводящий в точку A' (рис. 14). При этом заметим, что мы рассматриваем симметрию путей, только начиная с момента попадания на прямую $x = s$. Проведенные рассуждения показывают нам, что из путей, приводящих в точку A в идеализированном движении, мы должны отбросить при подсчете числа благоприятствующих случаев в реальном движении ровно столько, сколько путей ведет в точку A'. Отсюда, очевидно, следует, что

$$\overline{P}_n(m, s) = \mathsf{P}\{\mu = m\} - \mathsf{P}\{\mu = 2s - m\}.$$

В силу локальной теоремы Муавра–Лапласа имеем:

$$\overline{P}_n(m, s) \sim \frac{1}{\sqrt{2\pi n}}\left\{e^{-\frac{m^2}{2n}} - e^{-\frac{(2s-m)^2}{2n}}\right\}$$

Упражнения

1. Рабочий обслуживает 12 однотипных станков. Вероятность того, что станок потребует к себе внимания рабочего в течение промежутка времени длительности τ равна 1/3. Чему равна вероятность того, что

а) за время τ 4 станка потребуют к себе внимания рабочего;

б) число требований к рабочему со стороны станков за время τ будет между 3 и 6 (включая границы)?

2. В некотором семействе имеется 10 детей. Считая вероятности рождений мальчика и девочки равными 1/2, найти вероятность того, что в семействе

а) 5 мальчиков и 5 девочек;

б) число мальчиков заключается между 3 и 8.

3. В обществе, состоящим из 4 человек, дни рождений трех приходятся на один месяц, а четвертого – на один из остальных одиннадцати. Считая вероятность рождения в течение каждого из месяцев для каждого лица равной 1/12, найти вероятность того, что

а) указанные три лица родились в январе, а четвертое лицо в октябре;

б) три лица родились в каком-то одном месяце, а четвертое в каком-то из остальных одиннадцати.

4. При 14 400 бросаниях монеты герб выпал 7428 раз. Как вероятно столь большое или большее уклонение числа выпадений герба от np, если монета симметрична (т.е. вероятность выпадения герба в каждом испытании равна 1/2)?

5. К электросети подключено n приборов, каждый мощностью a киловатт и потребляет в данный момент энергию с вероятностью p. Найти вероятность того, что потребляемая в данный момент мощность

а) окажется меньше чем nap;

б) превзойдет $rnap$ ($r > 0$) при условии, что np велико.

6. В одном из учебных заведений обучаются 730 студентов. Вероятность того, что день рождения наудачу взятого по списку студента приходится на определенный день года, равна 1/365 для каждого из 365 дней. Найти

а) наиболее вероятное число студентов, родившихся 1 января.

б) вероятность того, что найдутся три студента, имеющие один и тот же день рождения.

7. Известно, что вероятность выпуска сверла повышенной хрупкости (брак) равна 0,02. Сверла укладываются в коробки по 100 штук. Чему равна вероятность того, что

а) в коробке не окажется бракованных сверл;

б) число бракованных сверл окажется не более 3;

в) сколько нужно класть в коробку сверл, чтобы с вероятностью, не меньшей 0,9, в ней было не менее 100 исправных?

Указание. Воспользоваться распределением Пуассона.

8. В страховом обществе застраховано 10000 лиц одного возраста и одной социальной группы. Вероятность смерти в течение года для каждого лица равна 0,006. Каждый застрахованный вносит 1 января 12 руб. страховых и в случае смерти его родственники получают от общества 1000 руб. Чему равна вероятность того, что

а) общество потерпит убытки;

б) получит прибыль, не меньшую 40000, 60 000, 80 000 руб.?

9. Доказать теорему: если P и P' – вероятности наиболее вероятного числа появлений события A в n и $n+1$ независимых испытаниях (в каждом из испытаний $\mathsf{P}(A)=p$), то $P' \leqslant P$; равенство исключается, если $(n+1)p$ – не целое число.

10. В схеме Бернулли $p = 1/2$. Доказать, что:

а) $$\frac{1}{2\sqrt{n}} \leqslant P_{2n}(n) \leqslant \frac{1}{\sqrt{2n+1}};$$

б) $$\lim_{n\to\infty} \frac{P_{2n}(n \pm h)}{P_{2n}(n)} = e^{-z^2},$$

если $z = h/\sqrt{n}$ $(0 \leqslant z < +\infty)$.

11. Доказать, что при $npq \geqslant 25$

$$P_n(m) = \frac{1}{\sqrt{2npq}}\, e^{-\frac{z^2}{2}} \left[1 + \frac{(q-p)(z^3-3z)}{6\sqrt{npq}}\right] + \Delta,$$

где

$$z = \frac{m-np}{\sqrt{npq}}, \qquad |\Delta| < \frac{0{,}15 + 0{,}25\,|p-q|}{\sqrt{(npq)^3}}\,|z|\,e^{-\frac{3}{2}\sqrt{npq}}$$

12. Произведено n независимых испытаний. Вероятность появления события A и i-м испытании равна p_i; $P_n(m)$ – вероятность m-кратного появления события A в n испытаниях. Доказать, что

а) $$\frac{P_n(1)}{P_n(0)} \geqslant \frac{P_n(2)}{P_n(1)} \geqslant \ldots \geqslant \frac{P_n(n)}{P_n(n-1)};$$

б) $P_n(m)$ сначала возрастает, а затем убывает (если только $P_n(0)$ или $P_n(n)$ сами не являются максимальными).

13. Доказать, что при $x > 0$ функция $\int_x^\infty e^{-\frac{z^2}{2}}\,dz$ удовлетворяет неравенствам

$$\frac{x}{1+x^2}\, e^{-\frac{1}{2}x^2} \leqslant \int_x^\infty e^{-\frac{1}{2}z^2}\,dz \leqslant \frac{1}{x}\, e^{-\frac{1}{2}x^2}.$$

14. Задача Банаха. Некий математик носит с собой две коробки спичек. Каждый раз, когда он хочет достать спичку, он выбирает наугад одну из коробок. Найти вероятность того, что когда математик вынет пустую коробку, в другой коробке окажется r спичек ($r = 0, 1, 2, \ldots, n$; n – число спичек, бывших первоначально в каждой из коробок).

15. К линии электропередачи подключено n механизмов. Вероятность того, что механизм, потребляющий энергию в момент времени t прекратит ее потребление до момента $t + \Delta t$, равна $\alpha \Delta t + o(\Delta t)$. Если в момент t механизм не потребляет энергии, то вероятность того, что он станет ее потреблять до момента $t + \Delta t$ равна $\beta \Delta t + v(\Delta t)$ независимо от работы других механизмов. Составить дифференциальные уравнения, которым удовлетворяют вероятности $P_r(t)$ того, что в момент t энергию потребляют r механизмов.

Примечание. Легко указать конкретные осуществления условий этой задачи: движение трамваев, электросварка, потребление энергии станками с автоматическим выключением и пр.

16. Один рабочий обслуживает n однотипных станков-автоматов. Если в момент t станок работает, то вероятность того, что он потребует обслуживания до момента $t + \Delta t$ равна $\alpha \Delta t + o(\Delta t)$. Если в момент t рабочий обслуживает какой-нибудь станок, то вероятность того, что он закончит обслуживание до момента $t + \Delta t$, равна $\beta \Delta t + o(\Delta t)$. Составить дифференциальные уравнения, которым удовлетворяют вероятности $P_r(t)$ того, что в момент t работают $n - r$ станков, один обслуживается и $r - 1$ ожидают очереди на обслуживание ($P_0(t)$ – вероятность того, что все станки работают).

Примечание. Нетрудно аналогичным путем составить дифференциальные уравнения для более сложной задачи, когда N станков обслуживает бригада из k рабочих. Для практических целей важно сравнить экономичность той и другой системы организации труда. С этой целью следует изучить установившийся режим, т.е. рассмотреть вероятности $P_r(t)$ при $t \to \infty$.

Оказывается, работа бригады, обслуживающей kn станков выгоднее как в смысле лучшего использования рабочего времени станка, так и рабочего времени рабочего, чем обслуживание одним рабочим n станков.

ГЛАВА 3

ЦЕПИ МАРКОВА

§ 15. Определение цепи Маркова

Непосредственным обобщением схемы независимых испытаний является схема так называемых цепей Маркова, впервые систематически изученная известным русским математиком А.А. Марковым. Мы ограничимся изложением элементов его теории.

Представим себе, что производится последовательность испытаний, в каждом из которых может осуществиться одно и только одно из k несовместимых событий $A_1^{(s)}, A_2^{(s)}, \ldots, A_k^{(s)}$ (верхний индекс, как и в предыдущей главе, обозначает номер испытания). Мы скажем, что последовательность испытаний образует *цепь Маркова*, точнее, *простую цепь Маркова*, если *условная вероятность в* $s+1$*-м испытании* ($s = 1, 2, 3, \ldots$) *осуществиться событию* $A_i^{(s+1)}$ ($i = 1, 2, \ldots, k$) *зависит только от того, какое событие произошло при* s*-м испытании и не изменяется от добавочных сведений о том, какие события происходили в более ранних испытаниях.*

Часто при изложении теории цепей Маркова придерживаются иной терминологии и говорят о некоторой физической системе S, которая в каждый момент времени может находиться в одном из состояний $A_1, A_2, \ldots, A_k$ и меняет свое состояние только в моменты $t_1, t_2, \ldots, t_n, \ldots$ Для цепей Маркова вероятность перейти в какое-либо состояние A_i ($i = 1, 2, \ldots, k$) в момент t_s зависит только от A_i и того, в каком состоянии система находилась в момент t ($t_{s-1} < t < t_s$), и не изменяется от того, что становятся известными ее состояния в более ранние моменты.

Для иллюстрации рассмотрим два схематических примера.

Пример 1. Представим себе, что частица, находящаяся на прямой, движется по этой прямой под влиянием случайных толчков, происходящих в моменты $t_1, t_2, t_3, \ldots$ Частица может находиться в точках с целочисленными координатами a, $a+1$, $a+2, \ldots, b$; в точках a и b находятся отражающие стенки. Каждый толчок перемещает частицу вправо с вероятностью p и влево с вероятностью $q = 1 - p$, если только частица не находится у стенки. Если же частица находится у стенки, то любой толчок переводит ее на единицу внутрь промежутка между стенками. Мы видим, что проведенный пример блуждания частицы представлят собой типичную цепь Маркова. Точно также можно бы было рассмотреть случай, когда частица прилипает к одной из стенок или к обеим из них.

Пример 2. В модели Бора атома водорода электрон может находиться на одной из допустимых орбит. Обозначим через A_i событие, состоящее в том, что электрон находится на i-й орбите. Предположим далее, что изменение состояния атома может наступать только в моменты $t_1, t_2, t_3, \ldots$ (в действительности эти моменты представляют собой случайные величины). Вероятность перехода с i-й орбиты на j-ю в момент t_s зависит только от i и j (разность $j - i$ зависит от количества энергии, на которую изменился заряд атома в момент t_s) и не зависит от того, на каких орбитах находился электрон в прошлом.

Последний пример представляет собой цепь Маркова с бесконечным (правда, только в принципе) числом состояний; этот пример был бы несравненно ближе к реальной обстановке, если бы моменты перехода нашей системы в новое состояние могли меняться непрерывно.

§ 16. Матрица перехода

Мы ограничимся далее изложением простейших фактов для *однородных цепей Маркова,* в которых условия вероятность появления события $A_j^{(s+1)}$ в $(s + 1)$-м испытании при условии, что в s-м испытании осуществилось событие $A_i^{(s)}$, не зависит от номера испытания. Мы назовем эту вероятность *вероятностью перехода* и обозначим буквой p_{ij}; в этом обозначении первый индекс всегда будет обозначать результат предшествующего испытания, а второй индекс указывает, в какое состояние перейдет система в последующий момент времени.

Полная вероятностная картина возможных изменений, осуществляющихся при переходе от одного испытания к непосредственно следующему, задается матрицей

$$\pi_1 = \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1k} \\ p_{21} & p_{22} & \ldots & p_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ p_{k1} & p_{k2} & \ldots & p_{kk} \end{pmatrix},$$

составленной из вероятностей перехода, которую мы будем называть *матрицей перехода.*

Отметим, каким условиям должны удовлетворять элементы этой матрицы. Прежде всего они, как вероятности, должны быть неотрицательными числами, т.е. при всех i и j

$$0 \leqslant p_{ij} \leqslant 1.$$

Далее из того, что при переходе из состояния $A_i^{(s)}$ в s-м испытании система обязательно переходит в одно и только в одно из состояний $A_j^{(s+1)}$, в $(s + 1)$-м испытании вытекает равенство

$$\sum_{j=1}^{k} p_{ij} = 1 \quad (i = 1, 2, \ldots, k).$$

Таким образом, сумма элементов каждой строки матрицы перехода равна единице.

Наша первая задача в теории цепей Маркова состоит в определении вероятности перехода из состояния $A_i^{(s)}$ в s-м испытании в состояние $A_j^{(s+n)}$ через n испытаний. Обозначим эту вероятность знаком $P_{ij}(n)$.

Рассмотрим какое-нибудь промежуточное испытание с номером $s + m$. В этом испытании осуществится какое-то одно из возможных событий $A_r^{(s+m)}$ $(1 \leqslant r \leqslant k)$. Вероятность такого перехода, согласно с только что введенными обозначениями, равна $P_{ir}(m)$. Вероятность же перехода из состояния $A_r^{(s+m)}$ в состояние $A_j^{(s+n)}$ равна $P_{rj}(n - m)$. По формуле полной вероятности

$$P_{ij}(n) = \sum_{r=1}^{k} P_{ir}(m) \cdot P_{rj}(n - m). \tag{1}$$

Обозначим через π_n матрицу перехода через n испытаний

$$\pi_n = \begin{pmatrix} P_{11}(n) P_{12}(n) & \ldots & P_{1k}(n) \\ \ldots\ldots & \ldots & \ldots\ldots \\ P_{k1}(n) P_{k2}(n) & \ldots & P_{kk}(n) \end{pmatrix}.$$

Согласно (1) между матрицами π_s с различными индексами существует соотношение

$$\pi_n = \pi_m \cdot \pi_{n-m} \qquad (0 < m < n).$$

В частности, при $n = 2$ находим, что

$$\pi_2 = \pi_1 \cdot \pi_1 = \pi_1^2;$$

при $n = 3$

$$\pi_3 = \pi_1 \cdot \pi_2 = \pi_2 \cdot \pi_1 = \pi_1^3;$$

и вообще при любом n

$$\pi_n = \pi_1^n.$$

Отметим частный случай формулы (1): при $m = 1$

$$P_{ij}(n) = \sum_{r=1}^{n} p_{ir} P_{rj}(n-1).$$

В качестве упражнения предлагается читателю написать матрицу перехода для первого примера предыдущего параграфа.

§ 17. Теорема о предельных вероятностях

Т е о р е м а. *Если при некотором $s > 0$ все элементы матрицы перехода π_s положительны, то существуют такие постоянные числа p_j $(j = 1, 2, \ldots, k)$, что независимо от индекса i имеют место равенства*

$$\lim_{n\to\infty} P_{ij}(n) = p_j.$$

Д о к а з а т е л ь с т в о. Идея доказательств этой теоремы весьма проста: сначала устанавливается, что наибольшая из вероятностей $P_{ij}(n)$ с ростом n не может возрастать, а наименьшая не может убывать; далее показывается, что максимум разности $P_{ij}(n) - P_{lj}(n)$ $(i, l = 1, 2, \ldots, k)$ стремится к нулю, когда $n \to \infty$. Этим доказательство теоремы, очевидно, завершается. Действительно, в силу известной теоремы о пределе монотонной последовательности мы заключаем из первых двух указанных свойств вероятностей $P_{ij}(n)$, что существуют

$$\lim_{n\to\infty} \min_{1\leqslant i\leqslant k} P_{ij}(n) = \overline{p}_j$$

и

$$\lim_{n\to\infty} \max_{1\leqslant i\leqslant k} P_{ij}(n) = \overline{\overline{p}}_j.$$

А так как в силу третьего из указанных свойств

$$\lim_{n\to\infty} \max_{1\leqslant i,l\leqslant k} |P_{ij}(n) - P_{lj}(n)| = 0,$$

то

$$\overline{p}_j = \overline{\overline{p}}_j = p_j.$$

Мы перейдем теперь к осуществлению намеченного плана. Заметим прежде всего, что при $n > 1$ имеет место неравенство

$$P_{ij}(n) = \sum_{l=1}^{k} p_{il} P_{lj}(n-1) \geqslant \min_{1\leqslant l\leqslant k} P_{lj}(n-1) \sum_{l=1}^{k} p_{il} =$$

$$= \min_{1\leqslant l\leqslant k} P_{lj}(n-1).$$

Это неравенство имеет место при каждом i, в частности при том, при котором

$$P_{ij}(n) = \min_{1 \leqslant l \leqslant k} P_{lj}(n).$$

Таким образом,

$$\min_{1 \leqslant i \leqslant k} P_{ij}(n) \geqslant \min_{1 \leqslant i \leqslant k} P_{ij}(n-1).$$

Подобным же путем легко обнаружить, что

$$\max_{1 \leqslant i \leqslant k} P_{ij}(n) \leqslant \max_{1 \leqslant i \leqslant k} P_{ij}(n-1).$$

Мы можем считать, что $n > s$, и поэтому имеем право записать по формуле (1), что

$$P_{ij}(n) = \sum_{r=1}^{k} P_{ir}(s) \cdot P_{rj}(n-s).$$

Рассмотрим разность

$$P_{ij}(n) - P_{lj}(n) = \sum_{r=1}^{k} P_{ir}(s) P_{rj}(n-s) - \sum_{r=1}^{k} P_{lr}(s) \cdot P_{rj}(n-s) =$$

$$= \sum_{r=1}^{k} [P_{ir}(s) - P_{lr}(s)]\, P_{rj}(n-s).$$

Обозначим положительные разности $P_{ir}(s) - P_{lr}(s)$ символом $\beta_{il}^{(r)}$, а неположительные разности через $-\beta_{il}'^{(r)}$. Так как

$$\sum_{r=1}^{k} P_{ir}(s) = \sum_{r=1}^{k} P_{lr}(s) = 1,$$

то

$$\sum_{r=1}^{k} [P_{ir}(s) - P_{lr}(s)] = \sum_{(r)} \beta_{il}^{(r)} - \sum_{(r)} \beta_{il}'^{(r)} = 0. \tag{2}$$

Из этого равенства заключаем, что

$$h_{il} = \sum_{(r)} \beta_{il}^{(r)} = \sum_{(r)} \beta_{il}'^{(r)}.$$

Так как по предположению при всех i и r $(i, r = 1, 2, 3, \ldots, k)$ $P_{ir}(s) > 0$, то

$$\sum_{(r)} \beta_{il}^{(r)} < \sum_{r=1}^{k} P_{il}(s) = 1.$$

Таким образом,

$$0 \leqslant h_{il} < 1.$$

Пусть

$$h = \max_{1 \leqslant i, l \leqslant k} h_{il}.$$

Так как число возможных исходов конечно, то наряду с величинами h_{il} величина h удовлетворяет неравенствам

$$0 \leqslant h < 1. \tag{3}$$

Из (1) находим, что при любых i и l $(i, l = 1, 2, \ldots, k)$

$$|P_{ij}(n) - P_{lj}(n)| = |\sum_{(r)} \beta_{il}^{(r)} P_{rj}(n-s) - \sum_{(r)} \beta_{il}'^{(r)} P_{rj}(n-s)| \leqslant$$

$$\leqslant |\max_{1 \leqslant r \leqslant k} P_{rj}(n-s) \sum_{(r)} \beta_{il}^{(r)} - \min_{1 \leqslant r \leqslant k} P_{rj}(n-s) \sum_{(r)} \beta_{il}'^{(r)}| \leqslant$$

$$\leqslant h\, |\max_{1 \leqslant r \leqslant k} P_{rj}(n-s) - \min_{1 \leqslant r \leqslant k} P_{rj}(n-s)| \leqslant$$

$$\leqslant h \max_{1 \leqslant i, l \leqslant k} |P_{ij}(n-s) - P_{lj}(n-s)|$$

и, следовательно, также

$$\max_{1 \leqslant i, l \leqslant k} |P_{ij}(n) - P_{lj}(n)| \leqslant h \max_{1 \leqslant i, l \leqslant k} |P_{ij}(n-s) - P_{lj}(n-s)|.$$

Применив это неравенство $\left[\dfrac{n}{s}\right]$ раз, найдем, что

$$\max_{1 \leqslant i, l \leqslant k} |P_{ij}(n) - P_{lj}(n)| \leqslant h^{\left[\frac{n}{s}\right]} \max_{1 \leqslant i, l \leqslant k} \left| P_{ij}\left(n - \left[\frac{n}{s}\right] s\right) - \right.$$

$$\left. - P_{lj}\left(n - \left[\frac{n}{s}\right] s\right)\right|.$$

Так как всегда $|P_{ij}(m) - P_{lj}(m)| \leqslant 1$, то ясно, что

$$\max_{1 \leqslant i, l \leqslant k} |P_{ij}(n) - P_{lj}(n)| \leqslant h^{\left[\frac{n}{s}\right]}.$$

При $n \to \infty$ также $\left[\frac{n}{s}\right] \to \infty$, поэтому в силу (3) отсюда следует, что

$$\lim_{n\to\infty} \max_{1\leqslant i,l\leqslant k} |P_{ij}(n) - P_{lj}(n)| = 0.$$

Из доказанного заключаем также, что

$$\sum_{j=1}^{k} p_j = 1.$$

Действительно,

$$\sum_{j=1}^{k} p_j = \lim_{n\to\infty} \sum_{j=1}^{k} P_{ij}(n) = \lim_{n\to\infty} 1 = 1.$$

Таким образом, на величины p_j можно смотреть как на вероятности появления исхода $A_j^{(n)}$ при n-м испытании, когда n велико.

Физический смысл доказанной теоремы ясен: вероятность системе находиться в состоянии A_j практически не зависит от того, в каком состоянии она находилась в далеком прошлом.

Только что обнаруженная теорема была впервые доказана творцом теории цепных зависимостей А.А. Марковым; она явилась первым строго доказанным результатом среди так называемых эргодических теорем, играющих важную роль в современной физике и инженерном деле.

Упражнения

1. Вероятности перехода даются матрицей

$$\pi_1 = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \end{pmatrix}.$$

Чему равно число состояний? Найти вероятности перехода из состояния в состояние за два шага.

2. Электрон может находиться на одной из счетного множества орбит в зависимости от наличной энергии. Переход с i-й орбиты на j-ю происходит за одну секунду с вероятностью $c_i e^{-\alpha|i-j|}$.

Найти: а) вероятности перехода за две секунды, б) постоянные c_i.

3. Вероятности перехода даются матрицей

$$\pi_1 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Применима ли в данном случае эргодическая теорема Маркова? Если да, то найти предельные вероятности.

ГЛАВА 4

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

§ 18. Основные свойства функций распределения

Одним из основных понятий теории вероятностей является понятие случайной величины. Прежде чем переходить к формальному его определению, мы остановимся на рассмотрении примеров.

Число космических частиц, попадающих на определенный участок земной поверхности в течение промежутка времени определенной длины, подвержено значительным колебаниям в зависимости от многих случайных обстоятельств.

Число вызовов, поступивших от абонентов на телефонную станцию в течение определенного промежутка времени, не остается постоянным, а подвержено значительным случайным колебаниям.

Размер уклонения точки падения снаряда от центра цели определяется большим количеством разнообразных причин, носящих случайный характер. В результате в теории стрельбы вынуждены считаться с явлением рассеивания снарядов около центра цели как со случайным явлением и рассматривать указанные уклонения как случайные величины.

Скорость молекулы газа не остается неизменной, а меняется в зависимости от столкновений с другими молекулами. Этих столкновений очень много даже в течение короткого промежутка времени. Зная скорость молекулы в данный момент, нельзя с полной определенностью указать ее значение, скажем, через 0,01 или 0,001 секунды. Изменение скорости молекулы носит случайный характер.

Приведенные примеры показывают с достаточной определенностью, что со случайными величинами приходится иметь дело в самых разнообразных областях науки и техники. Возникает естественная и притом весьма важная задача создания методов изучения случайных величин.

Несмотря на всю разнородность конкретного содержания приведенных нами примеров, все они с точки зрения математики представляют одну и ту же картину. А именно, в каждом примере мы имеем дело с величиной, так или иначе характеризующей исследуемое явление. Каждая из этих величин под влиянием случайных обстоятельств способна принимать различные значения. Заранее предсказать, какое значение примет эта величина, нельзя, так как оно меняется случайным образом от испытания к испытанию.

Таким образом, для того чтобы знать случайную величину, прежде всего необходимо знать те значения, которые она может принимать. Однако одного перечня значений случайной величины еще недостаточно, чтобы по ним можно было делать какие-либо существенные выводы. Действительно, если в третьем примере рассмотреть газ при разных температурах, то возможные значения скоростей молекул останутся теми же самыми, тогда как состояния газа будут различны. Таким образом, для задания случайной величины необходимо знать не только, какие значения может она принимать, но и как часто, т.е. с какой вероятностью она принимает эти значения.

Разнообразие случайных величин весьма велико. Число принимаемых ими значений может быть конечным, счетным или несчетным; значения могут быть расположены дискретно или заполнять интервалы сплошь или же не заполнять интервалы, но располагаться всюду плотно. Для того чтобы задавать вероятности значений случайных величин, столь различных по своей природе, и притом задавать их одним и тем же способом, в теории вероятностей вводят понятие функции распределения случайной величины.

Пусть ξ – случайная величина и x - произвольное действительное число. Вероятность того, что ξ примет значение, меньшее чем x, называется *функцией распределения вероятностей* случайной величины ξ:

$$F(x) = \mathsf{P}\{\xi < x\}.$$

Условимся в дальнейшем, как правило, случайные величины обозначать греческими буквами, а принимаемые ими значения – строчными латинскими.

Резюмируем сказанное: *случайной величиной* называется величина, значения которой зависят от случая и для которой определена функция распределения вероятностей*).

Рассмотрим примеры функций распределения.

Пример 1. Обозначим через μ число появлений события A в последовательности n независимых испытаний, в каждом из которых вероятность его появления постоянна и равна p. В зависимости от случая μ может принимать все целочисленные значения от 0 до n включительно. Согласно результатам главы 2

$$P_n(m) = \mathsf{P}\{\mu = m\} = C_n^m p^m q^{n-m}$$

Само собой разумеется, что этими словами мы не дали математического определения новому понятию, а только описали то общее представление, которое складывается у человека, знакомящегося с реальными примерами случайных величин.

---

*) На с. 119 дано формализованное определение случайной величины.

Функция распределения величины μ определяется следующим способом:

$$F(x)=\begin{cases}0 & \text{для} \quad x\leqslant 0,\\ \sum\limits_{k<x} P_n(k) & \text{для} \quad 0<x\leqslant n,\\ 1 & \text{для} \quad x>n.\end{cases}$$

Функция распределения представляет собой ступенчатую линию со скачками в точках $x=0,1,2,\ldots,n$; скачок в точке $x=k$ равен $P_n(k)$.

Рассмотренный пример показывает, что так называемая схема Бернулли может быть включена в общую теорию случайных величин.

Пример 2. Пусть случайная величина ξ принимает значения $0, 1, 2, \ldots$ с вероятностями

$$p_n = \mathsf{P}\{\xi=n\}=\frac{\lambda^n e^{-\lambda}}{n!} \qquad (n=0,1,2,\ldots),$$

где $\lambda>0$ – постоянная. Функция распределения величины ξ представляет собой как бы лестницу с бесконечным числом ступенек, со скачками во всех неотрицательных целочисленных точках. Величина скачка в точке $x=n$ равна p_n; при $x\leqslant 0$ имеем $F(x)=0$. Про случайную величину, рассмотренную в настоящем примере, говорят, что она распределена по *закону Пуассона.*

Пример 3. Мы скажем, что случайная величина *нормально распределена,* если ее функция распределения имеет вид

$$\Phi(x)=C\int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}}\,dz,$$

где $C>0$, $\sigma>0$, a – постоянные. Впоследствии мы установим связь между постоянными σ и C и выясним теоретико-вероятностный смысл параметров a и σ. Нормально распределенные случайные величины играют особо важную роль в теории вероятностей и ее приложениях, в дальнейшем у нас будет много поводов убедиться в этом.

Заметим, что если в двух первых рассмотренных нами примерах случайные величины могли принимать только конечное или счетное множество значений (*дискретные* величины), то случайные величины, распределенные по нормальному закону могут принимать значения из любого интервала. Действительно, как мы увидим ниже, вероятность нормально распределенной случайной величине принять значение, заключающееся в интервале $x_1\leqslant\xi<x_2$, равна

$$\Phi(x_2)-\Phi(x_1)=C\int_{x_1}^{x_2} e^{-\frac{(z-a)^2}{2\sigma^2}}\,dz$$

и, следовательно, при любых x_1 и x_2 $(x_1\neq x_2)$ положительна.

После сделанных нами замечаний интуитивного характера можно перейти и к изложению принятого теперь строго формального определения случайной величины.

В соответствии с общим аксиоматическим понятием случайного события, мы будем исходить из множества элементарных событий Ω. Каждому элементарному событию ω поставим в соответствие некоторое число

$$\xi = f(\omega).$$

Мы скажем, что ξ есть случайная величина, если функция $f(\omega)$ измерима относительно введенной в рассматриваемом множестве Ω вероятности. Иначе говоря, мы требуем, чтобы для каждого измеримого по Борелю множества A_ξ значений ξ множество A_ω тех ω, для которых $f(\omega) \subset A_\xi$ принадлежало множеству F случайных событий и, следовательно, для него была бы определена вероятность

$$\mathsf{P}\{\xi \subset A_\xi\} = \mathsf{P}\{A_\omega\}.$$

В частности, если множество A_ξ совпадает с полупрямой $\xi < x$, то вероятность $\mathsf{P}\{A_\omega\}$ есть функция переменного x

$$\mathsf{P}\{\xi < x\} = \mathsf{P}\{A_\omega\} = F(x),$$

которую мы назвали функцией распределения случайной величины ξ.

Пример 4. Рассмотрим последовательность n независимых испытаний, в каждом из которых вероятность появления события A постоянна и равна p. В этом примере элементарные события состоят из последовательностей появлений и непоявлений события A в n испытаниях. Так, одним из элементарных событий будет появление события A в каждом из испытаний. Всего, как нетрудно подсчитать, будет 2^n элементарных событий.

Определим функцию $\mu = f(\omega)$ от элементарного события ω так: она равна числу появлений события A в элементарном событии ω. Согласно результатам главы 2

$$\mathsf{P}\{\mu = k\} = P_n(k) = C_n^k p^k q^{n-k}.$$

Измеримость функции $\mu = f(\omega)$ в поле вероятностей непосредственно очевидна. Отсюда, согласно определению, заключаем, что μ есть случайная величина.

Пример 5. Произведены три наблюдения за положением молекулы, двигающейся по прямой линии. Множество элементарных событий состоит из точек трехмерного евклидова пространства R_3. Множество случайных событий F состоит из всевозможных борелевских множеств пространства R_3.

Для каждого случайного события A определим вероятность $\mathsf{P}\{A\}$ посредством равенства

$$\mathsf{P}\{A\} = \frac{1}{(\sigma\sqrt{2\pi})^3} \iiint_A e^{-\frac{1}{2\sigma^2}[(x_1-a)^2+(x_2-a)^2+(x_3-a)^2]}\, dx_1 dx_2 dx_3.$$

Рассмотрим теперь функцию $\xi = f(\omega)$ элементарного события, определенную посредством равенства

$$\xi = \frac{1}{3}(x_1 + x_2 + x_3).$$

Эта функция измерима относительно введенной нами вероятности, поэтому ξ является случайной величиной. Для нее функция распределения равна

$$F(x) = \mathsf{P}\{\xi < x\} = \frac{1}{(\sigma\sqrt{2\pi})^3} \iiint_{x_1+x_2+x_3<3x} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{3}(x_k-a)^2} \times$$
$$\times\, dx_1 dx_2 dx_3 = \frac{1}{\sigma\sqrt{\frac{2}{3}\pi}} \int_{-\infty}^{x} e^{-\frac{3(z-a)^2}{2\sigma^2}}\, dz.$$

С только что развитой точки зрения действия над случайными величинами сводятся к известным операциям над функциями. Так, если ξ_1 и ξ_2 являются случайными величинами, т.е. измеримыми относительно введенной вероятности функциями

$$\xi_1 = f_1(\omega), \qquad \xi_2 = f_2(\omega),$$

то любая борелевская функция от них также является случайной величиной. Для примера

$$\zeta = \xi_1 + \xi_2$$

измерима относительно введенной вероятности и потому является случайной величиной.

Позднее мы разовьем только что сделанное замечание и получим ряд важных для применений результатов. В частности, мы выведем формулы для функции распределения суммы и частного двух случайных величин по распределению слагаемых.

При помощи функции распределения случайной величины ξ можно определить вероятность неравенства $x_1 \leqslant \xi < x_2$ при любых x_1 и x_2. В самом деле, если через A обозначить событие, состоящее в том, что ξ примет значение, меньшее чем x_2, через B – событие, состоящее в том, что $\xi < x_1$, и, наконец, через C – событие $x_1 \leqslant \xi < x_2$, то, очевидно, имеет место следующее равенство:

$$A = B + C.$$

Так как события B и C несовместимы, то

$$\mathsf{P}(A) = \mathsf{P}(B) + \mathsf{P}(C).$$

Но

$$\mathsf{P}(A) = F(x_2), \quad \mathsf{P}(B) = F(x_1), \quad \mathsf{P}(C) = \mathsf{P}\{x_1 \leqslant \xi < x_2\},$$

поэтому

$$\mathsf{P}\{x_1 \leqslant \xi < x_2\} = F(x_2) - F(x_1). \tag{1}$$

Так как, по определению вероятность есть неотрицательное число, то из равенства (1) следует, что при любых x_1 и x_2 $(x_2 > x_1)$ имеет место неравенство

$$F(x_2) \geqslant F(x_1),$$

т.е. что *функция распределения любой случайной величины есть неубывающая функция.*

Очевидно, далее, что функция распределения $F(x)$ при любом x удовлетворяет неравенству

$$0 \leqslant F(x) \leqslant 1. \tag{2}$$

Мы скажем, что функция распределения $F(x)$ имеет при $x = x_0$ *скачок*, если

$$F(x_0 + 0) - F(x_0 - 0) = C_0 > 0.$$

Функция распределения может иметь не более чем счетное множество скачков. В самом деле, скачков размера, большего 1/2, функция распределения может иметь не более одного; скачков размера от одной четвертой до половины $(1/4 < C_0 \leqslant 1/2)$ – не более трех. Вообще скачков размером от $1/2^n$ до $1/2^{n-1}$ может быть не более чем $2^n - 1$. Совершенно ясно, что мы можем пронумеровать все скачки, расположив их по величине, начиная с больших значений и повторяя равные значения столько раз, сколько скачков этой величины имеет функция $F(x)$.

Установим еще несколько общих свойств функций распределения. Определим $F(-\infty)$ и $F(+\infty)$ равенствами

$$F(-\infty) = \lim_{n \to +\infty} F(-n), \quad F(+\infty) = \lim_{n \to \infty} F(+n)$$

и докажем, что

$$F(-\infty) = 0, \quad F(+\infty) = 1.$$

Действительно, так как неравенство $\xi < +\infty$ достоверно, то

$$\mathsf{P}\{\xi < +\infty\} = 1.$$

Обозначим через Q_k событие, состоящее в том, что $k - 1 \leqslant \xi < k$. Так как событие $\xi < +\infty$ эквивалентно сумме событий Q_k, то на основании расширенной аксиомы сложения

$$\mathsf{P}\{\xi < +\infty\} = \sum_{k=-\infty}^{\infty} \mathsf{P}\{Q_k\}.$$

Следовательно, при $n \to \infty$

$$\sum_{k=1-n}^{n} \mathsf{P}\{Q_k\} = \sum_{k=1-n}^{n} [F(k) - F(k-1)] = F(n) - F(-n) \to 1.$$

Отсюда, принимая во внимание неравенства (2), заключаем, что при $n \to \infty$

$$F(-n) \to 0, \quad F(+n) \to 1.$$

Функция распределения непрерывна слева.

Выберем какую-нибудь возрастающую последовательность $x_0 < x_1 < < x_2 < \ldots < x_n < \ldots$, сходящуюся к x.

Обозначим через A_n событие $\{x_n \leqslant \xi < x\}$. Тогда ясно, что $A_i \subset A_j$, при $i > j$, и произведение всех событий A_n есть невозможное событие. По аксиоме непрерывности должно быть

$$\lim_{n \to \infty} \mathsf{P}(A_n) = \lim_{n \to \infty} \{F(x) - F(x_n)\} =$$

$$= F(x) - \lim_{n \to \infty} F(x_n) = F(x) - F(x-0) = 0,$$

что и требовалось доказать.

Точно так же можно доказать, что

$$\mathsf{P}\{\xi \leqslant x\} = F(x+0).$$

Мы видим, таким образом, что *каждая функция распределения является неубывающей, непрерывной слева и удовлетворяющей условиям $F(-\infty) = 0$ и $F(+\infty) = 1$ функцией.* Верно и обратное: *каждая функция, удовлетворяющая перечисленным условиям, может рассматриваться как функция распределения некоторой случайной величины.*

Заметим, что в то время как каждая случайная величина однозначно определяет свою функцию распределения, существует сколько угодно различных случайных величин, имеющих одну и ту же функцию распределения. Так, если ξ принимает два значения -1 и $+1$, каждое с вероятностью $1/2$ и $\eta = -\xi$, то ясно, что всегда ξ отлична от η. Тем не менее обе эти случайные величины имеют одну и ту же функцию распределения

$$F(x) = \begin{cases} 0 & \text{при} \quad x \leqslant -1, \\ 1/2 & \text{при} \quad -1 < x \leqslant 1, \\ 1 & \text{при} \quad x > 1. \end{cases}$$

§ 19. Непрерывные и дискретные распределения

Иногда поведение случайной величины характеризуют не заданием ее функции распределения, а каким-либо иным способом. Всякая такая характеристика носит название *закона распределения* случайной величины, если только по определенным правилам можно получить из нее функцию распределения. Так, законом распределения будет функция интервала $P\{x_1, x_2\}$, представляющая собой вероятность неравенства $x_1 \leqslant \xi < x_2$. Действительно, зная $P\{x_1, x_2\}$, мы можем найти функцию распределения по формуле

$$F(x) = P\{-\infty, x\}.$$

Мы уже знаем, что и по $F(x)$ можно найти для любых x_1 и x_2 функцию $P\{x_1, x_2\}$: $P\{x_1, x_2\} = F(x_2) - F(x_1)$.

Часто в качестве закона распределения полезно брать функцию множества $P\{E\}$, определенную для всех борелевских множеств и представляющую собой вероятность того, что случайная величина ξ примет значение, принадлежащее множеству E. Вероятность $P\{E\}$, в силу расширенной аксиомы сложения, есть вполне аддитивная функция множества, т.е. для любого множества E, представляющего собой сумму конечного или счетного числа непересекающихся множеств E_k:

$$P\{E\} = \sum P\{E_k\}.$$

Из всевозможных случайных величин мы выделим прежде всего те, которые могут принимать только конечное или счетное множество значений. Такие величины мы будем называть *дискретными*. Для полной вероятностной характеристики дискретной случайной величины ξ, принимающей с положительными вероятностями значения $x_1, x_2, x_3, \ldots$, достаточно знать вероятности $p_k = \mathbf{P}\{\xi = x_k\}$ *). Очевидно, что по совокупности вероятностей p_k можно определить функцию распределения $F(x)$ посредством равенства

$$F(x) = \sum p_k,$$

в котором суммирование распространяется на все те индексы, для которых $x_k < x$.

Функция распределения любой дискретной величины разрывна, возрастает скачками при тех значениях x, которые являются возможными

---

*) Эти и только эти значения x_n мы назовем ***возможными значениями*** дискретной случайной величины ξ.

значениями ξ. Величина скачков функции $F(x)$ в точке x, как мы выяснили ранее, равна разности $F(x+0) - F(x)$.

Если два возможных значения величины ξ разделены интервалом, в котором других возможных значений ξ нет, то в этом интервале функция распределения $F(x)$ постоянна. Если возможных значений ξ конечное число, например n, то функция распределения $F(x)$ представляет собой ступенчатую кривую с $n+1$ интервалом постоянства. Если же возможных значений ξ имеется счетное множество, то это последнее может быть и всюду плотным, так что интервалов постоянства у функции распределения дискретной случайной величины может и не быть. Пусть для примера возможными значениями ξ будут все рациональные числа и только они. Пусть эти числа занумерованы каким-нибудь способом: $r_1\ r_2, \ldots$ и вероятности $\mathsf{P}\{\xi = r_k\} = p_k$ определены посредством равенства $p_k = 1/2^k$. В нашем примере все рациональные точки являются точками разрыва функции распределения.

В качестве другого важного класса случайных величин мы выделим те из них, для которых существует неотрицательная функция $p(x)$, удовлетворяющая при любых x равенству

$$F(x) = \int_{-\infty}^{x} p(z)dz.$$

Случайные величины, обладающие этим свойством, называются *непрерывными*: функция $p(x)$ называется *плотностью распределения вероятностей*.

Отметим, что плотность распределения вероятностей обладает следующими свойствами:

1. $p(x) \geqslant 0$.
2. При любых x_1 и x_2 удовлетворяет равенству

$$\mathsf{P}\{x_1 \leqslant \xi < x_2\} = \int_{x_1}^{x_2} p(x)dx.$$

В частности, если $p(x)$ непрерывна в точке x, то с точностью до бесконечно малых высших порядков $\mathsf{P}\{x \leqslant \xi < x+dx\} = p(x)dx$.

3. $\int p(x)dx = 1$.

Величины, распределенные по нормальному или равномерному закону*), дают нам примеры непрерывных случайных величин.

П р и м е р. Рассмотрим ближе нормальный закон распределения. Для него плотность распределения вероятностей равна

$$p(x) = C\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$$

*) Так называется закон с функцией распределения, линейно изменяющейся от 0 до 1 в некотором интервале (a, b) и равной нулю левее точки a и единице правее b.

Постоянное C определяется, исходя из свойства 3. Действительно,

$$C\int e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx = 1.$$

Заменой переменных $\dfrac{x-a}{\sigma} = z$ это равенство приводится к виду

$$C\sigma\int e^{-z^2/2}dz = 1.$$

Интеграл, стоящий в правой части этого равенства, известен под именем интеграла Пуассона, причем

$$\int e^{-z^2/2}dz = \sqrt{2\pi}.$$

Таким образом, находим, что

$$C = \frac{1}{\sigma\sqrt{2\pi}}$$

и, значит, для нормального распределения

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Функция $p(x)$ достигает максимума при $x = a$, имеет точки перегиба при $x = a \pm \sigma$; ось абсцисс служит для нее асимптотой при $x \to \pm\infty$. Для иллюстрации влияния параметра σ на форму графика плотности нормального распределения мы приводим на рис. 15 графики $p(x)$ при $a = 0$ и 1) $\sigma^2 = 1/4$, 2) $\sigma^2 = 1$, 3) $\sigma^2 = 4$. Мы видим, что чем меньше значение σ, тем кривая $p(x)$ имеет большее значение максимума и падает

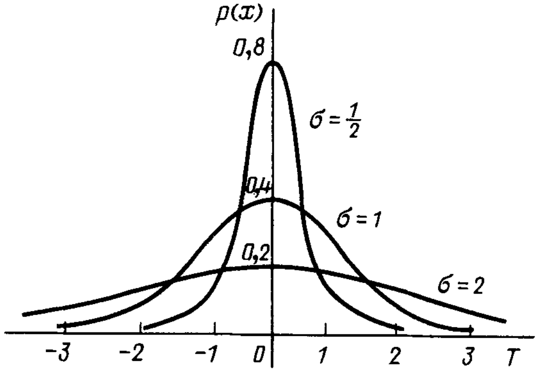

Рис. 15

круче. Это означает, в частности, что вероятность попасть в интервал $(-\alpha, \alpha)$ больше для той случайной величины, распределенной по нормальному закону (с параметром $a = 0$), для которой величина σ меньше. Мы, следовательно, можем считать σ характеристикой разбросанности значений величины ξ. При $a \neq 0$ кривые плотностей имеют ту же форму, но сдвинуты вправо ($a > 0$) или влево ($a < 0$) в зависимости от знака параметра a.

Помимо дискретных и непрерывных случайных величин существуют, разумеется, и другие случайные величины. Кроме величин, которые ведут себя в одних интервалах как непрерывные, а в других как дискретные, имеются величины, не являющиеся ни в одном интервале ни дискретными, ни непрерывными. К таким случайным величинам относятся, например, все те, функции распределения которых непрерывны, но при этом возрастают только на множестве лебеговской меры нуль. В качестве примера такой случайной величины приведем величину, имеющую функцией распределения известную кривую Кантора. Напомним построение этой кривой. Величина ξ принимает только значения между нулем и единицей. Следовательно, ее функция распределения удовлетворяет равенствам

$$F(x) = 0 \text{ при } x \leqslant 0, \quad F(x) = 1 \text{ при } x > 1.$$

Внутри интервала $(0, 1)$ ξ принимает значения только в первой и последней его третях, в каждой с вероятностью $1/2$. Таким образом,

$$F(x) = 1/2 \text{ при } 1/3 < x \leqslant 2/3.$$

В интервалах $(0, 1/3)$ и $(2/3, 1)$ ξ снова может принимать значения только в первой и последней трети каждого из них, в каждой с вероятностью $1/4$. Этим определяются значения $F(x)$ еще в двух интервалах:

$$F(x) = 1/4 \text{ при } 1/9 < x \leqslant 2/9,$$

$$F(x) = 3/4 \text{ при } 7/9 < x \leqslant 8/9.$$

Далее в каждом из оставшихся интервалов повторяется то же построение и этот процесс продолжается до бесконечности. В результате функция $F(x)$ оказывается определенной на счетном множестве интервалов, являющихся интервалами смежности некоторого нигде не плотного совершенного множества меры нуль. На этом множестве доопределяем функцию $F(x)$ по непрерывности. Величина ξ с таким образом определенной функцией распределения не дискретна, так как ее функция распределения непрерывна, но в то же время ξ не непрерывна, так как ее функция распределения не является интегралом от своей производной.

Все введенные нами определения переносятся легко на случай условных вероятностей. Так, например, функцию $F(x|B) = \mathsf{P}\{\xi < x|B\}$ мы будем называть *условной функцией распределения* случайной величины ξ при условии B. Очевидно, что $F(x|B)$ обладает всеми свойствами обычной функции распределения.

§ 20. Многомерные функции распределения

Для дальнейшего нам необходимо не только понятие случайной величины, но и понятие случайного вектора или, как часто говорят, многомерной случайной величины.

Рассмотрим вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, на котором определены n случайных величин

$$\xi_1 = f_1(\omega), \quad \xi_2 = f_2(\omega), \ldots, \quad \xi_n = f_n(\omega)$$

(функции $f_i(\omega)$ измеримы). Вектор $(\xi_1, \xi_2, \ldots, \xi_n)$ называется *случайным вектором* или *n-мерной случайной величиной*. Пусть $(\xi_1, \xi_2, \ldots, \xi_n)$ случайный вектор. Обозначим через $\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\}$ множество тех элементарных событий ω, для которых одновременно выполняются все неравенства

$$f_1(\omega) < x_1, \quad f_2(\omega) < x_2, \ldots, \quad f_n(\omega) < x_n.$$

Поскольку это событие является произведением событий $\{f_k(\omega) < x_k\}$ $(1 \leqslant k \leqslant n)$, оно принадлежит множеству $\mathfrak{F}$, т.е.

$$\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \quad \xi_n < x_n\} \in \mathfrak{F}.$$

Таким образом, при любом наборе чисел $x_1, x_2, \ldots, x_n$ определена вероятность $F(x_1, x_2, \ldots, x_n) = \mathsf{P}\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\}$. Эта функция n аргументов называется *n-мерной функцией распределения случайного вектора* $(\xi_1, \xi_2, \ldots, \xi_n)$.

В дальнейшем мы прибегнем к геометрической иллюстрации и станем рассматривать величины $\xi_1, \xi_2, \ldots, \xi_n$ как координаты точек n-мерного евклидова пространства. Очевидно, что положение точки $(\xi_1, \xi_2, \ldots, \xi_n)$ зависит от случая и что функция $F(x_1, \ldots, x_n)$ при такой интерпретации дает вероятность попадания точки $(\xi_1, \ldots, \xi_n)$ в n-мерный параллелепипед $\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n$ с ребрами, параллельными осям координат.

С помощью функции распределения легко вычислить вероятность того, что точка $(\xi_1, \xi_2, \ldots, \xi_n)$ окажется внутри параллелепипеда

$$a_i \leqslant \xi_i < b_i \quad (i = 1, 2, \ldots, n),$$

где a_i и b_i – произвольные постоянные. Нетрудно подсчитать, что

$$\mathsf{P}\{a_1 \leqslant \xi_1 < b_1, \quad a_2 \leqslant \xi_2 < b_2, \ldots, a_n \leqslant \xi_n < b_n\} =$$
$$= F(b_1, b_2, \ldots, b_n) - \sum_{i=1}^{n} p_i + \sum_{i<j} p_{ij} \mp \ldots +$$
$$+ (-1)^n F(a_1, a_2, \ldots, a_n), \tag{1}$$

где через $p_{ij\ldots k}$ обозначено значение функции $F(c_1, c_2, \ldots, c_n)$ при $c_i = a_i$, $c_j = a_j, \ldots, c_k = a_k$ и при остальных c_s, равных b_s. Мы предоставляем доказательство этой формулы читателю. Заметим, в частности, что $F(x_1, \ldots, x_{k-1}, +\infty, x_{k+1}, \ldots, x_n)$ дает нам вероятность того, что будет выполнена следующая система неравенств:

$$\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_{k-1} < x_{k-1}, \xi_{k+1} < x_{k+1}, \ldots, \xi_n < x_n.$$

Так как по расширенной аксиоме сложения вероятностей

$$\mathsf{P}\{\xi_1 < x_1, \ldots, \xi_{k-1} < x_{k-1}, \xi_{k+1} < x_{k+1}, \ldots, \xi_n < x_n\} =$$
$$= \sum_{s=-\infty}^{\infty} \mathsf{P}\{\xi_1 < x_1, \ldots, \xi_{k-1} < x_{k-1}, \quad s \leqslant \xi_k < s+1, \ \xi_{k+1} <$$
$$< x_{k+1}, \ldots, \xi_n < x_n\} = F(x_1, \ldots, x_{k-1}, \infty, x_{k+1}, \ldots, x_n),$$

то $F(x_1, \ldots, x_{k-1}, \infty, x_{k+1}, \ldots, x_n)$ является функцией распределения $(n-1)$-мерной случайной величины $(\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_n)$. Продолжив этот процесс далее, мы можем определить k-мерные функции распределения любой группы из k величин $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k}$ по формуле

$$F_k(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = \mathsf{P}\{\xi_{i_1} < x_{i_1}, \ldots, \xi_{i_k} < x_{i_k}\} =$$
$$= F(c_1, c_2, \ldots, c_n),$$

где $c_s = x_s$, если $s = i_r$ $(1 \leqslant r \leqslant k)$ и $c_s = +\infty$ в иных случаях. В частности, функция распределения случайной величины ξ_k равна

$$F_k(x) = F(c_1, c_2, \ldots, c_n),$$

где все c_i $(i \neq k)$ равны $+\infty$, а $c_k = x$.

Подобно тому как поведение одномерной случайной величины можно характеризовать не только посредством функции распределения, но и другими способами, многомерные случайные величины могут быть определены, скажем, посредством неотрицательной вполне аддитивной функции множества $\Phi\{E\}$, определенной для любых борелевских множеств n-мерного пространства. Эту функцию мы определим как вероятность попадания точки $(\xi_1, \ldots, \xi_n)$ на множество E. Этот способ вероятностной характеристики n-мерной случайной величины следует признать наиболее естественным и с точки зрения теоретической наиболее удачным.

Рассмотрим примеры.

Пример 1. Случайный вектор $(\xi_1, \ldots, \xi_n)$ называется *равномерно распределенным* в параллелепипеде $a_i \leqslant \xi_i < b_i$ $(1 \leqslant i \leqslant n)$, если вероятность попадания точки $(\xi_1, \xi_2, \ldots, \xi_n)$ в любую внутреннюю область указанного параллелепипеда пропорциональна ее объему и вероятность попадания внутрь параллелепипеда является достоверным событием. Функция распределения искомой величины имеет вид

$$F(x_1, \ldots, x_n) = \begin{cases} 0, & \text{если } x_i \leqslant a_i \text{ хотя бы при одном } i, \\ \prod\limits_{i=1}^{n} \dfrac{c_i - a_i}{b_i - a_i}, & \text{где } c_i = x_i, \text{ если } a_i \leqslant x_i \leqslant b_i, \\ & \text{и } \quad c_i = b_i, \text{ если } x_i > b_i. \end{cases}$$

Пример 2. Двумерная случайная величина (ξ_1, ξ_2) распределена *нормально*, если для нее функция распределения равна

$$F(x_1, x_2) = C \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} e^{-Q(x,y)} dx\, dy,$$

где $Q(x, y)$ – положительно определенная квадратичная форма.

Известно, что положительно определенная квадратичная форма от x и y может быть записана в виде

$$Q(x, y) = \frac{(x-a)^2}{2A^2} - r\,\frac{(x-a)(y-b)}{AB} + \frac{(y-b)^2}{2B^2},$$

где A и B – положительные числа, а r, a и b – вещественные числа, причем r подчиняется условию $-1 < r < +1$.

Легко видеть, что при $r^2 \neq 1$ каждая из случайных величин ξ_1 и ξ_2 подчинена одномерному нормальному закону. Действительно,

$$F_1(x_1) = \mathsf{P}\{\xi_1 < x_1\} = F(x_1, +\infty) = C \int_{-\infty}^{x_1} \int e^{-Q(x,y)} dx\, dy =$$

$$= C \int_{-\infty}^{x_1} e^{-\frac{(x-a)^2}{2A^2}(1-r^2)} \int e^{-\frac{1}{2}\left[\frac{y-b}{B} - \frac{r(x-a)}{A}\right]^2} dy.$$

Так как

$$\int e^{-\frac{1}{2}\left[\frac{y-b}{B} - r\frac{x-a}{A}\right]^2} dy = B\sqrt{2\pi},$$

то

$$F_1(x_1) = BC\sqrt{2\pi}\int\limits_{-\infty}^{x_1} e^{-\frac{(x-a)^2}{2A^2}(1-r^2)}dx. \tag{2}$$

Постоянное C может быть выражено через A, B и r. Эту зависимость можно найти из условия $F_1(+\infty) = 1$. Имеем:

$$1 = BC\sqrt{2\pi}\int e^{-\frac{(x-a)^2}{2A^2}(1-r^2)}dx = \frac{ABC\sqrt{2\pi}}{\sqrt{1-r^2}}\int e^{-\frac{z^2}{2}}dz =$$

$$= \frac{2ABC\pi}{\sqrt{1-r^2}}.$$

Отсюда

$$C = \frac{\sqrt{1-r^2}}{2\pi AB}.$$

Если $r^2 \neq 1$, то мы положим

$$A = \sigma_1\sqrt{1-r^2}, \quad B = \sigma_2\sqrt{1-r^2}.$$

В этих новых обозначениях двумерный нормальный закон принимает такой вид:

$$F(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times$$

$$\times \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2} - 2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2}\right]}dx\,dy.$$

Теоретико-вероятностный смысл входящих в эту формулу параметров будет выяснен в следующей главе.

Подобно тому как это было сделано в одномерном случае, для многомерных функций распределения можно установить ряд свойств. Мы приведем их формулировки, предоставив читателю их доказательства. Функция распределения

1) есть неубывающая функция по каждому аргументу,
2) непрерывна слева по каждому аргументу,
3) удовлетворяет соотношениям

$$F(+\infty, +\infty, \ldots, +\infty) = 1,$$

$$\lim_{x_k \to -\infty} F(x_1, x_2, \ldots, x_n) = 0 \qquad (1 \leqslant k \leqslant n)$$

при произвольных значениях остальных аргументов.

В одномерном случае мы видели, что перечисленные свойства необходимы и достаточны, чтобы функция $F(x)$ была функцией распределения некоторой случайной величины. В многомерном случае, оказывается, этих свойств уже недостаточно. Для того чтобы функция $F(x_1, \ldots, x_n)$ была функцией распределения, помимо перечисленных трех свойств, нужно добавить еще следующее:

4) при любых a_i и b_i $(i = 1, 2, \ldots, n)$ выражение (1) не отрицательно.

Что это требование может быть не выполнено, несмотря на наличие у функции $F(x_1, \ldots, x_n)$ свойств 1 – 3, показывает следующий пример. Пусть

$$F(x, y) = \begin{cases} 0, & \text{если } x \leqslant 0, \text{ или } x + y \leqslant 1, \text{ или } y \leqslant 0, \\ 1 & \text{в остальной части плоскости.} \end{cases}$$

Эта функция удовлетворяет требованиям 1 – 3, но для нее

$$F(1, 1) - F(1, 1/2) - F(1/2, 1) + F(1/2, 1/2) = -1, \tag{3}$$

и, следовательно, четвертое требование не выполнено.

Функция $F(x, y)$ не может быть функцией распределения, так как разность (3) согласно соотношению (1) равна вероятности попадания точки (ξ_1, ξ_2) в прямоугольник $1/2 \leqslant \xi_1 < 1$, $1/2 \leqslant \xi_2 < 1$.

Если существует такая функция $p(x_1, x_2, \ldots, x_n)$, что при любых $x_1, x_2, \ldots, x_n$ имеет место равенство

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_n} p(z_1, z_2, \ldots, z_n)\, dz_n \ldots dz_2\, dz_1,$$

то эта функция называется *плотностью распределения вероятностей* случайного вектора $(\xi_1, \xi_2, \ldots, \xi_n)$. Легко видеть, что плотность распределения обладает следующими свойствами:

1) $p(x_1, x_2, \ldots, x_n) \geqslant 0$.

2) Вероятность попадания точки $(\xi_1, \xi_2, \ldots, \xi_n)$ в какую-нибудь область G равна

$$\int \ldots \int_G p(x_1, \ldots, x_n)\, dx_n \ldots dx_1 .$$

В частности, если функция $(x_1, x_2, \ldots, x_n)$ непрерывна в точке $(x_1, \ldots, x_n)$, то вероятность попадания точки $(\xi_1, \xi_2, \ldots \xi_n)$ в параллелепипед $x_k \leqslant \xi_k < x_k + dx_k$ $(k = 1, 2, \ldots, n)$ с точностью до бесконечно малых высших порядков равна

$$p(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n.$$

Пример 3. В качестве примера n-мерной случайной величины, имеющей плотность, приведем величину, равномерно распределенную в n-мерной области G. Если через V обозначим n-мерный объем области G, то плотность распределения будет равна

$$p(x_1, x_2, \ldots, x_n) = \begin{cases} 0, & \text{если} \quad (x_1, x_2, \ldots, x_n) \overline{\in} G, \\ 1/V, & \text{если} \quad (x_1, x_2, \ldots, x_n) \in G. \end{cases}$$

Пример 4. Плотность двумерного нормального закона дается формулой

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2} - 2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2}\right]}.$$

Заметим, что плотность нормального распределения сохраняет постоянное значение на эллипсах

$$\frac{(x-a)^2}{\sigma_1^2} - 2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} = \lambda^2, \tag{4}$$

где λ – постоянное; на этом основании эллипсы (4) носят название *эллипсов равных вероятностей*.

Найдем вероятность попадания точки (ξ_1, ξ_2) внутрь эллипса (4). По определению плотности

$$P(\lambda) = \iint\limits_{G(\lambda)} p(x, y)\,dx\,dy, \tag{5}$$

где через $G(\lambda)$ обозначена область, ограниченная эллипсом (4). Для вычисления этого интеграла введем полярные координаты

$$x - a = \rho\cos\theta, \quad y - b = \rho\sin\theta.$$

Интеграл (5) при этом принимает вид

$$P(\lambda) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int\limits_0^{2\pi\lambda/s} \int\limits_0^{\sqrt{1-r^2}} e^{-\frac{\rho^2}{2}s^2} \rho\,d\rho\,d\theta,$$

где для краткости обозначено

$$s^2 = \frac{1}{1-r^2}\left[\frac{\cos^2\theta}{\sigma_1^2} - 2r\frac{\cos\theta\sin\theta}{\sigma_1\sigma_2} + \frac{\sin^2\theta}{\sigma_2^2}\right].$$

Интегрирование по ρ дает:

$$P(\lambda) = \frac{1 - e^{-\frac{\lambda^2}{2(1-r^2)}}}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^{2\pi} \frac{d\theta}{s^2}.$$

Интегрирование по θ можно выполнить по правилам интегрирования тригонометрических функций, но в этом нет необходимости, так как оно автоматически производится с помощью вероятностных соображений. Действительно,

$$P(+\infty) = 1 = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^{2\pi} \frac{d\theta}{s^2}.$$

Отсюда

$$\int_0^{2\pi} \frac{d\theta}{s^2} = 2\pi\sigma_1\sigma_2\sqrt{1-r^2}$$

и, стало быть,

$$P(\lambda) = 1 - e^{-\frac{\lambda^2}{2(1-r^2)}}.$$

Нормальное распределение играет исключительно большую роль в различных прикладных вопросах. Распределение многих практически важных случайных величин оказывается подчиненным нормальному закону распределения. Так, например, огромная практика артиллерийских стрельб, произведенная в различных условиях, показала, что рассеивание снарядов на плоскости при стрельбе из одного орудия на определенном прицеле подчиняется нормальному закону. В главе 8 мы увидим, что эта "универсальность" нормального закона объясняется тем, что всякая случайная величина, являющаяся суммой очень большого числа независимых случайных величин, каждая из которых оказывает лишь незначительное влияние на сумму, распределена почти по нормальному закону.

Важнейшее понятие теории вероятностей – независимость событий – сохраняет свое значение и для случайных величин. В соответствии с определением независимости событий мы скажем, что *случайные величины* $\xi_1, \xi_2, \ldots, \xi_n$ *независимы*, если для любой группы $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k}$ этих величин имеет место равенство

$$\mathbf{P}\{\xi_{i_1} < x_{i_1}, \xi_{i_2} < x_{i_2}, \ldots, \xi_{i_k} < x_{i_k}\} =$$
$$= \mathbf{P}\{\xi_{i_1} < x_{i_1}\}\, \mathbf{P}\{\xi_{i_2} < x_{i_2}\} \ldots \mathbf{P}\{\xi_{i_k} < x_{i_k}\}$$

при произвольных $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ и любом $k (1 \leqslant k \leqslant n)$. В частности, для произвольных $x_1, x_2, \ldots, x_n$ выполняется равенство

$$\mathsf{P}\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\} =$$
$$= \mathsf{P}\{\xi_1 < x_1\}\, \mathsf{P}\{\xi_2 < x_2\} \ldots \mathsf{P}\{\xi_n < x_n\}$$

или в терминах функций распределения

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1) F_2(x_2) \ldots F_n(x_n),$$

где $F_k(x_k)$ означает функцию распределения величины ξ_k.

Легко видеть, что верно и обратное предложение: если функция распределения $F(x_1, x_2, \ldots, x_n)$ системы случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ имеет вид

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1) F_2(x_2) \ldots F_n(x_n),$$

где функции $F_k(x_k)$ удовлетворяют соотношениям

$$F_k(+\infty) = 1 \quad (k = 1, 2, \ldots, n),$$

то величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и функции $F_1(x_1)$, $F_2(x_2), \ldots$ $\ldots, F_n(x_n)$ являются их функциями распределения.

Проверку этого предложения мы предоставляем читателю.

Пример 5. Рассмотрим n-мерную случайную величину, компоненты которой $\xi_1, \xi_2, \ldots, \xi_n$ являются взаимно независимыми случайными величинами, распределенными по нормальным законам

$$F_k(x_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \int_{-\infty}^{x_k} e^{-\frac{(z-a_k)^2}{2\sigma_k^2}}\, dz.$$

В рассматриваемом примере функция распределения равна

$$F(x_1, x_2, \ldots, x_n) = (2\pi)^{-\frac{n}{2}} \prod_{k=1}^{n} \sigma_k^{-1} \int_{-\infty}^{x_k} e^{-\frac{(z-a_k)^2}{2\sigma_k^2}}\, dz.$$

Если независимые случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ имеют плотности распределения $p_1(x)$, $p_2(x), \ldots, p_n(x_n)$, то n-мерная величина $(\xi_1, \xi_2, \ldots$ $\ldots, \xi_n)$ имеет плотность распределения, равную

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \ldots p_n(x_n).$$

Пример 6. Если величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и имеют плотности распределения

$$p_k(x) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x-a_k)^2}{2\sigma_k^2}} \quad (1 \leqslant k \leqslant n),$$

то n-мерная плотность распределения величины $(\xi_1, \xi_2, \ldots, \xi_n)$ равна

$$p(x_1, x_2, \ldots, x_n) = \frac{(2\pi)^{-\frac{n}{2}}}{\sigma_1 \sigma_2 \ldots \sigma_n} e^{-\frac{1}{2} \sum_{k=1}^{n} \frac{(x_n - a_k)^2}{\sigma_k^2}} \qquad (6)$$

При $n = 2$ эта формула принимает вид

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x_1 - a_1)^2}{2\sigma_1^2} - \frac{(x_2 - a_2)^2}{2\sigma_2^2}}.$$

Сравнение этой функции с плотностью двумерного нормального закона (пример 4) показывает, что для независимых случайных величин ξ_1 и ξ_2 параметр r равен 0.

При $n = 3$ формула (6) может быть истолкована как плотность распределения вероятностей компонент ξ_1, ξ_2, ξ_3 скорости молекулы по осям координат (распределение Максвелла), если только предположить, что

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \frac{1}{hm},$$

где m – масса молекулы, а h – константа.

§ 21. Функции от случайных величин

Сведения, полученные нами о функциях распределения, позволяют нам приступить к решению следующей задачи: по функции распределения $F(x_1, x_2, \ldots, x_n)$ совокупности случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ определить функцию распределения $\Phi(y_1, y_2, \ldots, y_k)$ величин $\eta_1 = f_1(\xi_1, \xi_2, \ldots, \xi_n), \eta_2 = f_2(\xi_1, \ldots, \xi_n), \ldots, \eta_k = f_k(\xi_1, \ldots, \xi_n)$.

Общее решение этой задачи весьма просто, но требует расширения понятия интеграла. Чтобы не отвлекаться в сторону чисто аналитических вопросов, мы ограничимся рассмотрением важнейших частных случаев: дискретных и непрерывных случайных величин. В следующем параграфе будут изложены определение и основные свойства интеграла Стилтьеса; там мы дадим общую запись важнейших результатов настоящего параграфа.

Рассмотрим сначала случай, когда n-мерный вектор $(\xi_1, \ldots, \xi_n)$ имеет плотность распределения вероятностей $p(x_1, x_2, \ldots, x_n)$. Из предыдущего видно, что искомая функция распределения определяется равенством

$$\Phi(y_1, y_2, \ldots, y_k) = \int_D \ldots \int p(x_1, x_2, \ldots x_n) dx_1 dx_2 \ldots dx_n,$$

причем область интегрирования D определяется неравенствами

$$f_i(x_1, x_2, \ldots, x_n) < y_i \quad (i = 1, 2, \ldots, k).$$

В случае дискретных случайных величин решение, очевидно, дается с помощью n-мерной суммы, также распространенной на область D.

Мы применим теперь только что сделанное нами общее замечание относительно решения поставленной нами общей задачи к нескольким важным частным случаям.

Функция распределения суммы. Пусть требуется найти функцию распределения суммы

$$\eta = \xi_1 + \xi_2 + \ldots + \xi_n,$$

если $p(x_1, x_2, \ldots, x_n)$ — плотность распределения вероятностей вектора $(\xi_1, \xi_2, \ldots, \xi_n)$. Искомая функция равна вероятности попадания точки $(\xi_1, \xi_2, \ldots, \xi_n)$ в полупространство $\xi_1 + \xi_2 + \ldots + \xi_n < x$ и, следовательно,

$$\Phi(x) = \int_{\Sigma x_k < x} \ldots \int p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n.$$

Рассмотрим подробнее случай $n = 2$. Предыдущая формула принимает в этом случае такой вид:

$$\Phi(x) = \iint_{x_1 + x_2 < x} p(x_1, x_2)\, dx_1 dx_2 = \int \int_{-\infty}^{x - x_2} p(x_1, x_2)\, dx_1 dx_2. \tag{1}$$

Если величины ξ_1 и ξ_2 независимы, то $p(x_1, x_2) = p_1(x_1)\, p_2(x_2)$ и равенство (1) может быть записано в таком виде:

$$\Phi(x) = \int dx_1 \int_{-\infty}^{x - x_2} p_1(x_1) p_2(x_2)\, dx_2 = \int dx_1 \int_{-\infty}^{x} p_1(x_1) p_2(z - x_1)\, dz =$$

$$= \int_{-\infty}^{x} dz \left\{ \int p_1(x_1)\, p_2(z - x_1)\, dx_1 \right\}. \tag{2}$$

В общем случае формула (1) дает:

$$\Phi(x) = \int_{-\infty}^{x} dx_1 \int p(z, x_1 - z)\, dz. \tag{3}$$

Последние равенства доказывают, что если многомерное распределение слагаемых имеет плотность распределения вероятностей, то их сумма также имеет плотность распределения. Эта плотность в случае независимых слагаемых может быть записана в виде

$$p(x) = \int p_1(x - z) p_2(z)\, dz. \tag{4}$$

Рассмотрим примеры.

Пример 1. Пусть ξ_1 и ξ_2 независимы и равномерно распределены в интервале (a, b). Найти плотность распределения суммы $\eta = \xi_1 + \xi_2$.

Плотности распределения вероятностей ξ_1 и ξ_2 равны

$$p_1(x) = p_2(x) = \begin{cases} 0, & \text{если} \quad x \leqslant a \quad \text{или} \quad x > b, \\ \dfrac{1}{b-a}, & \text{если} \quad a < x \leqslant b. \end{cases}$$

По формуле (4) находим, что

$$p_\eta(x) = \int_a^b p_1(z)\, p_2(x-z)\, dz = \frac{1}{b-a} \int_a^b p_2(x-z)\, dz.$$

Из того, что при $x < 2a$

$$x - z < 2a - z < a,$$

а при $x > 2b$

$$x - z > 2b - z > b,$$

заключаем, что при $x < 2a$ и $x > 2b$

$$p_\eta(x) = 0.$$

Пусть теперь $2a < x < 2b$. Подинтегральная функция отлична от нуля только при тех значениях z, которые удовлетворяют неравенствам

$$a < x - z < b$$

или, что то же самое, неравенствам

$$x - b < z < x - a.$$

Так как $x > 2a$, то $x - a > a$. Очевидно, что $x - a \leqslant b$ при $x \leqslant a + b$. Следовательно, если $2a < x \leqslant a + b$, то

$$p_\eta(x) = \int_a^{x-a} \frac{dz}{(b-a)^2} = \frac{x-2a}{(b-a)^2}.$$

Точно так же при $a + b < x \leqslant 2b$

$$p_\eta(x) = \int_{x-b}^{b} \frac{dz}{(b-a)^2} = \frac{2b-x}{(b-a)^2}.$$

Собрав вместе полученные результаты, находим, что

$$p_\eta(x)=\begin{cases} 0 & \text{при } x\leqslant 2a \text{ и } x>2b,\\ \dfrac{x-2a}{(b-a)^2} & \text{при } 2a<x\leqslant a+b,\\ \dfrac{2b-x}{(b-a)^2} & \text{при } a+b<x\leqslant 2b.\end{cases} \tag{5}$$

Функция $p_\eta(x)$ носит название *закона распределения Симпсона.*

Вычисления в рассмотренном примере значительно упрощаются, если воспользоваться геометрическими соображениями. Изобразим, как обычно, ξ_1 и ξ_2 как прямоугольные координаты на плоскости. Тогда вероятность неравенства $\xi_1+\xi_2<x$ при $2a<x\leqslant a+b$ равна вероятности попадания в дважды заштрихованный прямоугольный треугольник (рис. 16). Эта вероятность, как легко подсчитать, равна

$$F_\eta(x)=\frac{(x-2a)^2}{2(a-b)^2}.$$

При $a+b<x\leqslant 2b$ вероятность неравенства $\xi_1+\xi_2<x$ равна вероятности попадания во всю заштрихованную фигуру. Эта вероятность равна

$$F_\eta(x)=1-\frac{(2b-x)^2}{2(b-a)^2}.$$

Дифференцирование по x приводит нас к формуле (5).

В связи с рассмотренным примером интересно заметить следующее.

Общие вопросы геометрии привели Н.И. Лобачевского к необходимости решения следующей задачи: имеется группа из n взаимно независимых случайных величин $\xi_1,\xi_2,\ldots,\xi_n$, найти распределение вероятностей их среднего арифметического.

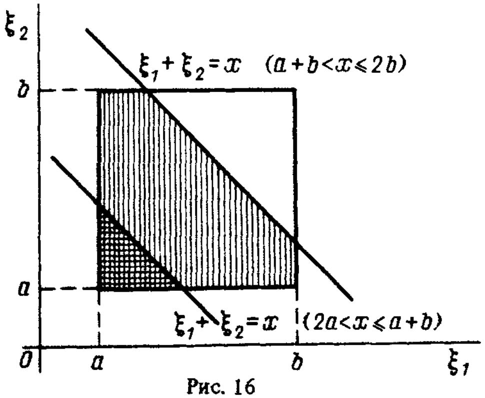

Рис. 16

Эта задача была им решена только для случая, когда все ошибки равномерно распределены в интервале $(-1,\ 1)$. При этом оказалось, что вероятность ошибке среднего арифметического заключаться в пределах от $-x$ до x равна

$$P_n(x)=\frac{1}{2^{n-1}}\times$$
$$\times\Sigma(-1)^r\frac{[n-nx-2r]^r}{(r!)^2(n-r)!},$$

где суммирование распространяется на все целые r от $r = 0$ до $r = \left[\dfrac{n - nx}{2}\right]$.

Пример 2. Двумерная случайная величина (ξ_1, ξ_2) распределена по нормальному закону

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{(x-a)^2}{\sigma_1^2} - 2r\frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2}\right)\right\}.$$

Найти функцию распределения суммы $\eta = \xi_1 + \xi_2$.

Согласно формуле (3)

$$p_\eta(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{(z-a)^2}{\sigma_1^2} - \right.\right.$$

$$\left.\left. - 2r\frac{(z-a)(x-z-b)}{\sigma_1\sigma_2} + \frac{(x-z-b)^2}{\sigma_2^2}\right)\right\} dz.$$

Обозначим для краткости $x - a - b$ через v и $z - a$ через u; тогда

$$p_\eta(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times$$

$$\times \int \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{u^2}{\sigma_1^2} - 2r\frac{u(v-u)}{\sigma_1\sigma_2} + \frac{(v-u)^2}{\sigma_2^2}\right)\right\} du.$$

Так как

$$\frac{u^2}{\sigma_1^2} - 2r\frac{u(v-u)}{\sigma_1\sigma_2} + \frac{(v-u)^2}{\sigma_2^2} = u^2\frac{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2}{\sigma_1^2\sigma_2^2} - 2uv\frac{\sigma_1 + r\sigma_2}{\sigma_1\sigma_2^2} + \frac{v^2}{\sigma_2} =$$

$$= \left[u\frac{\sqrt{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2}}{\sigma_1\sigma_2} - \frac{v}{\sigma_2}\frac{\sigma_1 + r\sigma_2}{\sqrt{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2}}\right]^2 +$$

$$+ \frac{v^2}{\sigma_2^2}\left(1 - \frac{(\sigma_1 + r\sigma_2)^2}{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2}\right) = \left[u\frac{\sqrt{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2}}{\sigma_1\sigma_2} - \right.$$

$$\left. - \frac{v}{\sigma_2}\frac{\sigma_1 + r\sigma_2}{\sqrt{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2}}\right]^2 + \frac{v^2(1-r^2)}{\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2},$$

то, введя обозначение

$$t = \frac{1}{\sqrt{1-r^2}}\left[u\,\frac{\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2} - \frac{v}{\sigma_2}\,\frac{\sigma_1+r\sigma_2}{\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}\right],$$

мы приведем выражение для $p_\eta(x)$ к виду

$$p_\eta(x) = \frac{\exp\left\{-\dfrac{v^2}{2(\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)}\right\}}{2\pi\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}\int e^{-\frac{t^2}{2}}\,dt.$$

Так как

$$v = x-a-b \quad \text{и} \quad \int e^{-\frac{t^2}{2}}\,dt = \sqrt{2\pi},$$

то

$$p_\eta(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)}}\,e^{-\frac{(x-a-b)^2}{2(\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)}}.$$

В частности, если случайные величины ξ_1 и ξ_2 независимы, то $r=0$ и формула для $p_\eta(x)$ принимает вид

$$p_\eta(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\,e^{-\frac{(x-a-b)^2}{2(\sigma_1^2+\sigma_2^2)}}.$$

Нами получен, таким образом, следующий результат: *сумма нормально распределенных случайных величин распределена по нормальному закону.*

Интересно заметить, что когда слагаемые независимы, имеет место и обратное предложение (теорема Г. Крамера): *если сумма двух независимых случайных величин распределена по нормальному закону, то каждое слагаемое также распределено по нормальному закону.* Мы не останавливаемся на доказательстве этого предложения, так как оно требует более сложного математического аппарата.

Пример 3. Распределение χ^2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ – независимые случайные величины, распределенные по одному и тому же нормальному закону с параметрами a и σ.

Функция распределения величины

$$\chi^2 = \frac{1}{\sigma^2}\sum_{k=1}^{n}(\xi_k - a)^2$$

носит название χ^2-*распределения.*

Это распределение играет важную роль в различных вопросах статистики.

Мы вычислим сейчас функцию распределения величины $\zeta = \chi/\sqrt{n}$. Она окажется независимой от a и σ.

Очевидно, что для отрицательных значений аргумента функция распределения $\Phi(y)$ величины ζ равна 0; для положительных значений y функция $\Phi(y)$ равна вероятности попадания точки $(\xi_1, \ldots, \xi_n)$ внутрь шара

$$\sum_{k=1}^{n} (x_k - a)^2 = y^2 \cdot n \cdot \sigma^2 .$$

Таким образом,

$$\Phi(y) = \int\limits_{\Sigma x_i^2 < y^2 n} \cdots \int \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\sum\limits_{i=1}^{n} \frac{x_i^2}{2}} dx_1 dx_2 \ldots dx_n .$$

Перейдем для вычисления этого интеграла к сферическим координам, т.е. сделаем замену

$$x_1 = \rho \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1},$$

$$x_2 = \rho \cos\theta_1 \cos\theta_2 \ldots \sin\theta_{n-1},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n = \rho \sin\theta_1 .$$

В результате этой замены

$$\Phi(y) = \int\limits_{-\pi/2}^{\pi/2} \ldots \int\limits_{-\pi/2}^{\pi/2} \int\limits_{0}^{y\sqrt{n}} \frac{1}{(\sqrt{2\pi})^n} e^{-\rho^2/2} \rho^{n-1} D(\theta_1 \ldots \theta_{n-1}) \times$$

$$\times d\rho\, d\theta_{n-1} \ldots d\theta_1 = C_n \int\limits_{0}^{y\sqrt{n}} e^{-\rho^2/2} \rho^{n-1} d\rho,$$

где постоянная

$$C_n = \frac{1}{(\sqrt{2\pi})^n} \int\limits_{-\pi/2}^{\pi/2} \ldots \int\limits_{-\pi/2}^{\pi/2} D(\theta_1 \ldots \theta_{n-1}) d\theta_{n-1} \ldots d\theta_1$$

зависит только от n.

Эту постоянную легко вычислить, пользуясь равенством

$$\Phi(+\infty) = 1 = C_n \int\limits_{0}^{\infty} e^{-\rho^2/2} \rho^{n-1} d\rho = C_n \Gamma\left(\frac{n}{2}\right) \cdot 2^{n/2-1} .$$

Отсюда находим, что

$$\Phi(y) = \frac{1}{2^{n/2-1}\Gamma(n/2)} \int_0^{y\sqrt{n}} \rho^{n-1} e^{-\rho^2/2}\, d\rho.$$

Плотность распределения случайной величины ζ при $y \geqslant 0$ равна

$$\varphi(y) = \frac{\sqrt{2n}}{\Gamma(n/2)} \left(\frac{y\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-ny^2/2} \qquad (6)$$

Отсюда, в частности, при $n = 1$ мы получим, естественно, плотность распределения, равную удвоенной плотности исходного нормального закона

$$\varphi(y) = \sqrt{2/\pi}\, e^{-y^2/2} \qquad (y \geqslant 0).$$

При $n = 3$ мы получаем известный закон Максвелла

$$\varphi(y) = \frac{3\sqrt{6}}{\sqrt{\pi}} y^2 e^{-3y^2/2}.$$

Из формулы (6) легко вывести плотность распределения величины χ^2. Эта плотность равна 0 при $x \leqslant 0$, а при $x > 0$

$$p_n(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)}.$$

Распределения величин, тесно связанных с χ^2 и часто используемых на практике, сведем в следующую таблицу:

Таблица 13

Величина	Плотность распределения при $x > 0$
$\chi^2 = \frac{1}{\sigma^2} \sum_{k=1}^{n} (\xi_k - a)^2$	$\frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)}$
$\frac{1}{n}\chi^2 = \frac{1}{n\sigma^2} \sum_{k=1}^{n} (\xi_k - a)^2$	$\frac{(n/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-nx/2}$
$\chi = \sqrt{\frac{1}{\sigma^2} \sum_{k=1}^{n} (\xi_k - a)^2}$	$\frac{2}{2^{n/2}\Gamma(n/2)} x^{n-1} e^{-x^2/2}$
$\zeta = \frac{\chi}{\sqrt{n}} = \sqrt{\frac{1}{n\sigma^2} \sum_{k=1}^{n} (\xi_k - a)^2}$	$\frac{\sqrt{2n}}{\Gamma(n/2)} \left(\frac{x\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-nx^2/2}$

П р и м е р 4. Ф у н к ц и я р а с п р е д е л е н и я ч а с т н о г о. Пусть плотность распределения вероятностей величины (ξ, η) равна $p(x, y)$. Требуется найти функцию распределения частного $\zeta = \xi/\eta$.

Согласно определению

$$F_\zeta(x) = \mathbf{P}\{\xi/\eta < x\}.$$

Если ξ и η изображают координаты точки на плоскости, то $F_\zeta(x)$ равна вероятность того, что точка (ξ, η) попадает в область, координаты точек

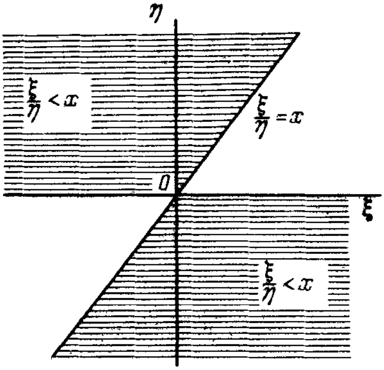

Рис. 17

которой удовлетворяют неравенству $\xi/\eta < x$. На рис. 17 эта область заштрихована.

Согласно общей формуле искомая вероятность равна

$$F_\zeta(x) = \int_0^\infty \int_{-\infty}^{zx} p(y, z)\,dy\,dz + \int_{-\infty}^0 \int_{zx}^\infty p(y, z)\,dy\,dz. \tag{7}$$

Отсюда вытекает, что если ξ и η независимы, а $p_1(x)$ и $p_2(x)$ – их плотности распределения, то

$$F_\zeta(x) = \int_0^\infty F_1(xz)p_2(z)\,dz + \int_{-\infty}^0 (1 - F_1(xz))\,p_2(z)\,dz. \tag{7'}$$

Продифференцировав (7), находим, что

$$p_\zeta(x) = \int_0^\infty zp(zx, z)\,dz - \int_{-\infty}^0 zp(zx, z)\,dz. \tag{8}$$

В частности, если ξ и η независимы, то

$$p_\zeta(x) = \int_0^\infty z p_1(zx) p_2(z) dz - \int_{-\infty}^0 z p_1(zx) p_2(z) dz. \qquad (8')$$

Пример 5. Случайная величина (ξ, η) распределена по нормальному закону

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{x^2}{\sigma_1^2} - 2r\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]\right\}.$$

Найти функцию распределения частного $\zeta = \xi/\eta$.

По формуле (8)

$$p_\zeta(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times$$

$$\times \left[\int_0^\infty - \int_{-\infty}^0\right] z \exp\left\{-\frac{z^2}{2(1-r^2)}\left[\frac{\sigma_2^2 x^2 - 2r\sigma_1\sigma_2 x + \sigma_1^2}{\sigma_1^2\sigma_2^2}\right]\right\} dz =$$

$$= \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^\infty z \exp\left\{-\frac{z^2}{2(1-r^2)} \cdot \frac{\sigma_2^2 x^2 - 2r\sigma_1\sigma_2 x + \sigma_1^2}{\sigma_1^2\sigma_2^2}\right\} dz.$$

Произведем под интегралом замену, положив

$$u = \frac{z^2}{2(1-r^2)} \frac{\sigma_2^2 x^2 - 2r\sigma_1\sigma_2 x + \sigma_1^2}{\sigma_1^2\sigma_2^2}.$$

Выражение для $p_\zeta(x)$ при этом принимает такой вид:

$$p_\zeta(x) = \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{\pi(\sigma_2^2 x^2 - 2r\sigma_1\sigma_2 x + \sigma_1^2)} \int_0^\infty e^{-u} du = \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{\pi(\sigma_2^2 x^2 - 2r\sigma_1\sigma_2 x + \sigma_1^2)};$$

если, в частности, величины ξ и η независимы, то

$$p_\zeta(x) = \frac{\sigma_1\sigma_2}{\pi(\sigma_1^2 + \sigma_2^2 x^2)}.$$

Плотность распределения величины ζ называется *законом Коши.*

Пример 6. Распределение Стьюдента. Найти функцию распределения частного $\zeta = \xi/\eta$, где ξ и η – независимые величины, причем ξ распределено по нормальному закону

$$p_\xi(x) = \sqrt{\frac{n}{2\pi}}\, e^{-\frac{nx^2}{2}},$$

а $\eta = \chi/\sqrt{n}$ (см. пример 3),

$$p_\eta(x) = \frac{\sqrt{2n}}{\Gamma(n/2)}\left(\frac{y\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-\frac{ny^2}{2}}.$$

Согласно формуле (8′)

$$p_\zeta(x) = \int\limits_0^\infty z\sqrt{\frac{n}{2\pi}}\, e^{-\frac{nz^2x^2}{2}} \frac{\sqrt{2n}}{\Gamma(n/2)}\left(\frac{z\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-\frac{nz^2}{2}}\,dz =$$

$$= \frac{1}{\sqrt{\pi}\,\Gamma(n/2)}\int\limits_0^\infty \left(\frac{z\sqrt{n}}{\sqrt{2}}\right)^{n-1} e^{-\frac{nz^2}{2}(x^2+1)} nz\,dz.$$

Сделав замену

$$u = \frac{nz^2}{2}(x^2+1),$$

находим, что

$$p_\zeta(x) = \frac{(x^2+1)^{-\frac{n+1}{2}}}{\sqrt{\pi}\,\Gamma(n/2)}\int\limits_0^\infty u^{\frac{n-1}{2}} e^{-u}\,du = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\,\Gamma(n/2)}(x^2+1)^{-\frac{n+1}{2}}.$$

Плотность распределения вероятностей

$$p_\zeta(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\,\Gamma(n/2)}(1+x^2)^{-\frac{n+1}{2}}$$

носит название закона Стьюдента (Стьюдент – псевдоним английского статистика Госсета, впервые нашедшего этот закон эмпирическим путем).

При $n = 1$ закон Стьюдента превращается в закон Коши.

Пример 7. Поворот осей координат. По функции распределения двумерной случайной величины (ξ, η) найти функцию распределения величин

$$\xi' = \xi \cos\alpha + \eta \sin\alpha, \tag{9}$$

$$\eta' = -\xi \sin\alpha + \eta \cos\alpha.$$

Обозначим через $F(x, y)$ и $\Phi(x, y)$ функции распределения величин (ξ, η) и (ξ', η'). Если мы станем изображать (ξ, η) и (ξ', η') как прямоугольные координаты точки на плоскости, то легко видеть, что система осей $\xi' O \eta'$ получается из системы $\xi O \eta$ путем поворота последней на угол α. Мы ограничимся случаем $0 < \alpha < \pi/2$, предоставив читателю вывод аналогичных формул для остальных значений α.

Обозначим через $p(x, y)$ плотность распределения вектора (ξ, η) и через $\pi(x, y)$ – вектора (ξ', η'). Из (9) находим, что

$$\xi' = \xi \cos\alpha - \eta \sin\alpha,$$

$$\eta' = \xi \sin\alpha + \eta \cos\alpha$$

и, следовательно,

$$\pi(x, y) = p(x \cos\alpha - y \sin\alpha,\ \ x \sin\alpha + y \cos\alpha). \tag{10}$$

Это равенство дает возможность получить формулу, связывающую функции распределения векторов (ξ, η) и (ξ', η'). При разыскании формулы следует учитывать геометрическую картину.

Пример 8. Двумерная случайная величина (ξ, η) распределена по нормальному закону

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{x^2}{\sigma_1^2} - 2r\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]\right\}.$$

Найти плотность распределения случайных величин

$$\xi' = \xi \cos\alpha + \eta \sin\alpha,$$

$$\eta' = -\xi \sin\alpha + \eta \cos\alpha.$$

Согласно равенству (10)

$$\pi(x', y') = p(x'\cos\alpha - y'\sin\alpha,\ x'\sin\alpha + y'\cos\alpha) =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{ -\frac{1}{2(1-r^2)} [Ax'^2 - 2Bx'y' + Cy'^2] \right\},$$

где обозначено

$$A = \frac{\cos^2\alpha}{\sigma_1^2} - 2r\frac{\cos\alpha\sin\alpha}{\sigma_1\sigma_2} + \frac{\sin^2\alpha}{\sigma_2^2},$$

$$B = \frac{\cos\alpha\sin\alpha}{\sigma_1^2} - r\frac{\sin^2\alpha - \cos^2\alpha}{\sigma_1\sigma_2} - \frac{\cos\alpha\sin\alpha}{\sigma_2^2},$$

$$C = \frac{\sin^2\alpha}{\sigma_1^2} + 2r\frac{\cos\alpha\sin\alpha}{\sigma_1\sigma_2} + \frac{\cos^2\alpha}{\sigma_2^2}.$$

Из полученной формулы мы заключаем, что поворот осей переводит нормальное распределение в нормальное.

Заметим, что если угол α выбран так, что

$$\operatorname{tg} 2\alpha = \frac{2r\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2},$$

то $B = 0$ и

$$\pi(x', y') = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{Ax'^2}{2(1-r^2)} - \frac{By'^2}{2(1-r^2)}}.$$

Это равенство означает, что любая нормально распределенная двумерная случайная величина путем поворота осей координат может быть приведена к системе двух нормально распределенных *независимых* случайных величин. Этот результат может быть перенесен на n-мерные случайные величины.

Можно доказать более сильное предложение, исчерпывающе характеризующее нормальное распределение вероятностей. Пусть на плоскости имеется *невырожденное* (т.е. не сосредоточенное на одной прямой) распределение вероятностей. Для того чтобы это распределение было нормальным, необходимо и достаточно, чтобы двумя различными способами можно было на плоскости выбрать оси координат $\xi_1 O \xi_2$ и $\eta_1 O \eta_2$ такие, что координаты ξ_1 и ξ_2, так же как η_1 и η_2, рассматриваемые как случайные величины с заданным распределением вероятностей, были бы независимы.

§ 22. Интеграл Стилтьеса

Дальнейшее изложение существенно использует понятие интеграла Стилтьеса. Для облегчения изучения последующих параграфов мы приводим здесь определение и основные свойства интеграла Стилтьеса, не останавливаясь при этом на доказательствах.

Предположим, что в интервале (a, b) определены функция $f(x)$ и неубывающая функция $F(x)$ с ограниченной вариацией. При этом мы для определенности будем предполагать, что функция $F(x)$ непрерывна слева. Если a и b конечны, то разделим интервал (a, b) точками $a = x_0 < x_1 < < x_2 < \ldots < x_n = b$ на конечное число частичных интервалов (x_i, x_{i+1}) и образуем сумму

$$\sum_{i=1}^{n} f(\tilde{x}_i)[F(x_i) - F(x_{i-1})],$$

где $\tilde{x}_i$ – произвольное число, выбранное в сегменте (x_{i-1}, x_i). Станем теперь увеличивать число точек подразделения, одновременно приближая длину максимального из частичных интервалов к нулю. Если при этом написанная выше сумма стремится к определенному пределу

$$J = \lim_{n \to \infty} \sum_{i=1}^{n} (f(\tilde{x}_i)[F(x_i) - F(x_{i-1})], \tag{1}$$

то этот предел называется интегралом Стилтьеса от функции $f(x)$ с интегрирующей функцией $F(x)$ и обозначается символом

$$J = \int_a^b f(x)dF(x). \tag{2}$$

Несобственный интеграл Стилтьеса, когда промежуток интегрирования бесконечен, определяется обычным путем: рассматривается интеграл по произвольному конечному интервалу (a, b); величины a и b произвольным образом заставляют стремиться к $-\infty$ и $+\infty$: если при этом существует предел

$$\lim_{\substack{a \to -\infty \\ b \to \infty}} \int_a^b f(x)dF(x),$$

то этот предел называется интегралом Стилтьеса от функции $f(x)$ по функции $F(x)$ в промежутке $(-\infty, \infty)$ и обозначается

$$\int f(x)dF(x).$$

Можно доказать, что если функция $f(x)$ непрерывна и ограничена, то предел суммы (1) существует, как в случае конечных, так и в случае бесконечных пределов интегрирования.

В некоторых случаях интеграл Стилтьеса существует и для неограниченных функций $f(x)$. Для теории вероятностей рассмотрение таких интегралов представляет значительный интерес (математическое ожидание, дисперсия, моменты и пр.).

Заметим, что всюду в дальнейшем *мы считаем, что интеграл от непрерывной функции* $f(x)$ *существует тогда и только тогда, когда существует интеграл от* $|f(x)|$ *с той же интегрирующей функцией* $F(x)$.

Для целей теории вероятностей важно распространить определение интеграла Стилтьеса на тот случай, когда функция $f(x)$ может иметь конечное или счетное множество точек разрыва. Можно доказать (см. В.И. Гливенко, Интеграл Стилтьеса, стр. 116), что всякая ограниченная функция, имеющая конечное или счетное множество точек разрыва, в частности, всякая функция ограниченной вариации, интегрируема при любой интегрирующей функции ограниченной вариации. При этом требуется несколько видоизменить само определение интеграла Стилтьеса, именно, при образовании предела (1) надо рассматривать только такие последовательности подразделений интервала интегрирования на части, что каждая точка разрыва $f(x)$ входит в число точек деления всех подразделений, за исключением, быть может, конечного числа их.

Заметим, что при установлении пределов интегрирования важно указывать, включается в промежуток интегрирования или нет тот или иной его конец. Действительно, из определения интеграла Стилтьеса мы получаем следующее равенство (символ $a-0$ означает, что a включено в промежуток интегрирования, а символ $a+0$ – что a исключено из него):

$$\int\limits_{a-0}^{b} f(x)dF(x) = \lim_{n\to\infty} \sum_{i=1}^{n} f(\widetilde{x}_i)[F(x_i) - F(x_{i-1})] =$$

$$= \lim_{n\to\infty} \sum_{i=2}^{n} f(\widetilde{x}_i)[F(x_i) - F(x_{i-1})] + \lim_{x_1\to x_0=a} f(\widetilde{x}_1)[F(x_1) -$$

$$- F(x_0)] = \int\limits_{a+0}^{b} f(x)dF(x) + f(a)[F(a+0) - F(a)].$$

Таким образом, если $f(a) \neq 0$ и функция $F(x)$ имеет скачок при $x = a$, то

$$\int_{a-0}^{b} f(x)dF(x) - \int_{a+0}^{b} f(x)dF(x) = f(a)[F(a+0) - F(a-0)].$$

Это обстоятельство указывает на то, что интеграл Стилтьеса, распространенный на промежуток, сводящийся к одной точке, может давать отличный от нуля результат. Мы условимся в дальнейшем, если не будет сделано особой оговорки, правый конец промежутка исключать, а левый включать в интервал интегрирования. Это условие позволяет написать следующее равенство:

$$\int_{a}^{b} dF(x) = F(b) - F(a).$$

В самом деле, по определению,

$$\int_{a}^{b} dF(x) = \lim_{n \to \infty} \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})] =$$

$$= \lim_{n \to \infty} [F(x_n) - F(x_0)] = F(b) - F(a)$$

(напомним, что $F(x)$, по определению, непрерывна слева и для нее, следовательно, $F(b) = \lim\limits_{s \to 0} F(b - \epsilon)$).

В частности, если $F(x)$ есть функция распределения случайной величины ξ, то

$$\int_{a}^{b} dF(x) = F(b) - F(a) = \mathsf{P}\{a \leqslant \xi < b\},$$

$$\int_{-\infty}^{b} dF(x) = F(b) = \mathsf{P}\{\xi < b\}.$$

Если $F(x)$ имеет производную и является интегралом от нее, то из того, что по формуле конечных приращений

$$F(x_i) - F(x_{i-1}) = p(\widetilde{x}_i)(x_i - x_{i-1}),$$

где $x_{i-1} < \tilde{x}_i < x_i$, следует равенство

$$\int_a^b f(x)\,dF(x) = \lim_{n\to\infty} \sum_{i=1}^{n} f(\tilde{x}_i)[F(x_i) - F(x_{i-1})] =$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} f(\tilde{x}_i)\,p(\tilde{x}_i)(x_i - x_{i-1}) = \int_a^b f(x)\,p(x)\,dx.$$

Мы видим, что в этом случае интеграл Стилтьеса сводится к обыкновенному интегралу.

Если $F(x)$ имеет скачок в точке $x = c$, то, выбрав подразделения так, что при некоторых значениях индекса $x_k < c < x_{k+1}$, имеем:

$$\int_a^b f(x)\,dF(x) = \lim_{n\to\infty} \sum_{i=1}^{k} f(\tilde{x}_i)[F(x_i) - F(x_{i-1})] +$$

$$+ f(c)[F(x_{k+1}) - F(x_k)] + \lim_{n\to\infty} \sum_{i=k+2}^{n} f(\tilde{x}_i)[F(x_i) - F(x_{i-1})] =$$

$$= \int_a^c f(x)\,dF(x) + \int_{c+0}^{b} f(x)\,dF(x) + f(c)[F(c+0) - F(c-0)].$$

В частности, если изменение функции $F(x)$ происходит только в точках $c_1, c_2, \ldots, c_n, \ldots$, то

$$\int_a^b f(x)\,dF(x) = \sum_{n=1}^{\infty} f(c_n)[F(c_n + 0) - F(c_n - 0)]$$

и интеграл Стилтьеса сводится к ряду.

Перечислим основные свойства интеграла Стилтьеса, которые нам потребуются в дальнейшем. Доказательства этих свойств легко могут быть проведены читателем, исходя из определения интеграла Стилтьеса и пользуясь рассуждениями, используемыми в теории обычного интеграла.

1. При $a < c_1 < c_2 < \ldots < c_n < b$

$$\int_a^b f(x)\,dF(x) = \sum_{i=0}^{n} \int_{c_i}^{c_{i+1}} f(x)\,dF(x) \qquad [a = c_0,\ b = c_{n+1}].$$

2. Постоянный множитель выносится за знак интеграла

$$\int_a^b cf(x)dF(x) = c\int_a^b f(x)dF(x).$$

3. Интеграл от суммы функций равен сумме их интегралов

$$\int_a^b \sum_{i=1}^n f_i(x)dF(x) = \sum_{i=1}^n \int_a^b f_i(x)dF(x).$$

4. Если $f(x) \geqslant 0$ и $b > a$, то

$$\int_a^b f(x)dF(x) \geqslant 0.$$

5. Если $F_1(x)$ и $F_2(x)$ – монотонные функции с ограниченным изменением, а c_1 и c_2 – произвольные постоянные, то

$$\int_a^b f(x)d[c_1F_1(x)+c_2F_2(x)] = c_1\int_a^b f(x)dF_1(x)+c_2\int_a^b f(x)dF_2(x).$$

6. Если $F(x) = \int_c^x g(u)dG(u)$, где c – постоянное, $g(u)$ – непрерывная функция и $G(u)$ – неубывающая функция с ограниченным изменением, то

$$\int_a^b f(x)dF(x) = \int_a^b f(x)g(x)dG(x).$$

Использовав понятие интеграла Стилтьеса, мы можем написать общие формулы для функции распределения суммы

$$F(x) = \int F_1(x-z)dF_2(z) = \int F_2(x-z)dF_1(z)$$

двух независимых слагаемых, а также частного $\frac{\xi_1}{\xi_2}$

$$F(x) = \int_0^\infty F_1(xz)dF_2(z) + \int_{-\infty}^0 [1-F_1(xz)]dF_2(z)$$

независимых случайных величин ξ_1 и ξ_2 в предположении, что $\mathbf{P}\{\xi_2 = 0\} = 0$.

Упражнения

1. Доказать, что если $F(x)$ – функция распределения, то при любом $h \neq 0$ функции

$$\Phi(x) = \frac{1}{h}\int\limits_{x}^{x+h} F(x)\,dx, \qquad \Psi(x) = \frac{1}{2h}\int\limits_{x-h}^{x+h} F(x)\,dx$$

также являются функциями распределения.

2. Случайная величина ξ имеет $F(x)$ своей функцией распределения ($p(x)$ – плотность распределения). Найти функцию распределения (плотность распределения) случайной величины:

а) $\eta = a\xi + b$, a и b – вещественные числа;

б) $\eta = \xi^{-1}$ $(\mathbf{P}\{\xi = 0\} = 0)$;

в) $\eta = \operatorname{tg} \xi$;

г) $\eta = \cos \xi$;

д) $\eta = f(\xi)$, где $f(x)$ – непрерывная монотонная функция без промежутков постоянства.

3. Из точки $(0, a)$ проведена прямая под углом φ к оси Oy.

Найти функцию распределения абсциссы точки пересечения этой прямой с осью Ox, если

а) угол φ равномерно распределен в промежутке $(0, \pi/2)$,

б) угол φ равномерно распределен в промежутке $(-\pi/2, \pi/2)$.

4. На окружность радиуса R с центром в начале координат наудачу брошена точка [иными словами, полярный угол точки попадания равномерно распределен в промежутке $(-\pi, \pi)$]. Найти плотность распределения

а) абсциссы точки попадания,

б) длины хорды, соединяющей точку попадания с точкой $(-R, 0)$.

5. На отрезок оси ординат между точками $(0, 0)$ и $(0, R)$ наудачу брошена точка [т.е. ордината этой точки равномерно распределена в промежутке $(0, R)$]. Через точку попадания проведена хорда окружности $x^2 + y^2 = R^2$, перпендикулярная к оси Oy. Найти распределение длины этой хорды.

6. Диаметр круга измерен приближенно. Считая, что его величина равномерно распределена в отрезке (a, b), найти распределение площади круга.

7. Плотность распределения случайной величины ξ дана равенством

$$p(x) = \frac{a}{e^{-x} + e^{x}}.$$

Найти:

а) постоянную a;

б) вероятность того, что в двух независимых наблюдениях ξ примет значения, меньшие 1.

8. Функция распределения случайного вектора (ξ, η) имеет вид:

а) $F(x, y) = F_1(x) F_2(y) + F_3(x)$;

б) $F(x, y) = F_1(x) F_2(y) + F_3(x) + F_4(y)$.

Могут ли функции $F_3(x)$ и $F_4(x)$ быть произвольными? Зависимы или независимы компоненты вектора (ξ, η)?

9. На отрезок $(0, a)$ наудачу брошены две точки [т.е. их абсциссы равномерно распределены в отрезке $(0, a)$]. Найти функцию распределения расстояния между ними.

10. На отрезок $(0, a)$ брошено n точек. Считая, что точки разбросаны случайно [т.е. каждая из них расположена независимо от других и распределена равномерно в $(0, a)$], найти:

а) плотность распределения абсциссы k-й слева точки;

б) совместную плотность распределения абсцисс k-й и m-й точек слева $(k < m)$.

11. Над случайной величиной ξ с непрерывной функцией распределения произведено n независимых испытаний, в результате которых были наблюдены следующие значения величины ξ: $x_1, x_2, \ldots, x_n$. Найти функции распределения случайных величин:

а) $\eta_n = \max(x_1, x_2, \ldots, x_n)$;

б) $\zeta_n = \min(x_1, x_2, \ldots, x_n)$;

в) k-го по величине результата наблюдения;

г) совместного распределения k-го и m-го по величине результатов наблюдения.

12. Функция распределения случайного вектора $(\xi_1, \xi_2, \ldots, \xi_n)$ равна $F(x_1, x_2, \ldots, x_n)$. В результате испытания компоненты вектора получили значения $(z_1, z_2, \ldots, z_n)$. Найти функцию распределения случайной величины:

а) $\eta_n = \max(z_1, z_2, \ldots, z_n)$;

б) $\zeta_n = \min(z_1, z_2, \ldots, z_n)$.

13. Случайная величина ξ имеет непрерывную функцию распределения $F(x)$. Как распределена случайная величина $\eta = F(\xi)$?

14. Случайные величины ξ и η независимы; их плотности распределения определяются равенствами

$$p_\xi(x) = p_\eta(x) = 0 \quad \text{при} \quad x \leqslant 0,$$

$$p_\xi(x) = c_1 x^\alpha e^{-\beta x}, \quad p_\eta(x) = c_2 x^\gamma e^{-\beta x} \quad \text{при} \quad x > 0.$$

Найти:

а) постоянные c_1 и c_2;

б) плотность распределения суммы $\xi + \eta$.

15. Найти функцию распределения суммы независимых случайных величин ξ и η, первая из которых равномерно распределена в сегменте $(-h, h)$, а вторая имеет функцию распределения $F(x)$.

16. Плотность распределения случайного вектора (ξ, η, ζ) равна

$$p(x, y, z) = \begin{cases} \dfrac{6}{(1 + x + y + z)^4} & \text{при } x > 0,\ y > 0,\ z > 0, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Найти распределение величины $\xi + \eta + \zeta$.

17. Найти распределение суммы независимых случайных величин ξ_1 и ξ_2, если их распределения заданы условиями:

а) $F_1(x) = F_2(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x$;

б) равномерно распределены соответственно в интервалах $(-5, 1)$; $(1, 5)$;

в) $p_1(x) = p_2(x) = \frac{1}{2\alpha} e^{-\frac{|x|}{\alpha}}$

18. Плотность распределения независимых случайных величин ξ и η равна:

а) $p_\xi(x) = p_\eta(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ ae^{-ax} & \text{при} \quad x > 0 \ (a > 0); \end{cases}$

б) $p_\xi(x) = p_\eta(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0,\ x > a. \\ 1/a & \text{при} \quad 0 < x \leqslant a. \end{cases}$

Найти плотность распределения величины $\zeta = \xi/\eta$.

19. Найти функцию распределения произведения независимых сомножителей ξ и η по их функциям распределения $F_1(x)$ и $F_2(x)$.

20. Случайные величины ξ и η независимы и распределены:

а) равномерно в интервале $(-a, a)$;

б) нормально с параметрами $a = 0$, $\sigma = 1$.

Найти функцию распределения их произведения.

21. Стороны ξ и η треугольника представляют собой независимые случайные величины. По их функциям распределения $F_\xi(x)$ и $F_\eta(x)$ найти функцию распределения третьей стороны, если угол между сторонами ξ и η равен постоянному числу α.

22. Доказать, что если величины ξ и η независимы и их плотность распределения равна

$$p_\xi(x) = p_\eta(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ e^{-x} & \text{при} \quad x > 0, \end{cases}$$

то величины $\xi + \eta$ и ξ/η также независимы.

23. Доказать, что если величины ξ и η независимы и нормально распределены с параметрами $a_1 = a_2 = 0$, $\sigma_1 = \sigma_2 = \sigma$, то величины

$$\zeta = \xi^2 + \eta^2, \quad \delta = \xi/\eta$$

также независимы.

24. Доказать, что если величины ξ и η независимы и распределены по закону χ^2 с параметрами m и n, то величины $\delta = \xi/\eta$ и $\zeta = \xi + \eta$ независимы.

25. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и имеют одну и ту же плотность распределения

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Найти двумерную плотность распределения величин $\eta = \sum_{k=1}^{n} \xi_k$ и $\zeta = \sum_{k=1}^{m} \xi_k$ $(m < n)$.

26. Доказать, что любая функция распределения обладает следующими свойствами,

$$\lim_{x \to \infty} x \int_x^{\infty} \frac{1}{z}\, dF(z) = 0, \qquad \lim_{x \to +0} x \int_x^{\infty} \frac{1}{z}\, dF(z) = 0,$$

$$\lim_{x \to -\infty} x \int_{-\infty}^{x} \frac{1}{z}\, dF(z) = 0, \qquad \lim_{x \to -0} x \int_{-\infty}^{x} \frac{1}{z}\, dF(z) = 0.$$

27. Над случайной величиной ξ, имеющей непрерывную функцию распределения $F(x)$, проведены две серии независимых испытаний, в результате которых ξ приняла значения, расположенные в порядке возрастания в каждой серии:

$$x_1 < x_2 < \ldots < x_M, \qquad y_1 < y_2 < \ldots < y_N.$$

Чему равна вероятность неравенств

$$y_\mu < x_{m+1} < y_{\mu+1},$$

где m и μ заданные числа $(0 < m < M,\ 0 < \mu < N)$?

28. Случайная величина ξ имеет непрерывную функцию распределения $F(x)$. В результате n независимых наблюдений над ξ получены следующие значения $x_1 < x_2 < \ldots < x_n$, упорядоченные по величине. Найти плотность распределения величины

$$\eta = \frac{F(x_n) - F(x_2)}{F(x_n) - F(x_1)}.$$

29. Случайные величины ξ и η независимы и одинаково распределены с плотностью распределения

$$p_\xi(x) = p_\eta(x) = \frac{C}{1 + x^4}.$$

Найти постоянную C и доказать, что величина ξ/η распределена по закону Коши.

30. Случайные величины ξ и η независимы, их плотности распределения соответственно равны

$$p_\xi(x) = \frac{1}{\pi\sqrt{1-x^2}} \quad (|x| < 1);$$

$$p_\eta(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0 \\ xe^{-x^2/2} & \text{при} \quad x > 0. \end{cases}$$

Доказать, что величина $\xi\eta$ нормально распределена.

31. Пусть ξ и ζ независимы и имеют плотности распределения

$$p_\xi(x) = p_\zeta(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ \lambda e^{-\lambda x} & \text{при} \quad x > 0. \end{cases}$$

Доказать, что отношение $\eta = \dfrac{\xi}{\xi + \eta}$ распределено равномерно на отрезке $(0, 1)$.

32. Случайные величины ξ и η независимы и равномерно распределены на отрезке $(-1, 1)$. Вычислить вероятность того, что корни уравнения $x^2 + \xi x + \eta = 0$ вещественны.

(Задачи 29–32) сообщены мне М.И. Ядренко.)

ГЛАВА 5

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

В предыдущей главе мы видели, что наиболее полная характеристика случайной величины дается ее функцией распределения. Действительно, функция распределения одновременно указывает на то, какие значения может принимать случайная величина и с какими вероятностями. Однако в ряде случаев о случайной величине требуется знать гораздо меньше, требуется получить о ней лишь некоторое суммарное представление. Для теории вероятностей и ее приложений большую роль играют некоторые постоянные числа, получаемые по определенным правилам из функций распределения случайных величин. Среди этих постоянных, служащих для получения общей количественной характеристики случайных величин, особенно важны математическое ожидание, дисперсия и моменты различных порядков.

§ 23. Математическое ожидание

Начнем изложение с рассмотрения следующего схематического примера: предположим, что при стрельбе из некоторого орудия для поражения некоторой цели требуется один снаряд с вероятностью p_1, два снаряда – с вероятностью p_2, три снаряда – с вероятностью p_3 и т.д. Кроме того, известно, что n снарядов заведомо достаточно для поражения этой цели. Таким образом, известно, что

$$p_1 + p_2 + \ldots + p_n = 1.$$

Спрашивается, сколько в среднем потребуется снарядов для поражения указанной цели.

Для ответа на поставленный вопрос будем рассуждать так. Предположим, что производится очень большое число стрельб в указанных выше условиях. Тогда на основании теоремы Бернулли мы можем утверждать, что относительное число стрельб, в которых для поражения цели потребовался только один снаряд, приблизительно равно p_1. Точно так же два снаряда потребовалось приблизительно в $100 p_2$ % стрельб и т.д. Таким образом, "в среднем" на поражение одной цели потребуется приблизительно

$$1 \cdot p_1 + 2 \cdot p_2 + \ldots + n \cdot p_n$$

снарядов.

Аналогичные задачи по подсчету среднего значения случайной величины возникают в самых разнообразных задачах. Вот почему в теории вероятностей вводят в рассмотрение особое постоянное, носящее название м а т е м а т и ч е с к о г о о ж и д а н и я. Мы сначала дадим определение для дискретных случайных величин, отправляясь от только что рассмотренного примера.

Пусть

$$x_1, x_2, \ldots, x_n, \ldots$$

обозначают возможные значения дискретной случайной величины ξ, а

$$p_1, p_2, \ldots, p_n, \ldots$$

— соответствующие им вероятности.

Если ряд $\sum\limits_{n=1}^{\infty} x_n p_n$ сходится а б с о л ю т н о, то его сумма называется *математическим ожиданием* случайной величины ξ и обозначается $\mathbf{M}\xi$.

Для непрерывных случайных величин естественным будет следующее определение: если случайная величина ξ непрерывна и $p(x)$ — ее плотность распределения, то математическим ожиданием величины ξ называется интеграл

$$\mathbf{M}\xi = \int xp(x)dx \tag{1}$$

в тех случаях, когда существует интеграл

$$\int |x| p(x)dx.$$

Для произвольной случайной величины ξ с функцией распределения $F(x)$ математическим ожиданием называется интеграл

$$\mathbf{M}\xi = \int xdF(x). \tag{2}$$

Пользуясь определением интеграла Стилтьеса, мы можем дать простое геометрическое истолкование понятию математического ожидания: математическое ожидание равно разности площадей, заключенных между осью ординат, прямой $y = 1$ и кривой $y = F(x)$ в интервале $(0, +\infty)$ и между осью абсцисс, кривой $y = F(x)$ и осью ординат в промежутке $(-\infty, 0)$. На рис. 18 соответствующие площади заштрихованы и указано с каким знаком следует взять в сумме каждую из площадей. Заметим кстати, что геометрическая иллюстрация позволяет математическое ожидание записать в таком виде:

$$\mathbf{M}\xi = -\int_{-\infty}^{0} F(x)dx + \int_{0}^{\infty} (1 - F(x))dx. \tag{3}$$

Сделанное замечание позволяет во многих случаях находить математическое ожидание почти без вычислений. Так, для случайной величины,

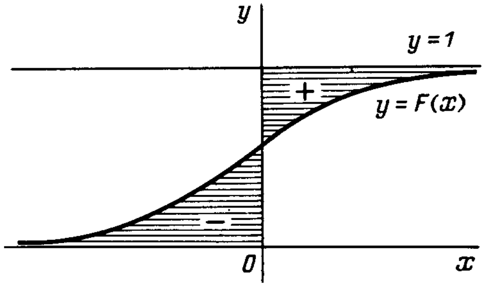

Рис. 18

распределенной по закону, указанному в конце § 19, математическое ожидание равно половине.

Заметим, что среди рассмотренных нами ранее случайных величин, случайная величина, распределенная по закону Коши (пример 5 § 21), не имеет математического ожидания.

Перейдем к рассмотрению примеров.

П р и м е р 1. Найти математическое ожидание случайной величины ξ, распределенной по нормальному закону

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

По формуле (2) находим, что

$$\mathbf{M}\xi = \int x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) dx.$$

Заменой $z = \frac{x - a}{\sigma}$ мы приводим вычисляемый интеграл к виду

$$\mathbf{M}\xi = \frac{1}{\sqrt{2\pi}} \int (\sigma z + a) e^{-z^2/2} dz =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int z e^{-z^2/2} dz + \frac{a}{\sqrt{2\pi}} \int e^{-z^2/2} dz.$$

Так как

$$\int e^{-z^2/2} dz = \sqrt{2\pi} \quad \text{и} \quad \int z e^{-z^2/2} dz = 0,$$

то

$$\mathbf{M}\xi = a.$$

Мы получили важный результат, вскрывающий вероятностный смысл одного из параметров, определяющих нормальный закон: *параметр a в нормальном законе распределения равен математическому ожиданию.*

Пример 2. Найти математическое ожидание случайной величины ξ, равномерно распределенной в интервале (a, b).

Имеем:

$$\mathbf{M}\xi = \int_a^b x \frac{dx}{b - a} = \frac{b^2 - a^2}{2(b - a)} = \frac{a + b}{2}.$$

Мы видим, что математическое ожидание совпадает с серединой интервала возможных значений случайной величины.

Пример 3. Найти математическое ожидание случайной величины ξ, распределенной по закону Пуассона

$$\mathbf{P}\ \xi = k\ = \frac{a^k e^{-a}}{k!} \qquad (k = 0, 1, 2, \ldots).$$

Имеем:

$$\mathbf{M}\xi = \sum_{k=0}^{\infty} k \cdot \frac{a^k e^{-a}}{k!} = \sum_{k=1}^{\infty} k \cdot \frac{a^k e^{-a}}{k!} = a e^{-a} \sum_{k=1}^{\infty} \frac{a^{k-1}}{(k-1)!} =$$

$$= a e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} = a.$$

Если $F(x \mid B)$ есть условная функция распределения для случайной величины ξ, то интеграл

$$\mathsf{M}(\xi \mid B) = \int x \, dF(x \mid B) \tag{4}$$

мы назовем *условным математическим ожиданием случайной величины* ξ *относительно события* B.

Пусть $B_1, B_2, \ldots, B_n$ — полная группа несовместимых событий и $F(x \mid B_1), F(x \mid B_2), \ldots, F(x \mid B_n)$ — соответствующие этим событиям условные функции распределения величины ξ. Обозначим через $F(x)$ безусловную функцию распределения величины ξ; по формуле полной вероятности находим, что

$$F(x) = \sum_{k=1}^{n} \mathsf{P}(B_k) F(x \mid B_k).$$

Это равенство совместно с (4) позволяет нам получить следующую формулу:

$$\mathsf{M}\xi = \sum_{k=1}^{n} \mathsf{P}(B_k) \mathsf{M}(\xi \mid B_k),$$

которая, очевидно, может быть записана и иначе:

$$\mathsf{M}\xi = \mathsf{M}\ \mathsf{M}(\xi \mid B_k) \ . \tag{5}$$

Только что найденная формула во многих случаях значительно упрощает вычисление математических ожиданий.

Пример 4. Рабочий обслуживает n однотипных станков, расположенных прямолинейно на расстоянии a друг от друга (рис. 19). Считая, что

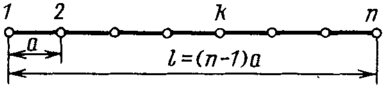

Рис. 19

рабочий обслуживает станки, подходя к ним в порядке очередности, найти средний переход (математическое ожидание величины перехода) между станками.

Пронумеруем станки слева направо от 1-го до n-го и обозначим через B_k событие, состоящее в том, что рабочий находится у станка с номером k. Так как все станки по условию задачи однотипны, то вероятность $p_i^{(k)}$ того, что следующим станком, потребующим внимания рабочего, будет станок с номером i, равна $1/n$ $(1 \leqslant i \leqslant n)$. Величина перехода λ в этом случае равна

$$\lambda_i^{(k)} = \begin{cases} (k-i)a & \text{при } k \geqslant i, \\ (i-k)a & \text{при } k < i. \end{cases}$$

По определению

$$\mathsf{M}(\lambda | B_k) = \frac{1}{n}\left(\sum_{i=1}^{k} (k-i)a + \sum_{i=k+1}^{n} (i-k)a\right) =$$

$$= \frac{a}{n}\left(\frac{k(k-1)}{2} + \frac{(n-k)(n-k+1)}{2}\right) =$$

$$= \frac{a}{2n}\,[2k^2 - 2(n+1)k + n(n+1)]\,.$$

Вероятность рабочему находиться у k-го станка равна $1/n$, поэтому по формуле (5) находим, что

$$\mathsf{M}\lambda = \sum_{k=1}^{n} \frac{a}{2n^2}\,[2k^2 - 2(n+1)k + n(n+1)]\,.$$

Известно, что

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}\,,$$

поэтому

$$\mathsf{M}\lambda = \frac{a(n^2-1)}{3n} = \frac{l}{3}\left(1 + \frac{1}{n}\right),$$

где $l = (n-1)a$ означает расстояние между крайними станками.

Математическим ожиданием n-мерной случайной величины $(\xi_1, \xi_2, \ldots, \xi_n)$ называется совокупность n интегралов

$$a_k = \int\int\ldots\int x_k\,dF(x_1, \ldots, x_k, \ldots, x_n) = \int x\,dF_k(x) = \mathsf{M}\xi_k,$$

где $F_k(x)$ – функция распределения величины ξ_k *).

---

*) Мы не даем формального определения n-мерного интеграла Стилтьеса, во-первых, потому что фактически будем рассматривать только дискретные и непрерывные случайные величины и, во-вторых, потому что, по существу, для теории вероятностей нужна не общая теория интегралов Стилтьеса, а теория абстрактного интеграла Лебега (см. об этом подробнее в гл. 1 монографии Гнеденко и Колмогорова "Предельные распределения для сумм независимых случайных величин", 1949 г.).

П р и м е р 5. Плотность распределения двумерной случайной величины (ξ_1, ξ_2) задана формулой (двумерное нормальное распределение)

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-a)^2}{\sigma_1^2} - \right.\right.$$

$$\left.\left. - \frac{2r(x_1-a)(x_2-b)}{\sigma_1\sigma_2} + \frac{(x_2-b)^2}{\sigma_2^2}\right]\right\};$$

найти ее математическое ожидание.

По определению

$$a_1 = \iint x_1 p(x_1, x_2)\, dx_1 dx_2 = \int x_1 p_1(x_1)\, dx_1$$

и

$$a_2 = \iint x_2 p(x_1, x_2)\, dx_1 dx_2 = \int x_2 p_2(x_2)\, dx_2.$$

В примере 2 § 22 мы видели, что

$$p_1(x_1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{(x_1-a)^2}{2\sigma_1^2}\right\},$$

$$p_2(x_2) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left\{-\frac{(x_2-b)^2}{2\sigma_2^2}\right\},$$

поэтому согласно результатам примера 1 настоящего параграфа находим, что

$$a_1 = a, \qquad a_2 = b.$$

Нам удалось выяснить вероятностный смысл параметров a и b также и для двумерного нормального распределения.

§ 24. Дисперсия

Дисперсией случайной величины ξ называется математическое ожидание квадрата уклонения ξ от $\mathbf{M}\xi$. Мы условимся дисперсию обозначать символом $\mathbf{D}\xi$. Таким образом, по определению

$$\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \int_0^\infty x\, dF_\eta(x), \tag{1}$$

где через $F_\eta(x)$ обозначена функция распределения случайной величины $\eta = (\xi - \mathbf{M}\xi)^2$. Найдем связь между функцией $F_\eta(x)$ и функцией распределения $F_\xi(x)$ величины ξ. Имеем:

$$F_\eta(x) = 0 \quad \text{при} \quad x \leqslant 0,$$

а при $x > 0$

$$F_\eta(x) = \mathbf{P}\{\eta < x\} = \mathbf{P}\{(\xi - \mathbf{M}\xi)^2 < x\} =$$

$$= \mathbf{P}\{-\sqrt{x} < \xi - \mathbf{M}\xi < \sqrt{x}\} = \mathbf{P}\{\mathbf{M}\xi - \sqrt{x} < \xi < \mathbf{M}\xi + \sqrt{x}\} =$$

$$= F_\xi(\mathbf{M}\xi + \sqrt{x}) - F_\xi(\mathbf{M}\xi - \sqrt{x} + 0).$$

Формула (1) переписывается так:

$$\mathbf{D}\xi = \int_0^\infty dx[F_\xi(\mathbf{M}\xi + \sqrt{x}) - F_\xi(\mathbf{M}\xi - \sqrt{x} + 0)] =$$

$$= \int_0^\infty x dF_\xi(\mathbf{M}\xi + \sqrt{x}) - \int_0^\infty x dF_\xi(\mathbf{M}\xi - \sqrt{x} + 0).$$

В первом интеграле произведем замену $z = \mathbf{M}\xi + \sqrt{x}$, а во втором – замену $z = \mathbf{M}\xi - \sqrt{x}$, в результате

$$\int_0^\infty x dF_\xi(\mathbf{M}\xi + \sqrt{x}) = \int_{\mathbf{M}\xi}^\infty (z - \mathbf{M}\xi)^2 dF_\xi(z),$$

$$\int_0^\infty x dF_\xi(\mathbf{M}\xi - \sqrt{x} + 0) = \int_{-\infty}^{\mathbf{M}\xi} (z - \mathbf{M}\xi)^2 dF_\xi(z).$$

Таким образом,

$$\mathbf{D}\xi = \int (z - \mathbf{M}\xi)^2 dF_\xi(z). \tag{2}$$

Так как

$$(z - \mathbf{M}\xi)^2 = z^2 - 2z\mathbf{M}\xi + (\mathbf{M}\xi)^2 \quad \text{и} \quad \mathbf{M}\xi = \int z dF_\xi(z),$$

то формула (2) может быть записана иначе

$$\mathbf{D}\xi = \int z^2 dF_\xi(z) - (\int z dF_\xi(z))^2 = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2. \tag{3}$$

Так как дисперсия является неотрицательной величиной, то из последнего соотношения мы выводим, что

$$\int z^2 dF_\xi(z) \geqslant (\int z dF_\xi(z))^2.$$

Это неравенство представляет собой частный случай известного неравенства Буняковского – Коши.

Подобно математическому ожиданию дисперсия существует не для всех случайных величин. Так, рассмотренный нами ранее (пример 5 § 21) закон Коши не имеет конечной дисперсии.

Рассмотрим примеры вычисления дисперсии.

Пример 1. Найти дисперсию случайной величины ξ, равномерно распределенной в интервале (a, b).

В нашем примере

$$\int x^2 dF_\xi(x) = \int_0^b \frac{x^2}{b-a}\, dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.$$

В предыдущем параграфе было найдено

$$\mathbf{M}\xi = \frac{a+b}{2}.$$

Таким образом,

$$\mathbf{D}\xi = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Мы видим, что дисперсия зависит только от длины интервала (a, b) и является возрастающей функцией длины. Чем больше интервал значений, принимаемых случайной величиной, т.е. чем больше рассеяны ее значения, тем больше дисперсия. Дисперсия, таким образом, играет роль меры рассеяния (разбросанности) значений случайной величины около математического ожидания.

Пример 2. Найти дисперсию случайной величины ξ, распределенной по нормальному закону

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Мы знаем, что $\mathbf{M}\xi = a$, поэтому

$$\mathbf{D}\xi = \int (x-a)^2 p(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx.$$

Произведем под интегралом замену переменных, положив

$$z = \frac{x-a}{\sigma};$$

при этом

$$\mathbf{D}\xi = \frac{\sigma^2}{\sqrt{2\pi}} \int z^2 e^{-z^2/2} dz.$$

Интегрированием по частям находим, что

$$\int z^2 e^{-z^2/2} dz = \int_{-\infty}^{\infty} - z e^{-z^2/2} + \int e^{-z^2/2} dz = \sqrt{2\pi}.$$

Таким образом, окончательно

$$\mathbf{D}\xi = \sigma^2.$$

Мы выяснили, таким образом, вероятностный смысл второго параметра, определяющего нормальный закон. Мы видим, что *нормальный закон распределения полностью определен математическим ожиданием и дисперсией*. Это обстоятельство широко используется в теоретических изысканиях.

Заметим, что и в случае нормально распределенной случайной величины дисперсия позволяет судить о рассеянии ее значений. Хотя при любых положительных значениях дисперсии нормально распределенные случайные величины могут принимать все вещественные значения, все же рассеяние значений случайной величины будет тем меньше, чем меньше дисперсия; при этом вероятности значений, близких к математическому ожиданию, будут больше. Это обстоятельство было отмечено нами в предыдущей главе при первоначальном знакомстве с нормальным законом.

Пример 3. Найти дисперсию случайной величины λ, рассмотренной в примере 4 § 23.

Сохранив обозначения примера 4, находим, что

$$\mathbf{M}(\lambda^2|B_k) = \frac{1}{n}\left(\sum_{i=1}^{k}(k-i)^2a^2 + \sum_{i=k+1}^{n}(i-k)^2a^2\right) =$$

$$= \frac{a^2}{6n}\left[(k-1)\cdot k(2k-1) + (n-k)(n-k+1)(2n-2k+1)\right] =$$

$$= \frac{a^2}{6}\left[6k^2 - 6(n+1)k + (2n+1)(n+1)\right]$$

и, следовательно,

$$\mathbf{M}(\lambda^2) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{M}(\lambda^2|B_k) =$$

$$= \frac{a^2}{6n}\left[n(n+1)(2n+1) - 3(n+1)^2n + n(n+1)(2n+1)\right] = \frac{a^2}{6}(n^2-1).$$

Отсюда следует, что

$$\mathbf{D}(\lambda) = \mathbf{M}(\lambda^2) - (\mathbf{M}\lambda)^2 = \frac{a^2}{6}(n^2-1) - \frac{a^2(n^2-1)^2}{9n^2} =$$

$$= \frac{a^2(n^2-1)(n^2+2)}{18n^2} = \frac{l^2}{18}\left(1 + \frac{2}{n-1} + \frac{1}{n(n-1)} + \frac{2}{n^2(n-1)}\right).$$

Дисперсией n-мерной случайной величины $(\xi_1, \xi_2, \ldots, \xi_n)$ называется совокупность n^2 постоянных, определяемых формулой

$$b_{jk} = \int\int \ldots \int (x_j - \mathbf{M}\xi_j)(x_k - \mathbf{M}\xi_k)\, dF(x_1, x_2, \ldots, x_n)$$
$$(1 \leqslant k \leqslant n, \quad 1 \leqslant j \leqslant n). \tag{4}$$

Так как при любых вещественных $t_j (1 \leqslant j \leqslant n)$

$$\int \ldots \int \{\sum_{j=1}^{n} t_j(x_j - \mathbf{M}x_j)\}^2 dF(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} b_{jk} t_j t_k \geqslant 0,$$

то, как известно из теории квадратичных форм, величины b_{jk} удовлетворяют неравенствам

$$\begin{vmatrix} b_{11} & b_{12} & \ldots & b_{1k} \\ b_{21} & b_{22} & \ldots & b_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ b_{k1} & b_{k2} & \ldots & b_{kk} \end{vmatrix} \geqslant 0 \quad \text{при} \quad k = 1, 2, \ldots, n.$$

Очевидно, что

$$b_{kk} = \mathbf{D}\xi_k.$$

Величины b_{jk} при $k \neq j$ называются *смешанными центральными моментами 2-го порядка* величин ξ_j и ξ_k; очевидно, что $b_{jk} = b_{kj}$.

Следующая функция от моментов второго порядка

$$r_{ij} = \frac{b_{ij}}{\sqrt{b_{ii} b_{jj}}}$$

носит название *коэффициента корреляции* между величинами ξ_i и ξ_j. Коэффициент корреляции является мерой силы связи (линейной связи) между величинами ξ_i и ξ_j. Величина коэффициента корреляции, как это следует из неравенства Буняковского, заключена в пределах $(-1, +1)$. Значения ± 1 достигаются только в случае, когда ξ и η связаны линейной зависимостью.

В дальнейшем мы увидим, что для независимых величин коэффициент корреляции равен нулю.

Пример 4. Найти дисперсию двумерной случайной величины (ξ_1, ξ_2), распределенной по невырожденному нормальному закону

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \times$$
$$\times \exp\left\{-\frac{1}{2(1 - r^2)}\left[\frac{(x - a)^2}{\sigma_1^2} - 2r\frac{(x - a)(y - b)}{\sigma_1\sigma_2} + \frac{(y - b)^2}{\sigma_2^2}\right]\right\}.$$

Согласно формуле (4) и результатам примера 2 настоящего параграфа и примера 1 § 23, находим, что

$$\mathsf{D}\xi_1 = \sigma_1^2, \quad \mathsf{D}\xi_2 = \sigma_2^2.$$

Далее,

$$b_{12} = b_{21} = \iint (x-a)(y-b)p(x,y)\,dx\,dy =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int e^{-\frac{(y-b)^2}{2\sigma_2^2}}\,dy \times$$

$$\times \int (x-a)(y-b)\exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x-a}{\sigma_1} - r\frac{y-b}{\sigma_2}\right)^2\right\}dx.$$

Заменой

$$z = \frac{1}{\sqrt{1-r^2}}\left(\frac{x-a}{\sigma_1} - r\frac{y-b}{\sigma_2}\right), \qquad t = \frac{y-b}{\sigma_2}$$

выражение для b_{12} приводится к виду

$$b_{12} = b_{21} = \frac{1}{2\pi}\iint (\sigma_1\sigma_2\sqrt{1-r^2}\,tz + r\sigma_1\sigma_2 t^2)e^{-\frac{t^2}{2}-\frac{z^2}{2}}\,dz\,dt =$$

$$= \frac{r\sigma_1\sigma_2}{2\pi}\int t^2 e^{-\frac{t^2}{2}}\,dt \int e^{-\frac{z^2}{2}}\,dz + \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{2\pi} \times$$

$$\times \int t e^{-\frac{t^2}{2}}\,dt \int z e^{-\frac{z^2}{2}}\,dz = r\sigma_1\sigma_2.$$

Отсюда находим, что

$$r = \frac{\iint (x-a)(y-b)p(x,y)\,dx\,dy}{\sigma_1\sigma_2} = \frac{\mathsf{M}(\xi_1 - \mathsf{M}\xi_1)(\xi_2 - \mathsf{M}\xi_2)}{\sqrt{\mathsf{D}\xi_1\,\mathsf{D}\xi_2}}.$$

Мы видим, что *двумерный нормальный закон* (так же, как и одномерный) *полностью определяется заданием математического ожидания и дисперсии*, т.е. определяется заданием пяти величин $\mathsf{M}\xi_1$, $\mathsf{M}\xi_2$, $\mathsf{D}\xi_1$, $\mathsf{D}\xi_2$ и r.

§ 25. Теоремы о математическом ожидании и дисперсии

Теорема 1. *Математическое ожидание постоянной равно этой постоянной.*

Доказательство. Постоянную C мы можем рассматривать как дискретную случайную величину, которая может принимать только одно значение C с вероятностью единица; поэтому

$$\mathsf{M}C = C \cdot 1 = C.$$

Теорема 2. *Математическое ожидание суммы случайных величин равно сумме их математических ожиданий:*

$$\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta.$$

Доказательство. Рассмотрим сначала случай дискретных случайных величин ξ и η. Пусть $a_1, a_2, \ldots, a_n, \ldots$ – возможные значения величины ξ и $p_1, p_2, \ldots, p_n, \ldots$ – их вероятности; $b_1, b_2, \ldots, b_k, \ldots$ – возможные значения величины η и $q_1, q_2, \ldots, q_k, \ldots$ – вероятности этих значений. Возможные значения величины $\xi + \eta$ имеют вид $a_n + b_k$ $(k, n = 1, 2, \ldots)$. Обозначим через p_{nk} вероятность того, что ξ примет значение a_n, а η – значение b_k. По определению математического ожидания

$$\mathsf{M}(\xi + \eta) = \sum_{n,k=1}^{\infty} (a_n + b_k)p_{nk} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} (a_n + b_k)p_{nk} =$$

$$= \sum_{n=1}^{\infty} a_n \Big(\sum_{k=1}^{\infty} p_{nk}\Big) + \sum_{k=1}^{\infty} b_k \Big(\sum_{n=1}^{\infty} p_{nk} \Big).$$

Так как по теореме о полной вероятности

$$\sum_{k=1}^{\infty} p_{nk} = p_n \quad \text{и} \quad \sum_{n=1}^{\infty} p_{nk} = q_k,$$

то

$$\sum_{n=1}^{\infty} a_n \sum_{k=1}^{\infty} p_{nk} = \sum_{n=1}^{\infty} a_n p_n = \mathsf{M}\xi$$

и

$$\sum_{k=1}^{\infty} b_k \Big(\sum_{n=1}^{\infty} p_{nk}\Big) = \sum_{k=1}^{\infty} b_k q_k = \mathsf{M}\eta.$$

Доказательство теоремы для случая дискретных слагаемых завершено.

Точно так же в случае, когда существует двумерная плотность распределения $p(x, y)$ случайной величины (ξ, η), по формуле (3) § 23 находим:

$$\mathsf{M}\zeta = \mathsf{M}(\xi + \eta) = \int x dF_\zeta(x) = \int x \Big(\int p(z, x - z)\, dz\Big)\, dx =$$

$$= \iint xp(z, x - z)\, dz\, dx = \iint (z + y)p(z, y)\, dz\, dy =$$

$$= \iint zp(z, y)\, dz\, dy + \iint yp(z, y)\, dz\, dy = \int zp_\xi(z)\, dz + \int yp_\eta(y)\, dy = \mathsf{M}\xi + \mathsf{M}\eta.$$

В общем случае теорема 2 будет доказана нами в дополнении 1.

Следствие 1. *Математическое ожидание суммы конечного числа случайных величин равно сумме их математических ожиданий.*

$$\mathsf{M}(\xi_1 + \xi_2 + \ldots + \xi_n) = \mathsf{M}\xi_1 + \mathsf{M}\xi_2 + \ldots + \mathsf{M}\xi_n.$$

Действительно, в силу только что доказанной теоремы

$$\mathsf{M}(\xi_1 + \xi_2 + \ldots + \xi_n) = \mathsf{M}\xi_1 + \mathsf{M}(\xi_2 + \xi_2 + \ldots + \xi_n) =$$
$$= \mathsf{M}\xi_1 + \mathsf{M}\xi_2 + \mathsf{M}(\xi_3 + \ldots + \xi_n) = \ldots = \mathsf{M}\xi_1 + \mathsf{M}\xi_2 + \ldots + \mathsf{M}\xi_n.$$

Следствие 2. *Математическое ожидание суммы*

$$\zeta_\mu = \xi_1 + \xi_2 + \ldots + \xi_\mu,$$

где μ – случайная величина, принимающая лишь целочисленные значения, случайные величины $\xi_1, \xi_2, \ldots$ не зависят от μ, математическое ожидание μ конечно и ряд

$$\sum_{k=1}^{\infty} \mathsf{M}|\xi_k| \mathsf{P}\{\mu \geqslant k\}$$

сходится, существует и равно

$$\mathsf{M}\zeta_\mu = \sum_{j=1}^{\infty} \mathsf{M}\xi_j \mathsf{P}\{\mu \geqslant j\}.$$

Доказательство. Действительно, условное математическое ожидание, при условии, что $\mu = k$, равно

$$\mathsf{M}\{\zeta_\mu | \mu = k\} = \mathsf{M}\xi_1 + \mathsf{M}\xi_2 + \ldots + \mathsf{M}\xi_k.$$

Безусловное математическое ожидание

$$\mathsf{M}\zeta_\mu = \sum_{k=1}^{\infty} \mathsf{M}\{\zeta_\mu | \mu = k\} \cdot \mathsf{P}\{\mu = k\} = \sum_{k=1}^{\infty} \mathsf{P}\{\mu = k\} \sum_{j=1}^{k} \mathsf{M}\xi_j =$$
$$= \sum_{j=1}^{\infty} \mathsf{M}\xi_j \sum_{k=j}^{\infty} \mathsf{P}\{\mu = k\} = \sum_{j=1}^{\infty} \mathsf{M}\xi_j \mathsf{P}\{\mu \geqslant j\}.$$

Если слагаемые $\xi_1, \xi_2, \xi_3, \ldots$ одинаково распределены, т.е. если $\mathsf{P}\{\xi_1 < x\} = \mathsf{P}\{\xi_2 < x\} = \ldots = F(x)$, то

$$\mathsf{M}\zeta_\mu = \mathsf{M}\zeta_1 \cdot \mathsf{M}\mu.$$

Действительно,

$$\mathsf{M}\zeta_\mu = \sum_{k=1}^{\infty} \mathsf{P}\{\mu = k\} \sum_{j=1}^{k} \mathsf{M}\xi_j = \mathsf{M}\xi_1 \sum_{k=1}^{\infty} k\mathsf{P}\{\mu = k\} = \mathsf{M}\xi_1 \cdot \mathsf{M}\mu.$$

Пример 1. Число космических частиц, попадающих на данную площадку, есть случайная величина μ, подчиненная закону Пуассона с параметром a, каждая из частиц несет энергию ξ, зависящую от случая. Найти среднюю энергию $\mathcal{E}$, получаемую площадкой в единицу времени.

Согласно следствию 2 имеем:

$$\mathsf{M}\mathcal{E} = \mathsf{M}\xi \cdot \mathsf{M}\mu = a\mathsf{M}\xi.$$

П р и м е р 2. По некоторой цели стрельба ведется до n-го попадания. Считая, что выстрелы производятся независимо друг от друга и вероятность попадания при каждом выстреле равна p, найти математическое ожидание расхода снарядов.

Обозначим через ξ_k число снарядов, потраченных от $(k-1)$-го до k-го попадания. Очевидно, что расход снарядов на n попаданий равен

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n$$

и, следовательно,

$$\mathbf{M}\xi = \mathbf{M}\xi_1 + \mathbf{M}\xi_2 + \ldots + \mathbf{M}\xi_n.$$

Но

$$\mathbf{M}\xi_1 = \mathbf{M}\xi_2 = \ldots = \mathbf{M}\xi_n$$

и

$$\mathbf{M}\xi_1 = \sum_{k=0}^{\infty} kq^{k-1}p = \frac{p}{(1-q)^2} = \frac{1}{p},$$

следовательно,

$$\mathbf{M}\xi = n/p.$$

Т е о р е м а 3. *Математическое ожидание произведения независимых случайных величин ξ и η равно произведению их математических ожиданий.*

Д о к а з а т е л ь с т в о. Если величины ξ и η дискретны; $a_1, a_2, \ldots, \ldots, a_k, \ldots$ — возможные значения ξ и $p_1, p_2, \ldots, p_k, \ldots$ — вероятности этих значений; $b_1, b_2, \ldots, b_n, \ldots$ — возможные значения η и $q_1, q_2, \ldots, \ldots, q_n, \ldots$ — вероятности этих значений, то вероятность того, что ξ примет значение a_k, а η — значение b_n, равна $p_k q_n$. По определению математического ожидания

$$\mathbf{M}\xi\eta = \sum_{k,n} a_k b_n p_k q_n = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_k b_n p_k q_n =$$

$$= \Big(\sum_{k=1}^{\infty} a_k p_k\Big)\Big(\sum_{n=1}^{\infty} b_n q_n\Big) = \mathbf{M}\xi\mathbf{M}\eta.$$

Лишь немногим сложнее доказательство для случая непрерывных величин, провести его мы предоставляем читателю.

С л е д с т в и е 1. *Постоянный множитель можно выносить за знак математического ожидания*

$$\mathbf{M}C\xi = C\mathbf{M}\xi.$$

Это утверждение очевидно, так как, каково бы ни было ξ, постоянное C и величину ξ можно рассматривать как независимые величины.

Т е о р е м а 4. *Дисперсия постоянного равна нулю.*

Д о к а з а т е л ь с т в о. Согласно теореме 1,

$$\mathbf{D}C = \mathbf{M}(C - \mathbf{M}C)^2 = \mathbf{M}(C - C)^2 = \mathbf{M}0 = 0.$$

Т е о р е м а 5. *Если c – постоянное, то*

$$\mathbf{D}c\xi = c^2\mathbf{D}\xi.$$

Д о к а з а т е л ь с т в о. В силу следствия из теоремы 3

$$\mathbf{D}c\xi = \mathbf{M}[c\xi - \mathbf{M}c\xi]^2 = \mathbf{M}[c\xi - c\mathbf{M}\xi]^2 =$$
$$= \mathbf{M}c^2[\xi - \mathbf{M}\xi]^2 = c^2\mathbf{M}[\xi - \mathbf{M}\xi]^2 = c^2\mathbf{D}\xi.$$

Т е о р е м а 6. *Дисперсия суммы независимых случайных величин ξ и η равна сумме их дисперсий*

$$\mathbf{D}(\xi + \eta) = \mathbf{D}\xi + \mathbf{D}\eta.$$

Д о к а з а т е л ь с т в о. Действительно,

$$\mathbf{D}(\xi + \eta) = \mathbf{M}[\xi + \eta - \mathbf{M}(\xi + \eta)]^2 = \mathbf{M}[(\xi - \mathbf{M}\xi) + (\eta - \mathbf{M}\eta)]^2 =$$
$$= \mathbf{D}\xi + \mathbf{D}\eta + 2\mathbf{M}(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta).$$

Величины ξ и η независимы, поэтому независимы также величины $\xi - \mathbf{M}\xi$ и $\eta - \mathbf{M}\eta$; отсюда

$$\mathbf{M}(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta) = \mathbf{M}(\xi - \mathbf{M}\xi) \cdot \mathbf{M}(\eta - \mathbf{M}\eta) = 0.$$

С л е д с т в и е 1. *Если $\xi_1, \xi_2, \ldots, \xi_n$ – случайные величины, каждая из которых независима от суммы предыдущих, то*

$$\mathbf{D}(\xi_1 + \xi_2 + \ldots + \xi_n) = \mathbf{D}\xi_1 + \mathbf{D}\xi_2 + \ldots + \mathbf{D}\xi_n.$$

С л е д с т в и е 2. *Дисперсия суммы конечного числа попарно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равна сумме их дисперсий.*

Д о к а з а т е л ь с т в о. Действительно

$$\mathbf{D}(\xi_1 + \xi_2 + \ldots + \xi_n) = \mathbf{M}\left(\sum_{k=1}^{n}(\xi_k - \mathbf{M}\xi_k)\right)^2 =$$
$$= \mathbf{M}\sum_{j=1}^{n}\sum_{k=1}^{n}(\xi_k - \mathbf{M}\xi_k)(\xi_j - \mathbf{M}\xi_j) = \sum_{k=1}^{n}\sum_{j=1}^{n}\mathbf{M}(\xi_k - \mathbf{M}\xi_k)(\xi_j - \mathbf{M}\xi_j) =$$
$$= \sum_{k=1}^{n}\mathbf{D}\xi_k + \sum_{k \neq j}\mathbf{M}(\xi_k - \mathbf{M}\xi_k)(\xi_j - \mathbf{M}\xi_j).$$

Из независимости любой пары величин ξ_k и ξ_j $(k \neq j)$ вытекает, что при $k \neq j$

$$\mathbf{M}(\xi_k - \mathbf{M}\xi_k)(\xi_j - \mathbf{M}\xi_j) = 0.$$

Этим, очевидно, доказательство завершено.

Пример 1. *Нормированным уклонением* случайной величины ξ называется отношение $\dfrac{\xi - \mathbf{M}\xi}{\sqrt{\mathbf{D}\xi}}$.

Доказать, что $\mathbf{D}\left(\dfrac{\xi - \mathbf{M}\xi}{\sqrt{\mathbf{D}\xi}}\right) = 1$.

Действительно, ξ и $\mathbf{M}\xi$, рассматриваемые как случайные величины, независимы, поэтому в силу теорем 5 и 6

$$\mathbf{D}\left(\frac{\xi - \mathbf{M}\xi}{\sqrt{\mathbf{D}\xi}}\right) = \frac{\mathbf{D}\xi + \mathbf{D}(-\mathbf{M}\xi)}{\mathbf{D}\xi} = \frac{\mathbf{D}\xi}{\mathbf{D}\xi} = 1.$$

Пример 2. Если ξ и η — независимые случайные величины, то

$$\mathbf{D}(\xi - \eta) = \mathbf{D}\xi + \mathbf{D}\eta.$$

Действительно, в силу теорем 5 и 7

$$\mathbf{D}(-\eta) = (-1)^2\ \mathbf{D}\eta = \mathbf{D}\eta$$

и

$$\mathbf{D}(\xi - \eta) = \mathbf{D}\xi + \mathbf{D}\eta.$$

Пример 3. Теоремы 2 и 6 позволяют весьма просто вычислять математическое ожидание и дисперсию числа μ наступлений события A при n независимых испытаниях.

Пусть p_k есть вероятность появления события A при k-м испытании. Обозначим через μ_k число появлений события A при k-м испытании. Очевидно, что μ_k есть случайная величина, принимающая значения 0 и 1 с вероятностями $q_k = 1 - p_k$ и p_k соответственно.

Величина μ, таким образом, может быть представлена в виде суммы

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n.$$

Так как

$$\mathbf{M}\mu_k = 0 \cdot q_k + 1 \cdot p_k = p_k$$

и

$$\mathbf{D}\mu_k = \mathbf{M}\mu_k^2 - (\mathbf{M}\mu_k)^2 = 0 \cdot q_k + 1 \cdot p_k - p_k^2 = p_k(1 - p_k) = p_k q_k,$$

то доказанные теоремы позволяют заключить, что

$$\mathbf{M}\mu = p_1 + p_2 + \ldots + p_n$$

и

$$\mathbf{D}\mu = p_1 q_1 + p_2 q_2 + \ldots + p_n q_n.$$

Для случая схемы Бернулли $p_k = p$ и, следовательно,

$$\mathsf{M}\mu = np \text{ и } \mathsf{D}\mu = npq.$$

Заметим, что отсюда

$$\mathsf{M}\frac{\mu}{n} = p; \qquad \mathsf{D}\frac{\mu}{n} = \frac{pq}{n}.$$

§ 26. Моменты

Моментом k-го порядка случайной величины ξ называется математическое ожидание величины $(\xi - a)^k$:

$$\nu_k(a) = \mathsf{M}(\xi - a)^k. \tag{1}$$

Если $a = 0$, то момент называется *начальным*. Легко видеть, что начальный момент первого порядка есть математическое ожидание величины ξ.

Если $a = \mathsf{M}\xi$, то момент называется *центральным*. Легко видеть, что центральный момент первого порядка равен нулю, а центральный момент второго порядка есть не что иное, как дисперсия.

Начальные моменты мы станем обозначать буквой ν_k, а центральные — буквой μ_k, указывая в обоих случаях нижним индексом порядок момента.

Между центральными и начальными моментами существует простая связь. Действительно,

$$\mu_n = \mathsf{M}(\xi - \mathsf{M}\xi)^n = \sum_{k=0}^{n} C_n^k (-\mathsf{M}\xi)^{n-k}\mathsf{M}\xi^k = \sum_{k=0}^{n} C_n^k (-\mathsf{M}\xi)^{n-k}\nu_k. \tag{2}$$

Так как $\nu_1 = \mathsf{M}\xi$, то

$$\mu_n = \sum_{k=2}^{n} (-1)^{n-k} C_n^k \nu_k \nu_1^{n-k} + (-1)^{n-1}(n-1)(\nu_1)^n. \tag{3}$$

Выпишем эту связь между моментами для первых четырех значений n:

$$\begin{aligned}
\mu_0 &= 1,\\
\mu_1 &= 0,\\
\mu_2 &= \nu_2 - \nu_1^2,\\
\mu_3 &= \nu_3 - 3\nu_2\nu_1 + 2\nu_1^2,\\
\mu_4 &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.
\end{aligned} \tag{3'}$$

Эти первые моменты играют особо важную роль в статистике.

Величина

$$m_k = \mathsf{M}\,|\,\xi - a\,|^k \tag{4}$$

носит название *абсолютного момента k-го порядка*.

Согласно определению математического ожидания $\mathbf{M}(\xi - a)^k$ должно вычисляться по формуле

$$\nu_k(a) = \int x dG(x), \tag{1'}$$

где $G(x)$ обозначает функцию распределения величины $(\xi - a)^k$. Однако при действительных расчетах предпочитают пользоваться другой формулой:

$$\nu_k(a) = \int (x - a)^k dF(x), \tag{5}$$

где $F(x)$ – функция распределения величины ξ. Для того чтобы формулы (1′) и (5) не противоречили друг другу, необходимо, чтобы имели место равенства

$$\int x dG(x) = \int (x - a)^k dF(x).$$

Докажем, что это действительно так.

Если k – нечетное число, то $(\xi - a)^k$ – неубывающая функция ξ и поэтому

$$G(x) = \mathbf{P}\{(\xi - a)^k < x\} = \mathbf{P}\{\xi - a < \sqrt[k]{x}\} =$$
$$= \mathbf{P}\{\xi < a + \sqrt[k]{x}\} = F(a + \sqrt[k]{x}).$$

При нечетном k, таким образом,

$$\mathbf{M}(\xi - a)^k = \int x dF(a + \sqrt[k]{x}).$$

Нетрудно подсчитать, что заменой $z = a + \sqrt[k]{x}$ мы приводим этот интеграл к виду (5).

Если же k – четное, то $(\xi - a)^k$ есть неотрицательная величина и, следовательно, $G(x) = 0$ при $x \leqslant 0$. При $x > 0$

$$G(x) = \mathbf{P}\{(\xi - a)^k < x\} = \mathbf{P}\{a - \sqrt[k]{x} < \xi < a + \sqrt[k]{x}\} =$$
$$= F(a + \sqrt[k]{x}) - F(a - \sqrt[k]{x} + 0).$$

Таким образом, при четном k

$$\mathbf{M}(\xi - a)^k = \int_0^\infty x dF(a + \sqrt[k]{x}) - \int_0^\infty x dF(a - \sqrt[k]{x} + 0).$$

Заменами $z = a + \sqrt[k]{x}$ в первом интеграле и $z = a - \sqrt[k]{x}$ – во втором мы приводим $\mathbf{M}(\xi - a)^k$ к виду (5).

Мы доказали частный случай следующей теоремы.

Теорема 1. *Если $F(x)$ – функция распределения величины ξ, $f(x)$ – непрерывная функция, то*

$$\mathbf{M}f(\xi) = \int f(x) dF(x).$$

Так как мы условились считать, что случайная величина ξ имеет математическое ожидание только в том случае, когда изображающий его интеграл абсолютно сходится, то ясно, что момент k-го порядка у величины ξ существует тогда и только тогда, когда сходится интеграл

$$\int |x|^k dF_\xi(x).$$

Из этого замечания следует, что если случайная величина ξ имеет момент k-го порядка, то она имеет также моменты всех положительных порядков, меньших чем k. В самом деле, так как при $r < k$ $|x|^k > |x|^r$, если $|x| > 1$, то

$$\int |x|^r dF_\xi(x) = \int_{|x|\leqslant 1} |x|^r dF_\xi(x) + \int_{|x|>1} |x|^r dF_\xi(x) \leqslant$$
$$\leqslant \int_{|x|\leqslant 1} |x|^r dF_\xi(x) + \int_{|x|>1} |x|^k dF_\xi(x).$$

Первый интеграл в правой части неравенства конечен в силу конечности пределов интегрирования и ограниченности подинтегральной функции, второй интеграл сходится в силу предположения.

Пример. Найти центральные и центральные абсолютные моменты случайной величины, распределенной по нормальному закону

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Имеем:

$$\mu_k = \frac{1}{\sigma\sqrt{2\pi}} \int (x-a)^k \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \frac{\sigma^k}{\sqrt{2\pi}} \int x^k e^{-\frac{x^2}{2}} dx.$$

При k нечетном, в силу нечетности подинтегральной функции,

$$\mu_k = 0.$$

При k четном

$$\mu_k = m_k = \sqrt{\frac{2}{\pi}}\, \sigma^k \int_0^\infty x^k e^{-\frac{x^2}{2}} dx.$$

Подстановкой $x^2 = 2z$ мы приводим этот интеграл к виду

$$\mu_k = m_k = \sqrt{\frac{2}{\pi}}\, \sigma^k 2^{\frac{k-1}{2}} \int_0^\infty z^{\frac{k-1}{2}} e^{-z} dz =$$

$$= \sqrt{\frac{2}{\pi}}\, \sigma^k 2^{\frac{k-1}{2}} \Gamma\left(\frac{k+1}{2}\right) = \sigma^k (k-1)(k-3)\ldots 1 = \sigma^k \frac{k!}{2^{k/2}(k/2)!}.$$

При k нечетном абсолютный момент равен

$$m_k = \sqrt{\frac{2}{\pi}}\,\sigma^k \int_0^\infty x^k e^{-\frac{x^2}{2}}\,dx = \sqrt{\frac{2}{\pi}}\,\sigma^k 2^{\frac{k-1}{2}}\,\Gamma\left(\frac{k+1}{2}\right) =$$

$$= \sqrt{\frac{2}{\pi}}\,2^{\frac{k-1}{2}}\left(\frac{k-1}{2}\right)!\,\sigma^k.$$

Моменты функций распределения не могут быть произвольными величинами. Действительно, каковы бы ни были постоянные $t_0, t_1, \ldots, t_n$, квадратичная форма

$$J_n = \int \left(\sum_{k=0}^{n} t_k (x-a)^k\right)^2 dF(x) = \sum_{j=1}^{n}\sum_{k=1}^{n} \nu_{k+j}(a) t_k t_j \geqslant 0$$

неотрицательна; поэтому первые $\nu_j(a)$ должны удовлетворять следующим неравенствам:

$$\begin{vmatrix} \nu_0(a) & \nu_1(a) & \ldots & \nu_k(a) \\ \nu_1(a) & \nu_2(a) & \ldots & \nu_{k+1}(a) \\ \ldots & \ldots & \ldots & \ldots \\ \nu_k(a) & \nu_{k+1}(a) & \ldots & \nu_{2k}(a) \end{vmatrix} \geqslant 0 \quad (k = 1, 2, \ldots, n).$$

Аналогичным неравенствам подчинены и абсолютные моменты.

Относительно абсолютных моментов мы докажем еще следующую теорему.

Т е о р е м а 2. *Если случайная величина ξ имеет абсолютный момент порядка k, то при любых t и τ* $(0 < t < \tau < k)$

$$\sqrt[t]{m_t} \leqslant \sqrt[\tau]{m_\tau} \leqslant \sqrt[k]{m_k},$$

где обозначено

$$m_t = \mathsf{M}|\xi - a|^t,$$

a – любое вещественное число.

Д о к а з а т е л ь с т в о. Докажем сначала теорему для того случая, когда t, τ и k – рациональные числа. Пусть для определенности

$$t = p/q, \qquad \tau = s/q, \qquad k = u/q,$$

причем по предположению теоремы

$$p < s < u.$$

Пусть теперь r — какое-нибудь целое положительное число, меньшее чем u. Рассмотрим неотрицательную квадратичную форму

$$m_{\frac{r-1}{q}} u^2 + 2m_{\frac{r}{q}} uv + m_{\frac{r+1}{q}} v^2 = \int [\, u|x|^{\frac{r-1}{2q}} + v|x|^{\frac{r+1}{2q}}]^2\, dF(x).$$

Условие ее неотрицательности состоит, как известно, в том, что

$$m^2_{\frac{r}{q}} \leqslant m_{\frac{r-1}{q}} \cdot m_{\frac{r+1}{q}}.$$

Это неравенство, очевидно, может быть записано и в таком виде:

$$m^{2r}_{\frac{r}{q}} \leqslant m^r_{\frac{r-1}{q}} \cdot m^r_{\frac{r+1}{q}}.$$

Если придать r последовательно значения от 1 до r, то мы получим последовательность неравенств

$$m^2_{\frac{1}{q}} \leqslant m_0 m_{\frac{2}{q}}, \quad m^{2\cdot 2}_{\frac{2}{q}} \leqslant m^2_{\frac{1}{q}} m^2_{\frac{3}{q}}, \ldots \quad m^{2r}_{\frac{r}{q}} \leqslant m^r_{\frac{r-1}{q}} m^r_{\frac{r+1}{q}}.$$

Заметив, что всегда $m_0 = 1$, перемножив выписанные неравенства и произведя сокращения, мы приходим к неравенству

$$m^{r+1}_{\frac{r}{q}} \leqslant m^r_{\frac{r+1}{q}}.$$

Таким образом, $m^{\frac{1}{r}}_{\frac{r}{q}} \leqslant m^{\frac{1}{r+1}}_{\frac{r+1}{q}}$, или же $m^{\frac{q}{r}}_{\frac{r}{q}} \leqslant m^{\frac{q}{r+1}}_{\frac{r+1}{q}}$.

Это неравенство доказывает, очевидно, теорему, в случае t, τ и k рациональных.

Так как функция m_t непрерывна относительно аргумента t в области $0 \leqslant t \leqslant k$, то предельным переходом мы убеждаемся в правильности теоремы при любых t, τ и k.

Заметим, что в только что доказанной теореме содержится следующее важное свойство моментов:

$$m_1 \leqslant m_2^{\frac{1}{2}} \leqslant m_3^{\frac{1}{3}} \leqslant \ldots \leqslant m_k^{\frac{1}{k}} \leqslant m_{k+1}^{\frac{1}{k+1}} \leqslant \ldots$$

В примерах предыдущих параграфов два первых момента случайной величины полностью определяли ее функцию распределения, если только заранее известен вид этой функции (так это имело место для распределений нормального, Пуассона, равномерного и др.). В математической статистике играют значительную роль законы распределения, зависящие от большего чем два числа параметров. Если заранее известно, что случайная величина подчинена закону вполне определенного вида, но неизвестны лишь значения параметров, то эти неизвестные параметры в важнейших случаях определяются через первые моменты. Если же нам неизвестно, к какому виду принадлежит функция распределения, то, вообще говоря, не только знание одних первых, но и знание всех целочисленных моментов не дает возможности определить неизвестную функцию распределения. Оказывается, можно построить примеры различных функций распределения с одинаковыми моментами всех целочисленных порядков. В связи с этим возникает вопрос (проблема моментов): дана последовательность постоянных чисел

$c_0 = 1,\ c_1, c_2, c_3, \ldots;$

1) при каких условиях существует такая функция распределения $F(x)$, для которой при всех n имеют место равенства

$c_n = \int x^n dF(x),$

2) когда эта функция единственна?

В настоящее время эта задача получила полное решение, мы не останавливаемся на нем, так как оно стоит в стороне от назначения нашей книги.

Среди прочих числовых характеристик наиболее существенную роль играют так называемые семиинварианты; их определение мы отложим до главы 7, сейчас же отметим только следующее. При сложении независимых случайных величин момент суммы, вообще говоря, не равен сумме моментов слагаемых. Для момента суммы независимых слагаемых ξ и η имеет место равенство

$$\mathsf{M}(\xi + \eta)^n = \sum_{k=0}^{n} C_n^k \mathsf{M}\xi^k \mathsf{M}\eta^{n-k}.$$

Семиинварианты различных порядков обладают тем свойством, что при сложении независимых слагаемых семиинвариант суммы равен сумме семиинвариантов слагаемых того же порядка. Оказывается, что семиинвариант любого порядка k есть рациональная функция моментов порядков, меньших или равных k.

Упражнения

1. Случайная величина ξ принимает целые неотрицательные значения с вероятностями

а) $\mathsf{P}\{\xi = k\} = \dfrac{a^k}{(1+a)^{k+1}}$, $a > 0$ – постоянная; это распределение носит название *распределения Паскаля*:

б) $p_k = \mathbf{P}\{\xi = k\} = \left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^k \frac{(1+\alpha)\ldots(1+(k-1)\alpha)}{k!} p_0$ при всех $k > 0$, где $\alpha > 0$, $\lambda > 0$ и $p_0 = \mathbf{P}\{\xi = 0\} = (1+\alpha\lambda)^{-1}$. Это распределение носит название *распределения Пойа*.

Найти $\mathbf{M}\xi$ и $\mathbf{D}\xi$.

2. Пусть μ – число появлений события A в n независимых испытаниях, в каждом из которых $\mathbf{P}(A) = p$. Найти

а) $\mathbf{M}\mu^3$, б) $\mathbf{M}\mu^4$, б) $\mathbf{M}\,|\mu - np|$.

3. Вероятность появления события A в k-м испытании равна p_k. Пусть μ – число появлений события A в n первых независимых испытаниях. Найти

а) $\mathbf{M}\mu$, б) $\mathbf{D}\mu$, в) $\mathbf{M}(\mu - \sum_{i=1}^{n} p_i)^3$, г) $\mathbf{M}(\mu - \sum_{i=1}^{n} p_i)^4$.

4. Доказать, что в условиях предыдущей задачи максимум $\mathbf{D}\mu$ достигается для данного значения $a = \frac{1}{n}\sum_{1}^{n} p_i$ при условии

$p_1 = p_2 = \ldots = p_n = a$.

5. Пусть μ – число появлений события A в n независимых испытаниях, в каждом из которых $\mathbf{P}(A) = p$. Пусть, далее, величина η равна 0 или 1 в зависимости от того, оказалось μ четным или нечетным. Найти $\mathbf{M}\eta$.

6. Плотность распределения случайной величины ξ равна

$$p(x) = \frac{1}{2\alpha} e^{-\frac{|x-a|}{\alpha}}$$

(*распределение Лапласа*). Найти $\mathbf{M}\xi$ и $\mathbf{D}\xi$.

7. Плотность распределения абсолютной величины скорости молекулы дается *распределением Максвелла*

$$p(x) = \frac{4x^2}{\alpha^3\sqrt{\pi}} e^{-\frac{x^2}{\alpha^2}} \quad \text{при} \quad x > 0$$

и $p(x) = 0$ при $x \leqslant 0$, $\alpha > 0$ - постоянная. Найти среднюю скорость молекулы, ее дисперсию, среднюю кинетическую энергию (масса молекулы равна m) и дисперсию кинетической энергии.

8. Плотность вероятностей молекуле, находящейся в броуновском движении, отстоять на расстоянии x от отталкивающей стенки в момент t, если в момент t_0 она отстояла на расстоянии x_0, дается формулой

$$p(x) = \begin{cases} \dfrac{1}{2\sqrt{\pi D t}} \left\{ e^{-\frac{(x+x_0)^2}{4Dt}} + e^{-\frac{(x-x_0)^2}{4Dt}} \right\} & \text{при} \quad x \geqslant 0, \\ 0 & \text{при} \quad x < 0. \end{cases}$$

Найти математическое ожидание и дисперсию величины перемещения молекулы за время от t_0 до t (D – постоянная).

9. Доказать, что для произвольной случайной величины ξ, возможные значения которой находятся в промежутке (a, b), выполняются следующие неравенства

$$a \leqslant \xi \leqslant b, \qquad \mathbf{D}\xi \leqslant (b-a)^2/4.$$

10. Пусть $x_1, x_2, \ldots, x_n$ – возможные значения случайной величины ξ. Доказать, что при $n \to \infty$

$$\text{а) } \mathbf{M}\xi^{n+1}/\mathbf{M}\xi^n \to \max_j x_j, \qquad \text{б) } \sqrt[n]{\mathbf{M}\xi^n} \to \max_j x_j.$$

11. Пусть $F(x)$ – функция распределения ξ. Доказать, что если $\mathbf{M}\xi$ существует, то

$$\mathbf{M}\xi = \int_0^\infty [1 - F(x) + F(-x)]\,dx$$

и для существования $\mathbf{M}\xi$ необходимо и достаточно, чтобы

$$\lim_{x\to-\infty} xF(x) = \lim_{x\to\infty} x[1 - F(x)] = 0.$$

12. На отрезок $(0, l)$ наудачу брошены две точки. Найти математическое ожидание, дисперсию и математическое ожидание n-й степени расстояния между ними.

13. Случайная величина ξ распределена *логарифмически нормально*, т.е. при $x > 0$ плотность распределения ξ равна

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(\ln x - a)^2}$$

$(p(x) = 0$ при $x < 0)$. Найти $\mathbf{M}\xi$ и $\mathbf{D}\xi$.

14. Случайная величина ξ нормально распределена с параметрами a и σ. Найти $\mathbf{M}\,|\,\xi - a\,|$.

15. В ящике содержится 2^n билетов; номер i $(i = 0, 1, 2, \ldots, n)$ обозначен на C_n^i из них. Наудачу вынимаются m билетов, s – сумма их номеров; найти $\mathbf{M}s$ и $\mathbf{D}s$.

16. Случайные величины $\xi_1, \xi_2, \ldots, \xi_{n+m}$ $(n > m)$ независимы, одинаково распределены и имеют конечную дисперсию. Найти коэффициент корреляции между суммами

$$s = \xi_1 + \xi_2 + \ldots + \xi_n \quad \text{и} \quad \sigma = \xi_{m+1} + \xi_{m+2} + \ldots + \xi_{m+n}.$$

17. Случайные величины ξ и η независимы и нормально распределены с одними и теми же параметрами a и σ. Найти коэффициент корреляции между $\alpha\xi + \beta\eta$ и $\alpha\xi - \beta\eta$, а также их совместное распределение (α и β – постоянные).

18. Случайный вектор (ξ, η) нормально распределен; $\mathbf{M}\xi = a$, $\mathbf{M}\eta = b$, $\mathbf{D}\xi = \sigma_1^2$, $\mathbf{D}\eta = \sigma_2^2$, R – коэффициент корреляции между ξ и η. Доказать, что $R = \cos q\pi$, где $q = \mathbf{P}\{(\xi - a)(\xi - b) < 0\}$.

19. Пусть x_1 и x_2 – результаты двух независимых наблюдений над нормально распределенной величиной ξ. Доказать, что $\mathbf{M}\max(x_1, x_2) = a + \sigma/\sqrt{\pi}$, где $a = \mathbf{M}\xi$, $\sigma^2 = \mathbf{D}\xi$.

20. Случайный вектор (ξ, η) нормально распределен; $\mathbf{M}\xi = \mathbf{M}\eta = 0$, $\mathbf{D}\xi = \mathbf{D}\eta = 1$, $\mathbf{M}\xi\eta = R$. Доказать, что

$$\mathbf{M}\max(\xi, \eta) = \sqrt{\frac{1-R}{\pi}}.$$

21. Неровнотой пряжи по длине волокна называется величина

$$\lambda = \frac{a'' - a'}{a},$$

где a есть математическое ожидание длины волокна, a'' – математическое ожидание длины тех волокон, которые больше a, a' – математическое ожидание длины тех волокон, которые меньше a. Найти связь между величинами (если ξ распределено нормально)

а) $\lambda, a, \mathbf{M}\,|\,\xi - a\,|$, б) λ, a, σ.

22. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n \ldots$ независимы и равномерно распределены в $(0, 1)$. Пусть ν – случайная величина, равная тому k, при котором впервые сумма

$$s_k = \xi_1 + \xi_2 + \ldots + \xi_k$$

превосходит 1. Доказать, что $\mathbf{M}\nu = e$.

23. Пусть ξ – случайная величина с плотностью распределения

$$p_\xi(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

Найти $\mathbf{M}\min(|\,\xi\,|, 1)$.

(Задачи 22 и 23 сообщены мне М.И. Ядренко.)

ГЛАВА 6

ЗАКОН БОЛЬШИХ ЧИСЕЛ

§ 27. Массовые явления и закон больших чисел

Огромный опыт, накопленный человечеством, учит нас, что явления, имеющие вероятность, весьма близкую к единице, почти обязательно происходят. Точно так же события, вероятность наступления которых очень мала (иными словами, очень близка к нулю), наступают очень редко. Это обстоятельство играет основную роль для всех практических выводов из теории вероятностей, так как указанный *опытный факт* дает право в практической деятельности считать мало вероятные события *практически невозможными*, а события, происходящие с вероятностями, весьма близкими к единице, *практически достоверными*. При этом на вполне естественный вопрос, какова должна быть вероятность, чтобы мы могли событие считать практически невозможным (практически достоверным), однозначного ответа дать нельзя. И это понятно, так как в практической деятельности необходимо учитывать важность тех событий с которыми приходится иметь дело. Так, например, если бы при измерении расстояния между двумя селениями оказалось, что оно равно 5340 *м* и ошибка этого измерения с вероятностью 0,02 равна или больше 20 *м*, то мы можем пренебречь возможностью такой ошибки и считать, что расстояние действительно равно 5340 *м*. Таким образом, в нашем примере мы считаем событие с вероятностью 0,02 практически несущественным и в своей практической деятельности его не учитываем. В то же время в других случаях пренебрегать вероятностями 0,02 и даже еще меньшими нельзя. Так, если при строительстве большой гидроэлектростанции, требующей огромных материальных затрат и человеческого труда, выяснилось, что вероятность катастрофического паводка в рассматриваемых условиях равна 0,02, то эта вероятность будет сочтена большой и при проектировании станции она должна быть учтена, а не отброшена, как это было сделано в предыдущем примере. Таким образом, только требования практики могут нам подсказать критерии, согласно которым мы будем считать те или иные события практически невозможными или практически достоверными.

В то же время необходимо заметить, что любое событие, имеющее положительную вероятность, как бы мала ни была эта вероятность, может произойти, и если число испытаний, в каждом из которых оно может произойти с одной и той же вероятностью, очень велико, то вероятность хотя бы однократного его появления может стать сколь угодно близкой к единице. Это обстоятельство постоянно следует иметь в виду. Однако если вероятность некоторого события очень мала, то чрезвычайно трудно ожидать его появления в каком-либо заранее определенном испытании. Так, если некто утверждает, что при первой же раздаче карт между четырьмя партнерами каждый получит карты только одной масти, то естественно заподозрить, что при раздаче руководствовались каким-то определенным соображением, например, расположили карты в определенном, известном сдающему, порядке. Эта наша уверенность основывается на том, что вероятность такой раскладки при хорошей тасовке равна $(9!)^4 4!/36! < 1{,}1 \cdot 10^{-18}$, т.е. чрезвычайно мала. Тем не менее однажды факт такой раскладки карт был зарегистрирован. Этот пример достаточно хорошо иллюстрирует различие между понятиями практической невозможности и невозможности, так сказать, категорической.

Из сказанного понятно, что в практической деятельности, да и в общетеоретических задачах, большое значение имеют события с вероятностями, близкими к единице или нулю. Отсюда становится ясным, что одной из основных задач теории вероятностей должно быть установление закономерностей, происходящих с вероятностями близкими к единице: при этом особую роль должны играть закономерности, возникающие в результате наложения большого числа независимых или слабо зависимых случайных факторов. Закон больших чисел является одним из таких предложений теории вероятностей и при том важнейшим.

Под законом больших чисел теперь было бы естественно понимать всю совокупность предложений, утверждающих с вероятностью, сколь угодно близкой к единице, что наступит некоторое событие, зависящее от неограниченно увеличивающегося числа случайных событий, каждое из которых оказывает на него лишь незначительное влияние. Это общее представление о теоремах типа закона больших чисел можно сформулировать и несколько определеннее: пусть дана последовательность случайных величин

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \tag{1}$$

Рассмотрим случайные величины ζ_n, являющиеся некоторыми заданными симметрическими функциями от первых n величин последовательности (1):

$$\zeta_n = f_n(\xi_1, \xi_2, \ldots, \xi_n).$$

Если существует такая последовательность постоянных $a_1, a_2, \ldots, a_n, \ldots$, что при любом $\epsilon > 0$

$$\lim_{n \to \infty} \mathsf{P}\{|\zeta_n - a_n| < \epsilon\} = 1, \tag{2}$$

то последовательность (1) подчиняется закону больших чисел с заданными функциями f_n.

Обычно, однако, в понятие закона больших чисел вкладывается гораздо более определенное содержание. А именно, ограничиваются тем случаем, когда f_n есть среднее арифметическое величин $\xi_1, \xi_2, \ldots, \xi_n$.

Если в соотношении (2) все величины a_n равны одной и той же величине a, то говорят, что случайные величины ζ_n *сходятся по вероятности* к a. В этих терминах соотношение (2) означает, что $\zeta_n - a_n$ сходится по вероятности к нулю.

Наблюдая единичные явления, мы наблюдаем их вместе со всеми их индивидуальными особенностями, затемняющими проявление тех закономерностей, которые имеют место при наблюдении большого числа аналогичных явлений. То, что факторы, не связанные с существом всего процесса в целом, а проявляющиеся только в единичных его осуществлениях, при рассмотрении среднего из большого числа наблюдений взаимно погашаются, было замечено еще давно.

Позднее этот эмпирический результат отмечался все чаще и чаще и притом, как правило, без попытки найти его теоретическое объяснение. Впрочем, такое объяснение многим авторам и не требовалось, так как наличие закономерностей как в явлениях природы, так и в общественных явлениях для них было не чем иным, как проявлением правил божественного порядка.

Некоторые авторы до сих пор обедняют содержание закона больших чисел и даже искажают его методологическое значение, сводя его попросту к наблюдающейся на опыте закономерности. На самом же деле непреходящая научная ценность исследований Чебышева, Маркова и других исследователей в области закона больших чисел состоит не в том, что они подметили эмпирическую устойчивость средних, а в том, что они нашли те общие условия, выполнение которых обязательно влечет за собой статистическую устойчивость средних.

Для иллюстрации действия закона больших чисел приведем следующий схематический пример. По современным физическим воззрениям любой газ состоит из огромного количества отдельных частиц, находящихся в непрестанном хаотическом движении. Про каждую отдельную молекулу нельзя предсказать, с какой скоростью она будет двигаться и в каком месте она будет находиться в каждый данный момент времени. Однако мы можем рассчитать при определенных условиях, в которых находится газ, долю тех молекул, которые будут двигаться с заданной скоростью, или долю тех из них, которые будут находиться в заданном объеме. Но, собственно, именно это и нужно знать физику, так как основные характеристики газа – давление, температура, вязкость и пр. – определяются не замысловатым поведением одной молекулы, а их совокупным действием. Так, давление газа равно суммарному воздействию молекул, ударившихся о пластинку площади единица за единицу времени. Число ударов и скорости ударившихся молекул меняются в зависимости от случая, однако в силу закона больших чисел (в форме Чебышева) давление должно быть почти постоянным. Это "уравнивающее" влияние закона больших чисел в физических явлениях обнаруживается с исключительной точностью. Достаточно вспомнить, что, скажем, в обычных условиях даже очень точные измерения с трудом позволяют отметить уклонения от закона Паскаля о давлении жидкости. Противникам молекулярного строения материи это чрезмерно хорошее совпадение результатов теории с опытом даже служило своеобразным аргументом: если бы материя имела молекулярное строение, то наблюдались бы и уклонения от закона Паскаля. Эти уклонения, так называемые флюктуации давления, действительно удалось наблюдать, когда научились изолировать сравнительно небольшие количества молекул, в результате чего влияние отдельных молекул еще не полностью нивелировалось и оставалось еще достаточно сильным.

§ 28. Закон больших чисел в форме Чебышева

Мы перейдем теперь к формулировке и доказательству теорем Чебышева, Маркова и др.; употребляемый при этом метод принадлежит Чебышеву.

Неравенство Чебышева. *Для любой случайной величины ξ, имеющей конечную дисперсию, при каждом $\epsilon > 0$ имеет место неравенство*

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| \geqslant \epsilon\} \leqslant \frac{\mathbf{D}\xi}{\epsilon^2}. \tag{1}$$

Доказательство. Если $F(x)$ обозначает функцию распределения случайной величины ξ, то ясно, что

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| \geqslant \epsilon\} = \int\limits_{|x - \mathbf{M}\xi| \geqslant \epsilon} dF(x).$$

Так как в области интегрирования $|x - \mathbf{M}\xi|/\epsilon \geqslant 1$, то

$$\int\limits_{|x - \mathbf{M}\xi| \geqslant \epsilon} dF(x) \leqslant \frac{1}{\epsilon^2} \int\limits_{|x - \mathbf{M}\xi| \geqslant \epsilon} (x - \mathbf{M}\xi)^2 dF(x).$$

Мы только усилим это неравенство, распространяя интегрирование на все значения x

$$\int\limits_{|x - \mathbf{M}\xi| \geqslant \epsilon} dF(x) \leqslant \frac{1}{\epsilon^2} \int (x - \mathbf{M}\xi)^2 dF(x) = \frac{\mathbf{D}\xi}{\epsilon^2}.$$

Неравенство Чебышева доказано.

Теорема Чебышева. *Если $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ – последовательность попарно независимых случайных величин, имеющих конечные диспер-*

сии, ограниченные одной и той же постоянной

$$\mathsf{D}\xi_1 \leqslant C, \quad \mathsf{D}\xi_2 \leqslant C, \ldots, \mathsf{D}\xi_n \leqslant C, \ldots,$$

то, каково бы ни было постоянное $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}\mathsf{M}\xi_k\right| < \epsilon\right\} = 1. \tag{2}$$

Доказательство. Мы знаем, что в условиях теоремы

$$\mathsf{D}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathsf{D}\xi_k$$

и, следовательно,

$$\mathsf{D}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) \leqslant \frac{C}{n}.$$

Согласно неравенству Чебышева

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}\mathsf{M}\xi_k\right| < \epsilon\right\} \geqslant 1 - \frac{\mathsf{D}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)}{\epsilon^2} \geqslant$$

$$\geqslant 1 - \frac{C}{n\epsilon^2}.$$

Переходя к пределу при $n \to \infty$, получаем, что

$$\lim_{n\to\infty} \mathsf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}\mathsf{M}\xi_k\right| < \epsilon\right\} \geqslant 1.$$

А так как вероятность не может быть больше единицы, то отсюда и следует утверждение теоремы.

Мы отметим некоторые важные частные случаи теоремы Чебышева.

1. Теорема Бернулли. *Пусть* μ – *число наступлений события A в n независимых испытаниях и p есть вероятность наступления события A в каждом из испытаний. Тогда, каково бы ни было* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}\left\{\left|\frac{\mu}{n} - p\right| < \epsilon\right\} = 1. \tag{3}$$

Доказательство. Действительно, введя случайные величины μ_k, равные числу наступлений события A при k-м испытании, имеем:

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n.$$

А так как

$$\mathsf{M}\mu_k = p, \quad \mathsf{D}\mu_k = pq \leqslant 1/4,$$

то теорема Бернулли является простейшим частным случаем теоремы Чебышева.

Так как на практике часто неизвестные вероятности приходится приближенно определять из опыта, то для проверки согласия теоремы Бернулли с опытом было проведено большое число опытов. При этом рассматривались события, вероятности которых можно считать по тем или иным соображениям известными, относительно которых легко проводить испытания и обеспечить независимость испытаний, а также постоянство вероятностей в каждом из испытаний. Все подробные опыты дали прекрасное совпадение с теорией. Мы приведем результаты нескольких таких легко воспроизводимых экспериментов.

В примере 5 § 3 мы рассмотрели результаты 100 разделений колоды карт на две равные части. Интересующее нас событие состояло в том, что в каждую полуколоду попадет одинаковое число красных и черных карт. В рассмотренном случае получилось довольно значительное окончательное (при $n = 100$) уклонение частоты от вероятности (приблизительно равное 0,02). По теореме Лапласа вероятность получить такое уклонение или еще большее равна

$$\mathsf{P}\left\{\left|\frac{\mu}{n}-p\right| \geqslant 0{,}02\right\}=\mathsf{P}\left\{\left|\frac{\mu-np}{\sqrt{npq}}\right| \geqslant 0{,}02\sqrt{\frac{n}{pq}}\right\} \approx 1-2\Phi\left(0{,}02\sqrt{\frac{n}{pq}}\right)=$$

$$=1-2\Phi\left(0{,}02\sqrt{\frac{100}{0{,}26\cdot 0{,}74}}\right)=1-2\Phi(0{,}455)\approx 0{,}65.$$

Таким образом, если повторить указанный эксперимент большое число раз, то приблизительно в двух третях случаев получится уклонение, не меньшее, чем полученное в нашем опыте.

Французский естествоиспытатель XVIII века Бюффон бросил монету 4040 раз, герб выпал при этом 2048 раз. Частота появления герба в опыте Бюффона приближенно равна 0,507.

Английский статистик К. Пирсон бросил монету 12 000 раз и при этом наблюдал 6019 выпадений герба. Частота выпадения герба в этом опыте Пирсона равна 0,5016.

В другой раз он бросил монету 24 000 раз, и герб при этом выпал 12 012 раз; частота выпадения герба при этом оказалась равной 0,5005. Во всех приведенных опытах частоты лишь немного уклонялись от вероятности – 0,5.

2. Теорема Пуассона. *Если в последовательности независимых испытаний вероятность появления события A в k-м испытании равна p_k, то*

$$\lim_{n\to\infty}\mathsf{P}\left\{\left|\frac{\mu}{n}-\frac{p_1+p_2+\ldots+p_n}{n}\right|<\epsilon\right\}=1,$$

где, как обычно, через μ обозначено число появлений события A в первых n испытаниях

Введя в рассмотрение случайные величины μ_k, равные числу появлений события A в k-м испытании, и заметив, что

$$\mathsf{M}\mu_k = p_k, \quad \mathsf{D}\mu_k = p_k q_k \leqslant 1/4,$$

мы убеждаемся, что теорема Пуассона является частным случаем теоремы Чебышева.

3. *Если последовательность попарно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ такова, что*

$$\mathsf{M}\xi_1 = \mathsf{M}\xi_2 = \ldots = \mathsf{M}\xi_n = \ldots = a$$

и

$$\mathsf{D}\xi_1 \leqslant C, \quad \mathsf{D}\xi_2 \leqslant C, \ldots, \mathsf{D}\xi_n \leqslant C, \ldots,$$

то, каково бы ни было постоянное $\epsilon > 0$,

$$\lim_{n \to \infty} \mathsf{P}\left\{\left| \frac{1}{n} \sum_{k=1}^{n} \xi_k - a \right| < \epsilon \right\} = 1.$$

Этот частный случай теоремы Чебышева дает основание правилу среднего арифметического, постоянно употребляющемуся в теории измерений. Предположим, что производится измерение некоторой физической величины a. Повторив измерения n раз в одинаковых условиях, наблюдатель получит не вполне совпадающие результаты $x_1, x_2, \ldots, x_n$. В качестве приближенного значения a принято брать среднее арифметическое из результатов наблюдений

$$a \sim \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Если измерения лишены систематической ошибки, т.е. если

$$\mathsf{M}x_1 = \mathsf{M}x_2 = \ldots = \mathsf{M}x_n = a,$$

и если сами наблюденные значения не обладают неопределенностью, то согласно закону больших чисел при достаточно больших значениях n с вероятностью, сколь угодно близкой к единице, мы указанным путем можем получить значение, сколь угодно близкое к искомой величине a. Сказанное мы должны пояснить следующим: если измерительный прибор устроен так, что он не может давать точности отсчета большей, чем некоторая величина δ, например, из-за того, что ширина деления шкалы, по которой производится отсчет, равна δ, то, понятно, нельзя и рассчитывать получить точность измерения, большую, чем $\pm\,\delta$. Каждое измерение в этом случае дает результат с неопределенностью δ; но ясно, что при этом и среднее арифметическое будет обладать той же неопределенностью, как и каждое измерение. Это замечание учит нас, что если приборы дают нам результаты измерений с некоторой неопределенностью δ, то стремиться посредством закона больших чисел получить значение a с большей степенью точности является заблуждением, а сами произведенные при этом вычисления превращаются в арифметическую забаву.

Мы ограничимся формулировкой теоремы Маркова; ее доказательство является очевидным следствием неравенства Чебышева.

Теорема Маркова. *Если последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ такова, что при $n \to \infty$*

$$\frac{1}{n^2} \mathsf{D} \left(\sum_{k=1}^{n} \xi_k \right) \to 0, \tag{4}$$

то, каково бы ни было положительное постоянное ϵ,

$$\lim_{n \to \infty} \mathsf{P} \left\{ \left| \frac{1}{n} \sum_{k=1}^{n} \xi_k - \frac{1}{n} \sum_{k=1}^{n} \mathsf{M}\xi_k \right| < \epsilon \right\} = 1.$$

Если случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ попарно независимы, то условие Маркова принимает вид: при $n \to \infty$

$$\frac{1}{n^2} \sum_{k=1}^{n} \mathsf{D}\xi_k \to 0.$$

Отсюда видно, что теорема Чебышева является частным случаем теоремы Маркова.

§ 29. Необходимое и достаточное условие для закона больших чисел

Мы уже указывали, что закон больших чисел является одним из основных предложений теории вероятностей. Отсюда становится понятным, почему так много усилий было положено на то, чтобы установить наиболее широкие условия, которым должны удовлетворять величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, чтобы для них имел место закон больших чисел.

История вопроса такова. В конце XVII – начале XVIII века Яков Бернулли нашел предложение, получившее его имя. Теорема Бернулли была впервые опубликована в 1713 г., после смерти автора, в трактате "Ars Conjectandi" (Искусство предположений). Затем в начале XIX века Пуассон доказал аналогичную теорему в более широких условиях. До середины XIX века не было достигнуто каких-либо новых успехов. В 1866 г. великий русский математик П.Л. Чебышев нашел метод, изложенный нами в предыдущем параграфе. Позднее А.А. Марков заметил, что рассуждения Чебышева позволяют получить более общий результат (см. § 27).

Дальнейшие усилия долго не приносили принципиальных успехов, и лишь в 1926 г. А.Н. Колмогоров получил условия, необходимые и достаточные для того, чтобы последовательность взаимно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ подчинялась закону больших чисел. В 1923 г. А.Я. Хинчин показал, что если случайные величины ξ_n не только независимы, но и одинаково распределены, то существование математического ожидания $\mathsf{M}\xi_n$ является достаточным условием для применимости закона больших чисел.

В последние годы много работ было посвящено определению условий, которые следует наложить на з а в и с и м ы е величины, чтобы для них выполнялся закон больших чисел. Теорема Маркова принадлежит к предложениям этого рода.

Используя метод Чебышева, легко получить условие, аналогичное условию Маркова, но уже не только достаточное, но и необходимое для применимости закона больших чисел к последовательности произвольных случайных величин.

Т е о р е м а. *Для того чтобы для последовательности*

$\xi_1, \xi_2, \xi_3, \ldots$

(как угодно зависимых) случайных величин при любом положительном ϵ *выполнялось соотношение*

$$\lim_{n\to\infty} \mathsf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_n - \frac{1}{n}\sum_{k=1}^{n}\mathsf{M}\xi_k\right| < \epsilon\right\} = 1, \tag{1}$$

необходимо и достаточно, чтобы при $n \to \infty$

$$\mathsf{M}\frac{\left(\sum_{k=1}^{n}(\xi_k - \mathsf{M}\xi_k)\right)^2}{n^2 + \left(\sum_{k=1}^{n}(\xi_k - \mathsf{M}\xi_k)\right)^2} \to 0. \tag{2}$$

Д о к а з а т е л ь с т в о. Предположим сначала, что (2) выполнено, и покажем, что в этом случае выполнено также (1). Обозначим через $\Phi_n(x)$ функцию распределения величины

$$\eta_n = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mathsf{M}\xi_k).$$

Легко проверить следующую цепочку соотношений:

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k-\mathbf{M}\xi_k)\right|\geqslant\epsilon\right\}=\mathbf{P}\{|\eta_n|\geqslant\epsilon\}=$$

$$=\int_{|x|\geqslant\epsilon}d\Phi_n(x)\leqslant\frac{1+\epsilon^2}{\epsilon^2}\int_{|x|\geqslant\epsilon}\frac{x^2}{1+x^2}d\Phi_n(x)\leqslant$$

$$\leqslant\frac{1+\epsilon^2}{\epsilon^2}\int\frac{x^2}{1+x^2}d\Phi_n(x)=\frac{1+\epsilon^2}{\epsilon^2}\mathbf{M}\frac{\eta_n^2}{1+\eta_n^2}\;{}^{*)}$$

Это неравенство доказывает достаточность условия теоремы.

Покажем теперь, что условие (2) необходимо. Легко видеть, что

$$\mathbf{P}\{|\eta_n|\geqslant\epsilon\}=\int_{|x|\geqslant\epsilon}d\Phi_n(x)\geqslant\int_{|x|\geqslant\epsilon}\frac{x^2}{1+x^2}d\Phi_n(x)=$$

$$=\int\frac{x^2}{1+x^2}d\Phi_n(x)-\int_{|x|<\epsilon}\frac{x^2}{1+x^2}d\Phi_n(x)\geqslant$$

$$\geqslant\int\frac{x^2}{1+x^2}d\Phi_n(x)-\epsilon^2=\mathbf{M}\frac{\eta_n^2}{1+\eta_n^2}-\epsilon^2. \qquad (3)$$

Таким образом,

$$0\leqslant\mathbf{M}\frac{\mu_n^2}{1+\mu_n^2}\leqslant\epsilon^2+\mathbf{P}\{|\mu_n|\geqslant\epsilon\}.$$

Выбирая сначала ϵ сколь угодно малым, а затем n достаточно большим, мы можем сделать правую часть последнего неравенства сколь угодно малой.

Отметим, что все теоремы, доказанные в предыдущем параграфе, легко вытекают из только что доказанного общего предложения. Действительно, так как при любом n и любых ξ_k имеет место неравенство

$$\frac{\eta_n^2}{1+\eta_n^2}\leqslant\eta_n^2=\left[\frac{1}{n}\sum_{k=1}^{n}(\xi_k-\mathbf{M}\xi_k\right]^2,$$

то в случае существования дисперсий отсюда вытекает неравенство

$$\mathbf{M}\frac{\eta_n^2}{1+\eta_n^2}\leqslant\frac{1}{n^2}\mathbf{D}\sum_{k=1}^{n}\xi_k.$$

---

*) Последнее равенство мы пишем на основании формулы

$$\mathbf{M}f(\xi)=\int f(x)\,dF_\xi(x)$$

(см. теорему 1 § 22).

Таким образом, если условие Маркова выполнено, то выполнено также условие (2) и, следовательно, последовательность $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ подчиняется закону больших чисел.

Все же мы должны заметить, что в более сложных случаях, когда у величин ξ_k не предполагается конечных дисперсий, доказанная теорема для фактической проверки применимости закона больших чисел весьма мало пригодна, так как условие (2) относится не к отдельным слагаемым, а к их суммам. Однако рассчитывать на то, что, не сделав никаких предложений о величинах ξ_k и о существующей между ними связи, удастся найти необходимые и достаточные условия, к тому же удобные для приложений, по-видимому, нельзя.

Если предположить, что величины $\xi_1, \xi_2, \xi_2, \ldots$ взаимно независимы, то можно показать, что условие (2) эквивалентно следующему: при $n \to \infty$

$$\sum_{k=1}^{n} \mathbf{M} \frac{\zeta_k^2}{n^2 + \zeta_k^2} \to 0,$$

где обозначено

$$\zeta_k = \xi_k - \mathbf{M}\xi_k.$$

Практическое использование только что доказанных теорем встречает одно принципиальное затруднение: можем ли мы считать, что изучаемое нами явление или производственный процесс протекают под воздействием независимых причин? Не противоречит ли само понятие независимости нашим основным представлениям о взаимосвязи явлений внешнего мира? При математическом изучении тех или иных явлений природы, технических процессов или тех или иных общественных явлений мы прежде всего должны выводить наши основные предпосылки, опираясь на глубокое изучение существа самого явления, качественных его особенностей. Мы должны учитывать изменение внешних условий, в которых протекает изучаемое нами явление и изменять математический аппарат и предпосылки, лежащие в основе его применения, как только обнаружится, что условия осуществления явления изменились.

Отбрасывая несущественные связи между причинами, под влиянием которых развивается изучаемое явление, мы приходим к возможности в качестве рабочего аппарата пользоваться независимыми случайными величинами. Насколько удачно мы произвели схематизацию явления, насколько удачно выбран нами математический аппарат для его изучения, мы можем судить по согласию созданной нами теории с практикой. Если наши теоретические результаты существенно расходятся с опытом, то мы должны пересмотреть предпосылки, в частности, если идет речь о применимости закона больших чисел, то быть может придется отказаться от предположения о полной независимости действующих причин и перейти к предположению об их зависимости, быть может и слабой.

Мы уже говорили, что накопленный опыт использования теорем о законе больших чисел показывает, что условие независимости удовлетворительно во многих важных задачах естествознания и техники.

§ 30. Усиленный закон больших чисел

Нередко из теоремы Бернулли делают совершенно необоснованный вывод, что частота события A при безграничном увеличении числа испытаний стремится к вероятности события A. На самом же деле теорема Бернулли устанавливает только тот факт, что для достаточно большого числа испытаний n вероятность одного единственного неравенства:

$$\left|\frac{\mu}{n} - p\right| < \epsilon$$

может быть сделана больше чем $1 - \eta$ при произвольном $\eta > 0$. В 1909 г. французский математик Э. Борель обнаружил более глубокое предложение, согласно которому при любых $\epsilon > 0$ и $\eta > 0$ можно указать такое n_0, что, каково бы ни было s, вероятность одновременного выполнения неравенств

$$\left|\frac{\mu}{n} - p\right| < \epsilon$$

для всех n, удовлетворяющих неравенствам $n_0 \leqslant n \leqslant n_0 + s$ больше чем $1 - \eta$.

Эту теорему мы выведем из теоремы Колмогорова об усиленном законе больших чисел.

Неравенство Колмогорова. *Если взаимно независимые случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ имеют конечные дисперсии, то вероятность совместного осуществления неравенств*

$$\left|\sum_{s=1}^{k} (\xi_s - \mathsf{M}\xi_s)\right| < \epsilon \quad (k = 1, 2, \ldots, n)$$

не меньше, чем

$$1 - \frac{1}{\epsilon^2} \sum_{k=1}^{n} \mathsf{D}\xi_k.$$

Доказательство. Введем обозначения

$$\eta_k = \xi_k - \mathsf{M}\xi_k, \qquad S_k = \sum_{j=1}^{k} \eta_j.$$

Пусть, далее, E_k обозначает событие, состоящее в том, что

$$|S_j| < \epsilon \text{ для } j \leqslant k - 1 \text{ и } |S_k| \geqslant \epsilon;$$

E_0 означает событие, состоящее в том, что $|S_j| < \epsilon$ для $j \leqslant n$.

Так как событие, состоящее в том, что хотя бы при одном $k\,(1 \leqslant k \leqslant n)$ будет выполнсно неравенство

$$|S_k| \geqslant \epsilon \qquad (k = 1, 2, \ldots, n)$$

(иными словами, что $\max\limits_{1 \leqslant k \leqslant n} |S_k| \geqslant \epsilon$) равносильно событию $\sum\limits_{k=1}^{n} E_k$, то в силу несовместимости событий E_k

$$\mathsf{P}\{\max_{1 \leqslant k \leqslant n} |S_k| \geqslant \epsilon\} = \sum_{k=1}^{n} \mathsf{P}(E_k).$$

Согласно равенству ((5) § 23)

$$\mathsf{D}S_n = \sum_{k=0}^{n} \mathsf{P}(E_k) \cdot \mathsf{M}(S_n^2 \mid E_k) \geqslant \sum_{k=1}^{n} \mathsf{P}(E_k) \cdot \mathsf{M}(S_n^2 \mid E_k).$$

Очевидно, далее, что

$$\mathsf{M}(S_n^2 \mid E_k) = \mathsf{M}\{S_k^2 + 2\sum_{j>k} S_k \eta_j + \sum_{j>k} \eta_j^2 + 2\sum_{j>h>k} \eta_j \eta_h \mid E_k\} \geqslant$$

$$\geqslant \mathsf{M}\{S_k^2 + 2\sum_{j>k} S_k \eta_j + 2\sum_{j>h>k} \eta_j \eta_h \mid E_k\}.$$

Так как осуществление события E_k налагает ограничение только на значения первых k из величин ξ_i, а последующие остаются при этом условии независимыми друг от друга и от S_k, то

$$\mathsf{M}(S_k \eta_j \mid E_k) = \mathsf{M}(S_k \mid E_k) \cdot \mathsf{M}(\eta_j \mid E_k) = 0$$

и

$$\mathsf{M}(\eta_j \eta_h \mid E_k) = 0 \qquad (h \neq j, \quad h > k, \quad j > k \geqslant 1).$$

Кроме того, согласно (1) имеет место неравенство

$$\mathsf{M}(S_k^2 \mid E_k) \geqslant \epsilon^2 \qquad (k \geqslant 1).$$

Мы можем написать поэтому, что

$$\mathsf{D}S_n \geqslant \epsilon^2 \sum_{k=1}^{n} \mathsf{P}\{E_k\}.$$

Отсюда

$$\sum_{k=1}^{n} \mathsf{P}\{E_k\} = \mathsf{P}\{\max_{1 \leqslant k \leqslant n} |S_k| \geqslant \epsilon\} \leqslant \frac{1}{\epsilon^2} \mathsf{D}S_n.$$

Неравенство Колмогорова доказано.

Мы скажем, что последовательность случайных величин

$$\xi_1, \xi_2, \xi_3, \ldots$$

подчиняется *усиленному закону больших чисел*, если, каковы бы ни были

$\epsilon > 0$ и $\eta > 0$, можно указать такое n_0, что для любого s и всех n, удовлетворяющих неравенствам $n_0 \leqslant n \leqslant n_0 + s$, вероятность неравенства

$$\max_{n_0 \leqslant n \leqslant n_0 + s} \left| \frac{1}{n} \sum_{k=1}^{n} \xi_k - \frac{1}{n} \sum_{k=1}^{n} \mathbf{M}\xi_k \right| < \epsilon$$

больше, чем $1 - \eta$.

Теорема Колмогорова. *Если последовательность взаимно независимых случайных величин $\xi_1, \xi_2, \ldots$ удовлетворяет условию*

$$\sum_{n=1}^{\infty} \frac{\mathbf{D}\xi_n}{n^2} < +\infty,$$

то она подчиняется усиленному закону больших чисел.

Доказательство. Положим

$$\zeta_n = \xi_n - \mathbf{M}\xi_n, \quad S_n = \sum_{k=1}^{n} \zeta_k, \quad v_n = \frac{1}{n} S_n.$$

Рассмотрим вероятность

$$P_m = \mathbf{P}\{\max | v_n | \geqslant \epsilon, \ 2^m \leqslant n < 2^{m+1}\}.$$

Так как

$$P_m \leqslant \mathbf{P}\{\max | v_n | \geqslant \epsilon, \quad 1 \leqslant n < 2^{m+1}\},$$

то согласно неравенству Колмогорова

$$P_m \leqslant \frac{1}{(2^m \epsilon)^2} \sum_{j < 2^{m+1}} \mathbf{D}\xi_j.$$

Так как, далее,

$$\mathbf{P}\{\max | v_n | \geqslant \epsilon \ \text{ для } \ n > \nu\} \leqslant \sum_{m=\rho}^{\infty} P_m,$$

где ρ определяется из неравенств $2^\rho \leqslant \nu < 2^{\rho+1}$, то ясно, что

$$\mathbf{P}\{\max | v_n | \geqslant \epsilon \ \text{ для } \ n > \nu\} \leqslant \frac{1}{\epsilon^2} \sum_{m=\rho}^{\infty} \frac{1}{2^{2m}} \sum_{j < 2^{m+1}} \mathbf{D}\xi_j$$

После перемены порядка суммирования в правой части последнего неравенства получим:

$$\sum_{m=\rho}^{\infty} \frac{1}{2^{2m}} \sum_{j < 2^{m+1}} \mathbf{D}\xi_j = \sum_{j=1}^{\infty} \mathbf{D}\xi_j \left(\sum\nolimits_j \frac{1}{2^{2m}} \right),$$

где сумма Σ_j распространена на те значения $m \geqslant \rho$, для которых $2^{m+1} > j$.

При $j \leqslant 2^{\rho+1}$ коэффициент при $\mathsf{D}\xi_j$ равен

$$\sum_{m \geqslant \rho} \frac{1}{4^m} = \frac{3}{4^{\rho-1}},$$

а при $2^{m_0+1} > j \geqslant 2^{m_0} \geqslant 2^{\rho+1}$ этот коэффициент равен $\dfrac{3}{4^{m_0-1}} \leqslant$
$\leqslant \dfrac{3 \cdot 16}{2^{2(m_0+1)}} \leqslant \dfrac{3 \cdot 16}{j^2}$.

Таким образом:

$$\sum_{m=\rho}^{\infty} \frac{1}{2^{2m}} \sum_{j<2^{m+1}} \mathsf{D}\xi_j \leqslant \frac{3}{4^{\rho-1}} \sum_{j=1}^{2^{\rho+1}} \mathsf{D}\xi_j + 3 \cdot 16 \sum_{j=2^{\rho+1}+1}^{\infty} \frac{\mathsf{D}\xi_j}{j^2} \leqslant$$

$$\leqslant \frac{3}{4^{\rho-1}} \sum_{j=1}^{\rho} \mathsf{D}\xi_j + 3 \cdot 4 \sum_{j=\rho+1}^{2^{\rho+1}} \frac{\mathsf{D}\xi_j}{2^{2(\rho+1)}} + 3 \cdot 16 \sum_{j=2^{\rho+1}+1}^{\infty} \frac{\mathsf{D}\xi_j}{j^2} \leqslant$$

$$\leqslant \frac{3}{4^{\rho-1}} \sum_{j=1}^{\rho} \mathsf{D}\xi_j + 3 \cdot 4 \sum_{j=\rho+1}^{2^{\rho+1}} \frac{\mathsf{D}\xi_j}{j^2} + 3 \cdot 16 \sum_{j=2^{\rho+1}+1}^{\infty} \frac{\mathsf{D}\xi_j}{j}.$$

В силу сходимости ряда $\sum \dfrac{\mathsf{D}\xi_n}{n^2}$:

1°. Две последние суммы в написанном выше неравенстве могут быть сделаны сколь угодно малыми при ρ достаточно большом.

2°. Существует такая постоянная C, что $\mathsf{D}\xi_n < Cn^2$, откуда следует, что

$$\frac{3}{4^{\rho+1}} \sum_{j=1}^{\rho} \mathsf{D}\xi_j \leqslant \frac{3 \cdot C\rho^3}{4^{\rho-1}}.$$

т.е. и первая сумма может быть сделана сколь угодно малой при достаточно большом ρ.

Из всего сказанного следует, что при n_0 достаточно большом

$$\mathsf{P}\{\max |v_n| \geqslant \epsilon \quad \text{для} \quad n > \nu\}$$

может быть сделана сколь угодно малой, что и требовалось доказать.

Следствие. *Если дисперсии случайных величин ξ_k ограничены одной и той же постоянной C, то последовательность взаимно независимых случайных величин $\xi_1, \xi_2, \xi_3, \ldots$ подчиняется усиленному закону больших чисел.*

Один окончательный результат, относящийся к усиленному закону больших чисел, был получен также А.Н. Колмогоровым для случая одинаково распределенных независимых слагаемых.

Т е о р е м а. *Существование математического ожидания является необходимым и достаточным условием для применимости усиленного закона больших чисел к последовательности одинаково распределенных и взаимно независимых случайных величин.*

Эту теорему мы можем вывести из уже доказанной нами теоремы Колмогорова.

Действительно, из существования математического ожидания следует конечность интеграла $\int |x|\, dF(x)$, где $F(x)$ – функция распределения случайных величин ξ_n.

Поэтому

$$\sum_{n=1}^{\infty} \mathsf{P}\{|\xi| > n\} = \sum_{n=1}^{\infty} \sum_{k \geqslant n} \mathsf{P}\{k < |\xi| \leqslant k+1\} =$$

$$= \sum_{n=1}^{\infty} k\mathsf{P}\{k < |\xi| \leqslant k+1\} \leqslant \sum_{k=0}^{\infty} \int\limits_{k<|x|\leqslant k+1} |x|\, dF(x) <$$

$$< \int |x|\, dF(x) < \infty. \tag{1}$$

Введем в рассмотрение случайные величины

$$\xi_n^* = \begin{cases} \xi_n & \text{при} \quad |\xi_n| \leqslant n, \\ 0 & \text{при} \quad |\xi_n| > n. \end{cases}$$

Тогда получим:

$$\mathsf{D}\xi_n^* \leqslant \mathsf{M}\xi_n^{*2} = \int_{-n}^{+n} x^2 dF(x) \leqslant \sum_{k=0}^{n} (k+1)^2 \mathsf{P}\{k < |\xi| \leqslant k+1\}$$

и

$$\sum_{n=1}^{\infty} \frac{\mathsf{D}\xi_n^*}{n^2} \leqslant \sum_{n=1}^{\infty} \sum_{k=0}^{n} \frac{(k+1)^2}{n^2} \mathsf{P}\{k < |\xi| \leqslant k+1\} \leqslant$$

$$\leqslant \sum_{k=0}^{\infty} \mathsf{P}\{k < |\xi| \leqslant k+1\}\, (k+1)^2 \sum_{n \geqslant k}^{\infty} \frac{1}{n^2}.$$

Так как

$$\sum_{n \geqslant k}^{\infty} \frac{1}{n^2} < \frac{1}{k^2} + \frac{1}{k} < \frac{2}{k},$$

то в силу (1) находим, что

$$\sum_{n=1}^{\infty} \frac{\mathsf{D}\xi_n^*}{n^2} < \infty,$$

т.е. ξ_n^* удовлетворяют усиленному закону больших чисел. Далее

$$\mathsf{P}\{\xi_n \neq \xi_n^* \text{ при каком-либо } n \geqslant N\} \leqslant \sum_{n \geqslant N} \mathsf{P}\{\xi_n \neq \xi_n^*\} =$$

$$= \sum_{n \geqslant N} \mathsf{P}\{|\xi_n| > n\} < \frac{\epsilon}{4} \tag{2}$$

при $N \geqslant N_0(\epsilon)$.

Выберем ν_0 столь большим, чтобы при $\nu \geqslant \nu_0(\epsilon, \eta)$

$$\mathsf{P}\left\{\left|\frac{\sum\limits_{k=1}^{N_0}(\xi_k - \mathsf{M}\xi_k)}{\nu}\right| \geqslant \frac{\eta}{3}\right\} \leqslant \frac{\epsilon}{4} \tag{3}$$

$$\mathsf{P}\left\{\left|\frac{\sum\limits_{k=1}^{N_0}(\xi_k^* - \mathsf{M}\xi_k^*)}{\nu}\right| \geqslant \frac{\eta}{3}\right\} \leqslant \frac{\epsilon}{4}\,. \tag{4}$$

Наконец, так как ξ_n^* удовлетворяет усиленному закону больших чисел, то

$$\mathsf{P}\left\{\max |v_n^*| \geqslant \frac{\eta}{3}\ ;\ n \geqslant \nu\right\} \leqslant \frac{\epsilon}{4} \quad \text{при} \quad \nu \geqslant \nu_1(\epsilon, \eta), \tag{5}$$

где $v_n^* = \frac{1}{n} \sum\limits_{k=1}^{n} (\xi_n^* - \mathsf{M}\xi_n^*)$.

Из (2), (3), (4) и (5) выводим

$$\mathsf{P}\{\max |v_n| \geqslant \eta;\ n \geqslant \nu\} \leqslant \epsilon$$

при $\nu \geqslant \max(\nu_0, \nu_1, N_0)$, т.е. и ξ_n удовлетворяют усиленному закону больших чисел.

Принципиальная роль усиленного закона больших чисел в теории вероятностей и в ее приложениях весьма велика. Действительно, предположим на минуту, что, скажем, в случае одинаково распределенных слагаемых, имеющих конечное математическое ожидание, усиленный закон не имеет места. Тогда с вероятностью, сколь угодно близкой к единице, можно утверждать, что будут повторяться моменты, когда средняя арифметическая результатов наблюдений будет далека от математического ожидания. И это бы случилось даже в тех случаях, когда наблюдения производятся без систематической ошибки и с полной определенностью (величина δ, о которой была речь в § 28, равна 0). Можно ли было бы в таких условиях считать, что среднее арифметическое из результатов наблюдений сближается с измеряемой величиной, могли ли бы мы в этих условиях считать, что среднее арифметическое можно считать за приближенное значение измеряемой величины? Сомнительно.

§ 31. Теорема В.И. Гливенко

Мы перейдем теперь к доказательству теоремы Гливенко, которая вскоре после ее обнаружения получила в математической литературе название основной теоремы математической статистики. Речь идет об оценке неизвестной функции распределения случайной величины ξ на основе результатов независимых испытаний. Пусть функция распределения случайной величины равна $F(x)$ а результаты последовательных независимых испытаний в неизменных условиях будут

$$x_1, x_2, \ldots, x_n \tag{1}$$

Последовательность результатов испытаний мы расположим в возрастающем порядке. Обозначив k-е по величине наблюденное значение через x_k^* мы можем последовательность (1) записать в следующем виде:

$$x_1^* \leqslant x_2^* \leqslant \ldots \leqslant x_n^* .$$

Эта последовательность, т.е. последовательность наблюденных значений исследуемой случайной величины, расположенных в возрастающем порядке, носит название *вариационного ряда.*

Эмпирической функцией распределения $F_n(x)$ мы назовем функцию, определенную следующими равенствами:

$$F_n(x) = \begin{cases} 0 & \text{при} \quad x \leqslant x_1^*, \\ k/n & \text{при} \quad x_k^* < x \leqslant x_{k+1}^*, \\ 1 & \text{при} \quad x > x_n^*. \end{cases}$$

Ясно, что эмпирическая функция распределения монотонна, непрерывна слева и имеет точки разрыва только при значениях аргумента, равных членам вариационного ряда. Величины скачков в точках разрыва являются целыми кратными от $1/n$. Для дальнейшего подчеркнем то обстоятельство, что при каждом значении x ордината $F_n(x)$ является случайной величиной, возможные значения которой будут $0,\ 1/n, \ldots, (n-1)/n,\ n/n = 1$. Вероятность равенства $F_n(x) = k/n$, как легко видеть, равна

$$\mathsf{P}\left\{F_n(x) = \frac{k}{n}\right\} = C_n^k \{F(x)\}^k \{1 - F(x)\}^{n-k}.$$

В простейшем частном случае, когда случайная величина ξ может принимать лишь конечное число значений $a_1, a_2, \ldots, a_s$ членами вариационного ряда обязательно будут только числа этой последовательности. Согласно закону больших чисел, если $m_1, m_2, \ldots, m_s$ $(m_1 + m_2 + \ldots + m_s = n)$ будут обозначать соответственно числа испытаний, при которых $\xi = a_1$, $\xi = a_2, \ldots, \xi = a_s$, то при достаточно большом значении n частоты будут представлять приближенные значения неизвестных нам вероятностей

$p_1 = \mathbf{P}\{\xi = a_1\}$, $p_2 = \mathbf{P}\{\xi = a_2\}, \ldots, p_s = \mathbf{P}\{\xi = a_s\}$. Более того, в нашем случае имеет место и усиленный закон больших чисел.

Прежде чем переходить к формулировке и доказательству теоремы, составляющей содержание настоящего параграфа, мы установим несколько вспомогательных предложений.

Рассмотрим некоторую последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$

Событие, заключающееся в том, что эта последовательность сходится к некоторой случайной величине ξ, имеет в силу принятых нами аксиом, как мы увидим при доказательстве леммы 1, определенную вероятность.

Если эта вероятность равна единице, то мы скажем, что последовательность $\{\xi_n\}$ сходится к ξ почти наверное*).

В другой форме утверждение о сходимости почти наверное можно выразить так: последовательность случайных величин

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

сходится почти наверное к случа..ной величине ξ, если с вероятностью единица для каждого целого положительного числа r найдется такое число n, что при всех $k > 0$ будут иметь место неравенства

$$|\xi_{n+k} - \xi| < 1/r.$$

Очевидно, что равенство

$$\mathbf{P}\{\xi_n \to \xi\} = 1 \tag{1}$$

мы можем записать и в иной форме:

$$\mathbf{P}\{\xi_n \not\to \xi\} = 0. \tag{2}$$

Это выражение означает, что вероятность того, что найдется такое число r, что при всех n и хотя бы при одном значении k имеет место неравенство

$$|\xi_{n+k} - \xi| > 1/r,$$

равна нулю.

Лемма 1. *Если при любом целом положительном* r

$$\sum_{n=1}^{\infty} \mathbf{P}\{|\xi_n - \xi| \geqslant 1/r\} < +\infty, \tag{3}$$

то имеет место (1) *или, что то же самое,* (2).

---

*) Это понятие в точности соответствует понятию сходимости почти всюду в теории функций.

Доказательство. Обозначим через E_n^r событие, состоящее в том, что выполняется неравенство

$$|\xi_n - \xi| \geqslant 1/r.$$

Положим, далее,

$$S_n^r = \sum_{k=1}^{\infty} E_{n+k}^r.$$

Из того, что

$$\mathsf{P}\{S_n^r\} \leqslant \sum_{k=1}^{\infty} \mathsf{P}\{E_{n+k}^r\} = \sum_{l=n+1}^{\infty} \mathsf{P}\{|\xi_l - \xi| \geqslant 1/r\},$$

мы в силу (3) выводим равенство

$$\lim_{n\to\infty} \mathsf{P}\{S_n^r\} = 0. \tag{4}$$

Пусть теперь

$$S^r = S_1^r S_2^r S_3^r \ldots$$

Из того, что событие S^r влечет за собой любое из событий S_n^r, в силу (4) получаем:

$$\mathsf{P}\{S^r\} = 0. \tag{5}$$

Положим, наконец,

$$S = S^1 + S^2 + S^3 + \ldots$$

Как нетрудно установить, это событие означает, что найдется такое r, что для каждого n ($n = 1, 2, 3, \ldots$) хотя бы при одном k [$k = k(n)$] будут выполняться неравенства

$$|\xi_{n+k} - \xi| \geqslant 1/r.$$

Так как

$$\mathsf{P}\{S\} \leqslant \sum_{r=1}^{\infty} \mathsf{P}\{S^r\},$$

то в силу (5)

$$\mathsf{P}\{S\} = 0,$$

что и требовалось доказать.

Лемма 2 (теорема Бореля). *Пусть μ – число наступлений события A при n независимых испытаниях, в каждом из которых событие A*

может появиться с вероятностью p. *Тогда при* $n \to \infty$

$$\mathsf{P}\left\{\frac{\mu}{n} \to p\right\} = 1.$$

Заметим, что теорема Бореля является простейшим частным случаем теоремы А.Н. Колмогорова об усиленном законе больших чисел; здесь же дана только иная формулировка этого частного случая, отличная от общей формулировки теоремы § 30.

Доказательство. События $\frac{\mu}{n} - p \to 0$ и $\left(\frac{\mu}{n} - p\right)^4 \to 0$, очевидно, эквивалентны. Введем, как это мы уже неоднократно делали, вспомогательные величины μ_i, равные числу появлений события A при i-м испытании. Находим, что

$$\mathsf{M}\left(\frac{\mu}{n} - p\right)^4 = \frac{1}{n^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathsf{M}(\mu_i - p)(\mu_j - p)(\mu_k - p)(\mu_l - p).$$

Элементарный подсчет показывает, что

$$\mathsf{M}\left(\frac{\mu}{n} - p\right)^4 = \frac{pq}{n^4}\,[n(p^3 + q^3) + 3pq(n^2 - n)] < \frac{1}{4n^2}.$$

Согласно лемме Чебышева

$$\mathsf{P}\left\{\left(\frac{\mu}{n} - p\right)^4 \geqslant \frac{1}{r}\right\} \leqslant r\,\mathsf{M}\left(\frac{\mu}{n} - p\right)^4 < \frac{r}{4n^2}.$$

Отсюда мы заключаем о сходимости ряда

$$\sum_{n=1}^{\infty} \mathsf{P}\left\{\left(\frac{\mu}{n} - p\right)^4 \geqslant \frac{1}{r}\right\}.$$

Применение предыдущей леммы доказывает наше утверждение.

Лемма 3. *Если событие* E *эквивалентно совместному осуществлению бесконечного числа событий* $E_1, E_2, \ldots$

$$E = E_1 E_2 \ldots$$

и каждое последующее событие E_{n+1} *влечет за собой предыдущее* E_n, *то*

$$\mathsf{P}\{E\} = \lim_{n \to \infty} \mathsf{P}\{E_n\}$$

Доказательство. Действительно, событие E_1 можно двумя следующими способами представить в виде суммы несовместимых событий:

$$E_1 = E_1\overline{E}_2 + E_2\overline{E}_3 + \ldots + E_{n-1}\overline{E}_n + E_n$$

и

$$E_1 = E_1\bar{E}_2 + E_2\bar{E}_3 + \ldots + E_{n-1}\bar{E}_n + E_n\bar{E}_{n+1} + \ldots + E.$$

Отсюда

$$\mathsf{P}\{E_1\} = \mathsf{P}\{E_1\bar{E}_2\} + \mathsf{P}\{E_2\bar{E}_3\} + \ldots + \mathsf{P}\{E_{n-1}\bar{E}_n\} + \mathsf{P}\{E_n\}$$

и

$$\mathsf{P}\{E_1\} = \mathsf{P}\{E_1\bar{E}_2\} + \mathsf{P}\{E_2\bar{E}_3\} + \ldots + \mathsf{P}\{E_{n-1}\bar{E}_n\} +$$

$$+ \mathsf{P}\{E_n\bar{E}_{n+1}\} + \ldots + \mathsf{P}\{E\}.$$

Сравнение последних двух равенств приводит нас к соотношению

$$\mathsf{P}\{E\} = \mathsf{P}\{E_n\} - \sum_{k=n}^{\infty} \mathsf{P}\{E_k\bar{E}_{k+1}\}.$$

Так как вычитаемое в правой части есть остаток сходящегося ряда, то

$$\mathsf{P}\{E\} = \lim_{n\to\infty} \mathsf{P}\{E_n\}.$$

Лемма 4. *Если каждое из событий конечной или бесконечной последовательности $E_1, E_2, \ldots, E_n, \ldots$ имеет вероятность, равную единице, то вероятность их совместного осуществления также равна единице.*

Доказательство. Рассмотрим сначала два события E_1 и E_2, для которых

$$\mathsf{P}\{E_1\} = \mathsf{P}\{E_2\} = 1.$$

Так как

$$\mathsf{P}\{E_1 + E_2\} = \mathsf{P}\{E_1\} + \mathsf{P}\{E_2\} - \mathsf{P}\{E_1E_2\}$$

и $\mathsf{P}\{E_1 + E_2\} = 1$, то

$$\mathsf{P}\{E_1E_2\} = 1.$$

Отсюда заключаем по индукции, что для любых n событий, для которых

$$\mathsf{P}\{E_1\} = \mathsf{P}\{E_2\} = \ldots = \mathsf{P}\{E_n\} = 1,$$

выполняется также равенство

$$\mathsf{P}\{E_1E_2 \ldots E_n\} = 1.$$

Пусть теперь имеется бесконечная последовательность событий $E_1, E_2, \ldots, E_n, \ldots$, для которых

$$\mathsf{P}\{E_1\} = \mathsf{P}\{E_2\} = \ldots = \mathsf{P}\{E_n\} = \ldots = 1.$$

Так как очевидно, что

$$E_1E_2E_3 \ldots = E_1(E_1E_2)(E_1E_2E_3) \ldots$$

и каждый последующий множитель в правой части равенства влечет за собой предыдущий, то согласно предыдущей лемме

$$\mathsf{P}\{E_1E_2E_3\ldots\} = \lim_{n\to\infty} \mathsf{P}\{E_1E_2\ldots E_n\}.$$

Это равенство доказывает лемму.

Теорема Гливенко. *Пусть $F(x)$ – функция распределения случайной величины ξ и $F_n(x)$ – эмпирическая функция распределения результатов n независимых наблюдений над величиной ξ. Тогда при $n\to\infty$*

$$\mathsf{P}\{\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0\} = 1.$$

Доказательство. Обозначим через $x_{r,k}$ наименьшее x, удовлетворяющее неравенствам

$$F(x-0) = F(x) \leqslant \frac{k}{r} \leqslant F(x+0) \quad (k=1,2,\ldots,r).$$

Пусть A означает событие, состоящее в том, что $\xi < x_{r,k}$. Ясно, что

$$\mathsf{P}\{A\} = F(x_{r,k}).$$

Так как частота появления события A равна $F_n(x_{r,k})$, то по теореме Бореля (лемма 2)

$$\mathsf{P}\{F_n(x_{r,k}) \underset{n\to\infty}{\to} F(x_{r,k})\} = 1. \tag{6}$$

Пусть теперь E_k^r есть событие, состоящее в том, что при $n\to\infty$

$$F_n(x_{r,k}) \to F(x_{r,k}) \quad (k=1,2,\ldots,r)$$

и

$$E^r = E_1^r E_2^r \ldots E_r^r.$$

Ясно, что событие E^r равносильно тому, что при $n\to\infty$

$$\max_{1\leqslant k\leqslant r} |F_n(x_{r,k}) - F(x_{r,k})| \to 0.$$

Так как согласно (6)

$$\mathsf{P}\{E_1^r\} = \mathsf{P}\{E_2^r\} = \ldots = \mathsf{P}\{E_r^r\} = 1,$$

то в силу леммы 4

$$\mathsf{P}\{E^r\} = 1.$$

Пусть далее

$$E = E^1E^2E^3\ldots$$

Согласно лемме 4

$$\mathbf{P}\{E\} = 1.$$

Обозначим, наконец, через S событие, состоящее в том, что при $n \to \infty$

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \to 0.$$

Для любого x, заключенного между $x_{r,k}$ и $x_{r,k+1}$, выполняются неравенства

$$F_n(x_{r,k} + 0) \leqslant F_n(x) \leqslant F_n(x_{r,k+1})$$

и

$$F(x_{r,k} + 0) \leqslant F(x) \leqslant F(x_{r,k+1}),$$

причем

$$0 \leqslant F(x_{r,k+1}) - F(x_{r,k} + 0) \leqslant 1/r.$$

Отсюда мы заключаем, что

$$F_n(x_{r,k} + 0) - F(x_{r,k+1}) \leqslant F_n(x) - F(x) \leqslant$$

$$\leqslant F_n(x_{r,k+1}) - F(x_{r,k} + 0),$$

т.е. что

$$|F_n(x) - F(x)| \leqslant \max_{1 \leqslant k \leqslant r} |F_n(x_{r,k}) - F(x_{r,k})| + 1/r$$

и что, следовательно,

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \leqslant \max_{1 \leqslant k \leqslant r} |F_n(x_{r,k}) - F(x_{r,k})| + 1/r.$$

Поскольку r произвольно, то из последнего неравенства вытекает, что $E \subset S$. Этим, очевидно, доказано, что

$$\mathbf{P}\{\sup_{-\infty < \infty < \infty} |F_n(x) - F(x)| \to 0\} = 1.$$

Упражнения

1. Доказать, что если случайная величина ξ такова, что $\mathbf{M}e^{a\xi}$ существует ($a > 0$ – постоянная), то

$$\mathbf{P}\{\xi \geqslant \epsilon\} \leqslant \frac{\mathbf{M}e^{a\xi}}{e^{a\epsilon}}.$$

2. Пусть $f(x) > 0$ – неубывающая функция. Доказать, что если существует $\mathbf{M}f(|\xi - \mathbf{M}\xi|)$, то

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| \geqslant \epsilon\} \leqslant \frac{\mathbf{M}f(|\xi - \mathbf{M}\xi|)}{f(\epsilon)}.$$

3. Последовательность независимых и одинаково распределенных случайных величин $\{\xi_i\}$ определена равенствами

а) $\mathbf{P}\{\xi_n = 2^{k-\ln k - 2\ln\ln k}\} = \dfrac{1}{2^k}$ $(k = 1, 2, 3, \ldots)$,

б) $\mathbf{P}\{\xi_n = k\} = \dfrac{c}{k^2 \ln^2 k}$ $\left(k \geqslant 2, \quad c^{-1} = \sum\limits_{k=2}^{\infty} \dfrac{1}{k^2 \ln^2 k}\right)$.

Доказать, что к указанным последовательностям закон больших чисел применим.

4. Доказать, что к последовательности независимых случайных величин $\{\xi_n\}$ таких, что

$$\mathbf{P}\{\xi_n = n^\alpha\} = \mathbf{P}\{\xi_n = -n^\alpha\} = 1/2,$$

закон больших чисел применим тогда и только тогда, когда $\alpha < 0{,}5$.

5. Доказать, что если независимые случайные величины $\{\xi_n\}$ таковы, что

$$\max_{1 \leqslant k \leqslant n} \int_{|x| \geqslant A} |x|\, dF_k(x) \to 0, \text{ когда } A \to \infty,$$

то к последовательности $\{\xi_n\}$ применим закон больших чисел.

6. Используя результат предыдущей задачи, доказать, что если для последовательности независимых случайных величин $\{\xi_n\}$ существуют такие числа $\alpha > 1$ и β, что $\mathbf{M}\,|\xi|^\alpha \leqslant \beta$, то к последовательности $\{\xi_k\}$ применим закон больших чисел (теоремы Маркова).

7. Дана последовательность случайных величин $\{\xi_k\}$, для которых $\mathbf{D}\xi_n \leqslant C$, $R_{ij} \to 0$ при $|i - j| \to \infty$ (R_{ij} – коэффициент корреляции между ξ_i и ξ_j). Доказать, что к данной последовательности применим закон больших чисел (теорема С.Н. Бернштейна).

ГЛАВА 7

ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ

Мы видели в предыдущих главах, что в теории вероятностей широко используются методы и аналитический аппарат различных отделов математического анализа. Простое решение весьма многих задач теории вероятностей, особенно тех из них, которые связаны с суммированием независимых случайных величин, удается получить с помощью характеристических функций, теория которых развита в анализе и известна под именем преобразований Фурье. Настоящая глава посвящена изложению основных свойств характеристических функций.

§ 32. Определение и простейшие свойства характеристических функций

Характеристической функцией случайной величины ξ называется математическое ожидание случайной величины $e^{it\xi}$ *). Если $F(x)$ есть функция распределения величины ξ то характеристическая функция равна по теореме 1 § 22

$$f(t) = \int e^{itx} dF(x). \tag{1}$$

Мы условимся обозначать в дальнейшем характеристическую функцию и соответствующую ей функцию распределения одними и теми же буквами, но только соответственно малой и большой.

Из того, что $|e^{itx}| = 1$ при всех вещественных t, следует существование интеграла (1) для всех функций распределения; следовательно, характеристическая функция может быть определена для каждой случайной величины

Теорема 1. *Характеристическая функция равномерно непрерывна на всей прямой и удовлетворяет следующим соотношениям*:

$$f(0) = 1, \quad |f(t)| \leqslant 1 \ (-\infty < t < \infty). \tag{2}$$

---

*) t – действительный параметр. Математическое ожидание для комплексной случайной величины $\xi + i\eta$ определяем как $\mathbf{M}\xi + i\mathbf{M}\eta$. Легко проверить, что теоремы 1,2 и 3 § 22 справедливы и в этом случае.

Доказательство. Соотношения (2) немедленно вытекают из определения характеристической функции. Действительно, по (1)

$$f(0) = \int 1 \cdot dF(x) = 1$$

и

$$|f(t)| = \left|\int e^{itx} dF(x)\right| \leqslant \int |e^{itx}|\, dF(x) = \int dF(x) = 1.$$

Нам остается доказать равномерную непрерывность функции $f(t)$. С этой целью рассмотрим разность

$$f(t+h) - f(t) = \int e^{itx}\,(e^{ixh} - 1)\, dF(x)$$

и оценим ее по модулю. Имеем:

$$|f(t+h) - f(t)| \leqslant \int |e^{ixh} - 1|\, dF(x).$$

Пусть $\epsilon > 0$ произвольно; выберем столь большое A, чтобы

$$\int_{|x|>A} dF(x) < \frac{\epsilon}{4},$$

и подберем столь малое h, чтобы для $|x| < A$

$$|e^{ixh} - 1| < \epsilon/2.$$

Тогда

$$|f(t+h) - f(t)| \leqslant \int_{-A}^{A} |e^{ixh} - 1|\, dF(x) + 2\int_{|x|\geqslant A} dF(x) \leqslant \epsilon.$$

Это неравенство доказывает теорему.

Теорема 2. *Если $\eta = a\xi + b$, где a и b – постоянные, то*

$$f_\eta(t) = f_\epsilon(at)\, e^{ibt},$$

где $f_\eta(t)$ и $f_\xi(t)$ означают характеристические функции величин η и ξ.

Доказательство. Действительно,

$$f_\eta(t) = \mathsf{M}\, e^{it\eta} = \mathsf{M}\, e^{it(a\xi+b)} = e^{itb}\mathsf{M} e^{ita\xi} = e^{itb} f_\xi(at).$$

Теорема 3. *Характеристическая функция суммы двух независимых случайных величин равна произведению их характеристических функций.*

Доказательство. Пусть ξ и η – независимые случайные величины и $\zeta = \xi + \eta$. Тогда очевидно, что вместе с ξ и η независимы также случайные величины $e^{it\xi}$ и $e^{it\eta}$. Отсюда вытекает, что

$$\mathsf{M}\, e^{it\zeta} = \mathsf{M} e^{it(\xi+\eta)} = \mathsf{M}\,(e^{it\xi} e^{it\eta}) = \mathsf{M}\, e^{it\xi} \mathsf{M} e^{it\eta}.$$

Это доказывает теорему.

Следствие. *Если*

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n,$$

причем каждое слагаемое независимо от суммы предыдущих, то характеристическая функция величины ξ равна произведению характеристических функций слагаемых.

Применение характеристических функций в значительной степени опирается на свойство, сформулированное в теореме 3. Сложение независимых случайных величин, как мы видели в § 21, приводит к весьма сложной операции — композиции функций распределения слагаемых. Для характеристических функций эта сложная операция заменяется весьма простой — простым умножением характеристических функций.

Теорема 4. *Если случайная величина ξ имеет абсолютный момент n-го порядка, то характеристическая функция величины ξ дифференцируема n раз и при $k \leqslant n$*

$$f^{(k)}(0) = i^k \mathbf{M}\, \xi^k. \tag{3}$$

Доказательство. Действительно, k-кратное ($k \leqslant n$) формальное дифференцирование характеристической функции приводит к равенству

$$f^{(k)}(t) = i^k \int x^k e^{itx} dF(x). \tag{4}$$

Но

$$\left| \int x^k e^{itx} dF(x) \right| \leqslant \int |x|^k dF(x)$$

и, следовательно, в силу предположения теоремы ограничен. Отсюда следуют существование интеграла (4) и законность дифференцирования. Положив в (4) $t = 0$, находим, что

$$f^{(k)}(0) = i^k \int x^k dF(x).$$

Математическое ожидание и дисперсия весьма просто выражаются при помощи производных от логарифма характеристической функции. В самом деле, положим

$$\psi(t) = \ln f(t).$$

Тогда

$$\psi'(t) = \frac{f'(t)}{f(t)}$$

и

$$\psi''(t) = \frac{f''(t) \cdot f(t) - |f'(t)|^2}{f^2(t)}.$$

Приняв во внимание, что $f(0) = 1$ и равенство (3), находим, что

$$\psi'(0) = f'(0) = i\,\mathbf{M}\xi$$

и

$$\psi''(0) = f''(0) - [f'(0)]^2 = i^2\mathbf{M}\xi^2 - [i\mathbf{M}\xi]^2 = -\mathbf{D}\,\xi.$$

Отсюда

$$\left.\begin{aligned} &\mathbf{M}\xi = \frac{1}{i}\,\psi'(0) \\ \text{и}\quad & \\ &\mathbf{D}\,\xi = -\psi''(0). \end{aligned}\right\} \qquad (5)$$

Производная k-го порядка логарифма характеристической функции в точке 0, умноженная на i^k, называется *семиинвариантом k-го порядка* случайной величины.

Как это непосредственно следует из теоремы 3, при сложении независимых случайных величин их семиинварианты складываются.

Мы только что видели, что первыми двумя семиинвариантами являются математическое ожидание и дисперсия, т.е. момент первого порядка и некоторая рациональная функция моментов первого и второго порядков. Путем вычислений легко убедиться, что семиинвариант любого порядка k есть (целая) рациональная функция первых k моментов. Для примера приведем явные выражения семиинвариантов третьего и четвертого порядков:

$$i^3\psi'''(0) = -\{\mathbf{M}\xi^3 - 3\mathbf{M}\xi^2\cdot\mathbf{M}\xi + 2[\mathbf{M}\xi]^3\},$$

$$i^4\psi^{\mathrm{IV}}(0) = \mathbf{M}\xi^4 - 4\mathbf{M}\xi^3\mathbf{M}\xi - 3[\mathbf{M}\xi^2]^2 + 12\mathbf{M}\xi^2[\mathbf{M}\xi]^2 - 6[\mathbf{M}\xi]^4.$$

Рассмотрим теперь несколько примеров характеристических функций.

П р и м е р 1. Случайная величина ξ распределена по нормальному закону с математическим ожиданием a и дисперсией σ^2. Характеристическая функция величины ξ равна

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}}\int e^{itx - \frac{(x-a)^2}{2\sigma^2}}\,dx.$$

Подстановкой

$$z = \frac{x-a}{\sigma} - it\sigma$$

$\varphi(t)$ приводится к виду

$$\varphi(t) = e^{iat - \sigma^2t^2/2}\,\frac{1}{\sqrt{2\pi}}\int_{-\infty - it\sigma}^{\infty - it\sigma} e^{-z^2/2}\,dz.$$

Известно, что при любом вещественном α

$$\int_{-\infty-i\alpha}^{\infty-i\alpha} e^{-z^2/2}\,dz = \sqrt{2\pi},$$

следовательно,

$$\varphi(t) = e^{iat-\sigma^2 t^2/2}.$$

Пользуясь теоремой 4, мы можем без труда вычислить центральные моменты для нормального распределения и тем самым другим путем получить результат примера, рассмотренного в § 26.

Пример 2. Найти характеристическую функцию случайной величины ξ, распределенной по закону Пуассона.

Согласно предположению величина ξ принимает только целочисленные значения, причем

$$\mathsf{P}\{\xi = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (k = 0, 1, 2, \ldots),$$

где $\lambda > 0$ – постоянная.

Характеристическая функция величины ξ равна

$$f(t) = \mathsf{M}\,e^{it\xi} = \sum_{k=0}^{\infty} e^{ikt}\,\mathsf{P}\{\xi = k\} = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda} =$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!} = e^{-\lambda+\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

Согласно (5) отсюда находим, что

$$\mathsf{M}\,\xi = \frac{1}{i}\psi'(0) = \lambda; \quad \mathsf{D}\xi = -\psi''(0) = \lambda.$$

Эти равенства были нами ранее (§ 23, пример 3) получены непосредственно.

Пример 3. Случайная величина ξ равномерно распределена в интервале $(-a, a)$. Характеристическая функция равна

$$f(t) = \int_{-a}^{a} e^{itx}\frac{dx}{2a} = \frac{\sin at}{at}.$$

Пример 4. Найти характеристическую функцию величины μ, равной числу появлений события A в n независимых испытаниях, в каждом из которых вероятность появления события A равна p.

Величина μ может быть представлена как сумма

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n$$

n независимых величин, каждая из которых принимает лишь два значения 0 и 1, соответственно с вероятностями $q = 1 - p$ и p. Величина μ_k принимает значение 1, если событие A происходит в k-м испытании, и значение 0, если событие A в k-м испытании не происходит.

Характеристическая функция величины μ_k равна

$$f_k(t) = \mathbf{M}\, e^{it\mu_k} = e^{it\cdot 0}q + e^{it\cdot 1}p = q + pe^{it}.$$

Согласно теореме 3 характеристическая функция величины μ равна

$$f(t) = \prod_{k=1}^{n} f_k(t) = (q + pe^{it})^n.$$

Найдем еще характеристическую функцию величины $\eta = \dfrac{\mu - np}{\sqrt{npq}}$. По теореме 2 она равна

$$f_\eta(t) = e^{-it\sqrt{\frac{np}{q}}}\, f\left(\frac{t}{\sqrt{npq}}\right) = e^{-it\sqrt{\frac{np}{q}}}\,(q + pe^{i\frac{t}{\sqrt{npq}}})^n =$$

$$= (qe^{-it\sqrt{\frac{p}{nq}}} + pe^{it\sqrt{\frac{q}{np}}})^n.$$

П р и м е р 5. Характеристические функции удовлетворяют равенству $f(-t) = \overline{f}(t)$.

Действительно,

$$f(-t) = \int e^{-itx}dF(x) = \overline{\int e^{itx}dF(x)} = \overline{f(t)}.$$

§ 33. Формула обращения и теорема единственности

Мы видели, что по функции распределения величины ξ всегда можно найти ее характеристическую функцию; для нас важно, что имеет место также обратное предложение: по характеристической функции функция распределения определяется однозначно.

Т е о р е м а 1. *Пусть $f(t)$ и $F(x)$ – характеристическая функция и функция распределения случайной величины ξ. Если x_1 и x_2 – точки непрерывности функции $F(x)$, то*

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{c\to\infty} \int_{-c}^{c} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t)\, dt. \qquad (1)$$

Д о к а з а т е л ь с т в о. Из определения характеристической функции следует, что интеграл

$$J_c = \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t)\, dt$$

равен

$$J_c = \frac{1}{2\pi} \int_{-c}^{c} \int \frac{1}{it} \left[e^{it(z-x_1)} - e^{it(z-x_2)}\right] dF(z)\, dt.$$

В последнем интеграле можно изменить порядок интегрирования, так как по z интеграл абсолютно сходится, а по t пределы интегрирования конечны. Таким образом,

$$J_c = \frac{1}{2\pi} \int \left[\int_{-c}^{c} \frac{e^{it(z-x_1)} - e^{it(z-x_2)}}{it} dt \right] dF(z) =$$

$$\frac{1}{2\pi} \int \left[\int_{0}^{c} \frac{e^{it(z-x_1)} - e^{-it(z-x_1)} - e^{it(z-x_2)} + e^{-it(z-x_2)}}{it} dt \right] dF(z) =$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{0}^{c} \left[\frac{\sin t(z-x_1)}{t} - \frac{\sin t(z-x_2)}{t} \right] dt\, dF(z).$$

Из анализа известно, что при $c \to \infty$

$$\frac{1}{\pi} \int_{0}^{c} \frac{\sin \alpha t}{t} dt \to \begin{cases} 1/2, & \text{если } \alpha > 0, \\ -1/2, & \text{если } \alpha < 0, \end{cases} \tag{2}$$

и эта сходимость равномерна относительно α в каждой области $\alpha > \delta > 0$ (соответственно $\alpha < -\delta$), и при $|\alpha| \leqslant \delta$, при всех c

$$\left| \frac{1}{\pi} \int_{0}^{c} \frac{\sin \alpha t}{t} dt \right| < 1. \tag{3}$$

Положим для определенности, что $x_2 > x_1$, и представим интеграл J_c в виде следующей суммы:

$$J_c = \int_{-\infty}^{x_1-\delta} + \int_{x_1-\delta}^{x_1+\delta} + \int_{x_1+\delta}^{x_2-\delta} + \int_{x_2-\delta}^{x_2+\delta} + \int_{x_2+\delta}^{\infty} \psi(c, z; x_1, x_2)\, dF(z),$$

где для краткости обозначено

$$\psi(c, z;\ x_1, x_2) = \frac{1}{\pi} \int_0^c \left\{ \frac{\sin t(z - x_1)}{t} - \frac{\sin t(z - x_2)}{t} \right\} dt$$

и $\delta > 0$ подобрано так, что $x_1 + \delta < x_2 - \delta$.

В области $-\infty < z < x_1 - \delta$ имеют место неравенства $z - x_1 < -\delta$ и $z - x_2 < -\delta$. Поэтому мы на основании (2) заключаем, что при $c \to \infty$

$$\int_{-\infty}^{x_1 - \delta} \psi(c, z;\ x_1, x_2)\, dF(z) \to 0.$$

Аналогично при $x_2 + \delta < z < +\infty$ и при $c \to \infty$

$$\int_{x_2 + \delta}^{\infty} \psi(c, z;\ x_1, x_2)\, dF(z) \to 0.$$

Далее, так как в области $x_1 + \delta < z < x_2 - \delta$ имеют место неравенства $z - x_1 > \delta$ и $z - x_2 < \delta$, то согласно (2) при $c \to \infty$

$$\int_{x_1 + \delta}^{x_2 - \delta} \psi(c, z;\ x_1, x_2)\, dF(z) \to \int_{x_1 + \delta}^{x_2 - \delta} dF(z) = F(x_2 - \delta) - F(x_1 + \delta).$$

Наконец, в силу (3) мы можем воспользоваться оценками

$$\left| \int_{x_1 - \delta}^{x_1 + \delta} \psi(c, z;\ x_1, x_2)\, dF(z) \right| < 2 \int_{x_1 - \delta}^{x_1 + \delta} dF(z) = 2[F(x_1 + \delta) - F(x_1 - \delta)]$$

и

$$\left| \int_{x_2 - \delta}^{x_2 + \delta} \psi(c, z;\ x_1, x_2)\, dF(z) \right| < 2 \int_{x_2 - \delta}^{x_2 + \delta} dF(z) = 2[F(x_2 + \delta) - F(x_2 - \delta)].$$

Таким образом, находим, что при любом $\delta > 0$

$$\overline{\lim_{c \to \infty}}\ J_c = F(x_2 - \delta) - F(x_1 + \delta) + R_1(\delta, x_1, x_2)$$

и

$$\underline{\lim_{c \to \infty}}\ J_c = F(x_2 - \delta) - F(x_1 + \delta) + R_2(\delta, x_1, x_2),$$

где

$$|R_i(\delta, x_1, x_2)| < 2\,\{F(x_1 + \delta) - F(x_1 - \delta) + F(x_2 + \delta) - F(x_2 - \delta)\}$$

$(i = 1, 2)$.

Пусть теперь $\delta \to 0$. При этом из того, что x_1 и x_2 являются точками непрерывности функции $F(x)$, следуют равенства

$$\lim_{\delta\to 0} F(x_1 + \delta) = \lim_{\delta\to 0} F(x_1 - \delta) = F(x_1)$$

и

$$\lim_{\delta\to 0} F(x_2 + \delta) = \lim_{\delta\to 0} F(x_2 - \delta) = F(x_2).$$

А так как J_c не зависит от δ, то

$$\lim_{c\to\infty} J_c = F(x_2) - F(x_1).$$

Равенство (1) носит название формулы обращения. Мы используем эту формулу для вывода следующего важного предложения (теорема единственности).

Теорема 2. *Функция распределения однозначно определяется своей характеристической функцией.*

Доказательство. Действительно из теоремы 1 непосредственно следует, что в каждой точке непрерывности функции $F(x)$ применима формула

$$F(x) = \frac{1}{2\pi} \lim_{y\to-\infty} \lim_{c\to\infty} \int_{-c}^{+c} \frac{e^{-ity} - e^{-itx}}{it} f(t)\,dt,$$

где предел по y берется по множеству точек y, являющихся точками непрерывности функции $F(x)$.

В качестве приложения последней теоремы мы докажем следующие предложения.

Пример 1. Если независимые случайные величины ξ_1 и ξ_2 распределены нормально, то их сумма $\xi = \xi_1 + \xi_2$ также распределена нормально.

Действительно, если

$$\mathbf{M}\xi_1 = a_1, \quad \mathbf{D}\xi_1 = \sigma_1^2; \quad \mathbf{M}\xi_2 = a_2, \quad \mathbf{D}\xi_2 = \sigma_2^2,$$

то характеристические функции величин ξ_1 и ξ_2 равны

$$f_1(t) = e^{ia_1 t - \frac{1}{2}\sigma_1^2 t^2} \qquad f_2(t) = e^{ia_2 t - \frac{1}{2}\sigma_2^2 t^2}$$

По теореме 3 § 32 характеристическая функция $f(t)$ суммы равна

$$f(t) = f_1(t) \cdot f_2(t) = e^{it(a_1 + a_2) - \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2}$$

Это - характеристическая функция нормального закона с математическим ожиданием $a = a_1 + a_2$ и дисперсией $\sigma^2 = \sigma_1^2 + \sigma_2^2$. На основании теоремы единственности заключаем, что функция распределения величины ξ нормальна.

П р и м е р 2. Независимые случайные величины ξ_1 и ξ_2 распределены по закону Пуассона, причем

$$\mathsf{P}\{\xi_1 = k\} = \frac{\lambda_1^k e^{-\lambda_1}}{k!}, \quad \mathsf{P}\{\xi_2 = k\} = \frac{\lambda_2^k e^{-\lambda_2}}{k!}.$$

Докажем, что случайная величина $\xi = \xi_1 + \xi_2$ распределена по закону Пуассона с параметром $\lambda = \lambda_1 + \lambda_2$.

Действительно, в примере 2 предыдущего параграфа мы нашли, что характеристические функции случайных величин ξ_1 и ξ_2 равны

$$f_1(t) = e^{\lambda_1 (e^{it} - 1)}, \quad f_2(t) = e^{\lambda_2 (e^{it} - 1)}.$$

В силу теоремы 3 предыдущего параграфа характеристическая функция суммы $\xi = \xi_1 + \xi_2$ равна

$$f(t) = f_1(t) \cdot f_2(t) = e^{(\lambda_1 + \lambda_2)(e^{it} - 1)},$$

т.е. является характеристической функцией некоторого закона Пуассона. Согласно теореме единственности единственное распределение, имеющее $f(t)$ своей характеристической функцией, есть закон Пуассона, для которой

$$\mathsf{P}\{\xi = k\} = \frac{(\lambda_1 + \lambda_2)^k e^{-(\lambda_1 + \lambda_2)}}{k!} \quad (k \geqslant 0).$$

Д.А. Райков доказал обратное более глубокое предложение: если сумма двух независимых случайных величин распределена по закону Пуассона, то каждое слагаемое также распределено по закону Пуассона.

П р и м е р 3. Характеристическая функция вещественна тогда и только тогда, когда соответствующая ей функция распределения симметрична, т.е. когда при любых x функция распределения удовлетворяет равенству

$$F(x) = 1 - F(-x + 0).$$

Если функция распределения симметрична, то ее характеристическая функция вещественна. Это доказывается несложным подсчетом:

$$f(t) = \int e^{itx} dF(x) =$$

$$= \int_0^\infty e^{-itx} dF(-x + 0) + \int_0^\infty e^{itx} dF(x) - F(+0) - F(-0) =$$

$$= \int_0^\infty (e^{-itx} + e^{itx}) dF(x) - F(+0) - F(-0) =$$

$$= 2 \int_0^\infty \cos tx \, dF(x) - F(+0) - F(-0) = \int \cos tx \, dF(x).$$

(Напомним здесь, что мы условились включать нижний предел в интервал интегрирования и не включать верхний).

Для доказательства обратного предложения рассмотрим случайную величину $\eta = -\xi$. Функция распределения величины η равна

$$G(x) = \mathsf{P}\{\eta < x\} = \mathsf{P}\{\xi > -x\} = 1 - F(-x+0).$$

Характеристические функции величин ξ и η связаны соотношением

$$g(t) = \mathsf{M}\, e^{it\eta} = \mathsf{M}\, e^{-it\xi} = \overline{\mathsf{M}\, e^{it\xi}} = \bar{f}(t).$$

Так как по условию $f(t)$ вещественна, то $\bar{f}(t) = f(t)$ и, значит,

$$g(t) = f(t).$$

Из теоремы единственности мы теперь заключаем, что функции распределения величин ξ и η совпадают, т.е. что

$$F(x) = 1 - F(-x+0),$$

что и требовалось доказать.

§ 34. Теоремы Хелли

В дальнейшем нам потребуются две теоремы чисто аналитического характера – первая и вторая теоремы Хелли.

Условимся говорить, что последовательность неубывающих функций

$$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots$$

сходится в основном к неубывающей функции $F(x)$, если при $n \to \infty$ она сходится к этой последней в каждой ее точке непрерывности.

Впоследствии мы всегда будем считать, что функции $F_n(x)$ удовлетворяют условию

$$F_n(-\infty) = 0,$$

и не станем далее оговаривать этого.

Отметим сразу же, что для сходимости в основном достаточно, чтобы последовательность функций сходилась к функции $F(x)$ на каком-нибудь всюду плотном множестве D. Действительно, пусть x – любая точка и x' и x'' – какие-нибудь две точки множества D, такие, что $x' \leqslant x \leqslant x''$. При этом также

$$F_n(x') \leqslant F_n(x) \leqslant F_n(x'').$$

Следовательно,

$$\lim_{n\to\infty} F_n(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant \lim_{n\to\infty} F_n(x'').$$

А так как по предположению

$$\lim_{n\to\infty} F_n(x') = F(x') \text{ и } \lim_{n\to\infty} F_n(x'') = F(x''),$$

то и

$$F(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x'').$$

Но средние члены в этих неравенствах не зависят от x' и x'', поэтому

$$F(x-0) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

Если функция $F(x)$ в точке x непрерывна, то

$$F(x-0) = F(x) = F(x+0).$$

Следовательно, в точках непрерывности функции $F(x)$

$$\lim_{n\to\infty} F_n(x) = F(x).$$

П е р в а я т е о р е м а Х е л л и. *Всякая последовательность ограниченных в совокупности неубывающих функций*

$$F_1(x),\ F_2(x),\ \ldots,\ F_n(x),\ \ldots \tag{1}$$

содержит по крайней мере одну подпоследовательность

$$F_{n_1}(x),\ F_{n_2}(x),\ \ldots, F_{n_k}(x), \ldots,$$

сходящуюся в основном к некоторой неубывающей функции $F(x)$.

Д о к а з а т е л ь с т в о. Пусть D – какое-нибудь счетное всюду плотное множество точек $x_1', x_2', \ldots, x_n', \ldots$ Возьмем значения функций последовательности (1) в точке x_1'

$$F_1(x_1'),\ F_2(x_1'), \ldots, F_n(x_1'),\ \ldots$$

Так как множество этих значений, по предположению, ограничено, то оно содержит по меньшей мере одну последовательность

$$F_{11}(x_1'),\ F_{12}(x_1'),\ \ldots,\ F_{1n}(x_1'),\ \ldots, \tag{2}$$

сходящуюся к некоторому предельному значению, которое мы обозначим через $G(x_1')$. Рассмотрим теперь множество чисел

$$F_{11}(x_2'),\ F_{12}(x_2'), \ldots,\ F_{1n}(x_2'),\ \ldots$$

Так как и это множество ограничено, то существует в нем последовательность, сходящаяся к некоторому предельному значению $G(x_2')$. Таким образом, из последовательности (2) мы можем выделить подпоследовательность

$$F_{21}(x),\ F_{22}(x), \ldots,\ F_{2n}(x),\ \ldots, \tag{3}$$

для которой одновременно $\lim\limits_{n\to\infty} F_{2n}(x_1') = G(x_1')$ и $\lim\limits_{n\to\infty} F_{2n}(x_2') = G(x_2')$.

Продолжим такое выделение подпоследовательностей

$$F_{k1}(x),\ F_{k2}(x),\ \ldots,\ F_{kn}(x),\ \ldots, \tag{4}$$

для которых одновременно имели бы место равенства $\lim\limits_{n\to\infty} F_{kn}(x_r') = G(x_r')$ при всех $r \leqslant k$. Составим теперь диагональную последовательность

$$F_{11}(x), F_{22}(x),\ \ldots,\ F_{nn}(x),\ \ldots \tag{5}$$

Вся она в конечном счете выделена из последовательности (1), поэтому для нее $\lim\limits_{n\to\infty} F_{nn}(x_1') = G(x_1')$. Далее, так как вся диагональная последовательность, за исключением лишь первого члена, выделена из последовательности (2), то $\lim\limits_{n\to\infty} F_{nn}(x_2') = G(x_2')$. Вообще вся диагональная последовательность, за исключением первых ее $k-1$ членов, выделена из последовательности (4); поэтому для нее также $\lim\limits_{n\to\infty} F_{nn}(x_k') = G(x_k')$ при каждом k. Полученный результат можно сформулировать так: последовательность (1) содержит по крайней мере одну подпоследовательность, которая во всех точках x_k' множества D сходится к некоторой функции $G(x)$, определенной на множестве D. При этом, так как функции $F_{nn}(x)$ не убывают и равномерно ограничены, то, очевидно, и функция $G(x)$ будет неубывающей и ограниченной.

Теперь ясно, что функцию $G(x)$, определенную на множестве D, можно продолжить так, что она будет определена на всей прямой $-\infty < x < \infty$, оставаясь неубывающей и ограниченной.

Последовательность (5) сходится к этой функции на всюду плотном множестве D; следовательно, она сходится к ней в основном, что и требовалось доказать. Заметим, что функция, полученная продолжением функции G, может оказаться не непрерывной слева. Но мы можем изменить ее значения в точках разрыва так, чтобы восстановить этой свойство. Подпоследовательность F_{nn} будет сходиться в основном и к таким образом "поправленной" функции.

Вторая теорема Хелли. *Пусть $f(x)$ – непрерывная функция и пусть последовательность неубывающих, ограниченных в совокупности функций*

$$F_1(x),\ F_2(x),\ \ldots,\ F_n(x), \ldots$$

сходится в основном к функции $F(x)$ на некотором конечном интервале

$a \leqslant x \leqslant b$, где a и b – точки непрерывности функции $F(x)$; тогда

$$\lim_{n\to\infty} \int_a^b f(x)\,dF_n(x) = \int_a^b f(x)\,dF(x).$$

Д о к а з а т е л ь с т в о. Из непрерывности функции $f(x)$ вытекает, что как бы мало ни было положительное постоянное ϵ, найдется подразделение интервала $a \leqslant x \leqslant b$ точками $x_0 = a, x_1, \ldots, x_N = b$ на частичные интервалы $x_k \leqslant x \leqslant x_{k+1}$ такое, что в каждом интервале (x_k, x_{k+1}) будет выполняться неравенство $|f(x) - f(x_k)| < \epsilon$. Пользуясь этим обстоятельством, мы можем ввести вспомогательную функцию $f_\epsilon(x)$, принимающую только конечное число значений, определив ее посредством равенств

$$f_\epsilon(x) = f(x_k) \text{ при } x_k \leqslant x < x_{k+1}.$$

Очевидно, что для всех x в интервале $a \leqslant x \leqslant b$ выполняется неравенство

$$|f(x) - f_\epsilon(x)| < \epsilon.$$

При этом мы можем заранее выбрать точки деления $x_1, x_2, \ldots, x_{N-1}$ так, чтобы они были точками непрерывности функции $F(x)$. В силу сходимости функций $F_1(x), F_2(x), F_3(x), \ldots$ к функции $F(x)$, при достаточно больших n во всех точках деления будут выполняться неравенства

$$|F(x_k) - F_n(x_k)| < \frac{\epsilon}{MN}, \tag{6}$$

где M – максимум модуля $f(x)$ в интервале $a \leqslant x \leqslant b$.

Без объяснений ясно, что

$$\left|\int_a^b f(x)dF(x) - \int_a^b f(x)dF_n(x)\right| \leqslant \left|\int_a^b f(x)dF(x) - \int_a^b f_\epsilon(x)dF(x)\right| +$$

$$+ \left|\int_a^b f_\epsilon(x)dF(x) - \int_a^b f_c(x)dF_n(x)\right| + \left|\int_a^b f_\epsilon(x)dF_n(x) -\right.$$

$$\left. - \int_a^b f(x)dF_n(x)\right|.$$

Нетрудно подсчитать, что первое слагаемое правой части не превосходит $\epsilon[F(b) - F(a)]$, а третье не превосходит $\epsilon[F_n(b) - F_n(a)]$. Что же касается второго слагаемого, то оно равно

$$\left|\sum_{k=0}^{N-1} f(x_k)[F(x_{k+1})-F(x_k)] - \sum_{k=0}^{N-1} f(x_k)[F_n(x_{k+1})-F_n(x_k)]\right| =$$

$$= \left|\sum_{k=1}^{N-1} f(x_k)[F(x_{k+1})-F_n(x_{k+1})] - \sum_{k=0}^{N-1} f(x_k)[F(x_k)-F_n(x_k)]\right|$$

и, следовательно, при достаточно больших n не превосходит 2ϵ, как это вытекает из неравенства (6). В силу ограниченности функций $F_n(x)$ в совокупности, сумма

$$\epsilon[F(b)-F(a)] + \epsilon[F_n(b)-F_n(a)] + 2\epsilon$$

может быть сделана сколь угодно малой вместе с ϵ.

Обобщенная вторая теорема Хелли. *Если функция $f(x)$ непрерывна и ограничена на всей прямой $-\infty < x < \infty$, последовательность ограниченных в совокупности неубывающих функций*

$$F_1(x), F_2(x), \ldots, F_n(x), \ldots$$

сходится в основном к функции $F(x)$ и

$$\lim_{n\to\infty} F_n(-\infty) = F(-\infty), \qquad \lim_{n\to\infty} F_n(+\infty) = F(+\infty),$$

то

$$\lim_{n\to\infty} \int f(x)dF_n(x) = \int f(x)dF(x).$$

Доказательство. Пусть $A<0$ и $B>0$; положим

$$J_1 = \left|\int_{-\infty}^{A} f(x)dF(x) - \int_{-\infty}^{A} f(x)dF_n(x)\right|,$$

$$J_2 = \left|\int_{A}^{B} f(x)dF(x) - \int_{A}^{B} f(x)dF_n(x)\right|,$$

$$J_3 = \left|\int_{B}^{\infty} f(x)dF(x) - \int_{B}^{\infty} f(x)dF_n(x)\right|.$$

Очевидно, что

$$\left|\int f(x)dF(x) - \int f(x)dF_n(x)\right| \leqslant J_1 + J_2 + J_3.$$

Величины J_1 и J_3 можно сделать сколь угодно малыми, если выбрать A и B достаточно большими по абсолютной величине и притом такими, чтобы точки A и B были точками непрерывности функции $F(x)$, а n выбрать достаточно большим. В самом деле, пусть M — верхняя грань

$|f(x)|$ при $-\infty < x < \infty$; тогда

$$J_1 \leqslant M[F(A)+F_n(A)],$$

$$J_3 \leqslant M[F(+\infty)-F(B)]+M[F_n(+\infty)-F_n(B)].$$

Но

$$\lim_{A\to-\infty} F(A) = 0, \qquad \lim_{B\to\infty} F(B) = F(+\infty).$$

А так как, по предположению,

$$\lim_{n\to\infty} F_n(A) = F(A), \qquad \lim_{n\to\infty} F_n(B) = F(B),$$

то наше утверждение об J_1 и J_3 доказано. Величина J_2 при достаточно большом n может быть сделана сколь угодно малой в силу теоремы Хелли для конечного интервала.

Теорема доказана.

§ 35. Предельные теоремы для характеристических функций

Важнейшими с точки зрения приложений характеристических функций к выводу асимптотических формул теории вероятностей являются две предельные теоремы — прямая и обратная. Эти теоремы устанавливают, что соответствие, существующее между функциями распределения и характеристическими функциями, не только взаимно однозначно, но и непрерывно.

Прямая предельная теорема. *Если последовательность функций распределения*

$$F_1(x), F_2(x), \ldots, F_n(x), \ldots$$

сходится в основном к функции распределения $F(x)$, то последовательность характеристических функций

$$f_1(t), f_2(t), \ldots, f_n(t), \ldots$$

сходится к характеристической функции $f(t)$. Эта сходимость равномерна в каждом конечном интервале t.

Доказательство. Так как

$$f_n(t) = \int e^{itx} dF_n(x), \quad f(t) = \int e^{itx} dF(x)$$

и функция e^{itx} непрерывна и ограничена на всей прямой $-\infty < t < \infty$, то согласно обобщенной второй теореме Хелли при $n\to\infty$

$$f_n(t) \to f(t).$$

Утверждение, что эта сходимость равномерна в каждом конечном интервале t, проверяется буквально теми же рассуждениями, какие мы провели при доказательстве второй теоремы Хелли.

Обратная предельная теорема. *Если последовательность характеристических функций*

$$f_1(t), f_2(t), \ldots, f_n(t), \ldots \tag{1}$$

сходится к непрерывной функции $f(t)$, то последовательность функций распределения

$$F_1(x),\ F_2(x), \ldots, F_n(x), \ldots \tag{2}$$

сходится в основном к некоторой функции распределения $F(x)$ (в силу прямой предельной теоремы $f(t) = \int e^{itx} dF(x)$).

Доказательство. На основании первой теоремы Хелли заключаем, что последовательность (2) непременно содержит подпоследовательность

$$F_{n_1}(x), F_{n_2}(x), \ldots, F_{n_k}(x), \ldots, \tag{3}$$

сходящуюся в основном к некоторой неубывающей функции $F(x)$. При этом понятно, что функцию $F(x)$ мы можем считать непрерывной слева:

$$\lim_{x' \to x - 0} F(x') = F(x).$$

Вообще говоря, функция $F(x)$ может и не быть функцией распределения, так как для этого должны удовлетворяться еще условия $F(-\infty) = 0$ и $F(+\infty) = 1$. Действительно, для последовательности функций

$$F_n(x) = \begin{cases} 0 & \text{при} \quad x \leqslant -n, \\ 1/2 & \text{при} \quad -n < x \leqslant n, \\ 1 & \text{при} \quad x > n, \end{cases}$$

предельная функция $F(x) \equiv 1/2$ и, следовательно, $F(-\infty)$ и $F(+\infty)$ также равны $1/2$. Однако в условиях нашей теоремы, как будет сейчас показано, обязательно будет $F(-\infty) = 0$ и $F(+\infty) = 1$.

В самом деле, если бы это было не так, то, приняв во внимание, что для предельной функции $F(x)$ должно быть $F(-\infty) \geqslant 0$ и $F(+\infty) \leqslant 1$, мы имели бы:

$$\delta = F(+\infty) - F(-\infty) < 1.$$

Возьмем теперь какое-нибудь положительное число ϵ, меньшее $1 - \delta$. Так как по условию теоремы последовательность характеристических функций (1) сходится к функции $f(t)$, то $f(0) = 1$. А так как, сверх того, функция $f(t)$ непрерывна, то можно выбрать достаточно малое положительное число τ такое, что будет иметь место неравенство

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)dt\right| > 1 - \epsilon/2 > \delta + \epsilon/2. \tag{4}$$

Но в то же время можно выбрать $X > 4/\tau\epsilon$ и настолько большое K, чтобы при $k > K$ было

$$\delta_k = F_{n_k}(X) - F_{n_k}(-X) < \delta + \epsilon/4.$$

Так как $f_{n_k}(t)$ есть характеристическая функция, то

$$\int_{-\tau}^{\tau} f_{n_k}(t)dt = \int \left[\int_{-\tau}^{\tau} e^{itx}dt\right] dF_{n_k}(x).$$

Интеграл, стоящий в правой части этого равенства, можно оценить следующим способом. С одной стороны, так как $|e^{itx}| = 1$, то

$$\left|\int_{-\tau}^{\tau} e^{itx}dt\right| \leqslant 2\tau.$$

С другой стороны,

$$\int_{-\tau}^{\tau} e^{itx}dt = \frac{2}{x}\sin\tau x,$$

и так как $|\sin\tau x| \leqslant 1$, то при $|x| > X$

$$\left|\int_{-\tau}^{\tau} e^{itx}dt\right| < \frac{2}{X}.$$

Отсюда, применив первую оценку при $|x| \leqslant X$ и вторую при $|x| > X$, получаем:

$$\left|\int_{-\tau}^{\tau} f_{n_k}(t)dt\right| \leqslant \left|\int_{|x| \leqslant X}\left(\int_{-\tau}^{\tau} e^{itx}dt\right) dF_{n_k}(x)\right| +$$

$$+\left|\int_{|x| > X}\left(\int_{-\tau}^{\tau} e^{itx}dt\right) dF_{n_k}(x)\right| < 2\tau\delta_k + \frac{2}{X}$$

и, следовательно,

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f_{n_k}(t)dt\right| < \delta + \frac{\epsilon}{2}.$$

Это неравенство сохраняется и в пределе

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)dt\right| \leqslant \delta + \frac{\epsilon}{2},$$

что, очевидно, противоречит неравенству (4).

Таким образом, функция $F(x)$, к которой сходится в основном последовательность $F_{n_k}(x)$, есть функция распределения; согласно прямой предельной теореме ее характеристическая функция равна $f(t)$. Чтобы закончить доказательство теоремы, нам остается доказать, что и вся последовательность (2) сходится в основном к функции $F(x)$. Предположим, что это не так. Тогда найдется подпоследовательность функций

$$F_{n'_1}(x),\ F_{n'_2}(x), \ldots, F_{n'_k}(x), \ldots, \tag{5}$$

сходящаяся в основном к некоторой функции $F^*(x)$, отличной от $F(x)$ по крайней мере в одной из точек непрерывности. По уже доказанному $F^*(x)$ должна быть функцией распределения с характеристической функцией $f(t)$. По теореме единственности должно быть

$$F^*(x) = F(x).$$

Это противоречит сделанному предположению.

Заметим, что условия теоремы выполнены в каждом из двух следующих случаев:

1) Последовательность характеристических функций $f_n(t)$ сходится к некоторой функции $f(t)$ равномерно в каждом конечном интервале t.

2) Последовательность характеристических функций $f_n(t)$ сходится к характеристической функции $f(t)$.

Пример. В качестве примера использования предельных теорем рассмотрим доказательство интегральной теоремы Муавра–Лапласа.

В примере 4 § 32 мы нашли характеристическую функцию случайной величины $\eta = \dfrac{\mu - np}{\sqrt{npq}}$:

$$f_n(t) = \left(qe^{-it\sqrt{\frac{p}{nq}}} + pe^{it\sqrt{\frac{q}{np}}} \right)^n.$$

Воспользовавшись разложением в ряд Маклорена, находим, что

$$qe^{-it\sqrt{\frac{p}{nq}}} + pe^{it\sqrt{\frac{q}{np}}} = 1 - \frac{t^2}{2n}(1 + R_n),$$

где

$$R_n = 2 \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{it}{\sqrt{n}} \right)^{k-2} \frac{pq^k + q(-p)^k}{\sqrt{(pq)^k}}.$$

Так как при $n \to \infty$

$$R_n \to 0,$$

то

$$f_n(t) = \left[1 - \frac{t^2}{2n}(1 + R_n)\right]^n \to e^{-t^2/2}.$$

В силу обратной предельной теоремы отсюда вытекает, что при любом x

$$\mathsf{P}\left\{\frac{\mu - np}{\sqrt{npq}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz,$$

когда $n \to \infty$.

Из непрерывности предельной функции легко вывести, что эта сходимость будет равномерна по x.

§ 36. Положительно определенные функции

Цель настоящего параграфа – дать исчерпывающее описание класса характеристических функций. Приводимая нами ниже основная теорема была одновременно найдена А.Я. Хинчиным и С. Бохнером и опубликована впервые С. Бохнером.

Для формулировки и доказательства этой теоремы нам нужно ввести новое понятие. Мы скажем, что непрерывная функция $f(t)$ вещественного аргумента t *положительно определена* в интервале $-\infty < t < \infty$, если каковы бы ни были вещественные числа $t_1, t_2, \ldots, t_n$, комплексные числа $\xi_1, \xi_2, \ldots, \xi_n$ и целое число n

$$\sum_{k=1}^{n} \sum_{j=1}^{n} f(t_j - t_k)\, \xi_j \bar{\xi}_k \geqslant 0. \tag{1}$$

Перечислим несколько простейших свойств положительно определенных функций.

1. $f(0) \geqslant 0$. В самом деле, положим $n = 1, t_1 = 0, \xi_1 = 1$; тогда из условия положительной определенности функции $f(t)$ находим, что

$$\sum_{k=1}^{n} \sum_{j=1}^{n} f(t_k - t_j)\, \xi_k \bar{\xi}_j = f(0) \geqslant 0.$$

2. При любых вещественных t

$$f(-t) = \overline{f(t)}.$$

Для доказательства положим в (1) $n = 2$, $t_1 = 0$, $t_2 = t$, ξ_1, ξ_2 произвольны. Имеем по предположению

$$0 \leqslant \sum_{k=1}^{2} \sum_{j=1}^{2} f(t_k - t_j)\,\xi_k \overline{\xi_j} = f(0-0)\,\xi_1\overline{\xi}_1 + f(0-t)\,\xi_1\overline{\xi}_2 +$$

$$+ f(t-0)\,\xi_2\overline{\xi}_1 + f(t-t)\,\xi_2\overline{\xi}_2 =$$

$$= f(0)(|\xi_1|^2 + |\xi_2|^2) + f(-t)\,\xi_1\overline{\xi}_2 + f(t)\,\overline{\xi}_1\xi_2, \tag{2}$$

поэтому величина

$$f(-t)\,\xi_1\overline{\xi}_2 + f(t)\,\overline{\xi}_1\xi_2$$

должна быть вещественной. Таким образом, если положить $f(-t) = \alpha_1 + i\beta_1$, $f(t) = \alpha_2 + i\beta_2$, $\xi_1\overline{\xi}_2 = \gamma + i\delta$, $\overline{\xi}_1\xi_2 = \gamma - i\delta$, то должно быть

$$\alpha_1\delta + \beta_1\gamma - \alpha_2\delta + \beta_2\gamma = 0.$$

Так как ξ_1 и ξ_2, а следовательно, γ и δ произвольны, то должно быть

$$\alpha_1 - \alpha_2 = 0, \quad \beta_1 + \beta_2 = 0.$$

Отсюда следует наше утверждение.

3. При любых вещественных t

$$|f(t)| \leqslant f(0).$$

Положим в неравенстве (2) $\xi_1 = f(t)$, $\xi_2 = -|f(t)|$; тогда согласно предыдущему

$$2f(0)\,|f(t)|^2 - |f(t)|^2\,|f(t)| - |f(t)|^2\,|f(t)| \geqslant 0.$$

Отсюда при $|f(t)| \neq 0$ получаем:

$$f(0) \geqslant |f(t)|.$$

Если же $|f(t)| = 0$, то опять-таки в силу свойства 1

$$f(0) \geqslant |f(t)|.$$

Из доказанного следует, между прочим, что если положительно определенная функция такова, что $f(0) = 0$, то $f(t) \equiv 0$.

Теорема Бохнера – Хинчина. *Для того чтобы непрерывная функция $f(t)$, удовлетворяющая условию $f(0) = 1$, была характеристической, необходимо и достаточно, чтобы она была положительно определенной.*

Доказательство. В одном направлении теорема тривиальна. Действительно, если

$$f(t) = \int e^{ixt}\,dF(x),$$

где $F(x)$ – некоторая функция распределения, то при любом целом n, произвольных вещественных $t_1, t_2, \ldots, t_n$ и комплексных числах $\xi_1, \xi_2, \ldots, \xi_n$ имеем:

$$\sum_{k=1}^{n} \sum_{j=1}^{n} f(t_k - t_j)\, \xi_k \overline{\xi_j} = \sum_{k=1}^{n} \sum_{j=1}^{n} \left\{ \int e^{ix(t_k - t_j)}\, dF(x) \right\} \xi_k \overline{\xi_j} =$$

$$= \int \sum_{k=1}^{n} \sum_{j=1}^{n} e^{ix(t_k - t_j)}\, \xi_k \overline{\xi_j}\, dF(x) =$$

$$= \int \left(\sum_{k=1}^{n} e^{it_k x}\, \xi_k \right) \left(\sum_{j=1}^{n} e^{-it_j x}\, \overline{\xi_j} \right) dF(x) =$$

$$= \int \left| \sum_{k=1}^{n} e^{it_k x}\, \xi_k \right|^2 dF(x) \geqslant 0.$$

Доказательство достаточности условий теоремы требует более сложных рассуждений.

Рассмотрим последовательность чисел $f\left(\dfrac{k}{n}\right)$, зависящую от целочисленного параметра n. В силу положительной определенности функции $f(x)$ мы имеем при любом N:

$$\mathcal{P}_N^{(n)}(x) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} f\left(\frac{k-j}{n}\right) e^{-i(k-j)x} \geqslant 0.$$

Легко подсчитать, что в этой сумме имеется $N - |r|$ слагаемых, для которых разность $k - j$ равна r. Далее очевидно, что число r может изменяться от $-N + 1$ до $N - 1$. Таким образом, имеет место равенство

$$\mathcal{P}_N^{(n)}(x) = \sum_{r=-N}^{N} \left(1 - \frac{|r|}{N}\right) f\left(\frac{r}{n}\right) e^{-irx}.$$

Умножим обе части полученного равенства на e^{isx} и проинтегрируем по x в пределах от $-\pi$ до π:

$$\int_{-\pi}^{\pi} e^{isx} \mathcal{P}_N^{(n)}(x)\,dx = \sum_{r=-N}^{N} \left(1 - \frac{|r|}{N}\right) f\left(\frac{r}{n}\right) \int_{-\pi}^{\pi} e^{-irx} e^{isx}\, dx,$$

Известно, что

$$\int_{-\pi}^{\pi} e^{-i(r-s)x}\, dx = \begin{cases} 0 \text{ для } r \neq s, \\ 2\pi \text{ для } r = s. \end{cases}$$

Поэтому

$$\left(1-\frac{|s|}{N}\right)f\left(\frac{s}{n}\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{P}_N^{(n)}(x)\,e^{isx}\,dx=\int_{-\pi}^{\pi}e^{isx}\,dF_N^{(n)}(x),$$

где

$$F_N^{(n)}(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{P}_N^{(n)}(x)\,dx$$

есть неубывающая функция с полным изменением, равным

$$F_N^{(n)}(\pi)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{P}_N^{(n)}(x)\,dx=f(0)=1,$$

т.е. является функцией распределения.

На основании первой теоремы Хелли мы можем найти последовательность $N_k\to\infty$ при $k\to\infty$, для которой функции $F_N^{(n)}(x)$ (n фиксировано!) сходятся к предельной; обозначим ее через $F^{(n)}(x)$. Функция $F^{(n)}(x)$ снова является функцией распределения, так как при любых N и произвольном $\epsilon>0$ $F_N^{(n)}(-\pi-\epsilon)=0$, $F_N^{(n)}(\pi+\epsilon)=1$ и, значит, при произвольном $\epsilon>0$ также

$$F^{(n)}(-\pi-\epsilon)=0,\quad F^{(n)}(\pi+\epsilon)=1.$$

Согласно второй теореме Хелли

$$\lim_{k\to\infty}\int_{-\pi}^{\pi}e^{isx}\,dF_{N_k}^{(n)}(x)=\int_{-\pi}^{\pi}e^{isx}\,dF^{(n)}(x).$$

Таким образом*),

$$f\left(\frac{s}{n}\right)=\int_{-\pi}^{\pi}e^{isx}\,dF^{(n)}(x)$$

при всех целочисленных s ($s=0,\pm1,\pm2,\ldots$).

---

*) Заметим, что попутно нами доказана следующая теорема Герглотца. *Если последовательность чисел c_n ($n=0,\pm1,\ldots$) обладает тем свойством, что при любом выборе комплексных чисел $\xi_1,\xi_2,\ldots,\xi_N$ и произвольном N*

$$\sum_{k=1}^{N}\sum_{j=1}^{N}c_{k-j}\,\xi_k\,\overline{\xi_j}\geqslant 0,$$

то последовательность c_n может быть записана в форме

$$c_n=\int_{-\pi}^{\pi}e^{inx}\,d\sigma(x),$$

где $\sigma(x)$ – неубывающая функция с ограниченной вариацией.

Рассмотрим теперь последовательность характеристических функций $f_n(t)$, определенных посредством равенства

$$f_n(t) = \int_{-\pi n}^{\pi n} e^{itx}\, dF_n(x),$$

где

$$F_n(x) = F^{(n)}\left(\frac{x}{n}\right).$$

Легко проверить, что при всех целочисленных k

$$f_n\left(\frac{k}{n}\right) = f\left(\frac{k}{n}\right). \tag{3}$$

Но каково бы ни было t, мы можем подобрать такую последовательность $k = k(n, t)$ *), что

$$0 \leqslant t - \frac{k}{n} < \frac{1}{n}.$$

Из непрерывности функций $f(t)$ следует, что

$$f(t) = \lim_{n\to\infty} f\left(\frac{k}{n}\right) = \lim_{n\to\infty} f_n\left(\frac{k}{n}\right). \tag{4}$$

Если мы докажем, что при всех вещественных t

$$f(t) = \lim_{n\to\infty} f_n(t), \tag{5}$$

то доказательство теоремы будет завершено, так как $f(t)$ – непрерывная функция и поэтому, в силу обратной предельной теоремы для характеристических функций, будет характеристической функцией.

С этой целью заметим, что из (3) и (4) следует равенство

$$\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty}\left\{\left[f_n(t) - f_n\left(\frac{k}{n}\right)\right] + f_n\left(\frac{k}{n}\right)\right\} =$$

$$= f(t) + \lim_{n\to\infty}\left[f_n(t) - f_n\left(\frac{k}{n}\right)\right]. \tag{6}$$

---

*) Всюду дальше под k мы понимаем числа $k(n, t)$.

Обозначим $\theta = t - \dfrac{k}{n}$. Согласно выбору величин k имеем $0 \leqslant \theta < 1/n$.

По определению функции $f_n(t)$

$$\left| f_n(t) - f_n\left(\frac{k}{n}\right) \right| = \left| \int_{-\pi n}^{\pi n} e^{i\frac{k}{n}x} (e^{i\theta x} - 1)\, dF_n(x) \right| \leqslant$$

$$\leqslant \int_{-\pi n}^{\pi n} | e^{i\theta x} - 1 |\, dF_n(x). \tag{7}$$

Воспользовавшись неравенством Буняковского, находим, что

$$\int_{-\pi n}^{\pi n} | e^{i\theta x} - 1 |\, dF_n(x) \leqslant \sqrt{\int_{-\pi n}^{\pi n} | e^{i\theta x} - 1 |^2\, dF_n(x)} =$$

$$= \sqrt{\int_{-\pi n}^{\pi n} 2(1 - \cos\theta x)\, dF_n(x)} = \sqrt{2(1 - Rf_n(\theta))}, \tag{8}$$

где символ $Rf_n(\theta)$ означает вещественную часть $f_n(\theta)$. Так как $\cos z \leqslant \cos\alpha z$ при $0 \leqslant \alpha < 1$ и $-\pi \leqslant z < \pi$, то

$$1 - Rf_n(\theta) = \int_{-\pi n}^{\pi n} (1 - \cos\theta x)\, dF_n(x) = \int_{-\pi}^{\pi} (1 - \cos\theta \cdot nz)\, dF_n(zn) \leqslant$$

$$\leqslant \int_{-\pi}^{\pi} (1 - \cos z)\, dF_n(zn).$$

А так как

$$F_n(zn) = F^{(n)}(z),$$

то

$$1 - Rf_n(\theta) \leqslant \int_{-\pi}^{\pi} (1 - \cos z)\, dF^{(n)}(z) = 1 - R \int_{-\pi}^{\pi} e^{iz}\, dF^{(n)}(z).$$

Отсюда в силу (3) находим, что

$$1 - Rf_n(\theta) \leqslant 1 - Rf\left(\frac{1}{n}\right). \tag{9}$$

Собрав вместе неравенства (7), (8) и (9), находим, что

$$\left| f_n(t) - f_n\left(\frac{k}{n}\right) \right| \leqslant \sqrt{2\left(1 - Rf\left(\frac{1}{n}\right)\right)}.$$

Из непрерывности функции $f(t)$ отсюда следует, что

$$\lim_{n\to\infty}\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]=0.$$

Соотношение (5), как это видно теперь из (6), доказано.

§ 37. Характеристические функции многомерных случайных величин

В настоящем параграфе мы излагаем без доказательств основные сведения о характеристических функциях многомерных случайных величин.

Характеристической функцией n-мерной случайной величины $(\xi_1, \xi_2, \dots, \xi_n)$ называется математическое ожидание величины $e^{i(t_1\xi_1+t_2\xi_2+\dots+t_n\xi_n)}$, где $t_1, t_2, \dots, t_n$ – вещественные переменные

$$f(t_1, t_2, \dots, t_n)=\mathbf{M}\exp\left(i\sum_{k=1}^{n}t_k\xi_k\right). \tag{1}$$

Если $F(x_1, x_2, \dots, x_n)$ есть функция распределения величины $(\xi_1, \xi_2, \dots, \xi_n)$, то, как мы знаем из предыдущего*),

$$f(t_1, t_2, \dots, t_n)=\int\dots\int\left(\exp i\sum_{k=1}^{n}t_k x_k\right)dF(x_1, \dots, x_n). \tag{2}$$

Подобно тому как и в одномерном случае, характеристическая функция n-мерной случайной величины равномерно непрерывна во всем пространстве $(-\infty < t_j < +\infty,\ 1\leqslant j\leqslant n)$ и удовлетворяет следующим соотношениям:

$$f(0, 0, \dots, 0)=1,$$

$$|f(t_1, t_2, \dots, t_n)|\leqslant 1\quad(-\infty < t_k < +\infty,\ k=1, 2, \dots),$$

$$f(-t_1, -t_2, \dots, -t_n)=\overline{f(t_1, t_2, \dots, t_n)}.$$

По характеристической функции $f(t_1, t_2, \dots, t_n)$ случайной величины $(\xi_1, \xi_2, \dots \xi_n)$ легко найти характеристическую функцию любой k-мерной $(k < n)$ величины $(\xi_{j1}, \xi_{j2}, \dots, \xi_{jk})$, компонентами которой являются величины ξ_s $(1\leqslant s\leqslant n)$. Для этого в формуле (2) нужно положить равными нулю все аргументы t_s при $s\neq j_r (1\leqslant r\leqslant k)$. Так, например, характеристическая функция величины ξ_1 равна

$$f_1(t_1)=f(t_1, 0, \dots, 0).$$

---

*) Ср. теорему 1 § 24 и замечание о многомерных интегралах Стилтьеса в § 23.

Из определения вытекает, что если компоненты величины $(\xi_1, \xi_2, \ldots, \xi_n)$ являются независимыми случайными величинами, то ее характеристическая функция равна произведению характеристических функций компонент

$$f(t_1, t_2, \ldots, t_n) = f(t_1) \cdot f(t_2) \cdot \ldots \cdot f(t_n).$$

Так же как и в одномерном случае, многомерные характеристические функции позволяют легко находить моменты различных порядков.

Так, например,

$$\mathbf{M}\, \xi_1^{k_1} \xi_2^{k_2} \ldots \xi_n^{k_n} = \iint \ldots \int x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}\, dF(x_1, x_2, \ldots, x_n) =$$

$$= (i)^{\sum\limits_1^n k_j} \left[\frac{\partial^{k_1 + k_2 + \ldots + k_n} f(t_1, t_2, \ldots, t_n)}{\partial t_1^{k_1} \partial t_2^{k_2} \ldots \partial t_n^{k_n}} \right]_{t_1 = t_2 = \ldots = t_n = 0}$$

Для вычисления характеристических функций полезно знать следующую теорему, доказательство которой без труда проведет читатель.

Теорема 1. *Если характеристическая функция величины* $(\xi_1, \xi_2, \ldots, \xi_n)$ *равна* $f(t_1, t_2, \ldots, t_n)$, *то характеристическая функция величины* $(\sigma_1 \xi_1 + a_1, \sigma_2 \xi_2 + a_2, \ldots, \sigma_n \xi_n + a_n)$, *где* a_i *и* σ_i $(1 \leqslant j \leqslant n)$ *– вещественные постоянные, равна*

$$\exp\left(i \sum_{n=1}^{n} a_k t_k\right) \cdot f(\sigma_1 t_1, \sigma_2 t_2, \ldots, \sigma_n t_n).$$

Пример 1. Вычислим характеристическую функцию двумерной случайной величины, распределенной по нормальному закону:

$$p(x, y) = \frac{1}{2\pi(1 - r^2)} \exp\left\{ - \frac{1}{2(1 - r^2)} [x^2 - 2rxy + y^2] \right\}. \tag{3}$$

По формуле (2)

$$f(t_1, t_2) = \frac{1}{2\pi(1 - r^2)} \iint e^{i(t_1 x + t_2 y)} p(x, y)\, dx\, dy.$$

Заменой переменных мы можем привести $f(t_1, t_2)$ к виду

$$f(t_1, t_2) = e^{-\frac{1}{2}(t_1^2 + 2rt_1t_2 + t_2^2)} \frac{1}{2\pi} \iint e^{-\frac{1}{2}(u^2 + v^2)}\, du\, dv =$$

$$= e^{-\frac{1}{2}(t_1^2 + 2rt_1t_2 + t_2^2)}$$

Пример 2. Применяя теорему 1, мы найдем характеристическую функцию величины (η_1, η_2), распределенной по нормальному закону:

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2(1-r^2)} \times$$
$$\times \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2} - 2r\frac{(x-a)(y-b)}{\sigma_1^2\sigma_2^2} + \frac{(y-b)^2}{\sigma_2^2}\right]\right\}. \quad (4)$$

Если мы положим $\eta_1 = \sigma_1\xi_1 + a$, $\eta_2 = \sigma_2\xi_2 + b$, то величина (ξ_1, ξ_2) будет распределена по закону (3). Согласно теореме 1 характеристическая функция величины (η_1, η_2) равна

$$\varphi(t_1, t_2) = \exp\left[iat_1 + iat_2 - \frac{1}{2}(\sigma_1^2 t_1^2 + 2\sigma_1\sigma_2 r t_1 t_2 + \sigma_2^2 t_2^2)\right].$$

Из определения характеристической функции вытекает следующая

Теорема 2. *Если* $f(t_1, t_2, \ldots, t_n)$ *есть характеристическая функция величины* $(\xi_1, \xi_2, \ldots, \xi_n)$, *то характеристическая функция суммы* $\xi_1 + \xi_2 + \ldots + \xi_n$ *равна*

$$f(t) = f(t, t, \ldots, t).$$

Примечание. Заметим, что

$$f(t) = f(tt_1, tt_2, \ldots, tt_n)$$

есть характеристическая функция суммы $t_1\xi_1 + t_2\xi_2 + \ldots + t_n\xi_n$.

Пример 3. Применим теорему 2 к определению функции распределения суммы $\eta_1 + \eta_2$, если (η_1, η_2) распределена по закону (4).

Согласно теореме 2 характеристическая функция суммы $\eta_1 + \eta_2$ равна

$$f(t) = \exp\left[it(a+b) - \frac{t^2}{2}(\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2)\right].$$

Мы знаем (пример 1 § 32), что это — характеристическая функция нормального закона с математическим ожиданием, равным $a + b$, и дисперсией, равной $\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2$. Этот результат был нами получен ранее непосредственно (§ 21, пример 2).

Важно заметить, что в многомерном случае сохраняется следующая теорема.

Теорема 3. *Функция распределения* $F(x_1, x_2, \ldots, x_n)$ *однозначно определяется своей характеристической функцией.*

Доказательство этого предложения основывается на формуле обращения.

Теорема 4. *Если* $f(t_1, t_2, \ldots, t_n)$ *– характеристическая функция, а* $F(x_1, x_2, \ldots, x_n)$ *– функция распределения случайной величины*

$(\xi_1, \xi_2, \ldots, \xi_n)$, *то*

$$\mathsf{P}\{a_k \leqslant \xi_k < b_k,\ k = 1, 2, \ldots, n\} =$$

$$= \lim_{T\to\infty} \frac{1}{(2\pi)^n} \int_{-T}^{T} \int_{-T}^{T} \ldots \frac{e^{it_k a_k} - e^{it_k b_k}}{it_k} \times$$

$$\times f(t_1, \ldots, t_n)\, dt_1 dt_2 \ldots dt_n,$$

где a_k и b_k – любые вещественные числа, удовлетворяющие единственному требованию: вероятность попадания на поверхность параллелепипеда $a_k \leqslant \xi_k < b_k$ $(k = 1, 2, \ldots, n)$ равна нулю.

Точно так же, как и в одномерном случае, имеют место прямая и обратная предельные теоремы для характеристических функций. Мы не будем на этом останавливаться.

П р и м е р 3. Говорят, что n-мерная случайная величина $(\xi_1, \xi_2, \ldots, \xi_n)$ имеет *невырожденное* (собственное) *n-мерное нормальное распределение,* если ее плотность распределения имеет вид

$$p(x_1, x_2, \ldots, x_n) = Ce^{-\frac{1}{2} Q(x_1, x_2, \ldots, x_n)},$$

где

$$Q(x_1, x_2, \ldots, x_n) = \sum_{i,j} b_{ij}(x_i - a_i)(x_j - a_j)$$

– положительно определенная квадратичная форма, C, a_i и b_{ij} – действительные постоянные.

Несложные подсчеты показывают*), что

$$C = (\sqrt{2\pi})^{-n} \sqrt{D},$$

где

$$D = \begin{vmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{vmatrix}.$$

---

*) Обычный прием при подобного рода подсчетах заключается в том, что заменой переменных приводят форму Q к сумме квадратов и все вычисления производят в новых переменных.

Обозначим через D_{ij} минор D, соответствующий элементу b_{ij}, тогда

$$\mathbf{M}\,\xi_j = a_j,\quad \sigma_j^2 = \mathbf{D}\,\xi_j = \frac{D_{jj}}{D}\quad (j = 1, 2, \ldots, n),$$

$$r_{ij} = \frac{\mathbf{M}\,(\xi_i - a_i)(\xi_j - a_j)}{\sigma_i\,\sigma_j} = \frac{D_{ij}}{\sqrt{D_{ii}D_{jj}}}\quad (i, j = 1, 2, \ldots, n).$$

Определитель D и главные миноры его положительны.

Обычными подсчетами легко проверить, что характеристическая функция величины $(\xi_1, \xi_2, \ldots, \xi_n)$ равна

$$f(t_1, t_2, \ldots, t_n) = e^{\,i\sum\limits_{j=1}^{n} a_j t_j - \frac{1}{2}\sum\limits_{k=1}^{n}\sum\limits_{j=1}^{n}\sigma_j\sigma_k r_{jk} t_j t_k}$$

Таким образом, *n-мерное нормальное распределение вполне определяется заданием математического ожидания и дисперсии.*

Из выражения для характеристической функции n-мерной нормально распределенной случайной величины мы видим, что распределение величины

$$(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k})$$

при любых $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ будет k-мерным нормальным распределением.

§ 38. Преобразование Лапласа – Стилтьеса

Для случайных величин, которые принимают только неотрицательные значения, во многих случаях предпочтительнее, чем характеристическими функциями, пользоваться преобразованиями Лапласа – Стилтьеса. Пусть случайная величина ξ – неотрицательная и $F(x)$ – ее функция распределения. *Преобразованием Лапласа – Стилтьеса* для $F(x)$ называется интеграл

$$f(s) = \int_{-0}^{\infty} e^{-sx}dF(x). \tag{1}$$

Преобразование Лапласа – Стилтьеса обладает рядом полезных свойств, в значительной мере повторяющих свойства характеристических функций.

1. Преобразование Лапласа – Стилтьеса является аналитической функцией в правой полуплоскости; для него нечетные производные отрицательны, а четные – положительны;

2. $f(0) = 1$,

3. Функция распределения однозначно определяется по своему преобразованию Лапласа – Стилтьеса;

4. Чтобы последовательность функций распределения сходилась в каждой точке непрерывности, необходимо и достаточно чтобы последовательность их преобразований Лапласа – Стилтьеса сходилась равномерно в каждом конечном отрезке аргумента;

5. Преобразование Лапласа – Стилтьеса суммы независимых случайных величин равно произведению преобразований Лапласа – Стилтьеса слагаемых;

6. $f^{(n)}(0) = (-1)^n \int\limits_{-0}^{\infty} x^n dF(x)$.

Приведем преобразования Лапласа – Стилтьеса для некоторых распределений.

1. *Единичное распределение*: $F(x) = 0$ при $x \leqslant a$ и $F(x) = 1$ при $x > a$;

$f(t) = e^{-as}$.

2. *Показательное распределение*: $F(x) = 1 - e^{-\lambda x}$ при $x \geqslant 0$;

$$f(s) = \frac{\lambda}{\lambda + s}.$$

3. *Гамма-распределение*: $F'(x) = \dfrac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}$, α и β – положительные постоянные;

$$f(s) = \frac{\beta^\alpha}{(\beta + s)^\alpha}.$$

4. *Распределение Пуассона*: $p_k = \dfrac{\lambda^k e^{-\lambda}}{k!}$, $k = 0, 1, 2, \ldots$

$f(s) = e^{-\lambda(1 - e^{-s})}$.

5. *Геометрическое распределение*: $p_k = \mathsf{P}\{\xi = k\} = qp^k$, $k = 0, 1, \ldots$

$$f(s) = \frac{q}{1 - pe^{-s}}, \quad q = 1 - p.$$

Проиллюстрируем использование теории преобразований Лапласа – Стилтьеса на примере одной технической задачи.

Мы начнем изложение с одной прикладной задачи. Она позволит нам понять важность развития теории суммирования случайного числа слагаемых для прикладных областей знания, а, следовательно, и для теоретической науки.

В современном инженерном деле одним из важнейших свойств технических систем принято считать их высокую надежность, т.е. способность длительное время выполнять без ущерба для дела положенные рабочие функции. Для увеличения надежности технических устройств используется широкий спектр мер, в том числе так называемое резервирование с восстановлением.

Предположим, что A_1 представляет собой или техническую систему в целом или какой-нибудь ее элемент. Для увеличения ее надежности мы подключаем точно такое же устройство A_2, которое принимает на себя рабочие функции в момент, когда A_1 исчерпывает свой рабочий ресурс (отказ элемента A_1). В момент отказа A_1 немедленно 1) A_1 отправляется на ремонт (восстановление рабочих функций) и 2) подключается в работу элемент A_2, который и берет на себя всю рабочую нагрузку. Спрашивается, какой эффект дает резервирование (введение дополнительного элемента A_2) и восстановление отказавших элементов, если после ремонта восстановленный элемент немедленно передается в резерв? Для решения этой задачи нам придется сделать некоторые предположения. А именно мы предположим, что

1) длительность безотказной работы представляет собой случайную величину с функцией распределения $F(x)$;

2) оба элемента A_1 и A_2 обладают одинаковыми техническими данными и после ремонта полностью восстанавливают свои рабочие свойства:

3) длительность восстановления η является случайной величиной с распределением $G(x)$;

4) отказавший элемент после отказа немедленно начинает ремонтироваться; после восстановления элемент немедленно переходит в резерв.

Обозначим через ζ длительность безотказной работы пары элементов A_1 и A_2 и через $\Phi(x)$ ее функцию распределения. Заметим, что под безотказной работой пары элементов мы понимаем период от начала работы до момента, когда оба элемента окажутся в состоянии отказа. Наша ближайшая задача состоит в том, чтобы вывести уравнения для неизвестной функции $\Phi(x)$. Нам будет удобно искать не функцию $\Phi(x)$, а функцию $\overline{\Phi}(x) = 1 - \Phi(x)$. И вообще для функции распределения $F_\xi(x)$ введем обозначение $\overline{F}_\xi(x) = 1 - F_\xi(x) = \mathsf{P}\{\xi \geqslant x\}$.

Событие $\zeta \geqslant x$ может наступить двумя принципиально различными способами: во-первых, может случиться, что элемент A_1 сам проработает время большее, чем x, а, во-вторых, A_1 может отказать до момента x, но система, состоящая из A_1 (направленного на ремонт) и A_2 (взявшего на себя нагрузку), проработает безотказно оставшееся до момента x время. Обозначим через $\overline{\omega}(u)$ вероятность того, что только что описанная система проработает безотказно время, большее u. Собрав все вместе, мы получаем следующее равенство:

$$\bar{\Phi}(x) = \bar{F}(x) + \int_0^x \bar{\omega}(x-z)\,dF(z). \tag{2}$$

Полученного равенства для решения стоящей перед нами задачи недостаточно, поскольку при его составлении мы ввели дополнительную неизвестную вероятность ω. Для ее определения составим новое уравнение. Найдем $\bar{\omega}(x)$. Здесь также может случиться, что интересующее нас событие может наступить двумя различными путями: описанная нами система проработает больше, чем время x, поскольку элемент A_2 проработает время большее x; элемент A_2 откажет до момента x, но до этого момента будет восстановлен элемент A_1 и система, состоящая из отказавшего элемента A_2 и вступившего в работу отремонтированного элемента A_1, проработает по меньшей мере до момента x. Сказанное приводит нас к равенству

$$\bar{\omega}(x) = \bar{F}(x) + \int_0^x \bar{\omega}(x-z)\,G(z)\,dF(z). \tag{3}$$

Для решения полученных уравнений мы воспользуемся преобразованиями Лапласа – Стилтьеса. С этой целью введем в рассмотрение преобразования

$$\varphi(s) = \int_0^\infty e^{-sx}\,d\Phi(x),\quad f(s) = \int_0^\infty e^{-sx}\,dF(x),$$

$$\tilde{\omega}(s) = \int_0^\infty e^{-sx}\,d\omega(x),\quad g(s) = \int_0^\infty e^{-sx}\,G(x)\,dF(x).$$

В терминах этих преобразований уравнения (2) и (3) принимают следующий вид:

$$\varphi(s) = f(s)\,\tilde{\omega}(s),\qquad \tilde{\omega}(s) = f(s) - g(s) + \tilde{\omega}(s)\,g(s)$$

Из них находим

$$\varphi(s) = f(s)\,\frac{f(s)-g(s)}{1-g(s)}\;. \tag{4}$$

Формально задача решена, поскольку мы нашли преобразование Лапласа – Стилтьеса распределения $\Phi(x)$ и, значит, по формуле обращения можем найти и саму функцию $\Phi(x)$. Однако полученный результат позволяет получить многочисленные следствия, на которых мы остановимся

Мы сейчас ограничимся определением средней длительности безотказной работы системы с резервом. Из формулы (4) находим

$$\frac{\varphi'(s)}{\varphi(s)} = \frac{f'(s)}{f(s)} + \frac{f'(s)-g'(s)}{f(s)-g(s)} + \frac{g'(s)}{1-g(s)} .$$

Отсюда, положив $s = 0$, приходим к равенству

$$\varphi'(0) = -f'(0)\left(1 + \frac{1}{1-g(0)}\right).$$

Но мы знаем, что

$$T = \mathbf{M}\,\zeta = -\varphi'(0), \quad a = \int_0^\infty x\,dF(x) = -f'(0)$$

и что

$$\alpha = 1 - g(0) = \int_0^\infty \bar{G}(x)\,dF(x)$$

есть ни что иное как вероятность того, что длительность восстановления окажется больше, чем длительность безотказной работы элемента. Таким образом,

$$T = a\left(1 + \frac{1}{\alpha}\right). \tag{5}$$

Из этой формулы мы видим, что среднюю длительность безотказной работы дублированной системы можно повысить двумя различными путями:

1) увеличивать среднюю длительность безотказной работы элементов,

2) уменьшать продолжительность ремонта.

Как правило, первый способ достается большой ценой, тогда как второй – нуждается только в хорошей организации работ. К тому же он приносит, как это видно из формулы, более ощутимые результаты; особенно, если $\alpha \approx 0$, т.е. если длительность восстановления оказывается, как правило, меньше длительности безотказной работы элемента.

Предположим теперь, что мы совершенствуем процесс восстановления и на n-й стадии достигаем функции распределения $G_n(x)$ такой, что $\alpha_n \to 0$, не обращаясь при этом в 0.

Докажем, что если математическое ожидание длительности безотказной работы конечно и равно a, а $\alpha_n \to 0$, то имеет место следующее предельное соотношение:

$$\mathbf{P}\left\{\frac{\zeta_n}{T_n} > x\right\} \to e^{-x} \quad \text{при} \quad x \geqslant 0 \qquad (n \to \infty).$$

Иными словами мы докажем, что асимптотически длительность безотказной работы рассматриваемой системы двух элементов имеет показательное распределение.

Согласно сформулированным нами свойствам преобразований Лапласа – Стилтьеса нам следует показать, что преобразование Лапласа – Стилтьеса для величины $\frac{\zeta_n}{T_n}$ стремится к $\frac{1}{1+s}$. С этой целью преобразуем выражение

$$\varphi_n\left(\frac{s}{T_n}\right) = f\left(\frac{s}{T_n}\right)\frac{f\left(\frac{s}{T_n}\right) - g_n\left(\frac{s}{T_n}\right)}{1 - g_n\left(\frac{s}{T_n}\right)}$$

следующим образом:

$$\varphi_n\left(\frac{s}{T_n}\right) = f\left(\frac{s}{T_n}\right)\frac{1}{1 + \frac{1 - f\left(\frac{s}{T_n}\right)}{\frac{s}{T_n}} \cdot \frac{s}{T_n\left(f\left(\frac{s}{T_n}\right) - g\left(\frac{s}{T_n}\right)\right)}}.$$

Но очевидно, что

$$\frac{1 - f\left(\frac{s}{T_n}\right)}{\frac{s}{T_n}} = -\frac{f\left(\frac{s}{T_n}\right) - f(0)}{\frac{s}{T_n}} \to -f'(0) = a \qquad (n \to \infty).$$

Ясно, что

$$T_n\left[f\left(\frac{s}{T_n}\right) - g\left(\frac{s}{T_n}\right)\right] \sim \frac{a}{\alpha_n}\left[f\left(\frac{s}{T_n}\right) - g\left(\frac{s}{T_n}\right)\right] =$$

$$= a\left[1 - 1 + \frac{f\left(\frac{s}{T_n}\right) - g\left(\frac{s}{T_n}\right)}{\alpha_n}\right].$$

Теперь

$$1-\frac{f\left(\frac{s}{T_n}\right)-g\left(\frac{s}{T_n}\right)}{\alpha_n}=\frac{1}{\alpha_n}\int_0^\infty \left(1-e^{-\frac{sx}{T_n}}\right)(1-G_n(x))\,dF(x).$$

Оценим интеграл, стоящий в правой части последнего соотношения. С этой целью разобьем его на два слагаемых:

$$\left(\int_0^{\sqrt{T_n}}+\int_{\sqrt{T_n}}^\infty\right)\left(1-e^{-\frac{sx}{T_n}}\right)(1-G_n(x))\,dF(x).$$

В первом слагаемом воспользуемся неравенством $1-e^{-x}\leqslant x$, а во втором – неравенством $1-e^{-x}\leqslant 1$. В результате получим, что

$$\int_0^{\sqrt{T_n}}\left(1-e^{-\frac{sx}{T_n}}\right)(1-G_n(x))\,dF(x)\leqslant$$

$$\leqslant\frac{s\sqrt{T_n}}{T_n}\int_0^\infty (1-G_n(x))\,dF(x)=o(\alpha_n),$$

$$\int_{\sqrt{T_n}}^\infty\left(1-e^{-\frac{sx}{T_n}}\right)(1-G_n(x))\,dF(x)\leqslant\int_{\sqrt{T_n}}^\infty (1-G_n(x))\,dF(x)=o(\alpha_n).$$

Собрав все оценки вместе, окончательно получаем

$$\varphi_n\left(\frac{s}{T_n}\right)=\frac{1}{1+s}(1+o(1)),$$

что и требовалось доказать.

Упражнения

1. Доказать, что функции

$$f_1(t)=\sum_{k=0}^\infty a_k\cos kt\; f_2(t)=\sum_{k=0}^\infty a_k e^{i\lambda_k t},$$

где $a_k\geqslant 0$, $\sum\limits_{k=0}^\infty a_k=1$, являются характеристическими. Определить соответствующие распределения вероятностей.

2. Найти характеристические функции для следующих плотностей вероятностей:

а) $p(x) = \frac{a}{2} e^{-a|x|}$;

б) $p(x) = \frac{a}{\pi(a^2 + x^2)}$;

в) $p(x) = \begin{cases} 0 & \text{при} \quad |x| \geqslant a, \\ \frac{a - |x|}{a^2} & \text{при} \quad |x| < a; \end{cases}$

г) $p(x) = \frac{2\sin^2 \frac{ax}{2}}{\pi a x^2}$.

Замечание. Внимательный читатель заметит, что примеры а) и б), а также в) и г) являются, так сказать, обратными.

3. Доказать, что функции

$$\varphi_1(t) = \frac{1}{\operatorname{ch} t}, \quad \varphi_2(t) = \frac{1}{\operatorname{sh} t}, \quad \varphi_3(t) = \frac{1}{\operatorname{ch}^2 t}$$

являются характеристическими соответственно для плотностей распределения

$$p_1(x) = \frac{1}{2\operatorname{ch}\frac{\pi x}{2}}, \quad p_2(x) = \frac{\pi}{4\operatorname{ch}^2\frac{\pi x}{2}}, \quad p_3(x) = \frac{x}{2\operatorname{sh}\frac{\pi x}{2}}$$

4. Найти распределения вероятностей случайных величин, характеристические функции которых равны

а) $\cos t$, б) $\cos^2 t$, в) $\frac{a}{a + it}$, г) $\frac{\sin at}{at}$

5. Доказать, что функция, определяемая равенствами

$$f(t) = f(-t), \quad f(t + 2a) = f(t),$$

$$f(t) = \frac{a - t}{a} \text{ при } \quad 0 \leqslant t \leqslant a,$$

является характеристической.

Замечание. Характеристические функции примеров 2г и 5 обладают следующим замечательным свойством:

$f_2(t) = f_5(t)$ при $|t| \leqslant a$,

$f_2(t) \neq f_5(t)$ при $|t| > a$ и $t \neq \pm 2a, \ldots$

Таким образом, существуют характеристические функции, значения которых совпадают в сколь угодно большом сегменте $(-a, a)$ и не равные тождественно. Первый пример таких двух характеристических функций был указан Б.В. Гнеденко; затем М.Г. Крейн указал необходимые и достаточные условия, при которых из равенства двух характеристических функций в каком-либо сегменте $(-a, a)$ следует их тождественное равенство.

6. Доказать, что можно найти такие независимые случайные величины ξ_1, ξ_2, ξ_3, что распределения ξ_2 и ξ_3 различны, а функции распределения сумм $\xi_1 + \xi_2$ и $\xi_1 + \xi_3$ одинаковы.

Указание. Воспользоваться результатами примеров 2а и 5.

7. Доказать, что если $f(t)$ является характеристической функцией, то функция

$$f(t) = \begin{cases} f(t) & \text{при} \quad |t| \leqslant a, \\ f(t+2a) & \text{при} \quad -\infty < t \leqslant \infty \end{cases}$$

также является характеристической.

Указание: Воспользоваться теоремой Бохнера – Хинчина.

8. Доказать, что если $f(t)$ является характеристической функцией, то функция

$$\varphi(t) = e^{f(t) - 1}$$

также является характеристической.

9. Доказать, что если функция $f(t)$ является характеристической, то функция

$$\varphi(t) = \frac{1}{t} \int_0^t f(t)\,dt$$

также является характеристической.

10. Доказать, что для любой вещественной характеристической функции $f(t)$ имеет место неравенство

$$1 - f(2t) \leqslant 4(1 - f(t)),$$

а, значит, для любой характеристической функции – неравенство

$$1 - |f(2t)|^2 \leqslant 4(1 - |f(t)|^2).$$

11. Доказать, что для любой вещественной характеристической функции имеет место неравенство

$$1 + f(2t) \geqslant 2[f(t)]^2.$$

12. Доказать, что если $F(x)$ – функция распределения, а $f(t)$ – ее характеристическая функция, то при любом значении x верно равенство

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)\, e^{-itx}\,dt = F(x+0) - F(x-0).$$

13. Доказать, что есть $F(x)$ – функция распределения, а $f(t)$ – ее характеристическая функция и x_k – абсциссы скачков $F(x)$, то

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |f(t)|^2\,dt = \sum_k [F(x_k+0) - F(x_k-0)].$$

14. Доказать, что если случайная величина имеет плотность распределения, то ее характеристическая функция при $t \to \infty$ стремится к 0.

15. Случайная величина ξ распределена по закону Пуассона; $\mathbf{M}\xi = \lambda$. Доказать, что при $\lambda \to \infty$ распределение величины $\dfrac{\xi - \lambda}{\sqrt{\lambda}}$ стремится к нормальному с параметрами $a = 0$ и $\sigma^2 = 1$.

16. Случайная величина ξ имеет плотность распределения

$$p(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ \dfrac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-\beta x} & \text{при} \quad x > 0. \end{cases}$$

Доказать, что при $\alpha \to \infty$ распределение величины $\dfrac{\beta\xi - \alpha}{\sqrt{\alpha}}$ сходится к нормальному с параметрами $a = 0$, $\sigma^2 = 1$.

З а м е ч а н и е. Результаты упр. 15 и 16 позволяют при вычислении вероятностей $\mathbf{P}\{a \leqslant \xi < b\}$ для больших значений λ соответственно α) использовать таблицы нормального распределения. В частности, оказывается, что для распределения χ^2 уже при $n \geqslant 30$ указанное предельное распределение дает прекрасную точность. Это последнее замечание постоянно используется в статистике.

17. Доказать, что если $\varphi(t)$ – характеристическая функция и функция $\psi(t)$ такова, что для некоторой последовательности $\{h_n\}$ ($h_n \to \infty$ при $n \to \infty$) произведения $f_n(t) = \varphi(t)\,\psi(h_n t)$ также являются характеристическими функциями, то функция $\psi(t)$ – характеристическая.

ГЛАВА 8

КЛАССИЧЕСКАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

§ 39. Постановка задачи

Интегральная предельная теорема Муавра–Лапласа, доказанная нами в главе 2, послужила источником большого цикла исследований, имеющих фундаментальное значение как для самой теории вероятностей, так и для ее многочисленных приложений в естествознании, технических и экономических науках. Для того, чтобы составить себе представление о направлении этих исследований, мы придадим теореме Муавра–Лапласа несколько иную форму. А именно, если, как это мы неоднократно делали, через μ_k обозначить число появлений события A в k-м испытании, то число появлений события A в n последовательных испытаниях равно $\sum_{k=1}^{n} \mu_k$. Далее, в примере 3 § 25 мы подсчитали, что $\mathbf{M} \sum_{k=1}^{n} \mu_k = np$ и $\mathbf{D} \sum_{k=1}^{n} \mu_k = npq$. Поэтому теорема Муавра–Лапласа может быть записана в таком виде: при $n \to \infty$

$$\mathbf{P}\left\{a \leqslant \frac{\sum\limits_{k=1}^{n} (\mu_k - \mathbf{M}\mu_k)}{\sqrt{\sum\limits_{k=1}^{n} \mathbf{D}\mu_k}} < b\right\} \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2}\, dz \tag{1}$$

и словами сформулирована так: вероятность того, что сумма уклонений независимых случайных величин, принимающих два значения 0 и 1 с вероятностями, соответственно равными q и $p = 1 - q$ $(0 < p < 1)$, от их математических ожиданий, деленная на квадратный корень из суммы дисперсий слагаемых, будет заключаться в пределах от a до b, при увеличении числа слагаемых до бесконечности, равномерно относительно a и b стремится к интегралу $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2}\, dz$.

Естественно возникает вопрос: насколько тесно связано соотношение (1) со специальным выбором слагаемых μ_k, не будет ли оно иметь место и при более слабых ограничениях, наложенных на функции распределения слагаемых? Постановка этой задачи, а также ее решение являются в основном делом П.Л. Чебышева и его учеников А.А. Маркова и А.М. Ляпунова. Их исследования показали, что на слагаемые следует наложить лишь самые общие ограничения, смысл которых состоит в том, что отдельные слагаемые должны оказывать незначительное влияние на сумму. В следующем параграфе мы дадим точную формулировку этого условия. Причины, в силу которых эти результаты приобрели огромное значение в приложениях, лежат в самом существе массовых явлений, изучение закономерностей которых, как мы говорили ранее, и составляет предмет теории вероятностей.

Одной из важнейших схем, по которой идет использование результатов теории вероятностей в естествознании и технике, состоит в следующем. Считают, что процесс протекает под влиянием большого числа независимо действующих случайных факторов, каждый из которых лишь ничтожно мало изменяет течение явления или процесса. Исследователь, интересующийся изучением процесса в целом, а не действием отдельных факторов, наблюдает лишь суммарное действие этих факторов. Приведем два типичных примера.

Пример 1. Пусть производится некоторое измерение. На результат неизбежно действует большое количество факторов, порождающих ошибки в измерении. Сюда относятся ошибки, вызванные состоянием измерительного прибора, которое может нечувствительно изменяться под влиянием различных атмосферных или механических причин. Сюда относятся личные ошибки наблюдателя, вызванные особенностями его зрения или слуха и также могущие незначительно изменяться в зависимости от психического или физического состояния наблюдателя, и т.д. Каждый из этих факторов породил бы ничтожную ошибку. Но на измерении сказываются сразу все эти ошибки, наблюдается "суммарная ошибка". Иначе говоря, фактически наблюдаемая ошибка измерения будет случайной величиной, являющейся суммой огромного числа ничтожных по величине и независимых между собой случайных величин. И хотя эти последние неизвестны, так же как неизвестны их функции распределения, их влияние на результаты измерений заметно и поэтому должно быть подвергнуто изучению.

Пример 2. В процессе массового производства, существующего во многих отраслях промышленности, изготовляются большие партии одинаковых предметов. Обратим внимание на какую-нибудь числовую характеристику интересующего нас продукта. Поскольку это изделие находится в соответствии с техническими нормами, существует некоторая нормальная величина избранной нами характеристики. В действительности же всегда наблюдается некоторое отклонение от этой нормальной величины. При правильно поставленном процессе производства такие отклонения могут вызываться лишь случайными причинами, каждая из которых производит лишь незаметный эффект. Суммарное же их действие производит заметное уклонение от нормы.

Подобных примеров можно привести сколько угодно.

Таким образом, возникает задача изучения закономерностей, свойственных суммам большого числа независимых случайных величин, каждая из которых оказывает лишь малое влияние на сумму. Этому последнему требованию мы придадим позднее более точный смысл. Вместо того чтобы изучать суммы очень большого, но конечного числа слагаемых, мы будем рассматривать последовательность сумм со все большим и большим числом слагаемых и считать, что решения интересующих нас задач даются предельными функциями распределения для последовательности функций распределения сумм. Такого рода переход от конечной постановки задачи к предельной является обычным как для современной математики, так и для многих отделов естествознания.

Итак, мы пришли к рассмотрению следующей задачи: дана последовательность взаимно независимых случайных величин

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots,$$

о которых мы предположим, что они имеют конечные математические ожидания и дисперсии. В дальнейшем мы станем придерживаться следующих обозначений:

$$a_k = \mathbf{M}\xi_k, \quad b_k^2 = \mathbf{D}\xi_k, \quad B_n^2 = \sum_{k=1}^{n} b_k^2 = \mathbf{D} \sum_{k=1}^{n} \xi_k.$$

Спрашивается, какие условия нужно наложить на величины ξ_k, чтобы функции распределения сумм

$$\frac{1}{B_n} \sum_{k=1}^{n} (\xi_k - a_k) \tag{2}$$

сходились к нормальному закону распределения? В следующем параграфе мы увидим, что для этого достаточно выполнения *условия Линдеберга:* при любом $\tau > 0$

$$\lim_{n \to \infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-a_k| > \tau B_n} (x - a_k)^2 dF_k(x) = 0,$$

где $F_k(x)$ обозначает функцию распределения величины ξ_k.

Выясним смысл этого условия.

Обозначим через A_k событие, состоящее в том, что

$$|\xi_k - a_k| > \tau B_n \quad (k = 1, 2, \ldots, n)$$

и оценим вероятность

$$\mathsf{P}\{\max_{1\leqslant k\leqslant n}|\xi_k - a_k| > \tau B_n\}.$$

Так как

$$\mathsf{P}\{\max_{1\leqslant k\leqslant n}|\xi_k - a_k| > \tau B_n\} = \mathsf{P}\{A_1 + A_2 + \ldots + A_n\}$$

и

$$\mathsf{P}\{A_1 + A_2 + \ldots + A_n\} \leqslant \sum_{k=1}^{n} \mathsf{P}\{A_k\},$$

то, заметив, что

$$\mathsf{P}\{A_k\} = \int_{|x-a_k|>\tau B_n} dF_k(x) \leqslant \frac{1}{(\tau B_n)^2}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_k(x)$$

находим неравенство

$$\mathsf{P}\{\max_{1\leqslant k\leqslant n}|\xi_k - a_k| \geqslant \tau B_n\} \leqslant \frac{1}{\tau^2 B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_k(x).$$

В силу условия Линдеберга, каково бы ни было постоянное $\tau > 0$, последняя сумма при $n \to \infty$ стремится к нулю. Таким образом, условие Линдеберга представляет собой своеобразное требование равномерной малости слагаемых $\frac{1}{B_n}(\xi_k - a_k)$ в сумме (2).

Отметим еще раз, что смысл условий, достаточных для сходимости функции распределения сумм (2) к нормальному закону, был вполне выяснен уже исследованиями А.А. Маркова и А.М. Ляпунова.

§ 40. Теорема Линдеберга

Мы начнем с доказательства достаточности условия Линдеберга.

Теорема. *Если последовательность взаимно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ при любом постоянном $\tau > 0$ удовлетворяет условию Линдеберга*

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_k(x) = 0, \tag{1}$$

то при $n \to \infty$ равномерно относительно x

$$\mathbf{P}\left\{\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k - a_k) < x\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}\,dz.$$

Д о к а з а т е л ь с т в о. Для краткости введем обозначения

$$\xi_{nk} = \frac{\xi_k - a_k}{B_n}, \quad F_{nk}(x) = \mathbf{P}\{\xi_{nk} < x\}.$$

Очевидно, что

$$\mathbf{M}\xi_{nk} = 0, \quad \mathbf{D}\xi_{nk} = \frac{1}{B_n^2}\mathbf{D}\xi_k$$

и, следовательно,

$$\sum_{k=1}^{n}\mathbf{D}\xi_{nk} = 1. \tag{2'}$$

Легко убедиться, что условие Линдеберга в этих обозначениях принимает следующий вид:

$$\lim_{n\to\infty}\sum_{k=1}^{n}\int_{|x|>\tau} x^2\,dF_{nk}(x) = 0. \tag{1'}$$

Характеристическая функция суммы

$$\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k - a_k) = \sum_{k=1}^{n}\xi_{nk}$$

равна

$$\varphi_n(t) = \prod_{k=1}^{n} f_{nk}(t).$$

Нам нужно доказать, что

$$\lim_{n\to\infty}\varphi_n(t) = e^{-t^2/2}.$$

С этой целью мы установим прежде всего, что множители $f_{nk}(t)$ при $n\to\infty$ равномерно относительно $k\,(1 \leqslant k \leqslant n)$ стремятся к 1. Действительно, принимая во внимание равенство $\mathbf{M}\,\xi_{nk} = 0$, находим, что

$$f_{nk}(t) - 1 = \int (e^{itx} - 1 - itx)\,dF_{nk}(x).$$

Так как при любом вещественном α*)

$$|e^{i\alpha} - 1 - i\alpha| \leqslant \alpha^2/2, \tag{3}$$

то

$$|f_{nk}(t) - 1| \leqslant \frac{t^2}{2} \int x^2 dF_{nk}(x).$$

Пусть ϵ — произвольное положительное число; тогда очевидно, что

$$\int x^2 dF_{nk}(x) = \int\limits_{|x| \leqslant \epsilon} x^2 dF_{nk}(x) + \int\limits_{|x| > \epsilon} x^2 dF_{nk}(x) \leqslant \epsilon^2 + \int\limits_{|x| > \epsilon} x^2 dF_{nk}(x).$$

Последнее слагаемое может быть, согласно (1'), при достаточно больших n сделано меньше, чем ϵ^2. Таким образом, для всех достаточно больших n равномерно относительно $k (1 \leqslant k \leqslant n)$ и t в любом конечном интервале $|t| \leqslant T$

$$|f_{nk}(t) - 1| \leqslant \epsilon^2 T^2.$$

Отсюда мы заключаем, что равномерно относительно $k (1 \leqslant k \leqslant n)$

$$\lim_{n \to \infty} f_{nk}(t) = 1 \tag{4}$$

и что для всех достаточно больших n при t, лежащих в произвольном конечном интервале $|t| \leqslant T$, выполняется неравенство

$$|f_{nk}(t) - 1| < 1/2. \tag{5}$$

Мы можем, следовательно, в интервале $|t| \leqslant T$ написать разложение

---

*) Это неравенство и целую серию ему подобных можно вывести хотя бы следующим путем. Из того, что

$$|e^{i\alpha} - 1| = \left|\int_0^\alpha e^{ix} dx\right| \leqslant \alpha \qquad (\alpha > 0)$$

вытекает неравенство

$$|e^{i\alpha} - 1 - i\alpha| = \left|\int_0^\alpha (e^{ix} - 1) dx\right| \leqslant \frac{\alpha^2}{2}.$$

Из последнего неравенства далее следует, что

$$\left|e^{i\alpha} - 1 - i\alpha + \frac{\alpha^2}{2}\right| = \left|\int_0^\alpha (e^{ix} - 1 - ix) dx\right| \leqslant \int_0^\alpha |e^{ix} - 1 - ix| dx \leqslant \int_0^\alpha \frac{x^2}{2} dx = \frac{\alpha^3}{6}, \tag{3'}$$

и т.д.

(ln обозначает *главное значение* логарифма).

$$\ln \varphi_n(t) = \sum_{k=1}^{n} \ln f_{nk}(t) = \sum_{k=1}^{n} \ln [1 + (f_{nk}(t) - 1)] =$$

$$= \sum_{k=1}^{n} (f_{nk}(t) - 1) + R_n, \tag{6}$$

где

$$R_n = \sum_{k=1}^{n} \sum_{s=2}^{\infty} \frac{(-1)^{s-1}}{s} (f_{nk}(t) - 1)^s.$$

В силу (5)

$$|R_n| \leqslant \sum_{k=1}^{n} \sum_{s=2}^{\infty} \frac{1}{2} |f_{nk}(t) - 1|^s =$$

$$= \frac{1}{2} \sum_{k=1}^{n} \frac{|t_{nk}(t) - 1|^2}{1 - |f_{nk}(t) - 1|} \leqslant \sum_{k=1}^{n} |f_{nk}(t) - 1|^2.$$

Так как

$$\sum_{k=1}^{n} |f_{nk}(t) - 1| = \sum_{k=1}^{n} \left| \int (e^{itx} - 1 - itx)\, dF_{nk}(x) \right| \leqslant$$

$$\leqslant \frac{t^2}{2} \sum_{k=1}^{n} \int x^2 dF_{nk}(x) = \frac{t^2}{2},$$

то

$$|R_n| \leqslant \frac{t^2}{2} \max_{1 \leqslant k \leqslant n} |f_{nk}(t) - 1|.$$

Из (4) вытекает, что равномерно относительно t в произвольном конечном интервале $|t| \leqslant T$, при $n \to \infty$

$$R_n \to 0. \tag{7}$$

Но

$$\sum_{k=1}^{n} (f_{nk}(t) - 1) = -\frac{t^2}{2} + \rho_n, \tag{8}$$

где

$$\rho_n = \frac{t^2}{2} + \sum_{k=1}^{n} \int (e^{itx} - 1 - itx)\, dF_{nk}(x).$$

Пусть $\epsilon > 0$ произвольно; тогда в силу (2′)

$$\rho_n = \sum_{k=1}^{n} \int_{|x| \leqslant \epsilon} \left(e^{itx} - 1 - itx - \frac{(itx)^2}{2}\right) dF_{nk}(x) +$$

$$+ \sum_{k=1}^{n} \int_{|x| > \epsilon} \left(\frac{t^2 x^2}{2} + e^{itx} - 1 - itx\right) dF_{nk}(x).$$

Неравенства (3) и (3′) позволяют получить следующую оценку:

$$|\rho_n| \leqslant \frac{|t|^3}{6} \sum_{k=1}^{n} \int_{|x| \leqslant \epsilon} |x|^3 dF_{nk}(x) + t^2 \sum_{k=1}^{n} \int_{|x| > \epsilon} x^2 dF_{nk}(x) \leqslant$$

$$\leqslant \frac{|t|^3}{6} \cdot \epsilon \sum_{k=1}^{n} \int_{|x| \leqslant \epsilon} x^2 dF_{nk}(x) + t^2 \sum_{k=1}^{n} \int_{|x| > \epsilon} x^2 dF_{nk}(x) =$$

$$= \frac{|t|^3}{6} \cdot \epsilon + t^2 \left(1 - \frac{|t|}{6} \epsilon\right) \sum_{k=1}^{n} \int_{|x| > \epsilon} x^2 dF_{nk}(x).$$

Согласно условию (1) второе слагаемое при любом $\epsilon > 0$ может быть сделано меньше любого $\eta > 0$, лишь бы n было достаточно большим. А так как $\epsilon > 0$ произвольно, то мы можем его выбрать настолько малым, чтобы, каковы бы ни были $\eta > 0$ и T, для всех t, заключенных в интервале $|t| \leqslant T$, выполнялось неравенство

$$|\rho_n| < 2\eta \qquad (n \geqslant n_0(\epsilon, \eta, T)).$$

Это неравенство показывает, что равномерно в каждом конечном интервале значений t

$$\lim_{n \to \infty} \rho_n = 0. \tag{9}$$

Собрав вместе соотношения (6), (7), (8) и (9), мы получаем окончательно, что равномерно в каждом конечном интервале t

$$\lim_{n \to \infty} \ln \varphi_n(t) = -t^2/2.$$

Теорема доказана.

Следствие. *Если независимые случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ одинаково распределены и имеют конечную отличную от нуля дисперсию, то при $n \to \infty$ равномерно по x*

$$\mathbf{P}\left\{\frac{1}{B_n} \sum_{k=1}^{n} (\xi_k - \mathbf{M}\xi_k) < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} dz.$$

Доказательство. Нам достаточно проверить, что при сделанных предположениях выполнено условие Линдеберга. С этой целью заметим, что в нашем случае

$$B_n = b\sqrt{n},$$

где b обозначает дисперсию отдельного слагаемого. Положив $\mathbf{M}\xi_k = a$, мы можем написать следующие очевидные равенства:

$$\sum_{k=1}^{n} \frac{1}{B_n^2} \int_{|x-a|>\tau B_n} (x-a)^2 dF_k(x) =$$

$$= \frac{1}{nb^2}\, n \int_{|x-a|>\tau B_n} (x-a)^2 dF_1(x) = \frac{1}{b^2} \int_{|x-a|>\tau B_n} (x-a)^2 dF_1(x).$$

В силу предположения о конечности дисперсии и ее положительности заключаем, что интеграл, стоящий в правой части последнего равенства, стремится к нулю, когда $n \to \infty$.

Теорема Ляпунова. *Если для последовательности взаимно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ можно подобрать такое положительное число $\delta > 0$, что при $n \to \infty$*

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{M}\,|\xi_k - a_k|^{2+\delta} \to 0, \tag{10}$$

то при $n \to \infty$ равномерно по x

$$\mathbf{P}\left\{\frac{1}{B_n} \sum_{k=1}^{n} (\xi_k - a_k) < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz.$$

Доказательство. Нам снова достаточно проверить, что условие Ляпунова (условие (10)) влечет за собой выполнение условия Линдеберга. Но это ясно из следующей цепочки неравенств:

$$\frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 dF_k(x) \leqslant$$

$$\leqslant \frac{1}{B_n^2(\tau B_n)^\delta} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} |x-a_k|^{2+\delta} dF_k(x) \leqslant$$

$$\leqslant \frac{1}{\tau^\delta}\, \frac{\sum_{k=1}^{n} \int |x-a_k|^{2+\delta} dF_k(x)}{B_n^{2+\delta}}.$$

§ 41. Локальная предельная теорема

Мы приведем теперь достаточные условия для применения другой классической предельной теоремы — *локальной теоремы.* При этом мы ограничимся рассмотрением только случая взаимно независимых слагаемых, имеющих одно и то же распределение вероятностей.

Условимся говорить, что дискретная случайная величина ξ имеет *решетчатое распределение,* если существуют такие числа a и $h > 0$, что все возможные значения ξ могут быть представлены в виде $a + kh$, где параметр k может принимать любые целые значения $(-\infty < k < \infty)$.

К решетчатым относятся, например, распределения Пуассона, Бернулли и др.

Выразим теперь условие решетчатости распределения случайной величины ξ в терминах характеристических функций. С этой целью докажем следующую лемму.

Л е м м а. *Для того чтобы случайная величина ξ имела решетчатое распределение, необходимо и достаточно, чтобы при некотором $t \neq 0$ модуль ее характеристической функции был равен единице.*

Д о к а з а т е л ь с т в о. Действительно, если ξ распределена решетчато и p_k есть вероятность равенства $\xi = a + kh$, то характеристическая функция величины ξ равна

$$f(t) = \sum_{k=-\infty}^{\infty} p_k e^{it(a+kh)} = e^{iat} \sum_{k=-\infty}^{\infty} p_k e^{itkh}.$$

Отсюда находим, что

$$f\left(\frac{2\pi}{h}\right) = e^{2\pi i \frac{a}{h}} \sum_{k=-\infty}^{\infty} p_k e^{2\pi ik} = e^{2\pi i \frac{a}{h}}.$$

Мы видим, таким образом, что для каждого решетчатого распределения

$$\left| f\left(\frac{2\pi}{h}\right) \right| = 1.$$

Предположим теперь, что при некотором $t_1 \neq 0$

$$|f(t_1)| = 1,$$

и докажем, что при этом ξ имеет решетчатое распределение. Последнее равенство означает, что при некотором θ

$$f(t_1) = e^{i\theta}.$$

Таким образом,

$$\int e^{it_1 x} dF(x) = e^{i\theta}$$

и, следовательно,

$$\int e^{i(t_1 x-\theta)}dF(x)=1.$$

Отсюда вытекает, что

$$\int \cos(t_1 x-\theta)\,dF(x)=1.$$

Для того чтобы это равенство было возможно, необходимо, чтобы функция $F(x)$ могла расти только при тех значениях x, при которых

$$\cos(t_1 x-\theta)=1.$$

Это означает, что возможные значения ξ должны быть вида

$$x=\frac{\theta}{t_1}+k\frac{2\pi}{t_1},$$

что и требовалось доказать.

Число h мы будем называть *шагом распределения*.

Шаг распределения h *максимален*, если ни при каких $b(-\infty<b<\infty)$ и $h_1>h$ нельзя представить все возможные значения ξ в виде $b+kh_1$.

Для иллюстрации различия между понятиями шага распределения и максимального шага распределения рассмотрим такой пример. Пусть ξ может принимать в качестве своих значений все нечетные числа. Очевидно, что все значения ξ могут быть записаны в виде $a+kh$, где $a=0, h=1$. Шаг h, однако, не будет максимальным, так как все возможные значения ξ можем записать также в виде $b+kh_1$, где $b=1, h_1=2$.

Условия максимальности шага распределения можно выразить и в других терминах.

В о-п е р в ы х, шаг распределения будет максимальным тогда и только тогда, когда общий наибольший делитель попарных разностей возможных значений величины ξ, поделенных на h, равен единице.

В о-в т о р ы х, шаг распределения h максимален тогда и только тогда, когда модуль характеристической функции в промужетке $0<|t|<2\pi/h$ меньше единицы и при $t=2\pi/h$ равен единице.

Последнее утверждение немедленно вытекает из только что доказанной леммы. В самом деле, если при $0<t_1<2\pi/h$

$$|f(t_1)|=1,$$

то согласно доказанному величина $2\pi/t_1$ должна быть шагом распределения, а так как $h<2\pi/t_1$, то шаг h не может быть максимальным.

Отсюда мы можем сделать такой вывод: если h — максимальный шаг распределения, то для каждого $\epsilon>0$ найдется такое число $c_0>0$, что при всех t в интервале $\epsilon\leqslant|t|\leqslant 2\pi/h-\epsilon$ имеет место неравенство

$$|f(t)|\leqslant e^{-c_0}. \tag{1}$$

Пусть теперь случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ взаимно независимы, решетчато распределены и имеют одну и ту же функцию распределения $F(x)$. Рассмотрим сумму

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n.$$

Очевидно, что она также является решетчатой случайной величиной и возможные ее значения могут быть записаны в виде $na + kh$. Обозначим через $P_n(k)$ вероятность равенства

$$\zeta_n = na + kh;$$

в частности, $P_1(k) = \mathbf{P}\{\xi_1 = a + kh\} = p_k$.

Обозначим далее

$$z_{nk} = \frac{an + kh - A_n}{B_n},$$

где $A_n = \mathbf{M}\zeta_n$, $B_n^2 = \mathbf{D}\zeta_n = n\mathbf{D}\xi_1$.

Мы можем теперь доказать следующее предложение, очевидным образом обобщающее локальную предельную теорему Муавра–Лапласа.

Т е о р е м а. *Пусть независимые решетчатые случайные величины*

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

имеют одну и ту же функцию распределения $F(x)$ и их математические ожидания и дисперсии конечны. Тогда для того чтобы равномерно относительно k $(-\infty < k < \infty)$ при $n \to \infty$ имело место соотношение

$$\frac{B_n}{h} P_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z_{nk}^2}{2}} \to 0,$$

необходимо и достаточно, чтобы шаг распределения h был максимальным.

Д о к а з а т е л ь с т в о. Необходимость условия теоремы почти очевидна. Действительно, если шаг h не максимален, то возможные значения суммы $\zeta_n = \sum_{k=1}^{n} \xi_k$ будут содержать систематические пропуски: разность между ближайшими возможными значениями суммы не может быть меньше dh, где d есть общий наибольший делитель разностей возможных значений ζ_n, деленных на h. Если h – не максимальный шаг, то $d > 1$ при всех значениях n.

Доказательство достаточности условия теоремы требует несколько более сложных рассуждений.

Характеристическая функция величины $\xi_k (k = 1, 2, 3, \ldots)$ равна

$$f(t) = \sum_{k=-\infty}^{\infty} p_k e^{iat+itkh} = e^{iat} \sum_{k=-\infty}^{\infty} p_k e^{itkh},$$

а характеристическая функция суммы ζ_n есть

$$f^n(t) = e^{iant} \sum_{k=-\infty}^{\infty} P_n(k)\, e^{itkh}.$$

Умножив последнее равенство на $e^{-iant-itkh}$ и проинтегрировав его в пределах от $-\pi/h$ до π/h, находим, что

$$\frac{2\pi}{h} P_n(k) = \int_{-\pi/h}^{\pi/h} f^n(t)\, e^{-iant-itkh}\, dt.$$

Заметив, что

$$hk = B_n z_{nk} + A_n - an$$

(вместо z_{nk} мы будем дальше писать z), можем написать

$$\frac{2\pi}{h} P_n(k) = \int_{-\pi/h}^{\pi/h} f^{*n}(t)\, e^{-itzB_n}\, dt,$$

где обозначено

$$f^*(t) = e^{-\frac{itA_n}{n}} f(t).$$

Положив, наконец, $x = tB_n$, находим окончательно;

$$\frac{2\pi B_n}{h} P_n(k) = \int_{-\pi B_n/h}^{\pi B_n/h} e^{-izx} f^{*n}\left(\frac{x}{B_n}\right) dx.$$

Легко подсчитать, что

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \frac{1}{2\pi} \int e^{-izx-x^2/2}\, dx.$$

Представим разность

$$R_n = 2\pi \left[\frac{B_n}{h} P_n(k) - \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\right]$$

в виде суммы четырех интегралов

$$R_n = J_1 + J_2 + J_3 + J_4,$$

где

$$J_1 = \int_{-A}^{A} e^{-izx}\left[f^{*n}\left(\frac{x}{B_n}\right) - e^{-x^2/2}\right] dx,$$

$$J_2 = -\int_{|x|>A} e^{-izx-x^2/2}\, dx,$$

$$J_3 = \int_{\epsilon B_n \leqslant |x| \leqslant \frac{\pi B_n}{h}} e^{-izx} f^{*n}\left(\frac{x}{B_n}\right) dx,$$

$$J_4 = \int_{A \leqslant |x| < \epsilon B_n} e^{-izx} f^{*n}\left(\frac{x}{B_n}\right) dx,$$

где $A > 0$ – достаточно большое, а $\epsilon > 0$ – достаточно малое постоянные числа, более точные значения которых будут выбраны нами позднее.

В силу следствия из теоремы, доказанной в предыдущем параграфе, в любом конечном интервале значений t равномерно относительно t выполняется соотношением

$$f^{*n}\left(\frac{t}{B_n}\right) \to e^{-t^2/2} \quad (n \to \infty).$$

Но отсюда следует, что каково бы ни было постоянное A,

$$J_1 \to 0 \quad (n \to \infty).$$

Интеграл J_2 оценивается посредством неравенства

$$|J_2| \leqslant \int_{|x|>A} e^{-x^2/2}\, dx \leqslant \frac{2}{A}\int_A^\infty x e^{-x^2/2}\, dx = \frac{2}{A}\, e^{-A^2/2}.$$

Выбрав достаточно большое A, мы можем сделать J_2 сколь угодно малым.

Согласно неравенству (1) имеем

$$J_3 \leqslant \int_{\epsilon B_n \leqslant |x| \leqslant \frac{\pi B_n}{h}} \left| f^*\left(\frac{x}{B_n}\right)\right|^n dx \leqslant e^{-nc_0}\, 2B_n\left(\frac{\pi}{h} - \epsilon\right)$$

Отсюда ясно, что при $n \to \infty$

$$J_3 \to 0.$$

Для оценки интеграла J_4 мы заметим, что существование дисперсии влечет за собой существование второй производной у функции $f^*(t)$. Мы можем, следовательно, в окрестности точки $t = 0$ воспользоваться согласно (3) § 32 разложением

$$f^*(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2),$$

и при $|t| \leqslant \epsilon$, если ϵ достаточно мало, получим:

$$|f^*(t)| < 1 - \frac{\sigma^2 t^2}{4} < e^{-\frac{\sigma^2 t^2}{4}}$$

Тогда при $|x| \leqslant \epsilon B_n$

$$\left| f^*\left(\frac{x}{B_n}\right) \right|^n < e^{-\frac{n\sigma^2 t^2}{4B_n^2}} = e^{-\frac{t^2}{4}}.$$

Поэтому

$$|J_4| \leqslant 2 \int_A^{\epsilon B_n} e^{-\frac{t^2}{4}} dt < 2 \int_A^{\infty} e^{-\frac{t^2}{4}} dt.$$

Выбором достаточно большого A мы можем добиться, чтобы интеграл J_4 был сколь угодно мал. Теорема доказана.

Имеется еще другой случай, когда естественно ставить вопрос о локальном поведении функций распределения сумм. Это – случай непрерывных распределений.

Здесь ставится вопрос о том, когда плотности распределения нормированных сумм сходятся к плотности нормального распределения, если соответствующие функции распределения сходятся к нормальной.

Оказывается, что для случая одинаково распределенных независимых слагаемых с конечной дисперсией достаточным является условие интегрируемости плотности слагаемых в какой-либо степени $S > 1$.

Если отказаться от этого условия, то легко указать примеры случайных величин, имеющих плотности распределения вероятностей и принимающих значения только в ограниченном интервале, но для которых локальная теорема не имеет места (см. монографию Б.В. Гнеденко и А.Н. Колмогорова).

Упражнения

1. Доказать, что при $n \to \infty$

$$\frac{1}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\left(\frac{n}{2}\right)^n} \int_0^{1+t\sqrt{\frac{2}{n}}} z^{\frac{n}{2}-1} e^{-\frac{nz}{2}} dz \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}z^2} dz.$$

Указание. Применить теорему Ляпунова к распределению χ^2.

2. Случайные величины

$$\xi_n = \begin{cases} -n^{\alpha} & \text{с вероятностью } 0{,}5, \\ +n^{\alpha} & \text{с вероятностью } 0{,}5 \end{cases}$$

независимы. Доказать, что при $\alpha > -0{,}5$ к ним применима теорема Ляпунова.

3. Доказать, что при $n \to \infty$

$$e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \to \frac{1}{2}.$$

Указание. Применить теорему Ляпунова к сумме случайных величин с параметром $\lambda = 1$.

4. Вероятность появления события A в i-м испытании равна p_i, μ – число появлений события A в n независимых испытаниях. Доказать, что

$$\mathbf{P}\left\{\frac{\mu - \sum\limits_{k=1}^{n} p_k}{\sqrt{\sum\limits_{k=1}^{n} p_k q_k}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

тогда и только тогда, когда $\sum\limits_{i=1}^{\infty} p_i q_i = \infty$.

5. Доказать, что в условиях предыдущей задачи требование $\sum\limits_{i=1}^{\infty} p_i q_i = \infty$ достаточно не только для интегральной, но и для локальной теоремы.

ГЛАВА 9

ТЕОРИЯ БЕЗГРАНИЧНО ДЕЛИМЫХ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

Долгое время центральной задачей теории вероятностей считали отыскание наиболее общих условий, при выполнении которых функции распределения сумм независимых случайных величин сходятся к нормальному закону. Весьма общие условия, достаточные для этой сходимости, были найдены, как мы говорили об этом в главе 8, А.М. Ляпуновым.

Попытки расширить условия Ляпунова увенчались успехом лишь в тридцатые годы, когда были найдены условия, являющиеся не только достаточными, но и, при весьма естественных ограничениях, необходимыми.

Параллельно с завершением классической проблематики возникло и развилось новое направление в теории предельных теорем для сумм независимых случайных величин, тесно связанное с возникновением и развитием теории стохастических (случайных) процессов. В первую очередь возник вопрос о том, какие законы, помимо нормального закона, могут быть предельными для сумм независимых случайных величин.

Оказалось, что класс предельных законов далеко не исчерпывается нормальным законом. Затем возник вопрос об определении условий, которые следует наложить на слагаемые, чтобы функции распределения сумм сходились к тому или иному предельному закону.

В настоящей главе мы ставим своей целью изложение некоторых исследований, посвященных предельным теоремам для сумм независимых случайных величин. При этом мы ограничиваемся случаем, когда слагаемые имеют конечные дисперсии. Рассмотрение задачи без этого ограничения требует более громоздких вычислений; для ознакомления с ее решением мы отсылаем читателя к упоминавшейся монографии Гнеденко и Колмогорова. В качестве простого следствия излагаемых нами общих теорем мы получим упомянутое нами необходимое и достаточное условие сходимости функций распределения сумм к нормальному закону.

Последний параграф главы посвящен новому направлению исследований — предельным теоремам для сумм случайного числа случайных слагаемых.

§ 42. Безгранично делимые законы и их основные свойства

Закон $\Phi(x)$ называется *безгранично делимым*, если при любом n его характеристическая функция является n-й степенью некоторой другой характеристической функции.

Исследования последних лет показали, что безгранично делимые законы играют значительную роль в различных вопросах теории вероятностей. В частности, оказалось, что класс предельных законов для сумм независимых случайных величин совпадает с классом безгранично делимых законов.

Мы перейдем теперь к изложению необходимых нам для дальнейшего свойств безгранично-делимых законов. Это изложение мы начнем с доказательства того, что законы нормальный и Пуассона безгранично делимы. Действительно, характеристическая функция нормального закона, имеющего математическое ожидание a и дисперсию σ^2, равна

$$\varphi(t) = e^{iat - \frac{1}{2}\sigma^2 t^2}.$$

При любом n корень n-й степени из $\varphi(t)$ есть снова характеристическая функция нормального закона, только с математическим ожиданием a/n и дисперсией σ^2/n.

Мы несколько обобщим встречавшееся ранее понятие закона Пуассона и скажем, что случайная величина ξ распределена по закону Пуассона, если она может принимать только значения $ak + b$, где a и b – вещественные постоянные, а $k = 0, 1, 2, \ldots$ и

$$\mathsf{P}\{\xi = ak + b\} = \frac{e^{-\lambda}\lambda^k}{k!}, \tag{1}$$

где λ – положительное постоянное. Характеристическая функция для закона (1), как легко подсчитать, дается формулой

$$\varphi(t) = e^{\lambda(e^{iat} - 1) + ibt}.$$

Мы видим, что при любом n корень n-й степени из $\varphi(t)$ есть снова характеристическая функция закона Пуассона, но с другими параметрами: a, $\dfrac{\lambda}{n}$ и $\dfrac{1}{n}b$.

Теорема 1. *Характеристическая функция безгранично делимого закона не обращается в нуль.*

Доказательство. Пусть $\Phi(x)$ – безгранично делимый закон и $\varphi(t)$ – его характеристическая функция. Тогда, по определению, при любом n мы имеем равенство

$$\varphi(t) = \{\varphi_n(t)\}^n, \tag{2}$$

где $\varphi_n(t)$ – некоторая характеристическая функция. В силу непрерывности функции $\varphi(t)$ существует область значений аргумента $|t| \leqslant a$, в которой $\varphi(t) \neq 0$; понятно, что в этой же области $\varphi_n(t) \neq 0$. При достаточно большом n мы можем величину $|\varphi_n(t)| = \sqrt[n]{|\varphi(t)|}$ сделать сколь угодно близкой к единице равномерно по t $(|t| \leqslant a)$.

Возьмем теперь две взаимно независимые случайные величины η_1 и η_2, распределенные по некоторому закону $F(x)$, и рассмотрим их разность $\eta = \eta_1 - \eta_2$. Характеристическая функция величины η равна

$$f^*(t) = \mathbf{M}e^{it(\eta_1 - \eta_2)} = |\mathbf{M}e^{it\eta_1}|^2 = |f(t)|^2.$$

Мы видим, таким образом, что квадрат модуля любой характеристической функции является характеристической функцией.

Далее, так как вещественная характеристическая функция имеет вид

$$f(t) = \int \cos xt dF(x),$$

то, следовательно, мы можем написать неравенство

$$1 - f(2t) = \int (1 - \cos 2xt) dF(x) = 2\int \sin^2 xt\, dF(x) =$$

$$= 2\int (1 - \cos xt)(1 + \cos xt) dF(x) \leqslant 4\int (1 - \cos xt) dF(x) = 4(1 - f(t)).$$

Из сказанного мы видим, что функция $|\varphi_n(t)|^2$ удовлетворяет неравенству

$$1 - |\varphi_n(2t)|^2 \leqslant 4(1 - |\varphi_n(t)|^2).$$

Из этого неравенства следует, что если n столь велико, что $1 - |\varphi_n(t)| < \epsilon$ при $|t| \leqslant a$, то в той же области

$$1 - |\varphi_n(2t)| \leqslant 1 - |\varphi_n(2t)|^2 \leqslant 4(1 - |\varphi_n(t)|^2) \leqslant 8(1 - |\varphi_n(t)|) < 8\epsilon.$$

Итак, в области $|t| \leqslant 2a$

$$1 - |\varphi_n(t)| < 8\epsilon.$$

Таким образом, при достаточно больших n в области $|t| \leqslant 2a$, $\varphi_n(t)$, а вместе с тем и $\varphi(t)$ в нуль не обращаются.

Подобным же способом мы докажем, что $\varphi(t) \neq 0$ в области $|t| < 4a$, и т.д.

Это доказывает нашу теорему.

Т е о р е м а 2. *Функция распределения суммы независимых случайных величин, имеющих безгранично делимые функции распределения, также безгранично делима.*

Д о к а з а т е л ь с т в о. Очевидно, что для доказательства теоремы достаточно ограничиться случаем двух слагаемых. Если $\varphi(t)$ и $\psi(t)$ – характеристические функции слагаемых, то по условию теоремы при любом n имеем:

$$\varphi(t)=\{\varphi_n(t)\}^n, \qquad \psi(t)=\{\psi_n(t)\}^n,$$

где $\varphi_n(t)$ и $\psi_n(t)$ – характеристические функции. Поэтому характеристическая функция суммы при любом n удовлетворяет равенству

$$\chi(t)=\varphi(t)\psi(t)=\{\varphi_n(t)\cdot\psi_n(t)\}^n.$$

Теорема 3. *Функция распределения, предельная (в смысле сходимости в основном) для последовательности безгранично делимых функций распределения, сама является безгранично делимой.*

Доказательство. Пусть последовательность $\Phi^{(k)}(x)$ безгранично делимых функций распределения сходится в основном к функции распределения $\Phi(x)$. Тогда

$$\lim_{k\to\infty}\varphi^{(k)}(t)=\varphi(t) \tag{3}$$

равномерно в каждом конечном интервале t. По условию теоремы при любом n функции (под $\sqrt[n]{}$ понимается его главное значение)

$$\varphi_n^{(k)}(t)=\sqrt[n]{\varphi^{(k)}(t)} \tag{4}$$

являются характеристическими функциями. Из (3) заключаем, что при каждом n

$$\lim_{k\to\infty}\varphi_n^{(k)}(t)=\varphi_n(t). \tag{5}$$

Из непрерывности $\varphi_n^{(k)}(t)$ следует непрерывность $\varphi_n(t)$. В силу предельной теоремы для характеристических функций, $\varphi_n(t)$ есть характеристическая функция. Из (3), (4) и (5) находим, что при каждом n имеет место равенство

$$\varphi(t)=\{\varphi_n(t)\}^n,$$

что и требовалось доказать.

§ 43. Каноническое представление безгранично делимых законов

В дальнейшем мы ограничимся изучением безгранично делимых законов с конечной дисперсией. Целью настоящего параграфа является доказательство следующей теоремы, найденной в 1932 г. А.Н. Колмогоровым и дающей полную характеристику интересующего нас класса законов распределения.

Теорема 4. *Для того чтобы функция распределения $\Phi(x)$ с конечной дисперсией была безгранично делимой, необходимо и достаточно, что-*

бы логарифм ее характеристической функции имел вид

$$\ln\varphi(t) = i\gamma t + \int \{e^{itx} - 1 - itx\} \frac{1}{x^2}\, dG(x), \tag{1}$$

где γ – вещественная постоянная, а $G(x)$ – неубывающая функция ограниченной вариации. При $x = 0$ подынтегральная функция считается равной $-t^2/2$.

Доказательство. Предположим сначала, что $\Phi(x)$ – безгранично делимый закон и $\varphi(t)$ – его характеристическая функция. Тогда при любом n

$$\varphi(t) = \{\varphi_n(t)\}^n,$$

где $\varphi_n(t)$ – некоторая характеристическая функция. Так как $\varphi(t) \neq 0$, то это равенство эквивалентно следующему *):

$$\ln\varphi(t) = n\ln\varphi_n(t) = n\ln[1 + \varphi_n(t) - 1)].$$

Каково бы ни было T, при $n\to\infty$ равномерно в интервале $|t| < T$

$$\varphi_n(t)\to 1,$$

поэтому в любом конечном интервале значений t величина $|\varphi_n(t) - 1|$ может быть сделана меньше любого, наперед заданного числа, лишь бы n было достаточно велико. Мы можем, следовательно, воспользоваться равенством

$$\ln[1 + (\varphi_n(t) - 1] = (\varphi_n(t) - 1)(1 + o(1)),$$

которое дает:

$$\ln\varphi(t) = \lim_{n\to\infty} n(\varphi_n(t) - 1) = \lim_{n\to\infty} n\int(e^{itx} - 1)\,d\Phi_n(x), \tag{2}$$

где $\Phi_n(x)$ – функция распределения, имеющая $\varphi_n(t)$ своей характеристической функцией. Из определения математического ожидания и связи между функциями $\Phi_n(x)$ и $\Phi(x)$ следует, что

$$n\int x\,d\Phi_n(x) = \int x\,d\Phi(x).$$

Обозначим эту величину через γ; тогда равенство (2) мы можем переписать в следующем виде:

$$\ln\varphi(t) = i\gamma t + \lim_{n\to\infty} n\int\{e^{itx} - 1 - itx\}\,d\Phi_n(x).$$

Положим теперь

$$G_n(x) = n\int_{-\infty}^{x} u^2\,d\Phi_n(u).$$

---

*) Логарифм здесь понимается в смысле главного значения.

Очевидно, что функции $G_n(x)$ не убывают с возрастанием аргумента и $G_n(-\infty) = 0$. Кроме того, функции $G_n(x)$ ограничены в совокупности. Последнее утверждение вытекает из свойств дисперсии и связи между функциями $\Phi(x)$ и $\Phi_n(x)$. Действительно,

$$G_n(+\infty) = n\int u^2 d\Phi_n(u) =$$
$$= n\left[\int u^2 d\Phi_n(u) - \left(\int u d\Phi_n(u)\right)^2\right] + n\left(\int u d\Phi_n(u)\right)^2 = \sigma^2 + \frac{1}{n}\gamma^2, \tag{3}$$

где σ^2 – дисперсия закона $\Phi(x)$.

В новых обозначениях (см. свойство 6 интеграла Стилтьеса в § 22)

$$\lg\varphi(t) = i\gamma t + \lim_{n\to\infty}\int (e^{itx} - 1 - itx)\frac{1}{x^2}\,dG_n(x).$$

Согласно первой теореме Хелли, из последовательности функций $G_n(x)$ можно выбрать подпоследовательность, сходящуюся к некоторой предельной функции $G(x)$. Если $A < 0$ и $B > 0$ являются точками непрерывности функций $G(x)$, то в силу второй теоремы Хелли при $k \to \infty$

$$\int_A^B (e^{itx} - 1 - itx)\frac{1}{x^2}\,dG_{n_k}(x) \to \int_A^B (e^{itx} - 1 - itx)\frac{1}{x^2}\,dG(x). \tag{4}$$

Мы знаем, что

$$|e^{itx} - 1 - itx| \leqslant |e^{itx} - 1| + |tx| \leqslant |tx| + |tx| = 2|t|\cdot|x|,$$

поэтому

$$\left|\int_{-\infty}^{A} + \int_B^\infty (e^{itx} - 1 - itx)\frac{1}{x^2}\,dG_{n_k}(x)\right| \leqslant$$
$$\leqslant \int_{-\infty}^{A} + \int_B^\infty \frac{|e^{itx} - 1 - itx|}{x^2}\,dG_{n_k}(x) \leqslant$$
$$\leqslant 2|t|\left(\int_{-\infty}^{A} + \int_B^\infty \frac{1}{|x|}\,dG_{n_k}(x)\right) \leqslant \frac{2|t|}{\Gamma}\left(\int_{-\infty}^{A} + \int_B^\infty dG_{n_k}(x)\right) \leqslant$$
$$\leqslant \frac{2|t|}{\Gamma}\max_{1\leqslant k<\infty}\int dG_{n_k}(x),$$

где $\Gamma = \min(|A|, B)$. Так как вариации функций $G_{n_k}(u)$ равномерно ограничены, то, каково бы ни было $\epsilon > 0$, мы можем, выбирая A и B достаточно большими, добиться выполнения неравенства

$$\left|\int_{-\infty}^{A} + \int_B^\infty (e^{itx} - 1 - itx)\frac{1}{x^2}\,dG_{n_k}(x)\right| < \frac{\epsilon}{2} \tag{5}$$

для всех t, заключенных в каком-либо конечном интервале, и для всех k.

Из (4) и (5) следует, что, каково бы ни было $\epsilon > 0$, для всех t, заключенных в произвольном конечном интервале, при достаточно больших n имеет место неравенство

$$\left| \int (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG_{n_k}(x) - \int (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG(x) \right| < \epsilon,$$

т.е., иными словами,

$$\lim_{k\to\infty} \int (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG_{n_k}(x) = \int (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG(x).$$

Мы доказали, таким образом, что логарифм характеристической функции любого безгранично делимого закона может быть записан в виде (1). Нам предстоит теперь доказать обратное предложение, что всякая функция, логарифм которой представим по формуле (1), является характеристической функцией некоторого безгранично делимого закона.

Для любого ϵ $(0 < \epsilon < 1)$ интеграл

$$\int_{\epsilon}^{\frac{1}{\epsilon}} (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG(x) \tag{6}$$

по определению интеграла Стилтьеса является пределом сумм

$$\sum_{s=1}^{n} (e^{it\bar{x}_s} - 1 - it\bar{x}_s) \frac{1}{\bar{x}_s^2} (G(x_{s+1}) - G(x_s)),$$

где $x_1 = \epsilon$, $x_{n+1} = 1/\epsilon$, $x_s \leqslant \bar{x}_s \leqslant x_s + 1$, и $\max(x_{s+1} - x_s) \to 0$. Каждое слагаемое этой суммы является логарифмом характеристической функции некоторого закона Пуассона. Согласно теоремам 2 и 3 интеграл (6) является логарифмом характеристической функции некоторого безгранично делимого закона. Переходя к пределу при $\epsilon \to 0$, мы убеждаемся, что то же самое имеет место для интеграла

$$\int_{x>0} (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG(x). \tag{7}$$

Подобным же образом доказываем, что интеграл

$$\int_{x<0} (e^{itx} - 1 - itx) \frac{1}{x^2}\, dG(x) \tag{8}$$

есть логарифм характеристической функции некоторого безгранично делимого закона. Интеграл, стоящий в правой части формулы (1), равен сумме интегралов (7) и (8) и величины

$$i\gamma t - \frac{1}{2} t^2 (G(+0) - G(-0)).$$

Последнее слагаемое есть логарифм характеристической функции нормального закона. Из теоремы 2 следует, что функция $\varphi(t)$, представимая формулой (1), является характеристической функцией некоторого безгранично делимого закона*). Нам остается теперь убедиться, что представление $\ln\varphi(t)$ формулой (1) единственно, т.е. что функция $G(x)$ и постоянное γ однозначно определяются заданием $\varphi(t)$.

Путем дифференцирования формулы (1) находим, что

$$\frac{d^2}{dt^2}\ln\varphi(t) = -\int e^{itx}dG(x). \qquad (9)$$

Из теории характеристических функций мы знаем, что функция $G(x)$ в этой формуле однозначно определяется через $\dfrac{d^2}{dt^2}\ln\varphi(t)$. В процессе доказательства теоремы мы видели, что постоянное γ является математическим ожиданием и, значит, так же однозначно определяется посредством функции $\varphi(t)$.

Отметим, наконец, вероятностный смысл полной вариации функции $G(x)$. Мы знаем, что если случайная величина ξ распределена по закону $\Phi(x)$, то (см. (6) § 33)

$$\mathbf{D}\xi = -\left[\frac{d^2}{dt^2}\ln\varphi(t)\right]_{t=0};$$

из (9), следовательно, вытекает, что

$$\mathbf{D}\xi = \int dG(x) = G(+\infty).$$

В качестве примеров мы приведем каноническое представление нормального закона и закона Пуассона.

Для нормального закона с дисперсией σ^2 и математическим ожиданием a

$$\gamma = a \quad \text{и} \quad G(x) = \begin{cases} 0 & \text{для} \quad x < 0, \\ \sigma^2 & \text{для} \quad x > 0. \end{cases}$$

---

*) Мы только что доказали, что всякий безгранично делимый закон является либо композицией конечного числа законов Пуассона и нормального закона, либо пределом равномерно сходящейся последовательности таких законов. Таким образом, мы видим, что законы нормальный и Пуассона являются теми основными элементами, из которых составлен каждый безгранично делимый закон.

Действительно, эта функция и постоянное γ приводят к данному закону, так как

$$\int \{e^{itx} - 1 - itx\} \frac{1}{x^2} \, dG(x) =$$

$$= \lim_{u \to 0} \frac{e^{itu} - 1 - itu}{u^2} [G(+0) - G(-0)] = - \frac{t^2 \sigma^2}{2},$$

а в силу единственности канонического представления другие функции $G(x)$ не могут дать нормального закона.

Подобным же способом легко убедиться, что закону Пуассона с характеристической функцией

$$\varphi(t) = e^{\lambda(e^{ita} - 1) + ibt}$$

соответствует функция $G(x)$ с единственным скачком в точке a:

$$G(x) = \begin{cases} 0 & \text{при} \quad x < a, \\ a^2 \lambda & \text{при} \quad x > a \end{cases}$$

и $\gamma = b + a\lambda$.

§ 44. Предельная теорема для безгранично делимых законов

Мы уже знаем, что если последовательность безгранично делимых законов распределения сходится к предельному закону распределения, то этот предельный закон сам является безгранично делимым. Теперь мы укажем условия, при выполнении которых данная последовательность безгранично делимых функций распределения будет сходиться к предельной.

Теорема 5. *Для того чтобы последовательность $\{\Phi_n(x)\}$ безгранично делимых функций распределения сходилась при $n \to \infty$ к некоторой функции распределения $\Phi(x)$ и дисперсии их сходились к дисперсии предельного закона, необходимо и достаточно, чтобы существовали такие постоянное γ и функция $G(x)$, для которых при $n \to \infty$*

1) *$G_n(x)$ сходится в основном к $G(x)$ $(-\infty \leqslant x \leqslant +\infty)$,*
2) *$G_n(\infty) \to G(\infty)$,*
3) *$\gamma_n \to \gamma$,*

где γ_n и $G_n(x)$ определяются формулой (1) § 43 для закона $\Phi_n(x)$, а постоянное γ и функция $G(x)$ определяют по той же формуле предельный закон $\Phi(x)$.

Доказательство. Достаточность условий теоремы является непосредственным следствием второй теоремы Хелли. Действительно, из условий теоремы и формулы (1) § 43 следует, что при $n \to \infty$

$$\ln \varphi_n(t) \to \ln\varphi(t)$$

равномерно в каждом конечном интервале t.

В предыдущем параграфе мы видели, что интегралы

$$\int dG_n(u) \quad \text{и} \quad \int dG(u)$$

равны дисперсиям законов $\Phi_n(x)$ и $\Phi(x)$; поэтому второе условие теоремы есть не что иное, как требование сходимости дисперсий.

Пусть теперь нам известно, что при $n \to \infty$

$$\Phi_n(x) \to \Phi(x) \tag{1}$$

и дисперсии законов $\Phi_n(x)$ сходятся к дисперсии предельного закона $\Phi(x)$. Мы докажем, что эти требования влекут за собой выполнение условий теоремы. В отношении условия 2, как мы только что заметили, это не требует дополнительных рассуждений. Отсюда следует, что полные вариации функций $G_n(u)$ ограничены в совокупности. Мы можем, следовательно, воспользоваться первой теоремой Хелли и из последовательности функций $G_n(u)$ выбрать подпоследовательность $Gn_k(u)$, сходящуюся при $k \to \infty$ к некоторой предельной функции $G_\infty(u)$. Наша цель состоит в том, чтобы доказать равенство

$$G_\infty(u) = G(u).$$

Для этого установим сначала, что

$$J_k = \int \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_{n_k}(u) \to$$

$$\to J_\infty = \int \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_\infty(u) \tag{2}$$

при $k \to \infty$. Пусть $A < 0$ и $B > 0$ – точки непрерывности функций $G_\infty(u)$; тогда по второй теореме Хелли при $k \to \infty$

$$\int_A^B \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_{n_k}(u) \to \int_A^B \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_\infty(u). \tag{3}$$

С другой стороны, из неравенства

$$| e^{itx} - 1 - itx | \leqslant 2 | tx |$$

мы видим, что

$$L_k = \Big| \int\limits_{-\infty}^{A} + \int\limits_{B}^{\infty} \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_{n_k}(u) \Big| \leqslant$$

$$\leqslant 2|t| \Big| \int\limits_{-\infty}^{A} + \int\limits_{B}^{\infty} \frac{1}{|u|} dG_{n_k}(u) \Big| \leqslant$$

$$\leqslant \frac{2|t|}{2} \Big(\int\limits_{-\infty}^{A} + \int\limits_{B}^{\infty} dG_{n_k}(u) \Big) \leqslant \frac{2|t|}{\Gamma} \int dG_{n_k}(u),$$

где $\Gamma = \min(-A, B)$. В силу ограниченности вариаций функций $G_n(u)$ в совокупности для любого $\epsilon > 0$ можно подобрать столь большие по абсолютной величине A и B, что

$$L_k < \epsilon. \tag{4}$$

Точно так же при любом $\epsilon > 0$ для достаточно больших по абсолютной величине A и B имеет место неравенство

$$\Big| \int\limits_{-\infty}^{A} + \int\limits_{B}^{\infty} \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_{\infty}(u) \Big| < \epsilon. \tag{5}$$

Из соотношений (3), (4) и (5) выводим, что, каково бы ни было $\epsilon > 0$, при достаточно больших значениях k

$$|J_k - J_\infty| < 3\epsilon.$$

Соотношение (2), таким образом, доказано. Из (1) видим, что

$$\lim_{n\to\infty} \ln \varphi_n(t) = \lim_{n\to\infty} \Big(i\gamma_n t + \int \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_n(u) \Big) =$$

$$= \ln \varphi(t) = i\gamma t + \int \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG(u),$$

или

$$\lim_{n\to\infty} \Big(i\gamma_{n_k} + \int \{e^{itu} - 1 - itu\} \frac{1}{tu^2} dG_{n_k}(u) \Big) =$$

$$= i\gamma + \int \{e^{itu} - 1 - itu\} \frac{1}{tu^2} dG(u). \tag{6}$$

Из неравенства

$$|e^{itu} - 1 - itu| \leqslant t^2 u^2/2$$

и ограниченности в совокупности полных вариаций функций $G_{n_k}(u)$

заключаем, что при $t \to 0$

$$\left| \int (e^{itu} - 1 - itu) \frac{1}{tu^2} \, dG_{n_k}(u) \right| \leqslant \left| t \int (e^{itu} - 1 - itu) \frac{1}{t^2 u^2} \, dG(u) \right| \to 0$$

равномерно по n. Поэтому при $t \to 0$ (6) дает:

$$\lim_{k \to \infty} \gamma_{n_k} = \gamma, \tag{7}$$

а, с другой стороны, по (2) и (7)

$$\ln \varphi(t) = i\gamma t + \int \{e^{iut} - 1 - iut\} \frac{1}{u^2} \, dG_\infty(u).$$

В силу единственности представления безгранично делимых законов формулой (1) § 43 мы заключаем, что $G_\infty(u) = G(u)$. Итак, *любая* сходящаяся подпоследовательность функций $G_{n_k}(u)$ сходится к функции $G(u)$ и одновременно постоянные γ_{n_k} сходятся к γ. Теперь легко доказать, что *вся* последовательность $G_n(u)$ также сходится к $G(u)$ и, значит, одновременно $\lim\limits_{n \to \infty} \gamma_n = \gamma$. Если бы это было не так, то нашлась бы точка непрерывности функций $G(u)$, назовем ее c, и подпоследовательность функций $G_{n_k}(u)$, которая в точке $u = c$ при $k \to \infty$ сходится к числу, отличному от $G(c)$. По первой теореме Хелли мы можем из этой подпоследовательности выбрать сходящуюся подпоследовательность $G_{nk_r}(u)$.

Из предыдущего следует, что во всех точках непрерывности функции $G(u)$

$$\lim_{r \to \infty} G_{nk_r}(u) = G(u).$$

Это противоречит сделанному нами допущению. Таким образом, во всех точках непрерывности функции $G(u)$

$$\lim_{n \to \infty} G_n(u) = G(u);$$

как мы видели, отсюда немедленно следует, что

$$\lim_{n \to \infty} \gamma_n = \gamma.$$

Теорема доказана.

§ 45. Постановка задачи о предельных теоремах для сумм

Дана последовательность серий

$$\left.\begin{array}{llll} \xi_{11}, & \xi_{12}, \ldots, & \xi_{1k_1}, \\ \xi_{21}, & \xi_{22}, \ldots, & \xi_{2k_2}, \\ \ldots & \ldots & \ldots \\ \xi_{n1}, & \xi_{n2}, \ldots, & \xi_{nk_n}, \\ \ldots & \ldots & \ldots \end{array}\right\} \qquad (1)$$

независимых в каждой серии случайных величин. Спрашивается, к каким предельным функциям распределения могут сходиться функции распределения сумм

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$$

при $n \to \infty$ и каковы условия этой сходимости?

В дальнейшем мы ограничимся изучением *элементарных систем*, т.е. последовательностей серий (1), для которых выполнены следующие условия:

1) величины ξ_{nk} имеют конечные дисперсии,

2) дисперсии сумм ζ_n ограничены не зависящей от n константой C,

3) $\beta_n = \max\limits_{1 \leqslant k \leqslant k_n} \mathsf{D}\xi_{nk} \to 0$ при $n \to \infty$.

Последнее требование означает, что влияние отдельных слагаемых на сумму становится все меньше и меньше с возрастанием n.

Рассмотренные нами ранее предельные теоремы для сумм, очевидно, укладываются в эту общую схему. Так, в теоремах Муавра–Лапласа и Липунова мы имели следующую последовательность серий:

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn},$$

где

$$\xi_{nk} = \frac{\xi_k - \mathsf{M}\xi_k}{\sqrt{\sum\limits_{k=1}^{n} \mathsf{D}\xi_k}} \qquad (1 \leqslant k \leqslant n;\ n = 1, 2, \ldots).$$

В теоремах Бернулли, Чебышева и Маркова о законе больших чисел мы также имели дело с последовательностью серий, в которых в качестве ξ_{nk} взяты величины

$$\xi_{nk} = \frac{\xi_k - \mathsf{M}\xi_k}{n}.$$

§ 46. Предельные теоремы для сумм

Пусть имеется элементарная система; обозначим через $F_{nk}(x)$ функцию распределения случайной величины ξ_{nk} и через $\bar{F}_{nk}(x)$ – функцию распределения величины $\bar{\xi}_{nk} = \xi_{nk} - \mathbf{M}\xi_{nk}$; очевидно, что

$$\bar{F}_{nk}(x) = F_{nk}(x + \mathbf{M}\xi_{nk}).$$

Т е о р е м а 6. *Для того чтобы функции распределения сумм*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n} \tag{1}$$

при $n \to \infty$ сходились к предельной функции распределения, необходимо и достаточно, чтобы к предельному закону сходились безгранично делимые законы, логарифмы характеристических функций которых определяются формулой

$$\psi_n(t) = \sum_{k=1}^{k_n} \{it\mathbf{M}\xi_{nk} + \int (e^{itx} - 1)d\bar{F}_{nk}(x)\}\ ^{*)}. \tag{2}$$

Предельные законы для обеих последовательностей совпадают.

Д о к а з а т е л ь с т в о. Характеристическая функция суммы (1) равна

$$f_n(t) = \prod_{k=1}^{k_n} f_{nk}(t) = e^{it\sum_{k=1}^{k_n}\mathbf{M}\xi_{nk}} \prod_{k=1}^{k_n} \bar{f}_{nk}(t), \tag{3}$$

где $f_{nk}(t)$ – характеристическая функция случайной величины ξ_{nk}, а $\bar{f}_{nk}(t)$ – характеристическая функция величины $\bar{\xi}_{nk}$.

Мы знаем, что для сходимости функций распределения сумм (1) к предельной $\Phi(x)$ необходимо и достаточно, чтобы при $n \to \infty$

$$f_n(t) \to \varphi(t),$$

где $\varphi(t)$ – непрерывная функция; $\varphi(t)$ при этом оказывается характеристической функцией закона $\Phi(x)$.

---

*) Если ввести обозначения

$$\gamma_n = \sum_{k=1}^{k_n} \mathbf{M}\xi_{nk}, \qquad G_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^{u} x^2 d\bar{F}_{nk}(x)$$

и заметить, что $\int x d\bar{F}_{nk}(x) = 0$, то функции $\psi_n(t)$ могут быть записаны в виде

$$\psi_n(t) = i\gamma_n t + \int \{e^{itu} - 1 - itu\} \frac{1}{u^2} dG_n(u).$$

Как мы знаем, это означает, что $\psi_n(t)$ является логарифмом характеристической функции некоторого безгранично делимого закона.

Отметим, что дисперсии ζ_n и безгранично делимых законов (2) совпадают.

Положим

$$\alpha_{nk} = \bar{f}_{nk}(t) - 1.$$

Для величин ξ_{nk} равномерно в каждом конечном интервале t

$$\alpha_n = \max_{1 \leqslant k \leqslant k_n} |\alpha_{nk}| \to 0. \tag{4}$$

Действительно,

$$\alpha_{nk} = \int (e^{itx} - 1)\, d\bar{F}_{nk}(x) = \int (e^{itx} - 1 - itx)\, d\bar{F}_{nk}(x),$$

так как

$$\mathbf{M}\bar{\xi}_{nk} = \int x\, d\bar{F}_{nk}(x) = 0.$$

Мы знаем, что при всех вещественных α

$$|e^{i\alpha} - 1 - i\alpha| \leqslant \alpha^2/2;$$

поэтому

$$|\alpha_{nk}| \leqslant \frac{t^2}{2} \int x^2\, d\bar{F}_{nk}(x) = \frac{t^2}{2} \mathbf{D}\xi_{nk}. \tag{5}$$

Из (5) и третьего условия элементарности системы следует (4).

Из (4) мы прежде всего выводим, что при любом T мы можем считать, что для достаточно больших n и $|t| \leqslant T$

$$|\alpha_{nk}| < 1/2. \tag{6}$$

В силу этого мы можем воспользоваться разложением логарифма в ряд

$$\ln \bar{f}_{nk}(t) = \ln(1 + \alpha_{nk}) = \alpha_{nk} - \frac{\alpha_{nk}^2}{2} + \frac{\alpha_{nk}^3}{3} - \ldots = \alpha_{nk} + r_{nk}.$$

Очевидно, что

$$R_n = \left| \ln f_n(t) - \sum_{n=1}^{k_n} (i + \mathbf{M}\xi_{nk} + \alpha_{nk}) \right| = \left| \sum_{n=1}^{k_n} (\ln \bar{f}_{nk}(t) - \alpha_{nk}) \right| \leqslant$$

$$\leqslant \sum_{k=1}^{k_n} \sum_{n=1}^{\infty} \frac{|\alpha_{nk}|^s}{s} \leqslant \frac{1}{2} \sum_{k=1}^{k_n} \frac{|\alpha_{nk}|^2}{1 - |\alpha_{nk}|} \tag{7}$$

Формулы (5) и (6) приводят к неравенству

$$R_n \leqslant \max_{1 \leqslant k \leqslant k_n} |\alpha_{nk}| \sum_{k=1}^{k_n} |\alpha_{nk}| \leqslant \frac{Ct^2}{2} \max_{1 \leqslant k \leqslant k_n} |\alpha_{nk}|.$$

В силу (4) мы заключаем, что равномерно в каждом конечном интервале t при $n\to\infty$

$$|\ln f_n(t)-\psi_n(t)|\to 0. \tag{8}$$

Таким образом, мы установили, что *в каждой элементарной системе функции распределения сумм ζ_n и безгранично делимые функции распределения, определяемые формулой* (2), *неограниченно сближаются при* $n\to\infty$, чем собственно теорема 6 и доказана.

Доказанная теорема позволяет заменять исследование сумм (1) случайных величин с функциями распределения, вообще говоря, произвольными исследованием безгранично делимых законов. Последнее, как мы увидим, во многих случаях оказывается весьма простым.

Теорема 7. *Всякий закон распределения, предельный для функций распределения сумм элементарной системы, является безгранично делимым с конечной дисперсией и, обратно, каждый безгранично делимый закон с конечной дисперсией является предельным для функций распределения сумм некоторой элементарной системы.*

Доказательство. Из предыдущей теоремы мы знаем, что предельный закон для функций распределения сумм (1) является предельным для безгранично делимых законов и, значит, по теореме 3 является безгранично делимым; его дисперсия конечна, так как дисперсии сумм по второму условию элементарности системы ограничены в совокупности. Обратное предложение, что каждый безгранично делимый закон с конечной дисперсией является предельным для сумм, немедленно вытекает из определения безгранично делимых законов.

Теорема 8. *Для того чтобы функции распределения сумм* (1) *при $n\to\infty$ сходились к какой-нибудь предельной функции распределения и их дисперсии сходились к дисперсии предельного закона, необходимо и достаточно, чтобы существовали такие функция $G(u)$ и постоянное γ, что при* $n\to\infty$

1) $\sum\limits_{k=1}^{k_n}\int\limits_{-\infty}^{u}x^2d\overline{F}_{nk}(x)\to G(u)$

в точках непрерывности функций $G(u)$

2) $\sum\limits_{k=1}^{k_n}\int x^2d\overline{F}_{nk}(x)\to G(+\infty)$,

3) $\sum\limits_{k=1}^{k_n}\int xdF_{nk}(x)\to\gamma$.

Логарифм характеристической функции предельного закона определяется формулой (1) § 43 *с только что определенными функцией $G(u)$ и постоянной γ.*

Доказательство. Если ввести обозначения

$$G_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^{u} x^2 d\bar{F}_{nk}(x)$$

и

$$\gamma_n = \sum_{k=1}^{k_n} \int x dF_{nk}(x),$$

то мы придем к условиям теоремы 5. Теорема этим доказана.

Несколько видоизменив формулировку последней теоремы, мы можем получить не только условия существования предельного закона, но также и условия сходимости к каждому данному предельному закону.

Теорема 9. *Для того чтобы функции распределения сумм* (1) *при* $n \to \infty$ *сходились к данной функции распределения* $\Phi(x)$ *и дисперсии сумм сходились к дисперсии предельного закона, необходимо и достаточно, чтобы при* $n \to \infty$ *выполнялись следующие условия:*

1) $\sum_{k=1}^{k_n} \int_{-\infty}^{u} x^2 d\bar{F}_{nk}(x) \to G(u)$

в точках непрерывности функции $G(u)$

2) $\sum_{k=1}^{k_n} \int x^2 d\bar{F}_{nk}(x) \to G(\infty)$,

3) $\sum_{k=1}^{k_n} \int x dF_{nk}(x) \to \gamma$,

где функции $G(u)$ *и постоянное* γ *определяются формулой* (1) § 43 *для функции* $\Phi(x)$.

§ 47. Условия сходимости к законам нормальному и Пуассона

Мы применим результаты предыдущего параграфа к выводу условий сходимости функций распределения сумм к законам нормальному и Пуассона.

Теорема 10. *Пусть дана элементарная система независимых случайных величин. Для того чтобы функции распределения сумм*

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n} \tag{1}$$

при $n \to \infty$ *сходились к закону*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx,$$

необходимо и достаточно, чтобы при $n \to \infty$ были выполнены условия

1) $\sum_{k=1}^{k_n} \int x dF_{nk}(x) \to 0$,

2) $\sum_{k=1}^{k_n} \int_{|x|>\tau} x^2 d\bar{F}_{nk}(x) \to 0$,

3) $\sum_{k=1}^{k_n} \int_{|x|<\tau} x^2 d\bar{F}_{nk}(x) \to 1$.

где τ – любая положительная постоянная.

Доказательство. Из теоремы 9 следует, что искомые условия состоят в выполнении при $n \to \infty$ соотношений

$$\sum_{k=1}^{k_n} \int x dF_{nk}(x) \to 0,$$

$$\sum_{k=1}^{k_n} \int_{-\infty}^{u} x^2 d\bar{F}_{nk}(x) \to \begin{cases} 0 & \text{для } u < 0, \\ 1 & \text{для } u > 0, \end{cases}$$

$$\sum_{k=1}^{k_n} \int x^2 d\bar{F}_{nk}(x) \to 1.$$

Первое из них совпадает с первым условием теоремы, равносильность двух остальных второму и третьему условиям теоремы очевидна.

Особенно простую форму эта теорема принимает, если элементарная система, рассматриваемая нами, нормирована заранее условиями

$$\sum_{k=1}^{k_n} \int x^2 dF_{nk}(x) = 1,$$

$$\int x dF_{nk}(x) = 0 \qquad (1 \leqslant k \leqslant k_n;\ n = 1, 2, \ldots). \tag{2}$$

Теорема 11. *Если элементарная система нормирована соотношениями* (2), *то для сходимости функций распределения сумм* (1) *к нормальному закону необходимо и достаточно, чтобы для всех $\tau > 0$ при $n \to \infty$*

$$\sum_{k=1}^{k_n} \int_{|x|>\tau} x^2 dF_{nk}(x) \to 0. \tag{3}$$

Доказательство теоремы очевидно.

Требование (3) носит название условия Линдеберга, так как им в 1923 г. была доказана его достаточность для сходимости функций распределения сумм к нормальному закону. В 1935 г. В. Феллером была доказана необходимость этого условия.

В качестве другого примера использования общих теорем предыдущего параграфа мы рассмотрим сходимость функций распределения элементарных систем к закону Пуассона

$$P(x)=\begin{cases}0 & \text{для } x\leqslant 0,\\ \sum\limits_{0\leqslant k<x} e^{-\lambda}\dfrac{\lambda^k}{k!} & \text{для } x>0.\end{cases} \tag{4}$$

Если ξ – случайная величина, распределенная по закону (4), то, как мы знаем $\mathsf{M}\xi=\mathsf{D}\xi=\lambda$.

Мы ограничимся элементарными системами, для которых

$$\left.\begin{aligned}&\sum_{k=1}^{k_n}\mathsf{M}\xi_{nk}\to\lambda,\\ &\sum_{k=1}^{k_n}\mathsf{D}\xi_{nk}\to\lambda.\end{aligned}\right\} \tag{5}$$

Т е о р е м а 12. *Пусть дана элементарная система, подчиненная условиям* (5). *Функции распределения сумм*

$$\zeta_n=\xi_{n1}+\xi_{n2}+\ldots+\xi_{nk_n}$$

тогда и только тогда сходятся к закону (4), *когда при любом* $\tau>0$

$$\sum_{k=1}^{k_n}\int\limits_{|x-1|>\tau}x^2dF_{nk}(x+\mathsf{M}\xi_{nk})\to 0\qquad (n\to\infty).$$

Доказательство этой теоремы мы предоставляем читателю.

В § 13 нами была доказана теорема Пуассона. Легко убедиться, что при $np_n=\lambda$ она является частным случаем только что доказанного предложения. Действительно, пусть ξ_{nk} $(1\leqslant k\leqslant n)$ есть случайная величина, принимающая значения 0 или 1 в зависимости от того, появится или не появится при k-м испытании n-й серии наблюдаемое нами событие A. При этом

$$\mathsf{P}\{\xi_{nk}=1\}=\frac{\lambda}{n}\quad\text{и}\quad\mathsf{P}\{\xi_{nk}=0\}=1-\frac{\lambda}{n}.$$

Очевидно, что сумма

$$\mu_n=\xi_{n1}+\xi_{n2}+\ldots+\xi_{nn}$$

представляет собой число появлений события A в n-й серии испытаний.

Согласно теореме Пуассона функции распределения величин μ_n при $n\to\infty$ сходятся к закону Пуассона (4). Этот результат следует и из только что сформулированной теоремы, так как все ее требования в данном случае выполнены.

Общие теоремы о сближении функций распределения сумм (1) с некоторыми безгранично-делимыми функциями распределения, доказанные в более широких, чем у нас, предположениях, позволяют также получить необходимое и достаточное условие для закона больших чисел (в случае независимых слагаемых). См. об этом уже упоминавшуюся монографию Б.В. Гнеденко и А.Н. Колмогорова.

§ 48. Суммирование независимых случайных величин в случайном числе

В разнообразных задачах практики приходится сталкиваться с задачей суммирования случайных величин не в заранее заданном, а в случайном числе. Приведем примеры.

За смену в магазин приходит случайное число покупателей и сумма, на которую каждый из них делает покупки, также случайна. Выручка магазина за смену является суммой случайного числа случайных слагаемых. Заметим, что число покупателей и суммы, которые они оставляют в магазине, представляют собой независимые случайные величины.

Возвратимся к задаче, рассмотренной в § 38. Там речь шла о длительности безотказной работы дублированной системы с восстановлением. Внимательно разберемся в структуре величины ζ – длительности безотказной работы резервированной системы. Заметим, что ζ в зависимости от случая может принимать различные значения, а именно:

$\zeta = \xi_1 + \xi_2$ – сумма времен безотказной работы основного и резервного элементов, если $\eta_1 > \xi_2$ (восстановление основного элемента продолжалось дольше, чем проработал резервный элемент);

$\zeta = \xi_1 + \xi_2 + \xi_3$, если $\eta_1 < \xi_2$, но $\eta_2 > \xi_3$, т.е. восстановленный основной элемент проработал меньше, чем длился ремонт резервного элемента;

вообще при $n \geqslant 2$ будет $\zeta = \xi_1 + \xi_2 + \ldots + \xi_n$ с вероятностью $\alpha^{n-2}(1-\alpha)$ (первые $n-2$ раз элемент восстанавливался прежде, чем работающий элемент отказывал, в последний же раз ремонт занял больше времени, чем проработал вступивший в работу элемент).

Мы вновь видим, что длительность безотказной работы дублированной системы с восстановлением представляет собой сумму независимых случайных величин в случайном числе. Число слагаемых не зависит от значений, принимаемых слагаемыми и распределено по геометрическому закону.

Теорема о сходимости к показательному распределению после только что сделанного замечания должна нами восприниматься как предельная теорема о сходимости функции распределения последовательных сумм случайного числа случайных слагаемых. Вот какую формулировку ей следует придать.

Т е о р е м а. *Пусть случайные величины последовательности* $\{\xi_n\}$ *независимы, одинаково распределены с функцией распределения* $F(x)$

и математическим ожиданием $a > 0$. *Пусть, далее,* $\{\nu_n\}$ – *последовательность целочисленных случайных величин, причем* ν_n *имеет геометрическое распределение с параметром* α_n: $\mathsf{P}\{\nu_n = k\} = (1 - \alpha_n)\alpha_n^{k-1}$. *Если* $\alpha_n \to 1$, *то функции распределения сумм*

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_{\nu_n}$$

сходятся к показательному распределению.

Рассмотрим следующую задачу: счетчик Гейгера (см. задачу 4, § 8) начинает свою работу в момент времени 0. Найти время его работы до первого пропуска частицы. Мы сейчас несколько обобщим условия задачи по сравнению с тем как она была поставлена в гл. 1. Предположим, что промежутки времени ξ_k между поступлениями последовательных частиц в счетчик независимы и имеют распределение $F(x)$. Длительность разряда, вызываемого зарегистрированной частицей является случайной величиной η_k с распределением $G(x)$. Все рассматриваемые и только что упомянутые величины независимы между собой.

Обозначим через ζ_n длительность искомого промежутка времени. Легко подсчитать, что $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_\nu$, где ν – случайная величина, независимая от ξ_k и распределенная геометрически.

Рассмотрим теперь последовательность $\{\xi_{nk}\}$ независимых и одинаково распределенных при каждом n случайных величин. Введем обозначения

$$F_n(x) = \mathsf{P}\{\xi_{nk} < x\}, \qquad f_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x),$$

Рассмотрим далее неограниченно возрастающую последовательность целых положительных чисел $\{k_n\}$ и последовательность $\{\nu_n\}$ целочисленных положительных случайных величин, которые обладают тем свойством, что при каждом n величины ν_n независимы от ξ_{nk}, $1 \leqslant k < \infty$.

Наша цель состоит в выяснении условий, при которых сходимость функций распределения сумм (при $n \to \infty$)

$$s_n = \sum_{k=1}^{k_n} \xi_{nk}$$

влечет за собой сходимость функций распределения сумм $(n \to \infty)$

$$s_{\nu_n} = \sum_{k=1}^{\nu_n} \xi_{nk}.$$

Предложение, которое сейчас будет доказано, носит название теоремы переноса.

Теорема переноса. *Если при* $n \to \infty$

А) $\mathbf{P}\{\sum_{k=1}^{k_n} \xi_{nk} < x\} \to \Phi(x)$

и

Б) $\mathbf{P}\left\{\frac{\nu_n}{k_n} < x\right\} \to A(x)$

(*причем будем считать, что* $A(+0) = 0.$), *то*

В) $\mathbf{P}\{\sum_{k=1}^{\nu_n} \xi_{nk} < x\} \to \Psi(x).$

Функция распределения $\Psi(x)$ *определяется через свою характеристическую функцию* $\psi(t)$ *посредством формулы*

$$\psi(t) = \int_0^\infty \varphi^z(t)\ A(z),$$

где $\varphi(t)$ – *характеристическая функция для* $\Phi(x)$.

Доказательство. Характеристическая функция суммы $\sum_{k=1}^{\nu_n} \xi_{kn}$ равна

$$\psi_n(t) = \sum_{j=0}^{\infty} p_{nj}(f_n(t))^j,$$

где $p_{nj} = \mathbf{P}\{\nu_n = j\}$.

Положим $A_n(x) = \mathbf{P}\{\nu_n < x\}$; тогда очевидно, что

$$\psi_n(t) = \int_0^\infty f_n^x(t)\,dA_n(x).$$

Пусть теперь

$$\bar{A}_n(z) = \mathbf{P}\left\{\frac{\nu_n}{k_n} < x\right\} = A_n(k_n z),$$

тогда

$$\psi_n(t) = \int_0^\infty [f_n^{k_n}(t)]^z d\bar{A}_n(z).$$

В математическом анализе известна теорема *), согласно которой, если последовательность равностепенно ограниченных непрерывных функций $f_n(x)$ сходится к функции $f(x)$ во всех точках прямой, а монотонные ограниченные функции $A_n(x)$ при всех x сходятся к функции $A(x)$, то при $n \to \infty$

$$\int_0^\infty f_n(x) dA_n(x) \to \int_0^\infty f(x) dA(x).$$

В силу этой теоремы и условий А) и В) теоремы переноса при $n \to \infty$

$$\psi_n(t) \to \psi(t) = \int_0^\infty \varphi^z(t) dA(z).$$

Теорема доказана.

Укажем теперь некоторые следствия из теоремы переноса.

Следствие 1. *Пусть функция* $\Phi(x)$ – *безгранично делима и* $\varphi(t)$ – *ее характеристическая функция; тогда функция*

$$\psi(t) = \frac{1}{1 - \ln \varphi(t)}$$

является также характеристической (и, как позднее будет показано, также безгранично делимой).

Действительно, мы знаем, что для любой безгранично делимой функции $\Phi(x)$ можно найти такую последовательность $\{k_n\}$ и независимых величин ξ_{nk}, что выполняется условие А) теоремы переноса. Выберем теперь такие ν_n, чтобы $A(x) = 1 - e^{-x}$. Это можно сделать многими способами, например, выбрав ν_n распределенными геометрически с соответственным значением параметра. Этим самым выполнено и условие Б) теоремы переноса.

---

*) См. Дубровский В.М. О некоторых свойствах вполне аддитивных функций множества и их предельном переходе под знаком интеграла // Изв. АН СССР Серия мат. – 1945. – Т. 9. – С. 311 – 320; 1947. – Т. 11. – С. 101 – 104.

Тогда мы знаем, что, в силу В), функция

$$\psi(t) = \int_0^\infty [\varphi(t)]^z dA(z) = \int_0^\infty e^{-z(1-\ln\varphi(t))}dz = \frac{1}{1-\ln\varphi(t)}$$

является характеристической.

В частности, если функция $\Phi(x)$ является нормальной $(\varphi(t) = e^{-t^2/2})$, то $\psi(t) = \dfrac{2}{2+t^2}$. Соответствующая ей функция распределения $\Psi(x)$ определяется формулой

$$\Psi(x) = \begin{cases} 0{,}5\, e^{x\sqrt{2}} & \text{при} \quad x \leqslant 0, \\ 1 - 0{,}5\, e^{-x\sqrt{2}} & \text{при} \quad x \geqslant 0. \end{cases}$$

Обычно распределение Лапласа определяют, указывая его плотность распределения. В нашем случае она равна $p(x) = e^{-|x|\sqrt{2}}$.

Позднее мы увидим, что при $A(x) = 1 - e^{-x}$ распределение Лапласа при суммировании до случайного индекса играет такую же роль, как и нормальное распределение в классической постановке предельных теорем для сумм одинаково распределенных независимых слагаемых.

Следствие 2. *Условия, при которых функции распределения сумм*

$$s_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nk_n}$$

одинаково распределенных независимых слагаемых сходятся к предельному распределению $\Phi(x)$ достаточны для того, чтобы функции распределения сумм

$$s_{\nu_n} = \xi_{n1} + \xi_{n2} + \ldots + \xi_{n\nu_n}$$

сходились к распределению $\Psi(x)$.

Следствие непосредственно вытекает из теоремы переноса.

Следствие 3. *Пусть $\{\xi_n\}$ – последовательность одинаково распределенных независимых случайных величин и $\xi_{nk} = \dfrac{\xi_k - a}{B_n}$, где действительные постоянные a и $B_n > 0$ таковы, что функции распределения сумм*

$\bar{s}_n = \frac{1}{B_n} \sum_{k=1}^{k_n} (\xi_k - a)$ *сходятся к* $\Phi(x)$. *Пусть, далее, выполнено условие* Б) *теоремы переноса. Тогда при сделанном выборе постоянных* a *и* B_n *функции распределения сумм* $\bar{s}_{\nu_n} = \frac{1}{B_n} \sum_{k=1}^{\nu_n} (\xi_k - a)$ *также сходятся к предельной* $\Psi(x)$.

Следствие непосредственно вытекает из теоремы переноса.

Заслуживает внимания то обстоятельство, что при всех возможных предельных распределениях $A(x)$ нормирующие множители B_n и центрирующие коэффициенты a могут быть выбраны раз и навсегда одинаковыми.

Замечание В. Феллера. На с. 642 второго тома замечательной книги В.Феллера "Введение в теорию вероятностей и ее приложения" (М.: Мир, 1984) имеется такое замечание: если в формуле (1) функция $A(x)$ безгранично делима, то и функция $\Psi(x)$ безгранично делима.

Доказательство этого факта вытекает из того факта, что если случайная величина ν с функцией распределения $A(x)$ представима в виде суммы $\nu = \nu_1 + \nu_2$ независимых случайных величин ν_1 и ν_2, то $\psi(t) = \psi_1(t)\psi_2(t)$ из определения безграничной делимости.

Упражнения

1. Доказать, что распределения

а) Паскаля (упр. 1а к гл. 5),

б) Полиа (упр. 1б к гл. 5),

в) Коши (пример 5 § 24)

безгранично делимы.

2. Доказать, что случайная величина с плотностью распределения

$$p(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{при} \quad x > 0. \end{cases}$$

где $\alpha > 0$, $\beta > 0$ – постоянные, безгранично делима.

Примечание. Отсюда, в частности, следует, что безгранично делимы распределения Максвелла и распределение χ^2 (при любом значении n).

3. Доказать, что каковы бы ни были постоянные $\alpha > 0$ и $\beta > 0$,

$$\varphi(t) = \left(1 + \frac{t^2}{\beta^2}\right)^{-\alpha}$$

является безгранично делимой характеристической функцией.

П р и м е ч а н и е. Отсюда, в частности, следует, что распределение Лапласа (упр. 6 к гл. 5) безгранично делимо.

4. Найти функцию $G(x)$ и параметр γ в формуле Колмогорова для логарифма безгранично делимой характеристической функции для распределений:

а) примера 2,

б) Лапласа.

5. Доказать, что если сумма двух независимых безгранично делимых случайных величин распределена:

а) по закону Пуассона,

б) по нормальному закону,

то каждое слагаемое распределено в случае:

а) по закону Пуассона,

б) по нормальному закону.

6. Найти условия сходимости функций распределения сумм случайных величин, составляющих элементарную систему, к распределению:

а) примера 2,

б) Лапласа.

ГЛАВА 10

ТЕОРИЯ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ

§ 49. Вводные замечания

Совершенствование физической статистики, а также ряда отраслей техники, поставило перед теорией вероятностей большое число новых, не укладывающихся в рамки классической теории, задач. В то время как физика и техника интересовало изучение *процесса*, т. е. явления, протекающего во времени, теория вероятностей не имела ни общих приемов, ни разработанных частных схем для решения задач, возникающих при изучении таких явлений. Появилась настоятельная необходимость в разработке общей теории случайных процессов, т. е. теории, которая изучала бы случайные величины, зависящие от одного или нескольких непрерывно изменяющихся параметров.

Перечислим несколько задач, иллюстрирующих необходимость построения теории случайных процессов.

Представим себе, что мы задались целью проследить за движением какой-либо молекулы газа или жидкости. Эта молекула в случайные моменты времени сталкивается с другими молекулами и меняет при этом свои скорость и положение. Состояние молекулы, таким образом, подвержено случайным изменениям в каждый момент времени. Многие физические явления требуют для своего изучения умения вычислять вероятности того, что определенное число молекул успеет за тот или иной промежуток времени переместиться на то или иное расстояние. Так, например, если приведены в соприкосновение два газа или две жидкости, то начинается взаимное проникновение молекул одной жидкости в другую: происходит диффузия. Как быстро протекает процесс диффузии, по каким законам, когда образующаяся смесь становится практически однородной? На все эти и многие другие вопросы дает ответ статистическая теория диффузии, в основе которой лежит теория случайных процессов, или, как принято теперь говорить, *теория стохастических процессов*. Очевидно, что подобная же задача возникает в химии, когда изучают процесс химической реакции. Какая часть молекул уже вступила в реакцию, как реакция протекает во времени, когда практически реакция уже закончилась?

Весьма важный круг явлений протекает по принципу радиоактивного распада. Это явление состоит в том, что атомы радиоактивного вещества распадаются, превращаясь в атомы другого элемента. Распад каждого атома происходит мгновенно, подобно взрыву, с выделением некоторого количества энергии. Многочисленные наблюдения показывают, что распад различных атомов для наблюдателя происходит в случайно взятые моменты времени. При этом расположение этих моментов времени независимо друг от друга в смысле теории вероятностей. Для изучения процесса радиоактивного распада существенно определить, какова вероятность того, что за определенный промежуток времени распадется то или иное количество атомов? Формально, если задаваться только выяснением математической картины явлений, точно так же протекают и другие явления: число вызовов, поступающих на телефонную станцию за определенный промежуток времени (загрузка телефонной станции), обрывность нитей на ватере (ватер – прядильная машина) или изменение числа частиц, находящихся в броуновском движении, оказавшихся в какой-либо момент времени в заданной области пространства. Мы дадим в этой главе простое решение тех математических задач, к которым приводят указанные явления.

К тому, с чем мы только что познакомились, добавим следующее: первые задачи физического характера, являющиеся одновременно задачами теории случайных процессов, были рассмотрены выдающимися физиками начала нашего века. Изложим сейчас вкратце, как, исходя из рассмотрения весьма схематической проблемы блуждания по прямой, Максом Планком и Фоккером было получено дифференциальное уравнение теории диффузии. Пусть частица, находящаяся в момент времени $t = 0$ в точке $x = 0$, в моменты $k\tau$ $(k = 1, 2, \ldots)$ испытывает случайные толчки, в результате которых она каждый раз перемещается с вероятностью p на величину h вправо и с вероятностью $q = 1 - p$ также на величину h влево. Обозначим через $f(x, t)$ вероятность того, что частица в результате n толчков окажется в момент t $(t = n\tau)$ в положении x (ясно, что при четном числе толчков величина x может равняться только четному числу шагов h, а при n нечетном – только нечетному числу шагов h). Если через m обозначить число шагов, сделанных частицей вправо (соответственно $n - m$ есть число шагов, которые частица совершила влево), то согласно формуле Бернулли

$$f(x, t) = C_n^m p^m q^{n-m}.$$

Ясно, что величины m, n, x и h связаны равенством

$$m - (n - m) = x/h.$$

Легко убедиться непосредственным подсчетом, что функция $f(x, t)$ удовлетворяет разностному уравнению

$$f(x, t + \tau) = pf(x - h, t) + q(x + h, t) \tag{1}$$

и начальным условиям

$$f(0,0) = 1, \quad f(x,0) = 0 \quad \text{при} \quad x \neq 0.$$

Посмотрим, во что превратится написанное разностное уравнение, если заставить стремиться к 0 как h, так и τ. Физическая природа заставит, оказывается, наложить на h и τ некоторые ограничения. Точно также величины p и q не могут быть взяты произвольно. Несоблюдение условий, о которых пойдет речь, может привести к тому, что за конечный промежуток времени частица с вероятностью единица уйдет в бесконечность. Для того чтобы избежать такую возможность, наложим следующие требования: при $h \to \infty$

$$x = nh, \quad t = n\tau, \quad \frac{h^2}{\tau} \to 2D, \quad \frac{p-q}{h} \to \frac{c}{D}, \tag{2}$$

где c и D – некоторые постоянные. Величина c носит наименование *скорости течения*, а D – *коэффициента диффузии.*

Отнимем от обеих частей равенства (1) величину $f(x, t)$. В результате получаем

$$f(x, t+\tau) - f(x, t) = p[f(x-h, t) - f(x, t)] +$$
$$+ q[f(x+h, t) - f(x, t)]. \tag{3}$$

Предположим, что $f(x, t)$ дифференцируема по t и дважды дифференцируема по x. Тогда

$$f(x, t+\tau) - f(x, t) = \tau \frac{\partial f(x, t)}{\partial t} + o(\tau),$$

$$-f(x, t) + f(x-h, t) = -h \frac{\partial f(x, t)}{\partial x} + \frac{1}{2} h^2 \frac{\partial^2 f(x, t)}{\partial x^2} + o(h^2),$$

$$f(x+h, t) - f(x, t) = h \frac{\partial f(x, t)}{\partial x} + \frac{1}{2} h^2 \frac{\partial^2 f(x, t)}{\partial x^2} + o(h^2).$$

После подстановки этих равенств в (3) получаем

$$\tau \frac{\partial f(x, t)}{\partial t} + o(\tau) = -(p-q)h \frac{\partial f(x, t)}{\partial x} + \frac{h^2}{2} \frac{\partial^2 f(x, t)}{\partial x^2} + o(h^2).$$

Отсюда, в силу соотношений (2), находим, что в пределе

$$\frac{\partial f(x, t)}{\partial t} = -2c \frac{\partial f(x, t)}{\partial x} + D \frac{\partial^2 f(x, t)}{\partial t^2}.$$

Мы получили уравнение, носящее в теории диффузии наименование *уравнения Фоккера – Планка*.

Интересно отметить, что при довольно искусственной постановке задачи получен физически осмысленный результат, хорошо отражающий истинную картину процесса диффузии. Позднее мы дадим вывод общих уравнений, которым подчиняются распределения для случайных процессов при весьма общих предположениях об их протекании.

Начало общей теории стохастических процессов было положено фундаментальными работами советских математиков А.Н. Колмогорова и А.Я. Хинчина в начале тридцатых годов. В статье А.Н. Колмогорова "Об аналитических методах в теории вероятностей" было дано систематическое и строгое построение основ теории стохастических процессов без последействия или, как часто говорят, процессов марковского типа. В ряде работ А.Я. Хинчина была создана теория так называемых стационарных процессов.

Заметим, что прежде чем подвергнуть математическому изучению те или иные явления природы или технические процессы, нужно их схематизировать. Причина этой необходимости лежит в том, что математический анализ применим к исследованию процесса изменения некоторой системы только в том случае, если предположено, что каждое возможное состояние этой системы вполне определено посредством некоторого определенного математического аппарата. Понятно, что такая математически определимая система не есть сама действительность, но лишь схема, пригодная для ее описания. С такой картиной мы встречаемся, скажем, в механике, когда предполагаем, что реальные движения систем материальных точек полностью могут быть описаны для любого момента времени указанием этого момента времени и ее состояния в любой предыдущий момент времени t_0. Иными словами, схема, которая принимается в теоретической механике для описания движения, состоит в следующем: принимается, что для любого момента времени t состояние системы y полностью определяется ее состоянием x в любой предыдущий момент времени t_0. При этом под состоянием системы в механике понимается задание положения точек материальной системы и их скоростей.

Вне классической механики, собственно, во всей современной физике, приходится иметь дело с более сложным положением, когда знание состояния системы в какой-либо момент времени t_0 уже не определяет однозначно состояния системы в последующие моменты времени, а лишь определяет вероятность того, что система будет находиться в одном из состояний некоторого множества состояний системы. Если через x обозначить состояние системы в момент t_0, а через E – некоторое множество состояний системы, то для только что описанных процессов определена вероятность

$$\mathsf{P}\{t_0, x; t, E\}$$

системе, находящейся в момент t_0 в состоянии x, в момент t перейти в одно из состояний множества E.

Если дополнительное знание состояний системы в моменты $t < t_0$ не изменяет этой вероятности, то естественно назвать выделенный нами класс случайных процессов процессами без последействия или за их аналогию с цепями Маркова – процессами марковского типа.

Общее понятие случайного процесса, базирующееся на изложенной ранее аксиоматике теории вероятностей, может быть введено следующим образом. Пусть Ω – множество элементарных событий и t – непрерывный параметр. *Случайным процессом* называется функция двух аргументов

$$\xi(t) = \varphi(\omega, t) \qquad (\omega \in \Omega).$$

Для каждого значения параметра t функция $\varphi(\omega, t)$ является функцией только ω и, следовательно, представляет собой случайную величину. Для каждого фиксированного значения ω (т.е. для каждого заданного элементарного события) $\varphi(\omega, t)$ зависит только от t и является, таким образом, обычной функцией одного вещественного аргумента. Каждая такая функция называется *реализацией случайного процесса* $\xi(t)$. На случайный процесс можно смотреть либо как на совокупность случайных величин $\xi(t)$, зависящих от параметра t, либо как на совокупность реализаций процесса $\xi(t)$. Естественно, что при этом для определения случайного процесса необходимо задать вероятностную меру в пространстве реализаций процесса.

Почти вся настоящая глава будет посвящена изучению процессов без последействия и только в последнем параграфе мы дадим представление о стационарных процессах.

§ 50. Процесс Пуассона

Мы начнем краткое знакомство с некоторыми фактами теории случайных процессов с рассмотрения одного важного примера процесса без последействия, играющего большую роль в ряде вопросов физики, теории связи, теории надежности. По-видимому, впервые этот процесс был подвергнут исследованию в начале нашего столетия физиками А. Эйнштейном и М. Смолуховским в связи с задачами броуновского движения.

Предположим, что в случайные моменты времени происходит некоторое событие. Нас интересует число появлений этого события в промежуток времени от 0 до t. Обозначим это число через $\xi(t)$. Относительно процесса появления события мы предположим, что он: 1) стационарен, 2) без последействия и 3) ординарен. В перечисленные предположения вкладывается следующий смысл.

1. Стационарность означает, что для любой группы из конечного числа непересекающихся промежутков времени вероятность наступления определенного числа событий на протяжении каждого из них зависит только от этих чисел и от длительности промежутков времени, но не изменяется от сдвига всех отрезков времени на одну и ту же величину. В частности, вероятность появления k событий в течение промежутка времени от τ до $\tau + t$ не зависит от τ и является функцией только k и t.

2. Отсутствие последействия означает, что вероятность наступления k событий в течение промежутка времени $(\tau, \tau + t)$ не зависит от того, сколько раз и как появлялись события ранее. Это предположение означает, что условная вероятность появления событий за промежуток времени $(\tau, \tau + t)$ при любом предположении о наступлении событий до момента τ совпадает с безусловной вероятностью. В частности, отсутствие последействия означает взаимную независимость появления того или иного числа событий в непересекающиеся промежутки времени.

3. Ординарность выражает собой требование практической невозможности появления двух или нескольких событий за малый промежуток времени Δt. Обозначим через $P_{>1}(\Delta t)$ вероятность появления более чем одного события за промежуток времени Δt. Тогда условие ординарности в точном его выражении состоит в следующем:

$$P_{>1}(\Delta t) = o(\Delta t).$$

Наша ближайшая задача состоит в определении вероятностей $P_k(t)$ того, что за промежуток длительности t произойдут k событий. В силу сделанных предположений эти вероятности не зависят от того, где расположен этот отрезок времени. С этой целью обнаружим, что при малых Δt имеет место равенство

$$P_1(\Delta t) = \lambda \Delta t + o(\Delta t),$$

где λ – постоянное.

Действительно, рассмотрим промежуток времени длительности 1 и обозначим через p вероятность того, что за этот срок не наступит ни одно событие. Разобьем наш промежуток на n равных непересекающихся частей. В силу первого и второго предположений имеет место равенство $p = [P_0(1/n)]^n$, откуда $P_0(1/n) = p^{1/n}$. Отсюда при любом целом k

$$P_0(k/n) = p^{k/n}.$$

Пусть теперь t – некоторое неотрицательное число. При любом n можно найти такое k, что $(k-1)/n \leqslant t < k/n$. Так как вероятность $P_0(t)$ есть убывающая функция времени, то

$$P_0\left(\frac{k-1}{n}\right) \geqslant P_0(t) \geqslant P_0\left(\frac{k}{n}\right).$$

Таким образом, $P_0(t)$ удовлетворяет неравенствам

$$p^{(k-1)/n} \geqslant P_0(t) \geqslant p^{k/n}.$$

Пусть теперь k и n стремятся к бесконечности так, чтобы

$$\lim_{n \to \infty} \frac{k}{n} = t.$$

Из предыдущего ясно, что при любом t

$$P_0(t) = p^t.$$

Так как $P_0(t)$, как вероятность, удовлетворяет неравенствам $0 \leqslant P_0(t) \leqslant 1$, то могут представиться три следующих случая: 1) $p = 0$, 2) $p = 1$, 3) $0 < p < 1$. Первые два случая малоинтересны. В первом из них при любом t имеет место равенство $P_0(t) = 0$ и, значит, вероятность за промежуток времени любой длительности произойти хотя бы одному событию равна единице. Другими словами, с вероятностью единица за промежуток времени любой длительности происходит бесконечно много событий. Во втором случае $P_0(t) = 1$ и, следовательно, события не наступают. Интерес представляет лишь третий случай, в котором положим $p = e^{-\lambda}$, где λ – некоторое положительное число ($\lambda = -\ln p$).

Итак, из предположений стационарности и отсутствия последействия мы вывели, что при любом $t > 0$

$$P_0(t) = e^{-\lambda t}. \tag{1}$$

Понятно, что при любом t имеет место равенство

$$P_0(t) + P_1(t) + P_{>1}(t) = 1.$$

Из (1) вытекает, что при малых t

$$P_0(t) = 1 - \lambda t + o(t).$$

Следовательно, в силу условия ординарности,

$$P_1(t) = \lambda t + o(t). \tag{2}$$

Теперь мы можем перейти к выводу формул для вероятностей $P_k(t)$ при $k \geqslant 1$. С этой целью определим вероятность того, что за время $t + \Delta t$ событие наступит ровно k раз. Это может осуществиться $k + 1$ различными способами, а именно:

1) за промежуток времени длительности t произойдут все k событий, а за время Δt – ни одного;

2) за время t произойдут $k - 1$ событие, а за время Δt – одно; ...

$k + 1$) за время t событие не наступит ни раза, а за время Δt произойдет k раз.

По формуле полной вероятности

$$P_k(t + \Delta t) = \sum_{j=0}^{k} P_j(t) P_{k-j}(\Delta t)$$

(при этом принято во внимание как условие стационарности, так и условие отсутствия последействия). Положим

$$R_k = \sum_{j=0}^{k-2} P_j(t) P_{k-j}(\Delta t).$$

Очевидно, что

$$R_k \leqslant \sum_{j=0}^{k-2} P_{k-j}(\Delta t) = \sum_{s=2}^{k} P_s(\Delta t) \leqslant$$

$$\leqslant \sum_{s=2}^{\infty} P_s(\Delta t) = P_{>1}(\Delta t) = o(\Delta t),$$

согласно условию ординарности.

Таким образом,

$$P_k(t + \Delta t) = P_k(t) P_0(\Delta t) + P_{k-1}(t) P_1(\Delta t) + o(\Delta t).$$

Далее, согласно (2)

$$P_0(\Delta t) = e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(\Delta t), \quad P_1(\Delta t) = \lambda \Delta t + o(\Delta t),$$

поэтому

$$P_k(t + \Delta t) = (1 - \lambda \Delta t) P_k(t) + \lambda \Delta t P_{k-1}(t) + o(\Delta t).$$

Отсюда

$$\frac{P_0(t + \Delta t) - P_k(t)}{\Delta t} = -\lambda P_k(t) + \lambda P_{k-1}(t).$$

Поскольку при $\Delta t \to 0$ предел правой части равенства существует, существует и предел левой части. В результате получаем уравнение

$$\frac{dP_k(t)}{dt} = -\lambda P_k(t) + \lambda P_{k-1}(t) \tag{3}$$

для определения $P_k(t)$. Начальные условия мы выберем такие:

$$P_0(0) = 1, \quad P_k(0) = 0 \quad \text{при} \quad k \geqslant 1. \tag{4}$$

Решение системы уравнений (3) проще всего осуществить посредством замены

$$P_k(t) = e^{-\lambda t} v_k(t) \tag{5}$$

где $v_k(t)$ – новая искомая функция. Заметим, что, в силу (1), $v_0(t) = 1$. Соотношения (4) приводят нас к таким начальным условиям:

$$v_0(0) = 1 \quad \text{и} \quad v_k(0) = 0 \quad \text{при} \quad k \geqslant 1. \tag{6}$$

Подстановка (5) в (3) приводит нас к уравнению

$$\frac{dv_k(t)}{dt} = \lambda v_{k-1}(t). \tag{7}$$

В частности,

$$\frac{dv_1(t)}{dt} = \lambda. \tag{7'}$$

Последовательное решение уравнений (7′) и (7) приводит нас при учете начальных условий к равенствам:

$$v_1(t) = \lambda t, \quad v_2(t) = \frac{(\lambda t)^2}{2}, \quad v_3(t) = \frac{(\lambda t)^3}{3!}$$

и вообще

$$v_k(t) = \frac{(\lambda t)^k}{k!}.$$

Таким образом, окончательно

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{8}$$

при любом $k \geqslant 0$. Задача, стоявшая перед нами, решена.

Высказанные нами в начале параграфа условия с большой точностью выполняются в многочисленных естественнонаучных явлениях и технических процессах. Для примера укажем на число спонтанно распавшихся атомов радиоактивного вещества за тот или иной промежуток времени (когда этого вещества не слишком мало и не слишком много); на число космических частиц, попавших на определенную площадку за промежуток времени t. Если мы имеем дело с какой-нибудь сложной радиотехнической системой, состоящей из большого числа элементов, каждый из которых лишь с малой вероятностью может отказать в работе за единицу времени независимо от состояния других элементов, то число элементов, отказавших за промежуток времени $(0, t)$, представляет собой случайный процесс. Этот процесс во многих случаях будет хорошо описываться только что рассмотренным процессом Пуассона. Число подобных примеров можно увеличивать без всяких затруднений.

Промежуток времени, протекший между появлениями двух последовательных появлений интересующих нас событий, представляет собой случайную величину, которую мы обозначим через τ. Найдем распределение вероятностей τ. Так как очевидно, что событие $\tau > t$ эквивалентно тому, что за промежуток времени t событие не появится ни разу, то

$$\mathsf{P}\{\tau > t\} = e^{-\lambda t}.$$

Искомая функция распределения, таким образом, задается формулой

$$\mathsf{P}\{\tau < t\} = 1 - e^{-\lambda t}. \tag{9}$$

Полученный результат мы можем физически трактовать многими способами. Например, мы можем смотреть на него как на распределение времени свободного пробега молекулы или как на распределение времени, протекшее между двумя отказами элементов в сложной радиотехнической схеме.

Заметим, что теория, развитая в настоящем параграфе, может быть применена не только в предположении, что параметр t играет роль времени. С этой целью рассмотрим дополнительный пример.

П р и м е р. В пространстве разбросаны точки с соблюдением следующих требований:

1) вероятность k точкам оказаться в области G зависит только от объема v этой области, но не зависит ни от ее формы, ни от ее положения в пространстве; эту вероятность обозначим $p_k(t)$;

2) числа точек, попавших в неперекрывающиеся области, являются независимыми случайными величинами;

3) $\sum\limits_{k=2}^{\infty} p_k(\Delta v) = o(\Delta v)$.

Наложенные условия являются ничем иным, как условиями стационарности, отсутствия последействия и ординарности. Отсюда

$$p_k(v) = \frac{(av)^k}{k!} e^{-av}.$$

Если в жидкости взвешены мельчайшие частицы какого-либо вещества; то под влиянием ударов окружающих молекул эти частицы находятся в непрерывном хаотическом движении (броуновское движение). В результате в каждый момент времени мы имеем случайное распределение частиц в пространстве, о чем только что была речь. Согласно теории настоящего примера следует считать, что распределение частиц, попадающих в некоторую определенную область, будет подчинено закону Пуассона.

В таблице 14 сравниваются результаты опыта с частицами золота, взвешенными в воде, заимствованные нами из статьи Смолуховского, и результаты вычислений по закону Пуассона.

Таблица 14

Число частиц	Число наблюдавшихся случаев	Частота $\frac{m}{518}$	$\frac{\lambda^n e^{-\lambda}}{n!}$	Вычисленное число случаев
0	112	0,216	0,213	110
1	168	0,325	0,328	173
2	130	0,251	0,253	131
3	69	0,133	0,130	67
4	32	0,062	0,050	26
5	5	0,010	0,016	8
6	1	0,002	0,004	2
7	1	0,002	0,001	1

Постоянное $\lambda = av$, которым определяется закон Пуассона, выбрано равным среднему арифметическому из наблюдавшегося числа частиц, т.е.

$$\lambda \approx \frac{0.112 + 1.168 + 2.130 + 3.69 + 4.32 + 5.5 + 6.1 + 7.1}{518} \approx 1,54.$$

§ 51. Процессы гибели и размножения

В начале текущего столетия в связи с задачами биологии и телефонной связи возникла простая, но весьма полезная схема, получившая наименование процессов гибели и размножения. В качестве весьма частного случая она включает в себя задачу предшествующего параграфа о процессе Пуассона. Несмотря на узость исходных предположений процессов гибели и размножения, они находят широкое применение в ряде прикладных задач, позволяя получить не только схематическое представление о происходящих изменениях системы, но и расчетные формулы.

Представим себе, что интересующая нас система может находиться в одном из состояний E_0, E_1, E_2, ..., множество которых конечно или счетно. Со временем состояния системы изменяются, причем за промежуток длительности h система из состояния E_n переходит в состояние E_{n+1} с вероятностью $\lambda_n h + o(h)$ и в состояние E_{n-1} с вероятностью $\nu_n h + o(h)$. Вероятности того, что за промежуток $(t, t+h)$ система перейдет в состояние $E_{n\pm k}$ с $k > 1$, бесконечно малы по сравнению с h. Отсюда следует, что вероятность остаться в том же состоянии E_n за промежуток времени h равна $1 - \lambda_n h - \nu_n h + o(h)$. Постоянные λ_n и μ_n мы предполагаем зависящими от n, но не зависящими от t и от того, каким путем система пришла в это состояние. Последнее обстоятельство означает, что рассматриваемый процесс является марковским. Теория, которая будет здесь изложена, может быть распространена и на тот случай, когда λ_n и ν_n зависят так же и от t.

Случайный процесс, о котором только что шла речь, носит название *процесса гибели и размножения*. Если под E_n понимать событие, состоящее в том, что численность популяции равна n, то переход $E_n \to E_{n+1}$ означает, что численность популяции увеличивается на единицу. Точно так же на переход $E_n \to E_{n-1}$ следует смотреть как на гибель одного члена популяции.

Если при любом $n \geqslant 1$ имеют место равенство $\nu_n = 0$, т.е. если возможны только переходы $E_n \to E_n$ или $E_n \to E_{n+1}$ в момент изменения состояния, то процесс называется *процессом размножения* (иногда говорят о *процессе чистого размножения*; именно таким является процесс Пуассона). Если же все $\lambda_n = 0$, то говорят, что имеет место *процесс гибели*.

Обозначим через $p_k(t)$ вероятность того, что изучаемая нами система в момент t находится в состоянии E_k. Рассуждениями, подобными тем, которые мы провели в предыдущем параграфе, мы придем к системе уравнений, управляющей процессом гибели и размножения

$$p_0'(t) = -\lambda_0 p_0(t) + \nu_1 p_1(t) \tag{1}$$

и при $k \geqslant 1$

$$p_k'(t) = -(\lambda_k + \nu_k) p_k(t) + \lambda_{k-1} p_{k-1}(t) + \nu_{k+1} p_{k+1}(t). \tag{2}$$

Наши обозначения несколько неполны, поскольку мы не указываем, из какого состояния E_i начала изменяться система. Исчерпывающим было бы такое обозначение: $p_{ij}(t)$ — вероятность того, что система окажется в момент t в состоянии E_j, если в момент 0 она находилась в состоянии E_i. В задаче о процессе Пуассона мы предположили, что в начальный момент 0 система находилась в состоянии E_0.

Уравнения (1) и (2) принимают особенно простой вид для процессов чистой гибели и чистого размножения. Во втором случае, проведя последовательное интегрирование, получим (формулы написаны в предположении, что все λ_n различны)

$$p_0(t) = e^{-\lambda_0 t},$$

$$p_1(t) = \frac{\lambda_0}{\lambda_1 - \lambda_0}\,[e^{-\lambda_0 t} - e^{-\lambda_1 t}],$$

$$p_2(t) = \frac{\lambda_0\lambda_1}{\lambda_1 - \lambda_0}\left[\frac{1}{\lambda_2 - \lambda_0}(e^{-\lambda_0 t} - e^{-\lambda_2 t}) + \frac{1}{\lambda_2 - \lambda_1}(e^{-\lambda_1 t} - e^{-\lambda_2 t})\right].$$

Мы предположили при этом, что при $t = 0$ система находится в состоянии E_0. Без труда можно выписать и общее решение, убедившись при этом, что функции $p_k(t)$ неотрицательны при всех k и t. Однако, если λ_k растут слишком быстро при возрастании k, может случиться, что $\sum\limits_{k=0}^{\infty} p_k(t) < 1$.

Теорема В. Феллера. *Для того чтобы при всех значениях t решения $p_k(t)$ уравнений чистого размножения удовлетворяли соотношению*

$$\sum_{k=0}^{\infty} p_k(t) = 1, \tag{3}$$

необходимо и достаточно расходимости ряда

$$\sum_{k=0}^{\infty} \lambda_k^{-1}. \tag{4}$$

Доказательство. Рассмотрим частичную сумму ряда (3)

$$s_n(t) = p_0(t) + p_1(t) + \ldots + p_n(t). \tag{5}$$

Из уравнений размножения вытекает, что

$$s_n'(t) = -\lambda_n p_n(t).$$

Отсюда находим, что

$$1 - s_n(t) = \lambda_n \int_0^t p_n(t)\,dt \tag{6}$$

(если вместо начального условия $p_0(0) = 1$ взять другое, а именно $p_i(0) = 1$, то равенство (6) имеет место при $n \geqslant i$).

Так как все члены суммы (5) неотрицательны, то при каждом фиксированном значении t сумма $s_n(t)$ с возрастанием n не убывает. Следовательно существует предел

$$\lim_{n\to\infty}\,[1 - s_n(t)] = \mu(t). \tag{7}$$

В силу (6) мы заключаем, что $\lambda_n \int\limits_0^t p_n(t)dt \geqslant \mu(t)$. Отсюда ясно, что $\int\limits_0^1 s_n(z)dz \geqslant \mu(t)\left(\frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_n}\right)$. Так как при любых t и n имеет место неравенство $s_n(t) \leqslant 1$, то

$$t \geqslant \mu(t)\left(\frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_n}\right).$$

Если ряд (4) расходится, то из последнего неравенства вытекает, что при всех t должно быть $\mu(t) = 0$. Из (7) теперь следует, что расходимость ряда (4) приводит к (3).

Из (6) ясно, что $\lambda_n \int\limits_0^t p_n(t)dt \leqslant 1$ и, следовательно, $\int\limits_0^t s_n(z)dz \leqslant \frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_n}$. В пределе при $n \to \infty$ получаем $\int\limits_0^t [1 - \mu(z)]dz \leqslant \sum\limits_{n=0}^{\infty} \lambda_n^{-1}$.

Если $\mu(t) = 0$ при всех t, то левая часть неравенства равна t, а поскольку t произвольно, ряд, стоящий в правой части, расходится. Теорема доказана.

В предыдущем параграфе мы имели $\lambda_n = \lambda$. Следовательно, ряд (4) расходится и при всех t имеет место равенство $\sum\limits_{n=0}^{\infty} p_n(t) = 1$.

На сумму $\sum\limits_{n=0}^{\infty} p_n(t)$ можно смотреть как на вероятность того, что за время t произойдет лишь конечное число изменений состояний системы. Таким образом разность $1 - \sum\limits_{n=0}^{\infty} p_n(t)$ следует интерпретировать как вероятность бесконечного числа изменений состояний системы за время t. В явлениях радиоактивного распада такая возможность означает лавинный распад.

Пример 1. Резервирование без восстановления. Представим себе техническую систему, состоящую из одного основного элемента и n таких же резервных. Основной прибор за промежуток времени $(t, t+h)$ отказывает с вероятностью $\lambda h + o(h)$, а каждый из резервных приборов — с вероятностью $\lambda' h + o(h)$. На смену отказавшему прибору немедленно ставится прибор из резерва, отказавший же прибор дальнейшего участия в работе системы не принимает. Система в целом отказывает в момент, когда все элементы — основной и резервные — окажутся в состоянии отказа. Найти вероятности того, что в момент t в системе имеется k отказавших элементов (событие E_k).

Мы имеем дело со случаем чистого размножения. При этом

$$\lambda_k = \lambda + (n-k)\lambda' \quad \text{при} \quad 0 \leqslant k \leqslant n,$$

$$\lambda_{n+k} = 0, \quad k \geqslant 1.$$

Несложные вычисления приводят к равенствам

$$p_k(t) = \frac{\lambda_0\lambda_1 \ldots \lambda_{k-1}}{k!\lambda'^k} e^{-\lambda_k t}(1-e^{-\lambda' t})^k, \quad 0 \leqslant k < n,$$

и

$$p_{n+1}(t) = \frac{\lambda_0\lambda_1 \ldots \lambda_{n-1}\lambda}{n!\lambda'^n} \int_0^t e^{-\lambda z}(1-e^{-\lambda' z})^n dz.$$

В частности, если $\lambda' = 0$ (резерв называется *ненагруженным* или *холодным*; элементы в таком резерве не отказывают), то имеют место равенства

$$p_k(t) = \frac{\lambda^k t^k}{k!} e^{-\lambda t} \quad (0 \leqslant k \leqslant n), \quad p_{n+1}(t) = 1 - \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

При $\lambda' = \lambda$ (*нагруженный* или *горячий* резерв; в таком резерве все элементы находятся в том же состоянии, что и основной)

$$p_k(t) = C_{n+1}^k e^{-(n+1-k)\lambda t}(1-e^{-\lambda t})^k.$$

Обозначим через ξ_k длительность жизни k-го элемента в период работы. Для ненагруженного резервирования длительность жизни системы равна $\xi_0 + \xi_1 + \ldots + \xi_n$. Так как средний срок службы одного прибора равен $\int_0^\infty e^{-\lambda t}dt = 1/\lambda$, то средний срок службы системы при холодном резервировании равен $(n+1)/\lambda$, т.е. пропорционален общему числу элементов системы.

Среднюю продолжительность безотказной работы резервированной системы при нагруженном резервировании вычислим следующим способом: отметим моменты последовательных отказов элементов — $t_1, t_2, \ldots$ $\ldots, t_{n+1}$ и введем обозначения $\tau_1 = t_1$, $\tau_2 = t_2 - t_1$, $\tau_3 = t_3 - t_2, \ldots, \tau_{n+1} =$ $= t_{n+1} - t_n$. Поскольку в первом отрезке времени работают все приборы, вероятность того, что за время t не откажет ни один из них, равна $e^{-(n+1)\lambda t}$; вероятность того, что во втором интервале не откажет ни один из работоспособных элементов, равна $e^{-\lambda n t}$. Наконец, вероятность того, что за время t не будет отказов в последнем интервале, равна $e^{-\lambda t}$. Теперь легко подсчитать, что время работы резервированной системы до отказа равно

$$\sum_{k=1}^{n+1} \mathbf{M}\tau_k = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right).$$

Если n велико, то

$$1 + \frac{1}{2} + \ldots + \frac{1}{n} \sim \ln n + c,$$

где c – постоянная Эйлера, $c = 0{,}577215\ldots$

Пример 2. Система обслуживания с потерями. Мы рассмотрим теперь одну из задач новой прикладной математической дисциплины, получившей название теории массового обслуживания. Первые ее задачи были рассмотрены датским ученым Эрлангом – долголетним сотрудником лаборатории Копенгагенской телефонной компании.

Предположим, что на телефонную станцию поступают вызовы от абонентов. Если в момент поступления вызова аппарат вызываемого абонента свободен, то происходит мгновенное соединение и начинается разговор, который продолжается столько, сколько необходимо для его завершения. Если же вызываемый абонент занят, то вызывающий абонент получает отказ.

Нам важно подчеркнуть две особенности, с которыми необходимо считаться при рассмотрении возникающих здесь вопросов. Во-первых, вызовы на станцию поступают в случайные моменты времени, и предсказать заранее, когда поступит очередной вызов, нет возможности. Во-вторых, длительность разговора не постоянна, а меняется в зависимости от случая.

Мы предположим, что имеется n равноправных линий связи у каждого из абонентов и если хотя бы одна из них свободна, то соединение наступает мгновенно. Каждая линия доступна для любого требования, каждое требование обслуживается лишь одной из линий. Вероятность того, что вызов от какого-то абонента поступит в промежуток времени от t до $t + h$ равна $\lambda h + o(h)$. Если в момент t заняты k линий, то вероятность того, что к моменту $t + h$ освободится одна из них, равна $k\nu h + o(h)$.

Мы находимся в условиях теории процессов гибели и размножения. В нашем случае $\lambda_k = \lambda$, $\nu_k = k\nu$ при $1 \leqslant k \leqslant n$ и $\nu_k = 0$ при $k > n$. Система обслуживания может находиться лишь в состояниях $E_0, E_1, E_2, \ldots, E_n$.

Уравнения (1) и (2) для нашей задачи записываются в следующем виде:

$$p_0'(t) = -\lambda p_0(t) + \nu p_1(t), \tag{8}$$

при $1 \leqslant k \leqslant n$

$$p_k'(t) = -(\lambda + k\nu)p_k(t) + \lambda p_{k-1}(t) + (k+1)\nu p_{k+1}(t) \tag{9}$$

и при $k = n$

$$p'_n(t) = \lambda p_{n-1}(t) - n\nu p_n(t). \tag{10}$$

К этим уравнениям мы должны добавить еще одно

$$\sum_{k=0}^{n} p_k(t) = 1,$$

смысл которого прост: в любой момент времени возможны только события $E_0, E_1, \ldots, E_n$.

Обычно интересуются изучением установившегося процесса, т.е. рассматривают решение при $t \to \infty$. Как мы увидим позднее, в условиях нашей задачи существуют пределы

$$p_k = \lim_{k \to \infty} p_k(t)$$

и эти предельные вероятности удовлетворяют следующей системе алгебраических уравнений, получающихся из (8)–(10) путем замены функций $p_k(t)$ на постоянные p_k, а производных $p'_k(t)$ на нули:

$$\begin{aligned} &-\lambda p_0 + \nu p_1 = 0, \\ &\lambda p_{k-1} - (\lambda + k\nu) p_k + (k+1)\nu p_{k+1} = 0, \qquad 1 \leqslant k < n, \\ &\lambda p_{n-1} - n\nu p_n = 0, \\ &\sum_{k=0}^{n} p_k = 1. \end{aligned} \tag{11}$$

Обозначением $z_k = \lambda p_{k-1} - k\nu p_k$ мы приводим систему наших алгебраических уравнений к следующей:

$$z_1 = 0, \quad z_k - z_{k-1} = 0 \quad \text{при} \quad 1 \leqslant k < n, \quad z_n = 0$$

откуда находим, что

$$k\nu p_{k-1} = \lambda p_{k-1} \qquad k = 1, 2, \ldots, n.$$

Простые преобразования приводят нас к равенствам

$$p_k = \frac{\rho^k}{k!} p_0 \qquad (k \geqslant 1, \quad \rho = \lambda/\nu).$$

Теперь (11) позволяет найти нормирующий множитель p_0:

$$p_0 = \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} \right]^{-1}$$

Окончательно:

$$p_k = \frac{\rho^k}{k!} \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} \right]^{-1}, \qquad 0 \leqslant k \leqslant n.$$

Эти формулы были найдены Эрлангом и носят название *формул Эрланга*; они находят широкое применение в задачах телефонии. При $k = n$ мы получаем вероятность того, что все линии заняты и, следовательно, вероятность того, что вновь прибывшее требование будет потеряно. Таким образом, вероятность получить отказ равна

$$p_n = \frac{\rho^n}{n!}\left[\sum_{k=0}^{n}\frac{\rho^k}{k!}\right]^{-1}.$$

Для иллюстрации быстроты потерь с увеличением ρ $\left(\text{загрузка, приходящаяся на одну линию равна } \frac{\lambda}{n\nu}\right)$ приведем небольшие таблички. При этом мы ограничимся случаями $n = 2$ и $n = 4$ и такими значениями ρ, при которых в соответствующих колонках приходятся одинаковые загрузки на прибор.

Таблица 15

$n = 2$

ρ	0,1	0,3	0,5	1,0	2,0	3,0	4,0
p_n	0,0045	0,0335	0,0769	0,2000	0,4000	0,5294	0,6054

$n = 4$

ρ	0,2	0,6	1,0	2,0	4,0	6,0	8,0
p_n	0,0001	0,0030	0,0154	0,0952	0,3107	0,4696	0,5746

Из табличек замечаем, что при малых загрузках увеличение числа приборов существенно уменьшает вероятность потерь. Например, при $n = 2$ и $\rho = 1{,}0$ вероятность потери равна 0,20, а при $n = 4$ и $\rho = 2$ соответствующая вероятность будет только 0,09. По мере же возрастания загрузки на один прибор вероятности потерь постепенно выравниваются и, например, при $\rho = 4$ и $n = 2$ вероятность потери равна 0,6054, а при $n = 4$ и $\rho = 8$ эта вероятность равна уже 0,5746, т.е. различие наступает только во втором знаке.

Вернемся к некоторым общим результатам теории процессов гибели и размножения, но изложим их без доказательств. В случае процесса чистого размножения система уравнений (1)–(2) разрешалась очень просто путем последовательного интегрирования, поскольку дифференциальные уравнения имели вид простых рекуррентных соотношений. Общие уравнения имеют иную структуру и последовательное определение функций $p_k(t)$ уже невозможно. В настоящее время условия существования и единственности решений этой системы хорошо изучены в работах Феллера, Рейтера, Карлина и Мак-Грегора. Оказалось, что равенство

$$\sum_{k=0}^{\infty} p_k(t) = 1$$

имеет место при всех t, если расходится ряд

$$\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{\nu_i}{\lambda_i}. \tag{12}$$

Если вдобавок сходится ряд

$$\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\nu_i}, \tag{13}$$

то при всех t существуют пределы

$$p_k = \lim_{t \to \infty} p_k(t). \tag{14}$$

Это условие, в частности, выполнено во всех случаях, когда, начиная с некоторого k_0, выполняется неравенство

$$\lambda_k / \nu_{k+1} \leqslant \alpha < 1.$$

Интуитивно это условие ясно: оно означает, что скорость поступления требований в систему не должна превышать скорости их обслуживания.

Для вычисления пределов (13) действует следующее простое правило: нужно составить и решить систему алгебраических уравнений, получающуюся из системы (1)–(2) путем замены $p_k(t)$ на p_k и подстановки 0 вместо $p'_k(t)$. Эта система, следовательно, имеет следующий вид:

$$-\lambda_0 p_0 + \nu_1 p_1 = 0,$$

$$-(\lambda_k + \nu_k) p_k + \lambda_{k-1} p_{k-1} + \nu_{k+1} p_{k+1} = 0 \qquad (k \geqslant 1).$$

Обозначения

$$z_k = -\lambda_k p_k + \nu_{k+1} p_{k+1}, \qquad k = 0, 1, 2, \ldots$$

обращают записанную алгебраическую систему в следующую:

$$z_0 = 0, \quad z_{k-1} - z_k = 0 \qquad (\text{при } k \geqslant 1).$$

Из нее вытекает, что при всех k

$$z_k = 0.$$

Следовательно,

$$p_k = \frac{\lambda_{k-1}}{\nu_k} p_{k-1} = \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\nu_i} p_0. \tag{15}$$

Постоянное p_0 определяется из условия нормировки $\sum\limits_{k=0}^{\infty} p_k = 1$:

$$p_0 = \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\nu_i}\right]. \tag{16}$$

Очевидно, что в полученных формулах содержатся найденные нами ранее формулы Эрланга.

Пример 3. Обслуживание с очередью. На n одинаковых приборов поступает пуассоновский поток требований с параметром (интенсивностью) λ. Требование, поступившее на какой-либо прибор, требует для всего обслуживания случайного времени с распределением вероятностей $H(x) = 1 - e^{-\nu x}$. Если в момент поступления требования имеется хотя бы один свободный прибор, оно начинает обслуживаться немедленно. Если же все приборы заняты, то вновь поступающие требования становятся в очередь. Если имеется очередь, то после окончания обслуживания прибор немедленно переключается на обслуживания очередного требования из очереди. Требуется найти вероятности пребывания в системе того или иного числа требований.

Мы находимся в условиях теории процессов гибели и размножения. Для нашей задачи $\lambda_k = \lambda$ при всех k, $\nu_k = k\nu$ при $k \leqslant n$ и $\nu_k = n\nu$ при $k \geqslant n$.

Согласно формулам (15) и (16), стационарные решения для нашей задачи имеют такой вид: при $k \leqslant n$

$$p_k = \frac{\rho^k}{k!} p_0$$

и при $k \geqslant n$

$$p_k = \frac{\rho^k}{n! n^{k-n}} p_0,$$

где $\rho = \lambda/\nu$. Постоянное p_0 определяется равенством

$$p_0 = \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \sum_{k=n+1}^{\infty} \left(\frac{\rho}{n}\right)^{k-n}\right]^{-1}.$$

Если $\rho < n$, то

$$p_0 = \left[1 + \sum_{k=1}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)!}\right]^{-1}.$$

Если же $\rho \geqslant n$, то ряд, стоящий в скобке, расходится и $p_0 = 0$. Из только что написанных формул мы заключаем, что $p_k = 0$ при всех k. Этот результат очень важен; словами его можно сформулировать так: *если $\rho \geqslant n$, то очередь на обслуживание неограниченно растет со временем.*

Пример 4. Обслуживание станков бригадой рабочих. Бригада из r рабочих обслуживает n однотипных станков. Каждый из этих станков в случайные моменты времени может потребовать к себе внимания рабочего. Станки выходят из рабочего состояния независимо друг от друга; вероятность выхода из рабочего состояния за промежуток времени $(t, t+h)$ равна $\lambda h + o(h)$. Вероятность того, что за время $(t, t+h)$ будет завершено восстановление рабочего состояния станка равна $\nu h + o(h)$. Каждый рабочий одновременно может восстанавливать только один станок; каждый станок восстанавливается только одним рабочим. Найти вероятность того, что в установившемся процессе обслуживания в данный момент будет простаивать заданное число станков.

Обозначим через E_k событие, состоящее в том, что в данный момент неисправны k станков. Очевидно, что наша система может находиться только в состояниях $E_0, E_1, \ldots, E_n$. Легко понять, что мы имеем дело с процессом гибели и размножения, для которого $\lambda_k = (n-k)\lambda$ при $0 \leqslant k < n$, $\lambda_k = 0$ при $k \geqslant n$; $\nu_k = k\nu$ при $1 \leqslant k \leqslant r$ и $\nu_k = r\nu$ при $k \geqslant r$. Формулы (15) и (16) приводят к равенствам: при $1 \leqslant k \leqslant r$ $(\rho = \lambda/\nu)$

$$p_k = \frac{n!}{k!(n-k)!}\rho^k p_0,$$

при $r \leqslant k \leqslant n$

$$p_k = \frac{n!}{r^{n-k} r!(n-k)!}\rho^k p_0$$

и

$$p_0 = \left[\sum_{k=0}^{r} \frac{n!}{k!(n-k)!}\rho^k + \sum_{k=r+1}^{n} \frac{n!}{r^{n-k} r!(n-k)!}\rho^k\right]^{-1}.$$

В частности, при $r = 1$

$$p_k = \frac{n!}{(n-k)!}\rho^k p_0,$$

$$p_0 = \left[\sum_{k=0}^{n} \frac{n!}{(n-k)!}\rho^k\right]^{-1}.$$

Проиллюстрируем полученные формулы простым числовым расчетом. Пусть обслуживание 8 станков поручено двум рабочим. Как рациональнее организовать работу: поручить ли все станки бригаде из двух рабочих или же каждому из рабочих поручить по четыре определенных станка? Вычисления проведены в предположении $\rho = 0{,}2$. Результаты собраны в таблице.

Таблица 16

$n = 8,\ r = 2$

Число неработающих станков	Число станков, ожидающих обслуживания	Число свободных рабочих	
0	0	2	0,2048
1	0	1	0,3277
2	0	0	0,2294
3	1	0	0,1417
4	2	0	0,0687
5	3	0	0,0255
6	4	0	0,0083
7	5	0	0,0017
8	6	0	0,0002

Среднее число станков, простаивающих по той причине, что рабочие заняты восстановлением других станков, равно

$$\sum_{k=2}^{8} (k-2)p_k = 0{,}3045.$$

Среднее время простоя станков (восстановление и ожидание начала обслуживания) равно

$$\sum_{k=2}^{8} kp_k = 1{,}6875.$$

Средняя длительность свободного времени рабочих равна

$$2.0{,}2048 + 1.0{,}3277 = 07373.$$

Иными словами каждый рабочий свободен от работы в течение 0,3686 доли рабочего дня.

Среднее время непроизводительных простоев станков (ожидание начала восстановления)

$$1.0{,}1914 + 2.0{,}0760 + 3.0{,}0153 = 0{,}3893.$$

Вся группа из восьми станков потеряет при этой второй системе организации работы 0,7886 рабочих дня, т.е. потеря времени на ожидание ремонта возрастут более чем вдвое (в первой системе она равна 0,3045 рабочих дня). Общая потеря времени 4 станками на ожидание и ремонт равна

$$1.0{,}3189 + 2.0{,}1914 + 3.0{,}0760 + 4.0{,}0153 = 0{,}9909.$$

Таблица 17

$n = 4, \quad r = 1$

Число неработающих станков	Число станков, ожидающих обслуживания	Число свободных рабочих	
0	0	1	0,3984
1	0	0	0,3189
2	1	0	0,1914
3	2	0	0,0760
4	3	0	0,0153

Все восемь станков теряют, таким образом, 1,9818 рабочих дня (против 1,6875 при первой системе организации работы). Несмотря на то, что станки простаивают при второй системе организации труда больше, рабочий в среднем свободен от работы больше, а именно 0,3984 доли рабочего дня (было 0,3686 доли рабочего дня).

Приведенные примеры показывают, что развитая теория позволяет проводить полезные предварительные расчеты и выбирать более разумные приемы работы.

§ 52. Условные функции распределения и формула Байеса

Для дальнейших выводов нам необходимо обобщить понятие условной вероятности, введенное в первой главе, на случай бесконечного множества возможных условий. В частности, нам нужно ввести понятие условной функции распределения относительно случайной величины.

Рассмотрим некоторое событие B и случайную величину ξ с функцией распределения $F(x)$. Обозначим через $A_{\alpha\beta}$ событие, состоящее в том, что

$$x - \alpha \leqslant \xi < x + \beta.$$

В силу определений первой главы

$$\mathsf{P}\{BA_{\alpha\beta}\} = \mathsf{P}\{A_{\alpha\beta}\} \cdot \mathsf{P}\{B \mid A_{\alpha\beta}\} = [F(x+\beta) - F(x-\alpha)]\,\mathsf{P}\{B \mid A_{\alpha\beta}\},$$

откуда

$$\mathsf{P}\{B \mid A_{\alpha\beta}\} = \frac{\mathsf{P}\{BA_{\alpha\beta}\}}{F(x+\beta) - F(x-\alpha)}.$$

Предел

$$\lim_{\alpha,\beta \to 0} \frac{\mathsf{P}\{BA_{\alpha\beta}\}}{F(x+\beta) - F(x-\alpha)},$$

если он существует *), называется *условной вероятностью события B при условии, что $\xi = x$, и обозначается символом* $\mathsf{P}\{B \mid x\}$. Очевидно, что при x фиксированном $\mathsf{P}\{B \mid x\}$ будет конечно-аддитивной функцией события B, определенной на некотором поле событий.

При некоторых условиях, которые практически всегда оказываются выполненными, $\mathsf{P}\{B \mid x\}$ будет обладать всеми свойствами обычной вероятности, удовлетворяющей аксиомам 1 – 3 § 8.

Если η – случайная величина и B означает событие $\eta < y$, то функция $\Phi(y \mid x) = \mathsf{P}\{\eta < y \mid x\}$, которая, как легко видеть, будет функцией распределения, называется *условной функцией распределения величины η при условии, что $\xi = x$.*

Очевидно, что если $F(x, y)$ есть функция распределения пары случайных величин ξ и η, то

$$\Phi(y \mid x) = \lim_{\alpha, \beta \to 0} \frac{F(x+\beta, y) - F(x-\alpha, y)}{F(x+\beta, \infty) - F(x-\alpha, \infty)},$$

если только этот предел существует.

Если функция $\mathsf{P}\{B \mid x\}$ интегрируема относительно $F(x)$, то имеет место *формула полной вероятности*

$$\mathsf{P}\{B\} = \int \mathsf{P}\{B \mid x\}\, dF(x).$$

Для доказательства этой формулы мы разделим промежуток изменения величины ξ точками x_i $(i = 0, \pm 1, \pm 2, \ldots)$ на интервалы $x_i \leqslant \xi < x_{i+1}$. Обозначим через A_i событие $x_i \leqslant \xi < x_{i+1}$. В силу расширенной аксиомы сложения имеем:

$$\mathsf{P}\{B\} = \sum_{i=-\infty}^{\infty} \mathsf{P}\{BA_i\} = \sum_{i=-\infty}^{\infty} \mathsf{P}\{B \mid A_i\}\,[F(x_{i+1}) - F(x_i)].$$

Станем теперь подразделять интервалы (x_i, x_{i+1}) на более мелкие таким образом, чтобы максимальная длина получившихся интервалов стремилась к нулю. В силу определения условной вероятности и интеграла Стилтьеса отсюда получаем:

$$\mathsf{P}\{B\} = \int \mathsf{P}\{B \mid x\}\, dF(x).$$

В частности,

$$\Phi(y) = \mathsf{P}\{\eta < y\} = \int \Phi(y \mid x)\, dF(x). \tag{1}$$

Если существует плотность распределения вероятностей величины η, то

$$\varphi(y) = \int \varphi(y \mid x)\, dF(x), \tag{1'}$$

где $\varphi(y \mid x)$ – *условная плотность* распределения величины η.

---

*) Этот предел существует для почти всех значений x в смысле меры, определяемой функцией $F(x)$.

П р и м е р. В качестве примера использования формулы (1) рассмотрим следующую задачу теории стрельбы. При стрельбе по некоторой цели возможны ошибки двоякого рода: 1) в определении положения цели и 2) ошибки выстрела, происходящие от большого числа различных причин (колебания в величине заряда в снаряде, неправильности обточки стакана снаряда, ошибки в наводке, незначительные колебания атмосферных условий и т.д.). Ошибки второго рода носят название технического рассеивания.

Производится n независимых выстрелов по одному определению положения цели. Требуется определить вероятность хотя бы одного попадания в цель.

Ради простоты мы ограничимся рассмотрением одномерной цели размера 2α, а снаряд будем считать точкой. Обозначим через $f(x)$ плотность вероятностей положения цели и через $\varphi_i(x)$ плотность вероятностей для точек попадания i-го снаряда.

Если центр цели находится в точке z, то вероятность попадания в цель при i-м выстреле равна вероятности попадания в интервал $(z-\alpha, z+\alpha)$, т.е. равна *)

$$\int_{z-\alpha}^{z+\alpha} \varphi_i(x)dx.$$

Условная вероятность промаха при i-м выстреле при условии, что центр цели находится в точке z, равна

$$1-\int_{z-\alpha}^{z+\alpha} \varphi_i(x)dx.$$

Условная вероятность промаха при всех n выстрелах (при том же условии) равна

$$\prod_{i=1}^{n} (1-\int_{z-\alpha}^{z+\alpha} \varphi_i(x)dx).$$

Отсюда заключаем, что вероятность хотя бы одного попадания, при условии, что центр цели находится в точке z, равна

$$1-\prod_{i=1}^{n} (1-\int_{z-\alpha}^{z+\alpha} \varphi_i(x)dx).$$

Безусловная вероятность хотя бы одного попадания в цель (по формуле (1)), таким образом, равна

$$P=\int f(z)\,[1-\prod_{i=1}^{n} (1-\int_{z-\alpha}^{z+\alpha} \varphi_i(x)dx)]dz.$$

---

*) Мы полагаем при этом, что определение положения цели и техническое рассеивание независимы.

Если условия стрельбы не изменяются от выстрела к выстрелу, то $\varphi_i(x) = \varphi(x)$ $(i = 1, 2, \ldots, n)$ и, следовательно,

$$P = \int f(z)\,[1 - (1 - \int_{z-\alpha}^{z+\alpha} \varphi(x)dx)^n]\,dz.$$

Пусть попрежнему A_i обозначает событие $x_i \leqslant \xi < x_{i+1}$. Согласно классической теореме Байеса

$$\mathsf{P}\{A_i \mid B\} = \frac{\mathsf{P}\{A_i\}\mathsf{P}\{B \mid A_i\}}{\mathsf{P}\{B\}}.$$

Если $F(x) = \mathsf{P}\{\xi < x\}$ и $\mathsf{P}\{\xi < x \mid B\}$ имеют непрерывные производные по x, то, пользуясь теоремой Лагранжа, получаем:

$$\mathsf{P}\{A_i \mid B\} = p_\xi(\bar{x}_i \mid B)\,(x_{i+1} - x_i) = \frac{F'(\bar{x}_i')\,\mathsf{P}\{B \mid A_i\}}{\mathsf{P}\{B\}}\,(x_{i+1} - x_i),$$

где $x_i < \bar{x}_i < x_{i+1}$, $x_i < \bar{x}_i' < x_{i+1}$. В пределе, когда $x_i \to x$, $x_{i+1} \to x$, получаем

$$p_\xi(x \mid B) = \frac{p(x)\,\mathsf{P}\{B \mid x\}}{\mathsf{P}\{B\}},$$

или

$$p_\xi(x \mid B) = \frac{p(x)\,\mathsf{P}\{B \mid x\}}{\int \mathsf{P}\{B \mid x\}p(x)\,dx}. \tag{2}$$

Это равенство естественно назвать *формулой Байеса.*

Пусть теперь событие B состоит в том, что некоторая случайная величина η принимает значение между $y - \alpha$ и $y + \beta$ и условная функция распределения $\Phi(y \mid x)$ величины η имеет при каждом x непрерывную плотность $p_\eta(y \mid x)$. Тогда, как это следует из равенства (2), если $\dfrac{1}{\beta + \alpha}\,\mathsf{P}\{B \mid x\}$ при α и β, стремящихся к нулю, равномерно относительно x стремится к $p_\eta(y \mid x)$, то имеет место равенство

$$p_\xi(x \mid y) = \frac{p(x)\,p_\eta(y \mid x)}{\int p_\eta(y \mid x)\,p(x)\,dx}.$$

Эта формула будет нами широко использована в следующей главе.

§ 53. Обобщенное уравнение Маркова

Мы перейдем теперь к изучению случайных процессов без последействия, ограничиваясь при этом лишь *простейшими* задачами. В частности мы будем предполагать, что множество возможных состояний системы есть множество действительных чисел. Таким образом, для нас *случайным процессом* будет совокупность случайных величин $\xi(t)$, зависящих от одного действительного параметра t. Мы будем называть параметр t временем и говорить о состоянии системы в тот или иной момент времени.

Полную вероятностную характеристику процесса без последействия мы получим, задав функцию $F(t, x; \tau, y)$, равную вероятности того, что в момент τ случайная величина $\xi(\tau)$ примет значение, меньшее y, если известно, что в момент $t\,(t<\tau)$ имело место равенство $\xi(t)=x$. Дополнительное знание состояний системы в более ранние чем t моменты времени для процессов без последействия не изменяет функцию $F(t, x; \tau, y)$.

Отметим теперь некоторые условия, которым должна удовлетворять функция $F(t, x; \tau, y)$. Прежде всего для нее, как для функции распределения, должны быть при любых x, t и τ выполнены равенства:

1) $\lim\limits_{y\to-\infty} F(t, x; \tau, y)=0, \quad \lim\limits_{y\to+\infty} F(t, x; \tau, y)=1;$

2) функция $F(t, x; \tau, y)$ непрерывна слева относительно аргумента y.

Предположим теперь, что функция $F(t, x; \tau, y)$ непрерывна по t, τ и по x.

Рассмотрим моменты времени $t, s, \tau\ (t<s<\tau)$. Так как из состояния x в момент t система переходит в момент s в одно из состояний интервала $(z, z+dz)$ с вероятностью $d_zF(t, x; s, z)$, а из состояния z в момент s переходит в состояние, меньшее y, в момент τ с вероятностью $F(s, z; \tau, y)$, то согласно формуле (1) предыдущего параграфа находим, что

$$F(t, x; \tau, y)=\int F(s, z; \tau, y)\,d_zF(t, x; s, z).$$

Полученное равенство естественно назвать *обобщенным уравнением Маркова*, так как оно представляет собой распространение равенства (1), § 17 теории цепей Маркова на теорию случайных процессов и в этой теории играет столь же важную роль, как упомянутое тождество в теории цепей Маркова.

Вероятность $F(t, x; \tau, y)$ определена пока только для $\tau>t$. Дополним это определение, приняв

$$\lim_{\tau\to t+0} F(t, x; \tau, y)=\lim_{t\to\tau-0} F(t, x; \tau, y)=E(x, y)=\begin{cases}0 & \text{для } y\leqslant x,\\ 1 & \text{для } y>x.\end{cases}$$

Если существует плотность

$$f(t, x; \tau, y)=\frac{\partial}{\partial y}F(t, x; \tau, y),$$

то для нее выполняются следующие очевидные равенства:

$$\int_{-\infty}^{y} f(t, x; \tau, z)\,dz = F(t, x; \tau, y),$$

$$\int f(t, x; \tau, z)\,dz = 1.$$

Для этого случая обобщенное уравнение Маркова должно быть записано в таком виде:

$$f(t, x; \tau, y) = \int f(t', z; \tau, y) f(t, x; t', z)\,dz.$$

§ 54. Непрерывный случайный процесс. Уравнения Колмогорова

Мы скажем, что случайный процесс $\xi(t)$ *непрерывен*, если за малые промежутки времени лишь с малой вероятностью $\xi(t)$ может получить заметные по величине приращения. Более точно, случайный процесс $\xi(t)$ непрерывен, если, каково бы ни было постоянное δ $(\delta > 0)$, имеет место соотношение

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| \geq \delta} dF(t - \Delta t, x;\ t, y) = 0. \tag{1}$$

Наша ближайшая задача состоит в выводе дифференциальных уравнений, которым при выполнении некоторых условий удовлетворяет функция $F(t, x;\ \tau, y)$, управляющая непрерывным случайным процессом без последействия. Эти уравнения впервые строго были доказаны А.Н. Колмогоровым (хотя второе из них и встречалось до этого в работах физиков) и носят название *уравнений Колмогорова*.

Мы предположим, что

1) частные производные

$$\frac{\partial F(t, x; \tau, y)}{\partial x} \quad \text{и} \quad \frac{\partial^2 F(t, x;\ \tau, y)}{\partial x^2}$$

существуют и непрерывны при любых значениях t, x и $\tau > t$;

2) каково бы ни было $\delta > 0$, существуют предел *)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x| < \delta} (y - x)\,d_y F(t - \Delta t, x;\ t, y) = a(t, x) \tag{2}$$

---

*) При некоторых достаточно общих предположениях А.Н. Колмогоров доказал существование пределов $a(t, x)$ и $b(t, x)$.

Наглядный смысл функций a и b мы выясним в конце параграфа.

и предел

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|<\delta}(y-x)^2 d_y F(t-\Delta t, x; t, y) = b(t, x), \tag{3}$$

и эта сходимость равномерна относительно x.

Левые части равенств (2) и (3) зависят от δ. Эта зависимость, однако, в силу определения непрерывности процесса (т.е. в силу (1)) является лишь кажущейся.

П е р в о е у р а в н е н и е К о л м о г о р о в а. *Если только что сформулированные условия* 1) *и* 2) *выполнены, то функция* $F(t, x; \tau, y)$ *удовлетворяет уравнению.*

$$\frac{\partial F(t, x; \tau, y)}{\partial t} = -a(t, x)\frac{\partial F(t, x; \tau, y)}{\partial x} - \frac{b(t, x)}{2}\frac{\partial^2 F(t, x; \tau, y)}{\partial x^2}. \tag{4}$$

Д о к а з а т е л ь с т в о. Согласно обобщенному уравнению Маркова

$$F(t-\Delta t, x; \tau, y) = \int F(t, z; \tau, y)\, d_z F(t-\Delta t, x; t, z).$$

Кроме того, в силу свойств функции распределения,

$$F(t, x; \tau, y) = \int F(t, x; \tau, y)\, d_z F(t-\Delta t, x; t, z).$$

Из этих равенств заключаем, что

$$\frac{F(t-\Delta t, x; \tau, y) - F(t, x; \tau, y)}{\Delta t} =$$

$$= \frac{1}{\Delta t}\int [F(t, z; \tau, y) - F(t, x; \tau, y)]\, d_z F(t-\Delta t, x; t, z).$$

По формуле Тейлора при сделанных нами предположениях имеет место равенство

$$F(t, z; \tau, y) = F(t, x; \tau, y) + (z-x)\frac{\partial F(t, x; \tau, y)}{\partial x} +$$

$$+ \frac{1}{2}(z-x)^2\frac{\partial^2 F(t, x; \tau, y)}{\partial x^2} + o((z-x)^2).$$

Последующие аналитические преобразования не требуют пояснений:

$$\frac{F(t-\Delta t, x; \tau, y) - F(t, x; \tau, y)}{\Delta t} =$$

$$= \frac{1}{\Delta t} \int\limits_{|z-x| \geqslant \delta} [F(t, z; \tau, y) - F(t, x; \tau, y)]\, d_z F(t-\Delta t, x; t, z) +$$

$$+ \frac{1}{\Delta t} \int\limits_{|z-x| < \delta} [F(t, z; \tau, y) - F(t, x; \tau, y)]\, d_z F(t-\Delta t, x; t, z) =$$

$$= \frac{1}{\Delta t} \int\limits_{|z-x| \geqslant \delta} [F(t, z; \tau, y) - F(t, z; \tau, y)]\, d_z F(t-\Delta t, x; t, z) +$$

$$+ \frac{\partial F(t, x; \tau, y)}{\partial x} \cdot \frac{1}{\Delta t} \int\limits_{|z-x| < \delta} (z-x)\, d_z F(t-\Delta t, x; t, z) +$$

$$+ \frac{1}{2} \frac{\partial^2 F(t, x; \tau, y)}{\partial x^2} \cdot \frac{1}{\Delta t} \int\limits_{|z-x| < \delta} [(z-x)^2 + o(z-x)^2] \times$$

$$\times d_z F(t-\Delta t, x; t, z). \qquad (5)$$

Перейдем теперь к пределу, положив $\Delta t \to 0$. Первое слагаемое правой части в силу (1) имеет своим пределом 0. Второе слагаемое, согласно (2), в пределе равно $a(t, x) \dfrac{\partial F}{\partial x}$. Наконец, третье слагаемое может отличаться от $\dfrac{1}{2} b(t, x) \dfrac{\partial^2 F}{\partial x^2}$ только на слагаемое, стремящееся к нулю при $\delta \to 0$. Но так как левая часть последчего равенства от δ не зависит и только что указанные предельные значения от δ не зависят, то предел правой части существует и равен

$$a(t, x) \frac{\partial F(t, x; \tau, y)}{\partial x} + \frac{1}{2} b(t, x) \frac{\partial^2 F(t, x; \tau, y)}{\partial x^2}.$$

Отсюда мы заключаем о существовании предела:

$$\lim_{\Delta t \to 0} \frac{F(t-\Delta t, x; \tau, y) - F(t, x; \tau, y)}{\Delta t} = \frac{\partial F(t, x; \tau, y)}{\partial t}.$$

Равенство (5) приводит нас к уравнению (4).

Если предположить, что существует плотность распределения

$$f(t, x; \tau, y) = \frac{\partial}{\partial y} F(t, x; \tau, y),$$

то простое дифференцирование (4) показывает, что плотность $f(t, x; \tau, y)$ удовлетворяет уравнению

$$\frac{\partial f(t, x; \tau, y)}{\partial t} + a(t, x) \frac{\partial f(t, x; \tau, y)}{\partial x} + $$

$$+ \frac{1}{2} b(t, x) \frac{\partial^2 f(t, x; \tau, y)}{\partial x^2} = 0. \tag{4'}$$

Мы перейдем теперь к выводу второго уравнения Колмогорова. При этом мы не станем стремиться к наибольшей возможной общности и сделаем допущения, не вызываемые существом дела. Помимо уже сделанных предположений, мы наложим на функцию $F(t, x; \tau, y)$ еще такие ограничения

3) существует плотность распределения вероятностей

$$f(t, x; \tau, y) = \frac{\partial F(t, x; \tau, y)}{\partial y};$$

4) существуют непрерывные производные

$$\frac{\partial f(t, x; \tau, y)}{\partial \tau}, \quad \frac{\partial}{\partial y}[a(\tau, y) f(t, x; \tau, y)], \quad \frac{\partial^2}{\partial y^2}[b(\tau, y) f(t, x; \tau, y)].$$

Второе уравнение Колмогорова*). *Если выполнены условия* 1) – 4), *то для непрерывного случайного процесса без последействия плотность* $f(t, x; \tau, y)$ *удовлетворяет уравнению*

$$\frac{\partial f(t, x; \tau, y)}{\partial \tau} = -\frac{\partial}{\partial y}[a(\tau, y) f(t, x; \tau, y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(\tau, y) f(t, x; \tau, y)]. \tag{6}$$

Доказательство. Пусть a и b $(a < b)$ – некоторые числа и $R(y)$ – неотрицательная непрерывная функция, имеющая непрерывные производные до второго порядка включительно. Кроме того, мы потребуем, чтобы

$$R(y) = 0 \quad \text{при} \quad y < a \quad \text{и} \quad y > b.$$

---

*) Второе уравнение Колмогорова было получено раньше физиками Фоккером и Планком в связи с развитием теории диффузии.

Из условия непрерывности функции $R(y)$ и ее производных заключаем, что

$$R(a) = R(b) = R'(a) = R'(b) = R''(a) = R''(b) = 0. \tag{7}$$

Заметим прежде всего, что

$$\int_a^b \frac{\partial f(t,x;\tau,y)}{\partial \tau} R(y)\,dy = \frac{\partial}{\partial \tau}\int_a^b f(t,x;\tau,y)R(y)\,dy =$$

$$= \lim_{\Delta\tau\to 0}\int \frac{f(t,x;\tau+\Delta\tau,y)-f(t,x;\tau,y)}{\Delta\tau} R(y)\,dy.$$

Согласно обобщенному уравнению Маркова

$$f(t,x;\tau+\Delta\tau,y) = \int f(t,x;\tau,z) f(\tau,z;\tau+\Delta\tau,y)\,dz,$$

поэтому

$$\int_a^b \frac{\partial f(t,x;\tau,y)}{\partial \tau} R(y)\,dy =$$

$$= \lim_{\Delta\tau\to 0}\frac{1}{\Delta\tau}\left[\int\int f(t,x;\tau,z) f(\tau,z;\tau+\Delta\tau,y) R(y)\,dz\,dy -\right.$$

$$\left.- \int f(t,x;\tau,y)R(y)\,dy\right] =$$

$$= \lim_{\Delta\tau\to 0}\frac{1}{\Delta\tau}\left[\int f(t,x,\tau,z)\int f(\tau,z;\tau+\Delta\tau,y)R(y)\,dy\,dz -\right.$$

$$\left.- \int f(t,x;\tau,y)R(y)\,dy\right] =$$

$$= \lim_{\Delta t\to 0}\frac{1}{\Delta\tau}\int f(t,x;\tau,y)\left[\int f(\tau,y;\tau+\Delta\tau,z)R(z)\,dz - R(y)\right] dy.$$

Произведенные преобразования очевидны: первый раз мы поменяли порядок интегрирования, а второй раз изменили обозначения переменных интегрирования (y на z, а z на y).

По формуле Тейлора

$$R(z) = R(y) + (z-y)R'(y) + (z-y)^2 R''(y) + o[(z-y)^2].$$

Так как в силу ограниченности функции $R(z)$ и условия (1)

$$\int_{|y-z|\geq\delta} f(\tau,y;\tau+\Delta\tau,z)R(z)\,dz = o(\Delta\tau)$$

и

$$\int\limits_{|y-z|\leqslant\delta} f(\tau, y; \tau+\Delta\tau, z)\,dz = 1 + o(\Delta\tau),$$

то

$$\int f(\tau, y; \tau+\Delta\tau, z) R(z)\,dz - R(y) =$$

$$= R'(y) \int\limits_{|y-z|<\delta} (z-y) f(\tau, y; \tau+\Delta\tau, z)\,dz +$$

$$+ \frac{1}{2} R''(y) \int\limits_{|y-z|<\delta} [(z-y)^2 + o(z-y)^2]\, f(\tau, y; \tau+\Delta\tau, z)\,dz + o(\Delta\tau).$$

Таким образом,

$$\int\limits_a^b \frac{\partial f(t, x; \tau, y)}{\partial\tau} R(y)\,dy =$$

$$= \lim_{\Delta\tau\to 0} \int f(t, x; \tau, y) \left\{ R'(y) \int\limits_{|y-z|<\delta} (z-y) f(\tau, y; \tau+\Delta\tau, z)\,dz + \right.$$

$$+ \frac{1}{2} R''(y) \int\limits_{|y-z|<\delta} [(z-y)^2 + o(z-y)^2] \times$$

$$\left. \times\, f(\tau, y\tau+\Delta\tau, z)\,dz + o(\Delta\tau) \right\} dy.$$

Перейдем к пределу, положив $\Delta\tau \to 0$. В силу предположения о равномерной сходимости к пределам в (2) и (3), заключаем, что предыдущее равенство может быть записано в виде

$$\int\limits_a^b \frac{\partial f(t, x; \tau, y)}{\partial\tau} R(y)\,dy =$$

$$= \int f(t, x; \tau, y) \left[a(\tau, y) R'(y) + \frac{1}{2} b(\tau, y) R''(y) \right] dy.$$

Так как $R'(y) = R''(y) = 0$ для $y \leqslant a$ и $y \geqslant b$, то

$$\int\limits_a^b \frac{\partial f(t, x; \tau, y)}{\partial\tau} R(y)\,dy =$$

$$= \int\limits_a^b f(t, x; \tau, y) \left[a(\tau, y) R'(y) + \frac{1}{2} b(\tau, y) R''(y) \right] dy. \tag{8}$$

Воспользовавшись формулой интегрирования по частям и равенствами (7), находим, что

$$\int_a^b f(t,x;\tau,y)\,a(\tau,y)\,R'(y)\,dy = -\int_a^b R(y)\,\frac{\partial}{\partial y}\,[a(\tau,y)f(t,x;\tau,y)]\,dy,$$

$$\int_a^b f(t,x;\tau,y)\,b(\tau,y)\,R''(y)\,dy = \int_a^b R(y)\,\frac{\partial^2}{\partial y^2}\,[b(\tau,y)f(t,x;\tau,y)]\,dy.$$

В результате подстановки полученных выражений в (8) получаем:

$$\int_a^b \frac{\partial f(t,x;\tau,y)}{\partial\tau}\,R(y)\,dy =$$

$$= \int_a^b \left\{-\frac{\partial}{\partial y}\,[a(\tau,y)f(t,x;\tau,y)] + \right.$$

$$\left. + \frac{1}{2}\,\frac{\partial^2}{\partial y^2}[b(\tau,y)f(t,x;\tau,y)]\right\} R(y)\,dy.$$

Это равенство может быть записано, очевидно, в таком виде:

$$\int_a^b \left\{\frac{\partial f(t,x;\tau,y)}{\partial\tau} + \frac{\partial}{\partial y}\,[a(\tau,y)f(t,x;\tau,y)] - \right.$$

$$\left. - \frac{1}{2}\,\frac{\partial^2}{\partial y^2}\,[b(\tau,y)f(t,x;\tau,y)]\right\} R(y)\,dy = 0. \qquad (9)$$

Так как функция $R(y)$ произвольна, то из последнего тождества вытекает (6). Действительно, предположим, что это не так. Тогда существует такая четверка чисел $(t,x;\ \tau,\ y)$, при которой выражение, стоящее в (9) в фигурных скобках, отлично от нуля. В силу сделанных предположений это выражение представляет собой непрерывную функцию; следовательно, найдется интервал $\alpha < y < \beta$, где оно сохраняет знак. Если $a \leqslant \alpha$ и $b \geqslant \beta$, то мы полагаем $R(y) = 0$ при $y \leqslant \alpha$ и $y \geqslant \beta$ и $R(y) > 0$ при $\alpha < y < \beta$. При таком выборе $R(y)$ интеграл, стоящий в левой части равенства (9) должен быть отличен от нуля. Мы пришли к противоречию. Таким образом, сделанное нами предположение ошибочно и, следовательно, из (9) вытекает (6).

Естественно, что основная задача, которую приходится решать, состоит не в проверке того, что данная функция $f(t,\ x;\ \tau,\ y)$ удовлетворяет уравнениям Колмогорова, а в разыскании неизвестной функции $f(t,x;\ \tau,y)$ по этим уравнениям, в которых коэффициенты $a(t,x)$ и $b(t,x)$ предполагаются известными. При этом, конечно, разыскивается не какое-нибудь решение уравнений Колмогорова, а лишь те из них, которые удовлетворяют следующим требованиям:

$$\left.\begin{array}{l} 1.\ f(t, x; \tau, y) = 0 \quad \text{при всех} \quad t, x, \tau, y. \\ 2.\ \int f(t, x; \tau, y)\, dy = 1 \\ \text{и при любом } \delta > 0 \\ 3.\ \lim\limits_{\tau \to t} \int\limits_{|y-x| \geqslant \delta} f(t, x; \tau, y)\, dy = 0. \end{array}\right\} \qquad (10)$$

Мы не будем останавливаться на выяснении тех условий, которые нужно наложить на функции $a(t, x)$ и $b(t, x)$, чтобы существовало решение уравнений Колмогорова, удовлетворяющее перечисленным требованиям и было бы при этом единственным.

Мы несколько усилим требование непрерывности с тем, чтобы выяснить физический смысл коэффициентов $a(t, x)$ и $b(t, x)$. Именно, предположим вместо (1), что при любом $\delta > 0$ имеет место соотношение

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| > \delta} (y-x)^2\, d_y F(t - \Delta t, x; t, y) = 0. \qquad (1')$$

Легко видеть, что из $(1')$ следует (1). Требования 2 и 3 могут быть теперь записаны иначе, а именно,

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y-x)\, d_y F(t - \Delta t, x; t, y) = a(t, x) \qquad (2')$$

и

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y-x)^2 d_y F(t - \Delta t, x; t, y) = b(t, x). \qquad (3')$$

Остальные требования, а также окончательные выводы от замены (1) на $(1')$ не изменяются. Так как

$$\int (y-x)\, d_y F(t - \Delta t, x; t, y) = \mathbf{M}\, [\xi(t) - \xi(t - \Delta t)]$$

является математическим ожиданием изменения $\xi(t)$ за время Δt, а

$$\int (y-x)^2 d_y F(t - \Delta t, x; t, y) = \mathbf{M}\, [\xi(t) - \xi(t - \Delta t)]^2$$

есть математическое ожидание квадрата изменения $\xi(t)$ и, следовательно, пропорционально кинетической энергии (в предположении, что $\xi(t)$ есть координата движущейся под влиянием случайных воздействий точки), то из $(2')$ и $(3')$ ясно, что $a(t, x)$ есть средняя скорость изменения $\xi(t)$, а $b(t, x)$ пропорционально средней кинетической энергии изучаемой нами системы.

Мы заключим этот параграф рассмотрением частного случая уравнений Колмогорова, когда функция $f(t, x; \tau, y)$ зависит от t, τ и

$y-x$, но не от самих x и y. Физически это означает, что процесс протекает однородно в пространстве: вероятность получить приращение $\Delta = y - x$ не зависит от того, в каком положении x находилась система в момент времени t. Очевидно, что в этом случае функции $a(t, x)$ и $b(t, x)$ не зависят от x, а являются функциями только одного аргумента t:

$$a(t) = a(t, x); \quad b(t) = b(t, x).$$

Уравнения Колмогорова в рассматриваемом нами случае переписываются в таком виде:

$$\left.\begin{aligned} \frac{\partial f}{\partial t} &= -a(t)\frac{\partial f}{\partial x} - \frac{1}{2}b(t)\frac{\partial^2 f}{\partial x^2}, \\ \frac{\partial f}{\partial \tau} &= -a(\tau)\frac{\partial f}{\partial y} + \frac{1}{2}b(\tau)\frac{\partial^2 f}{\partial y^2}. \end{aligned}\right\} \tag{11}$$

Рассмотрим сначала частный случай, когда $a(t) = 0$ и $b(t) = 1$. Уравнения (11) при этом превращаются в уравнение теплопроводности

$$\frac{\partial f}{\partial \tau} = \frac{1}{2}\frac{\partial^2 f}{\partial y^2}$$

и ему сопряженное (12)

$$\frac{\partial f}{\partial t} = -\frac{1}{2}\frac{\partial^2 f}{\partial x^2}.$$

Из общей теории уравнения теплопроводности известно, что единственное решение этих уравнений, удовлетворяющее условиям (10), дается функцией

$$f(t, x; \tau, y) = \frac{1}{\sqrt{2\pi(\tau - t)}}\, e^{-\frac{(y-x)^2}{2(\tau - t)}}.$$

Заменой переменных

$$x' = x - \int_a^t a(z)\,dz, \quad y' = y - \int_a^\tau b(z)\,dz,$$

$$t' = \int_a^t b(z)\,dz, \quad \tau' = \int_a^\tau b(z)\,dz$$

уравнения (11) сводятся к уравнениям (12). Это дает возможность искомое решение уравнений (11) записать в виде

$$f(t, x; \tau, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y - x - A)^2}{2\sigma^2}}$$

где обозначено

$$A = \int_t^\tau a(z)\,dz, \quad \sigma^2 = \int_t^\tau b(z)\,dz.$$

§ 55. Чисто разрывный процесс. Уравнения Колмогорова–Феллера

В современном естествознании большую роль играют процессы, в которых изменение системы происходит не непрерывно, а скачками. Примеры такого рода задач приведены во вводном к настоящей главе параграфе.

Мы будем говорить, что случайный процесс $\xi(t)$ чисто разрывен, если в течение любого промежутка времени $(t, t + \Delta t)$ величина $\xi(t)$ остается неизменной и равной x с вероятностью $1 - p(t,x)\Delta t + o(\Delta t)$ и лишь с вероятностью $p(t, x)\Delta t + o(\Delta t)$ может претерпеть изменение (при этом мы считаем, что вероятность более чем одного изменения $\xi(t)$ за промежуток времени Δt есть $o(\Delta t)$). Естественно, что поскольку мы ограничиваемся рассмотрением процессов без последействия, функция распределения дальнейших после скачка изменений $\xi(t)$ уже не зависит от того, какое значение имело $\xi(t)$ в моменты, предшествующие скачку.

Обозначим через $P(t, x, y)$ условную функцию распределения $\xi(t)$ при условии, что в момент t произошел скачок и непосредственно до скачка $\xi(t)$ было равно x (т.е. $\xi(t - 0) = x$).

Функция распределения $F(t, x; \tau, y)$ легко может быть выражена через функции $p(t, x)$ и $P(t, x, y)$, а именно

$$F(t, x; \tau, y) = [1 - p(t, x)(\tau - t)]\, E(x, y) +$$

$$+ (\tau - t)\, p(t, x)\, P(t, x, y) + o(\tau - t). \qquad (1)$$

По смыслу определения функций $p(t, x)$ и $P(t, x, y)$ они неотрицательны, причем для $P(t, x, y)$, как для функции распределения, выполнены равенства

$$P(t, x, -\infty) = 0, \quad P(t, x, +\infty) = 1.$$

Кроме того, мы предположим, что $p(t, x)$ ограничена, обе функции $p(t, x)$ и $P(t, x, y)$ непрерывны относительно t и x (достаточно, на самом деле, предположить, что они измеримы по Борелю относительно x).

В отношении функции $F(t, x;\ \tau, y)$ мы не станем делать никаких предположений и лишь сохраним ее определение при $t = \tau$:

$$\lim_{\tau \to t+0} F(t, x;\ \tau, y) = \lim_{t \to \tau - 0} F(t, x;\ \tau, y) =$$

$$= E(x, y) = \begin{cases} 0 & \text{при} \quad y \leqslant x, \\ 1 & \text{при} \quad y > x. \end{cases}$$

Одна из задач настоящего параграфа состоит в доказательстве следующей теоремы.

Т е о р е м а. *Функция распределения $F(t, x;\ \tau, y)$ чисто разрывного процесса без последействия удовлетворяет двум следующим интегро-дифференциальным уравнениям*:

$$\frac{\partial F(t, x; \tau, y)}{\partial t} = p(t, x)\ [F(t, x; \tau, y) -$$

$$- \int F(t, z;\ \tau, y) d_z P(t, x, z)], \tag{2}$$

$$\frac{\partial F(t, x; \tau, y)}{\partial \tau} = - \int_{-\infty}^{y} p(t, z)\, d_z F(t, x;\ \tau, y) +$$

$$+ \int p(\tau, z) P(\tau, z, y)\, d_z F(t, x; \tau, z). \tag{3}$$

Уравнение (2) было получено А.Н. Колмогоровым в 1931 г.; в сделанных нами предположениях оба уравнения (2) и (3) были получены В. Феллером в 1937 г. Это обстоятельство приводит нас к естественному наименованию уравнений (2) и (3) *уравнениями Коломогорова–Феллера*.

Д о к а з а т е л ь с т в о. В силу обобщенного уравнения Маркова

$$F(t, x; \tau, y) = \int F(t + \Delta t, z; \tau, y)\, d_z F(t, x; t + \Delta t, z).$$

Подставив сюда значение $F(t, x; t + \Delta t, z)$ по формуле (1), находим, что

$$F(t, x; \tau, y) = \int F(t + \Delta t, z;\ \tau, y)\ d_z [1 - p(t, x)\, \Delta(t) +$$

$$+ o(\Delta t)\ ]\ E(x, z) + \int F(t + \Delta t, z; \tau, y)\, d_z\, [\, p(t, x)\Delta t + o(\Delta t)]\ P(t, x, z).$$

Так как

$$\int F(t + \Delta t, z; \tau, y)\, d_z E(x, z) = F(t + \Delta t, x;\ \tau, y),$$

то

$$F(t, x; \tau, y) = [1 - p(t,\ x)\, \Delta t]\ F(t + \Delta t, x;\ \tau, y) +$$

$$+ \Delta t p\,(t, x) \int F(t + \Delta t, z;\ \tau, y)\, d_z P(t, x, z) + o(\Delta t).$$

Отсюда

$$\frac{F(t+\Delta t, x; \tau, y) - F(t, x; \tau, y)}{\Delta t} = p(t, x)\, F(t+\Delta t, x; \tau, y) +$$

$$+\ p(t, x) \int F(t+\Delta t, z; \tau, y)\, d_z P(t, x, z) + o(1).$$

Переход к пределу приводит нас к (2).

Уравнение Маркова и (1), а также определение функции $E(x, z)$ позволяют написать следующую цепочку равенств:

$$F(t, x;\ \tau + \Delta\tau, y) = \int F(\tau, z;\ \tau+\Delta\tau, y)\, d_z F(t, x; \tau, z) =$$

$$= \int \{ [1 - p(\tau, z)\,\Delta\tau]\, E(z, y) + \Delta\tau\, p(\tau, z)\, P(\tau, z, y) +$$

$$+\ o(\Delta\tau)\}\ d_z F(t, x;\ \tau, z) =$$

$$= \int_{-\infty}^{y} d_z F(t, x;\ \tau, z) - \Delta\tau \int_{-\infty}^{y} p(\tau, z)\, d_z F(t, x;\ \tau, z) +$$

$$+\ \Delta t \int p(\tau, z)\ P(\tau, z, y)\, d_z F(t, x;\ \tau, z) + o(\Delta\tau).$$

Обычным путем отсюда следует существование производной $\dfrac{\partial F}{\partial \tau}$ и равенство (3).

Мы решим еще одну важную для приложений задачу: с какой вероятностью в течение промежутка времени от t до τ $(\tau > t)$ система может изменить свое состояние то или иное число n раз $(n = 0, 1, 2, \ldots)$?

Обозначим через $p_n(t, x, \tau)$ вероятность того, что отправляясь от состояния x в момент t, система n раз изменит свое состояние до момента τ. Решение задачи начнем со случая $n = 0$.

С этой целью запишем следующее равенство:

$$p_0(t, x, \tau) = p_0(t, x,\ \tau + \Delta\tau) + p_0(t, x, \tau)\,[1 - p_0(\tau, x, \tau+\Delta\tau)], \tag{4}$$

которое означает, что отсутствие изменений состояния системы в промежуток времени (t, τ) может произойти двумя несовместимыми путями: 1) система не изменила состояния за больший промежуток времени $(t, \tau + \Delta\tau)$, 2) система не меняла состояния до момента τ, но в промежуток времени $(\tau, \tau + \Delta\tau)$ состояние ее изменилось. Так как по определению чисто разрывного процесса

$$p_0(\tau, x,\ \tau + \Delta\tau) = 1 - p(\tau, x)\,\Delta\tau + o(\Delta\tau),$$

то уравнение (4) может быть записано иначе:

$$\frac{p_0(t, x, \tau+\Delta\tau) - p_0(t, x, \tau)}{\Delta\tau} = -p_0(t, x, \tau)\, p(\tau, x) + o(1).$$

Отсюда, положив $\Delta\tau \to 0$, находим, что существует производная $\dfrac{\partial p_0(t, x, \tau)}{\partial \tau}$ и что

$$\frac{\partial p_0(t, x, \tau)}{\partial \tau} = -p_0(t, x, \tau)\, p(\tau, x).$$

Проинтегрировав это уравнение, находим

$$p_0(t, x, \tau) = Ce^{-\int\limits_t^\tau p(u,\, x)\, du}$$

Так как

$$p_0(\tau, x, \tau) = 1,$$

то $C = 1$ и

$$p_0(t, x, \tau) = e^{-\int\limits_t^\tau p(u,\, x) du} \tag{5}$$

Теперь мы увидим, что, зная $p_0(t, x, \tau)$, а также функцию $P(t, x, y)$, определенную раньше, мы можем подсчитать любую вероятность $p_n(t, x, \tau)$. В самом деле, n-кратное изменение состояния происходит следующим образом: 1) до момента $s\,(t < s < \tau)$ система не меняет состояния (вероятность этого события равна $p_0(t, x, s)$), 2) в промежуток $(s, s + \Delta s)$ система меняет состояние (вероятность этого равна $p_1(s, x, s + \Delta s) = p(s, x)\Delta s + o(\Delta s)$), 3) вероятность того, что новое состояние, в котором окажется система, будет заключаться между y и $y + \Delta y$, равна $P(s, x, y + \Delta y) - P(s, x, y) = \Delta_y P(s, x, y)$, 4) наконец, за время $(s + \Delta s, \tau)$ система изменит свое состояние $n - 1$ раз (вероятность этого события равна $p_{n-1}(s + \Delta s, y, \tau)$).

Вероятность того, что произойдут все четыре перечисленных события, в силу теоремы умножения, равна

$$p_0(t, x, s)\,[p(s, x) + o(1)]\,\Delta s \cdot \Delta_y P(s, x, y) \cdot p_{n-1}(s + \Delta s, y, \tau).$$

Так как s и y могут быть произвольными ($t < s < \tau$ и $-\infty < y < \infty$), то, в силу формулы полной вероятности,

$$p_n(t, x, \tau) = \int\limits_t^\tau \int p_0(t, x, s)\, p(s, x)\, p_{n-1}(s, y, \tau)\, d_y P(s, x, y)\, ds =$$

$$= \int\limits_t^\tau p_0(t, x, s)\, p(s, x) \int p_{n-1}(s, y, \tau)\, d_y P(s, x, y)\, ds. \tag{6}$$

Отсюда, в частности,

$$p_1(t, x, \tau) = \int\limits_t^\tau p_0(t, x, s)\, p(s, x) \int p_0(s, y, \tau)\, d_y P(s, x, y)\, ds. \qquad (7)$$

Процесс определения $p_n(t, x, \tau)$ очевиден: по формуле (5) находим $p_0(t, x, \tau)$, по формуле (7) вычисляем $p_1(t, x, \tau)$ и затем последовательно $p_2(t, x, \tau), p_3(t, x, \tau)$ и, наконец, $p_n(t, x, \tau)$.

Пример 1. Пусть интересующая нас величина $\xi(t)$ есть число изменений состояния за время от 0 до t. В предположении $p(t, x) = a$, где $a > 0$ – постоянное, найти $p_n(t, x, \tau)$.

Возможными состояниями системы будут в нашем случае все неотрицательные целые числа ($x = 0, 1, 2, \ldots$) и только они. Так как при каждом изменении состояния величина $\xi(t)$ увеличивается ровно на 1, то

$$P(t, x, y) = \begin{cases} 0 & \text{при } y \leqslant x, \\ 1 & \text{при } y > x. \end{cases}$$

По формуле (5) имеем:

$$p_0(t, x, \tau) = e^{-a(\tau - t)}.$$

Согласно (7)

$$p_1(t, x, \tau) = \int\limits_t^\tau p_0(t, x, s)\, p(s, x)\, p_0(s, x + 1, \tau)\, ds =$$

$$= a \int\limits_t^\tau e^{-(s - t)a} e^{-(\tau - s)a}\, ds = a(\tau - t)\, e^{-a(\tau - t)}.$$

По формуле (7)

$$p_2(t, x, \tau) = \int\limits_t^\tau p_0(t, x, s)\, p(s, x)\, p_1(s, x + 1, \tau)\, ds =$$

$$= \frac{[a(\tau - t)]^2}{2!} e^{-a(\tau - t)}.$$

Предположим теперь, что

$$p_{n-1}(t, x, \tau) = \frac{[a(\tau - t)]^{n-1}}{(n-1)!} e^{-a(\tau - t)}.$$

По формуле (7)

$$p_n(t, x, \tau) = \int_t^\tau p_0(t, x, s)\, p(s, x)\, p_{n-1}(s, x+1, \tau)\, ds =$$

$$= \int_t^\tau \frac{a[a(\tau - s)]^{n-1}}{(n-1)!} e^{-a(\tau - t)}\, ds = \frac{[a(\tau - t)]^n}{n!} e^{-a(\tau - t)}.$$

Этим доказано, что при любом целом $n \geqslant 0$

$$p_n(t, x, \tau) = \frac{[a(\tau - t)]^n}{n!} e^{-a(\tau - t)}.$$

Решением нашей задачи является, таким образом, закон Пуассона. В частности,

$$p_n(0, 0, \tau) = \frac{(a\tau)^n}{n!} e^{-a\tau}.$$

Легко сообразить, что функция

$$F(t, x; \tau, y) = \begin{cases} 0 & \text{при } y \leqslant 0, \\ \sum\limits_{n<y} \dfrac{[a(\tau - t)]^n}{n!} e^{-a(\tau - t)} & \text{при } y > 0 \end{cases}$$

является решением интегродифференциальных уравнений (2) и (3).

Пример 2. В момент $t = 0$ имеется N радиоактивных атомов. Вероятность распада атома в промежуток времени $(t, t + \Delta t)$ равна $aN(t)\Delta t + o(\Delta t)$, где $a > 0$ – постоянное, а $N(t)$ – число атомов, не распавшихся до момента t. Найти вероятность того, что за время от t до τ произойдет n распадов*).

Мы имеем типичный чисто разрывный случайный процесс. Величина $\xi(t)$, понятно, может принимать только значения $0, 1, 2, \ldots, N(t)$.

По условию задачи

$$p(t, x) = \begin{cases} 0 & \text{при } x \leqslant 0 \text{ и } x \geqslant N, \\ a(N - x) & \text{при } 0 < x \leqslant N, \end{cases}$$

а

$$P(t, x, y) = \begin{cases} 0 & \text{при } y \leqslant x, \\ 1 & \text{при } y > x. \end{cases}$$

---

*) Мы предполагаем при этом, что продукты распада атома сами уже не распадаются и во всяком случае не воздействуют на еще нераспавшиеся атомы.

Оценим прежде всего вероятность того, что за время от 0 до t произойдет n распадов.

По формуле (5)

$$p_0(0,0,\tau) = e^{-\int\limits_0^\tau p(t,\,0)dt} = e^{-aN\tau}.$$

Точно так же

$$p_0(t, k\,\tau) = e^{-a(N-k)(\tau-t)}.$$

Далее, по формуле (7)

$$p_1(0,0,\tau) = \int\limits_0^\tau p_0(0,0,s)\,p(s,0)\,p_0(s,1,\tau)\,ds =$$

$$= \int\limits_0^\tau e^{-aNs}\,aNe^{-a(N-1)(\tau-s)}\,ds =$$

$$= Ne^{-aN\tau}\int\limits_0^\tau ae^{a(\tau-s)}\,ds = Ne^{-aN\tau}[e^{a\tau}-1]. \tag{8}$$

По формуле (7) легко последовательно найти $p_2(0,0,\tau)$, $p_3(0,0,\tau)$ и т.д. и доказать, что

$$p_n(0,0,\tau) = C_N^n\,e^{-aN\tau}\,[e^{a\tau}-1]^n. \tag{9}$$

Это мы предоставляем читателю.

Очевидно, что при $0 \leqslant n \leqslant N-k$ имеет место равенство

$$p_n(t,k,\tau) = C_{N-k}^n\;e^{-a(N-k)(\tau-t)}\,[e^{a(\tau-t)}-1]^n. \tag{9'}$$

Теперь мы можем перейти к определению интересующей нас вероятности, которую мы обозначим через $p_n(t,\tau)$. По формуле полной вероятности, используя затем (9) и (9′), находим, что

$$p_n(t,\tau) = \sum_{k=0}^{N-n} p_k(0,0,t)\cdot p_n(t,k,\tau) =$$

$$= \sum_{k=0}^{N-n} C_N^k\,e^{-aNt}\,[e^{at}-1]^k\,C_{N-k}^n\,e^{-a(N-k)(\tau-t)}[e^{a(\tau-t)}-1]^n =$$

$$= e^{-aN\tau}\,[e^{a(\tau-t)}-1]^n\sum_{k=0}^{N-n} C_N^k\,C_{N-k}^n\,e^{ak(\tau-t)}\,[e^{at}-1]^k.$$

Так как

$$C_N^k\,C_{N-k}^n = C_N^n\,C_{N-n}^k$$

и

$$\sum_{k=0}^{N-n} C_{N-n}^{k}\,[e^{a(\tau-t)}\,(e^{a\tau}-1)]^{k} = [1+e^{a(2\tau-t)}-e^{a(\tau-t)}]^{N-n},$$

то окончательно

$$p_n(t,\tau) = C_N^n\,[e^{-at}-e^{-a\tau}]^n\,[e^{-a\tau}+e^{a(\tau-t)}-e^{-at}]^{N-n}.$$

Легко понять, что функция

$$F(t,x;\ \tau,y) = \begin{cases} 0 & \text{при } \ y \leqslant x, \\ \sum\limits_{n<y} p_n(t,x,\tau) & \text{при } \ y < N-x, \\ 1 & \text{при } \ y > N-x \end{cases}$$

является решением интегродифференциальных уравнений (2) и (3).

§ 56. Однородные случайные процессы с независимыми приращениями

Мы рассмотрим теперь важный класс случайных процессов, полная характеристика которых будет дана в терминах характеристических функций.

Под *однородным случайным процессом с независимыми приращениями* понимается совокупность случайных величин $\xi(t)$, зависящих от одного действительного параметра t и удовлетворяющих двум следующим условиям:

1) функция распределения величины $\xi(t+t_0)-\xi(t_0)$ не зависит от t_0 (однородность процесса по времени);

2) для любых неперекрывающихся промежутков $(a,\ b)$ параметра t приращения величины $\xi(t)$, т.е. разности $\xi(b)-\xi(a)$ взаимно независимы (независимость приращений).

Прежде чем переходить к получению конкретных результатов, мы рассмотрим несколько примеров. В этих примерах условия, о которых только что шла речь, могут быть приняты в качестве рабочей гипотезы. Естественно, что их допустимость оправдывается только согласием выводов с опытом.

Пример 1. Диффузия газов. Рассмотрим молекулу некоторого газа, движущуюся среди других молекул того же газа при условиях постоянных температуры и плотности. Введем в пространстве декартовы координаты и станем следить, как изменяется с течением времени одна из координат избранной молекулы, скажем, координата x.

Вследствие случайных столкновений данной молекулы с другими молекулами эта координата будет изменяться во времени, получая случайные приращения. Требование постоянства условий, в которых находится газ, очевидно, означает собой однородность изучаемого процесса во времени. Ввиду большого числа движущихся молекул и слабой зависимости их движения процесс оказывается с независимыми приращениями.

Пример 2. Скорости молекул. Рассмотрим снова молекулу некоторого газа, движущуюся в объеме, наполненном молекулами того или иного газа постоянной плотности и температуры. Отнесем снова все пространство к декартовым осям координат и будем следить, как изменяется со временем компонента скорости по одной из осей координат. В своем движении молекула будет подвергаться случайным столкновениям с другими молекулами. Вследствие этих столкновений компонента скорости будет получать случайные приращения. Мы снова имеем однородный случайный процесс с независимыми приращениями.

Пример 3. Радиоактивный распад. Известно, что радиоактивность вещества состоит в том, что его атомы превращаются в атомы другого вещества, выделяя при этом значительное количество энергии. Наблюдения над сравнительно большими массами радиоактивного вещества показывают, что распад различных атомов происходит независимо друг от друга, так что числа распадов атомов в неперекрывающиеся промежутки времени независимы между собой. Кроме того, вероятности того, что за промежуток времени определенной длины произойдет некоторое число распадов, зависят от длины этого промежутка и практически не зависят от того, где во времени он расположен. В действительности, конечно, по мере уменьшения массы вещества его радиоактивность постепенно убывает. Однако для сравнительно небольших промежутков времени (и не слишком больших количеств вещества) это изменение настолько незначительно, что им вполне можно пренебрегать.

Легко привести большое число других примеров, где интересующее нас явление природы или технический процесс может рассматриваться как однородный процесс с независимыми приращениями. Укажем дополнительно на такие примеры: космическое излучение (число космических частиц, попавших за определенный промежуток времени на определенную площадку), обрывность пряжи на ватере, загрузка телефонистки (число вызовов абонентов, поступающих за определенный промежуток времени) и пр.

Перейдем теперь к выяснению характеристического свойства однородных случайных процессов с независимыми приращениями.

Обозначим функцию распределения приращения величины $\xi(t)$ за промежуток времени τ через $F(x, \tau)$. Тогда, если промежутки времени τ_1 и τ_2 не пересекаются, то

$$F(x;\ \tau_1 + \tau_2) = \int F(x - y;\ \tau_1)\, d_y F(y, \tau_2). \qquad (1)$$

Если $f(z, \tau)$ – характеристическая функция, т.е. если

$$f(z, \tau) = \int e^{izx} d_x F(x; \tau),$$

то равенство (1) в терминах характеристических функций принимает следующий вид:

$$f(z; \tau_1 + \tau_2) = f(z, \tau_1) \cdot f(z, \tau_2). \tag{1'}$$

Вообще, если интервалы времени $\tau_1, \tau_2, \ldots, \tau_n$ не пересекаются, то

$$f(z; \sum_{k=1}^{n} \tau_k) = \prod_{k=1}^{n} f(z; \tau_k).$$

В частности, если $\tau_1 = \tau_2 \ldots = \tau_n$ и $\sum_{k=1}^{n} \tau_k = \tau$, то

$$f(z, \tau) = [f(z; \tau/n)]^n.$$

Таким образом, *функция распределения любого однородного случайного процесса с независимыми приращениями безгранично делима.*

Нужно отметить, что к рассмотрению безгранично делимых законов распределения в теории вероятностей пришли благодаря изучению однородных процессов с независимыми приращениями. Мы видели, что теория безгранично делимых законов распределения оказала решающее влияние на развитие классических задач теории вероятностей по суммированию случайных величин. Если раньше, как мы указывали, интересы исследователей были сосредоточены на определении наиболее широких условий, при которых имеют место закон больших чисел и сходимость нормированных сумм к нормальному закону, то после того как А.Н. Колмогоровым был полностью охарактеризован класс законов, управляющих однородными случайными процессами без последействия, естественно возникли те общие задачи, которые были рассмотрены в предыдущей главе. Оказалось при этом, что основные законы распределения, которые раньше получались как асимптотические, в теории случайных процессов играют роль точных решений соответствующих функциональных уравнений. Более того, эта новая точка зрения позволила выяснить причины, в силу которых в классической теории вероятностей рассматривались только две предельные функции распределения – нормальный закон и закон Пуассона.

Поскольку при произвольном $\tau > 0$ для однородных процессов с независимыми приращениями

$$f(z, \tau) = [f(z, 1)]^\tau,$$

то они полностью определяются заданием характеристической функции величины $\xi(1) - \xi(0)$. В § 43 мы видели, что для безграничноделимых законов с конечной дисперсией

$$\ln \varphi(z, 1) = i\gamma z + \int \{e^{izu} - 1 - izu\} \frac{1}{u^2} \, dG(u), \tag{2}$$

где γ – действительное постоянное, а $G(u)$ – неубывающая функция с ограниченным изменением. Мы ограничимся рассмотрением этого частного случая однородных процессов.

Введем в формуле (2) такие обозначения:

$$M(u) = \int_{-\infty}^{u} \frac{1}{x^2} \, dG(x) \quad \text{для} \quad u < 0,$$

$$N(u) = \int_{u}^{\infty} \frac{1}{x^2} \, dG(x) \quad \text{для} \quad u > 0,$$

$$\sigma^2 = G(+0) - G(-0);$$

тогда она примет следующий вид:

$$\ln \varphi(z, 1) = i\gamma z - \frac{\sigma^2 z^2}{2} + \int_{-\infty}^{0} \{e^{izu} - 1 - izu\} \, dM(u) +$$

$$+ \int_{0}^{\infty} \{e^{izu} - 1 - izu\} \, dN(u). \tag{2'}$$

Выясним теперь теоретико-вероятностный смысл функций $M(u)$ и $N(u)$.

В § 43 при выводе формулы канонического представления безгранично-делимых законом мы ввели функцию

$$G_n(u) = n \int_{-\infty}^{u} x^2 \, d\Phi_n(x).$$

Положим

$$M_n(u) = \int_{-\infty}^{\infty} \frac{1}{x^2} \, dG_n(x) = n \, \Phi_n(u) \text{ для } u < 0$$

и

$$N_n(u) = \int_{u}^{\infty} \frac{1}{x^2} \, dG_n(x) = n[1 - \Phi_n(u)] \qquad \text{для } u > 0.$$

Из того, что при $n \to \infty$ в точках непрерывности функции $G(u)$

$$G_n(u) \to G(u)$$

мы по второй теореме Хелли делаем вывод, что в точках непрерывности функции $M(u)$

$$M_n(u) = n\,\Phi_n(u) \to M(u).$$

С точки зрения случайных процессов, $\Phi_n(x)$ $(x < 0)$ есть вероятность того, что величина $\xi(\tau)$ за промежуток $\left(\frac{k}{n}, \frac{k+1}{n}\right)$ изменения параметра τ получит отрицательное приращение по абсолютной величине, большее, чем x. Таким образом, $M_n(x)$ есть сумма по всем k от 0 до $n-1$ вероятностей того, что величина $\xi(t)$ получит отрицательное приращение скачками по абсолютной величине, большими, чем x, за промежутки $\left(\frac{k}{n}, \frac{k+1}{n}\right)$ изменения параметра τ. Поскольку $M(u)$ и $N(u)$ являются пределами при $n \to \infty$ соответственно функций $M_n(u)$ и $N_n(u)$, то они получили название функций скачков.

Если $M(u) \equiv 0$ (для $u < 0$) и $N(u) \equiv 0$ (для $u > 0$), т.е. функции скачков отсутствуют, то из формулы (2') видно, что в этом случае стохастический процесс управляется нормальным законом. Мы видим, что случайный процесс, управляемый нормальным законом, является непрерывным в смысле теории вероятностей. Мы докажем теперь более сильное утверждение.

Т е о р е м а. *Для того чтобы однородный случайный процесс с независимыми приращениями и конечной дисперсией*) управлялся нормальным законом**), необходимо и достаточно, чтобы при произвольном* $\epsilon > 0$ *вероятность того, что максимальное значение абсолютной величины приращений* $\xi(\tau)$ за промежутки $\left(\frac{k-1}{n}, \frac{k}{n}\right)$ $(k = 1, 2, \ldots, n)$ *превзойдет* ϵ, *стремилась к нулю вместе с* $1/n$***).

Д о к а з а т е л ь с т в о. Мы только что видели, что однородный случайный процесс с независимыми приращениями управляется нормальным законом тогда и только тогда, когда при $x > 0$

$$M(-x) \equiv N(x) \equiv 0. \tag{3}$$

Так как

$$M(u) = \lim_{n\to\infty} M_n(u) \quad \text{и} \quad N(u) = \lim_{n\to\infty} N_n(u),$$

---

*) Теорема верна и без допущения конечности дисперсии.

**) В частности, нормальным законом с дисперисей 0, т.е. законом вида $F(x) = 0$ при $x \leqslant a$, $F(x) = 1$ при $x > a$.

***) Таким образом, процессы, управляемые нормальным законом, и только они, являются "равномерно непрерывными" в смысле теории вероятностей.

то условие (3) равносильно следующему:

$$\lim_{n\to\infty} n\Phi_n(-u) = \lim_{n\to\infty} n[1-\Phi_n(u)] = 0. \tag{4}$$

Обозначим приращение $\xi(\tau)$ в интервале $\left(\frac{k-1}{n}, \frac{k}{n}\right)$ через ξ_{nk}; тогда

$$p_{nk} = \Phi_n(-x) + 1 - \Phi_n(x+0) = \mathsf{P}\{|\xi_{nk}| > x\}.$$

Очевидно, что соотношения (4) эквивалентны такому:

$$\lim_{n\to\infty} \sum_{k=1}^{n} p_{nk} = 0.$$

Из неравенств

$$1 - \sum_{k=1}^{n} p_{nk} \leqslant \prod_{k=1}^{n} (1-p_{nk}) \leqslant e^{-\sum_{k=1}^{n} p_{nk}} \leqslant 1,$$

мы видим, что соотношения (4) равносильны утверждению, что

$$\lim_{n\to\infty} \prod_{k=1}^{n} (1-p_{nk}) = 1,$$

которое означает, что вероятность осуществления неравенств $|\xi_{nk}| < \epsilon$ при всех $k(1 \leqslant k \leqslant n)$ при $n\to\infty$ стремится к единице. Иначе говоря, мы доказали, что соотношения (3) имеют место тогда и только тогда, когда при $n\to\infty$

$$\mathsf{P}\{\max_{1\leqslant k\leqslant n} |\xi_{nk}| \geqslant \epsilon\} \to 0,$$

что и требовалось доказать.

§ 57. Понятие стационарного случайного процесса. Теорема Хинчина о корреляционной функции

Процессы марковского типа или, иначе, процессы без последействия, изученные нами в предыдущих параграфах, ни в какой мере не исчерпывают всех запросов естествознания к теории вероятностей. В самом деле, во многих случаях прошлые состояния системы оказывают весьма сильное влияние на вероятности ее будущих состояний, и пренебрегать этим воздействием прошлого нельзя даже при приближенной трактовке вопроса. Принципиально положение может быть исправлено изменением понятия состояния системы путем введения новых параметров. Так, например, если бы изменение положения частицы в явлениях диффузии или броуновского движения мы стали рассматривать как процесс без последействия, то это означало бы, что мы при этом не принимаем в расчет инерцию частицы, которая, само собой разумеется, в этих явлениях играет существенную роль. Введение в понятие состояния помимо координат частицы ее скорости в приведенном примере исправило бы положение. Однако существуют случаи, когда такое исправление никакого облегчения при решении поставленных задач не дает. В первую очередь здесь следует указать на статистическую механику, в которой указание на положение точки в той или иной ячейке фазового пространства дает только вероятностное суждение о будущем ее состоянии. При этом ознакомление с предыдущими положениями точки существенно меняет наши суждения относительно ее будущего. В связи с этим А.Я. Хинчин выделил важный класс случайных процессов с последействием, так называемые стационарные процессы, однородно ведущие себя во времени.

Стохастический процесс $\xi(t)$ называется *стационарным*, если распределения вероятностей для двух конечных групп переменных $\xi(t_1), \xi(t_2), \ldots, \xi(t_n)$ и $\xi(t_1 + u), \xi(t_2 + u), \ldots, \xi(t_n + u)$ совпадают и, значит, не зависят от u. Числа n и u, а также моменты времени $t_1, t_2, \ldots, t_n$ могут быть при этом выбраны совершенно произвольно.

К стационарным процессам приводит, например, изучение ряда акустических явлений, в том числе встречающихся в радиотехнике (случайные шумы), а также разыскание скрытых периодичностей, интересующее астрономов, геофизиков и метеорологов.

Часто в установившемся технологическом процессе легко подметить явления, протекающие по схеме стационарных процессов. Для примера рассмотрим процесс прядения. Значительная неоднородность свойств прядильных материалов (длина волокон, их крепость, величина поперечного сечения и пр.), колебания в скорости и равномерности подачи продукта на машинах в различные этапы процесса прядения и многие другие причины приводят к тому, что свойства пряжи меняются от одного сечения к другому. При этом оказывается, что знание того или иного свойства пряжи в какой-либо одной части мотка не дает нам полного знания ее свойств в другой его части. Но поскольку процесс прядения можно считать установившимся, постольку вероятностные характеристики качества пряжи представляют собой стационарный процесс.

Понятно, что любая числовая характеристика стационарного процесса $\xi(t)$ не зависит от момента t и, например, если $\xi(t)$ имеет конечную дисперсию, то, очевидно, имеют место следующие равенства:

$$\mathbf{M}\xi(t+u) = \mathbf{M}\xi(t) = \mathbf{M}\xi(0) = a,$$

$$\mathbf{D}\xi(t+u) = \mathbf{D}\xi(t) = \mathbf{D}\xi(0) = \sigma^2,$$

$$\mathbf{M}\{\xi(t+u)\,\xi(t)\} = \mathbf{M}\{\xi(u)\,\xi(0)\}.$$

Это обстоятельство позволяет без ограничения общности дальнейших результатов считать $a = 0$ и $\sigma = 1$ (для этого, очевидно, достаточно вместо $\xi(t)$ рассматривать отношение $\left.\dfrac{\xi(t) - a}{\sigma}\right)$.

Мы ограничимся здесь только изучением важнейшей числовой характеристики $\xi(t)$ – ее *корреляционной* функции, т.е. коэффициента корреляции между величинами $\xi(t)$ и $\xi(t+u)$

$$R(u) = \frac{\mathsf{M}[\xi(t+u) - \mathsf{M}\xi(t+u)]\ [\xi(t) - \mathsf{M}\xi(t)]}{\sqrt{\mathsf{D}\xi(t) \cdot \mathsf{D}\xi(t+u)}}.$$

В силу сделанного предположения о том, что $a = 0$ и $\sigma = 1$, выражение для $R(u)$ принимает более простой вид

$$R(u) = \mathsf{M}\{\xi(u)\,\xi(0)\}.$$

Мы назовем стационарный процесс н е п р е р ы в н ы м, если

$$\lim_{u\to 0} R(u) = 1.$$

В случае непрерывного стационарного процесса $R(u)$ есть непрерывная функция от u. Действительно,

$$|R(u + \Delta u) - R(u)| = |\mathsf{M}\{\xi(u+\Delta u)\xi(0)\} - \mathsf{M}\{\xi(u)\,\xi(0)\}| =$$

$$= |\mathsf{M}\{\xi(0)\,[\xi(u+\Delta u) - \xi(u)]\}|.$$

Но в силу неравенства Коши-Буняковского

$$|\mathsf{M}\{\xi(0)\,[\xi(u+\Delta u) - \xi(u)]\}| \leqslant \sqrt{\mathsf{M}\,\xi^2(0)\cdot \mathsf{M}[\xi(u+\Delta u) - \xi(u)]^2}.$$

А так как

$$\mathsf{M}\,\xi^2(0) = 1$$

и

$$\mathsf{M}[\xi(u+\Delta u) - \xi(u)]^2 = 2[1 - R(\Delta u)],$$

то окончательно

$$|R(u+\Delta u) - R(u)| \leqslant \sqrt{2(1 - R(\Delta u))}.$$

Это неравенство доказывает наше утверждение.

В теореме, которая сейчас будет доказана, стационарность процесса $\xi(t)$ можно понимать в следующем более широком смысле: процесс $\xi(t)$ *стационарен в широком смысле,* если математическое ожидание и дисперсия $\xi(t)$ не зависят от t, а коэффициент корреляции между $\xi(t_1)$ и $\xi(t_2)$ является функцией только $|t_2 - t_1|$.

Т е о р е м а Х и н ч и н а. *Для того чтобы функция $R(u)$ представляла корреляционную функцию некоторого непрерывного стационарного про-*

цесса, необходимо и достаточно, чтобы ее можно было представить в виде

$$R(u) = \int \cos ux\, dF(x), \tag{1}$$

где $F(x)$ – некоторая функция распределения.

Доказательство. Условие теоремы необходимо. В самом деле, если $R(u)$ есть корреляционная функция непрерывного стационарного процесса, то она непрерывна и ограничена. Докажем, кроме того, что она положительно-определенна. Действительно, каковы бы ни были действительные числа $u_1, u_2, \ldots, u_n$, комплексные числа $\eta_1, \eta_2, \ldots, \eta_n$ и целое число n, имеет место следующее соотношение:

$$0 \leqslant \mathbf{M} \left| \sum_{k=1}^{n} \eta_k \xi(u_k) \right|^2 = \mathbf{M} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \overline{\eta_j}\, \xi(u_i)\, \xi(u_j) \right\} =$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} R(u_i - u_j)\, \eta_i \overline{\eta_j}.$$

В силу теоремы Бохнера–Хинчина (§ 36) отсюда следует, что $R(u)$ может быть представлена в виде

$$R(u) = \int e^{iux}\, dF(x).$$

где $F(x)$ – неубывающая функция с ограниченным изменением. В силу вещественности функции $R(u)$ отсюда получаем:

$$R(u) = \int \cos ux\, dF(x).$$

Наконец, приняв во внимание условие непрерывности процесса:

$$R(+0) = 1,$$

находим, что $F(+\infty) - F(-\infty) = 1$, т.е. что $F(x)$ есть некоторая функция распределения.

Условие достаточно. Нам дано, что $R(u)$ есть функция вида (1). Требуется доказать. что существует стационарный процесс $\xi(t)$, имеющий своей корреляционной функцией функцию $R(u)$. С этой целью для каждого целого n и каждой группы действительных чисел $t_1, t_2, \ldots, t_n$ рассматриваем n-мерный вектор $\xi(t_1), \xi(t_2), \ldots, \xi(t_n)$, распределенный нормально и обладающий свойствами

$$\mathbf{M}\xi(t_1) = \mathbf{M}\xi(t_2) = \ldots = \mathbf{M}\xi(t_n) = 0,$$

$$\mathbf{D}\,\xi(t_1) = \mathbf{D}\xi(t_2) = \ldots = \mathbf{D}\,\xi(t_n) = 1,$$

при любых i и j коэффициент корреляции между $\xi(t_i)$ и $\xi(t_j)$ равен $R(t_i - t_j)$, т.е.

$$\mathbf{M}\xi(t_i)\,\xi(t_j) = R(t_i - t_j).$$

Вид функции $R(u)$ обеспечивает положительную определенность квадратичной формы, стоящей в показателе n-мерного нормального закона. Определенный таким образом н о р м а л ь н ы й случайный процесс стационарен и в узком и в широком смысле слова.

Доказанная теорема играет основную роль в теории стационарных процессов и в ее физических приложениях. За подробностями отсылаем к специальной литературе, для начала к литературе, приведенной в конце книги.

П р и м е р 1. Пусть

$$\xi(t) = \xi \cos \lambda t + \eta \sin \lambda t,$$

где ξ и η – некоррелированные*) случайные величины, для которых $\mathbf{M}\xi = \mathbf{M}\eta = 0$, $\mathbf{D}\xi = \mathbf{D}\eta = 1$, а λ – постоянное. Так как

$$R(u) = \mathbf{M}\xi(t+u)\,\xi(t) =$$

$$= \mathbf{M}\,[\xi \cos \lambda(t+u) + \eta \sin \lambda(t+u)] \cdot [\xi \cos \lambda t + \eta \sin \lambda t] =$$

$$= \mathbf{M}\,[\xi^2 \cos \lambda t \cos \lambda(t+u) + \xi\eta(\sin \lambda(t+u) \cos \lambda t +$$

$$+ \cos \lambda(t+u) \sin \lambda t) + \eta^2 \sin \lambda t \sin \lambda(t+u) =$$

$$= \cos \lambda t \cos \lambda(t+u) + \sin \lambda t \sin \lambda(t+u) = \cos \lambda u,$$

то процесс $\xi(t)$ стационарен в широком смысле. Для него в формуле (3) мы должны положить

$$F(x) = \begin{cases} 0 & \text{при} \quad x \leqslant -\lambda, \\ 0{,}5 & \text{при} \quad -\lambda < x \leqslant \lambda, \\ 1 & \text{при} \quad x > \lambda. \end{cases}$$

П р и м е р 2. Пусть

$$\xi(t) = \sum_{k=1}^{n} b_k \xi_k(t),$$

где $\xi_k(t) = \xi_k \cos \lambda_k t + \eta_k \sin \lambda_k t$, λ_k – постоянные, $\sum\limits_{k=1}^{n} b_k^2 = 1$, случайные величины ξ_k и η_k удовлетворяют следующим условиям:

$$\mathbf{M}\xi_k = \mathbf{M}\eta_k = 0, \quad \mathbf{D}\xi_k = \mathbf{D}\eta_k = 1, \qquad (1 \leqslant k \leqslant n),$$

$$\mathbf{M}\xi_i \xi_j = \mathbf{M}\eta_i \eta_j = 0 \quad \text{при} \quad i \neq j,$$

$$\mathbf{M}\xi_i \eta_j = 0 \quad \text{при} \quad i, j = 1, 2, \ldots, n.$$

---

*) Случайные величины ξ и η называются некоррелированными, если $\mathbf{M}\xi\eta = \mathbf{M}\xi \cdot \mathbf{M}\eta$.

Легко подсчитать, что корреляционная функция для $\xi(t)$ равна

$$R(u) = \sum_{k=1}^{n} b_k^2 \cos \lambda_k u$$

и что, следовательно, процесс является стационарным в широком смысле. Функция $F(x)$ в формуле (3) растет только в точках и имеет в них скачки размера $0{,}5\, b_k^2$.

Случайные процессы, для которых $F(x)$ растет только скачками, называются *процессами с дискретным спектром.*

Легко видеть, что всякий процесс вида

$$\xi(t) = \sum_{k=1}^{\infty} b_k \xi_k(t) \tag{4}$$

где $\sum\limits_{k=1}^{\infty} b_k^2 < \infty$ и $\xi_k(t)$ сохраняют смысл, приданный им в примере 2, является стационарным в широком смысле и имеет дискретный спектр. Важно отметить, что Е.Е. Слуцкий обнаружил глубокое обратное предложение: *всякий стационарный процесс с дискретным спектром представим в виде* (4). Обобщение этой теоремы Слуцкого на случай произвольного спектра будет сформулировано в следующем параграфе.

Параллельно с развитием теории стационарных процессов развивалась теория стационарных последовательностей. Последовательность случайных величин

$$\ldots, \xi_{-2}, \xi_{-1}, \xi_0, \xi_1, \xi_2, \ldots$$

называется *стационарной в широком смысле,* если для всех членов последовательности математические ожидания и дисперсии являются постоянными числами, не зависящими от места в последовательности

$$\ldots = \mathsf{M}\xi_{-2} = \mathsf{M}\xi_{-1} = \mathsf{M}\xi_0 = \mathsf{M}\xi_1 = \mathsf{M}\xi_2 = \ldots = a,$$

$$\ldots = \mathsf{D}\xi_{-2} = \mathsf{D}\xi_{-1} = \mathsf{D}\xi_0 = \mathsf{D}\xi_1 = \mathsf{D}\xi_2 = \ldots = \sigma^2,$$

а коэффициент корреляции между ξ_i и ξ_j является функцией только $|i-j|$.

В качестве упражнения мы предлагаем читателю

1) вывести, используя результаты § 36, общий вид корреляционной функции для стационарной последовательности;

2) доказать теорему – если для стационарной последовательности

$$\lim_{s \to \infty} R(s) = 0,$$

где $R(s)$ – коэффициент корреляции между ξ_i и ξ_{i+s}, то для нее имеет место закон больших чисел, т.е. при $n \to \infty$

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| < \epsilon\right\} \to 1,$$

каково бы ни было постоянное $\epsilon > 0$.

§ 58. Понятие стохастического интеграла. Спектральное разложение стационарных процессов

Для дальнейшего нам необходимо ввести понятие стохастического интеграла. Пусть в сегменте $a \leqslant t \leqslant b$ заданы случайный процесс $\xi(t)$ и числовая функция $f(t)$. Разобьем сегмент $[a, b]$ точками $a = t_0 < t_1 < \ldots < t_n = b$ и рассмотрим сумму

$$J_n = \sum_{i=1}^{n} f(t_i)\xi(t_i)(t_i - t_{i-1}).$$

Если при $\max\limits_{1 \leqslant i \leqslant n}(t_i - t_{i-1}) \to 0$ эта сумма стремится к некоторому пределу (представляющему собой, вообще говоря, случайную величину), то этот предел называется *интегралом от случайного процесса* $\xi(t)$ и обозначается символом

$$J = \int_a^b f(t)\xi(t)\,dt.$$

Несобственный интеграл (при $a = -\infty, b = +\infty$) определяется обычным путем как предел собственных интегралов при $a \to -\infty$ $b \to \infty$.

Сходимость интегральных сумм J_n мы будем понимать в следующем смысле: существует случайная величина J такая, что при $n \to \infty$

$$\mathsf{M}(J_n - J)^2 \to 0. \tag{1}$$

Опираясь на известные теоремы теории функций действительного переменного легко доказать, что последовательность случайных величин J_n сходится к пределу J в смысле (1) тогда и только тогда, когда при $\min(m, n) \to \infty$

$$\mathsf{M}(J_m - J_n)^2 \to 0. \tag{2}$$

На доказательстве этого факта мы останавливаться не станем.

Теорема 1. *Для существования интеграла*

$$J = \int_a^b f(t)\,\xi(t)\,dt$$

достаточно, чтобы существовал интеграл

$$A = \int_a^b \int_a^b R(t - s) f(t) f(s)\, ds\, dt.$$

При этом

$$A = \mathsf{M}\left[\int_a^b f(t)\xi(t)\, dt\right]^2.$$

Доказательство. Действительно, для доказательства первой половины теоремы достаточно обнаружить, что если существует интеграл A, то имеет место соотношение (2). Имеем

$$\mathsf{M}(J_n - J_m)^2 = \mathsf{M}\left[\sum_{i=1}^{n} f(t_i)\xi(t_i)\Delta t_i\right]^2 -$$

$$- 2\mathsf{M} \sum_{i=1}^{n} \sum_{j=1}^{m} f(t_i) f(s_j)\xi(t_i)\xi(s_j)\Delta t_i\, \Delta s_j + \mathsf{M}\left[\sum_{j=1}^{m} f(s_j)\xi(s_j)\,\Delta s_j\right]^2 =$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{n} f(t_i) f(\tau_k) R(t_i - \tau_k)\Delta t_i \Delta\tau_k -$$

$$- 2 \sum_{i=1}^{n} \sum_{j=1}^{m} f(t_i) f(s_j) R(t_i - s_j)\,\Delta t_i\,\Delta s_j +$$

$$+ \sum_{j=1}^{m} \sum_{k=1}^{m} f(s_j) f(\sigma_k) R(s_j - \sigma_k)\, ds_j\,\Delta\sigma_k.$$

Здесь численные значения t_i и τ_i, s_j и σ_j совпадают.

В силу предположения о существовании интеграла A,

$$A = \lim \sum_{k=1}^{n} \sum_{i=1}^{n} f(t_k) f(\tau_i) R(t_k - \tau_i)\Delta t_i \Delta\tau_k =$$

$$= \lim \sum_{i=1}^{n} \sum_{j=1}^{m} f(t_i) f(s_j) R(t_i - s_j)\,\Delta t_i\,\Delta s_j =$$

$$= \lim \sum_{j=1}^{m} \sum_{k=1}^{m} f(s_j) f(\sigma_k) R(s_j - \sigma_k)\,\Delta s_j\,\Delta\sigma_k,$$

если только $\max(\Delta t_i, \Delta s_j) \to 0$. Таким образом, при $\min(m, n) \to \infty$

$$\mathsf{M}(J_m - J_n)^2 \to 0.$$

Для доказательства второй части теоремы заметим, что

$$\mathbf{M}\left[\sum_{i=1}^{n} f(t_i)\,\xi(t_i)\,\Delta t_i\right]^2 = \mathbf{M}\sum_{j=1}^{n}\sum_{i=1}^{n} f(t_i)f(\tau_j)\xi(t_i)\xi(\tau_j)\,\Delta t_i\,\Delta\tau_j =$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{n} f(t_i)f(\tau_j)\,R(t_i-\tau_j)\Delta t_i\,\Delta\tau_j;$$

при $\max \Delta t_i \to 0$ последняя часть равенств стремится к A.

Наряду с только что введенным понятием стохастического интеграла можно рассматривать также *стохастический интеграл Стилтьеса*, который мы определим как предел сумм

$$\sum_{k=1}^{n} f(t_k)[\xi(t_k)-\xi(t_{k-1})] \tag{3}$$

при $\max(t_i - t_{i-1}) \to 0$. Здесь по-прежнему $a = t_0 < t_1 < \ldots < t_n = b$ и предел понимается в смысле (1). Если предел сумм (3) существует, то мы станем обозначать его символом

$$\int_a^b f(t)\,d\xi(t).$$

В конце предыдущего параграфа мы сформулировали теорему Слуцкого, выясняющую связь между стационарными процессами с дискретным спектром и рядами Фурье со случайными некоррелированными коэффициентами. Можно доказать, что для каждого стационарного в широком смысле процесса имеет место следующее свойство: каковы бы ни были $\epsilon > 0$ и (сколь угодно большое) T, существуют такие попарно некоррелированные случайные величины ξ_k, η_k $(1 \leqslant k \leqslant n)$ и такие вещественные числа $\lambda_k\,(1 \leqslant k \leqslant n)$*), что при любом t из сегмента $-T \leqslant t \leqslant T$ выполняется неравенство

$$\mathbf{M}\left[\xi(t) - \sum_{k=1}^{n}(\xi_k\cos\lambda_k t + \eta_k \sin\lambda_k t)\right]^2 < \epsilon.$$

Отсюда, в частности, следует, что в указанных условиях

$$\mathbf{P}\left\{\left|\,\xi(t) - \sum_{k=1}^{n}(\xi_k\cos\lambda_k t + \eta_k\sin\lambda_k t)\right| > \eta\right\} \leqslant \epsilon/\eta^2$$

где η — наперед заданное положительное число.

---

*) Числа n и λ_k, а также величины ξ_k и η_k не зависят от ϵ и T.

Приведем без доказательства следующую важную теорему.

Теорема 2. *Всякий стационарный в широком смысле случайный процесс представим в виде*

$$\xi(t) = \int_0^\infty \cos \lambda t\, dZ_1(\lambda) + \int_0^\infty \sin \lambda t\, dZ_2(\lambda), \tag{4}$$

где случайные процессы $Z_1(\lambda)$ *и* $Z_2(\lambda)$ *обладают следующими свойствами:*

а) $\mathbf{M}\,[Z_i(\lambda_1 + \Delta\lambda_1) - Z_i(\lambda_1)] \cdot \overline{[Z_j(\lambda_j + \Delta\lambda_2) - Z_j(\lambda_2)]} = 0,$

$i, j = 1, 2,$

если $i \neq j$ *и для неперекрывающихся отрезков* $(\lambda_1, \lambda_1 + \Delta\lambda_1)$, $(\lambda_2, \lambda_2 + \Delta\lambda_2)$ *также при* $i = j$;

б) $\mathbf{M}[Z_1(\lambda + \Delta\lambda) - Z_1(\lambda)]^2 = \mathbf{M}[Z_2(\lambda + \Delta\lambda) - Z_2(\lambda)]^2.$

Формулу (4) естественно называть *спектральным разложением процесса* $\xi(t)$.

Случайные процессы $Z_1(\lambda)$ и $Z_2(\lambda)$ формулы (4) могут быть определены посредством равенств

$$Z_1(\lambda) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{-\sin \lambda t}{t}\, \xi(t)\, dt$$

и

$$Z_2(\lambda) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{1 - \cos \lambda t}{t}\, \xi(t)\, dt.$$

Легко доказать, что оба указанных интеграла существуют; можно также показать, что

$$F(\lambda + \Delta\lambda) - F(\lambda) = \mathbf{M}\,[Z_1(\lambda + \Delta\lambda) - Z_1(\lambda)]^2,$$

где $F(\lambda)$ определена теоремой Хинчина.

Возможность разложения (4) для произвольного стационарного в широком смысле случайного процесса была указана в 1940 г. А.Н. Колмогоровым. Этот результат им формулировался в терминах геометрии гильбертовских пространств и доказывался посредством спектральной теории операторов. Теоретико-вероятностному истолкованию и выводу этого разложения были посвящены впоследствии работы многих авторов – Г. Крамера, К. Карунена, М. Лоэва, Бланк-Лапьера и др.

§ 59. Эргодическая теорема Биркгофа – Хинчина

В 1931 г. американский математик Георг Биркгоф доказал одну общую теорему механики, которая, как показал через три года А.Я. Хинчин, допускает широкое теоретико-вероятностное обобщение. Эта теорема состоит в следующем: *если непрерывный стационарный процесс $\xi(t)$ имеет конечное математическое ожидание, то с вероятностью единица существует предел*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \xi(t)\,dt.$$

Стационарность процесса предположена здесь в узком, а не в широком смысле этого слова.

Так как это предложение представляет собой своеобразную форму усиленного закона больших чисел, то мы докажем ее с целью непосредственного продолжения формулировок главы 6 не для процессов, а для стационарных последовательностей.

Теорема. *Для стационарной последовательности случайных величин*

$$\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots,$$

для которых $\mathsf{M}\xi_j$ *конечно, последовательность средних арифметических*

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i$$

с вероятностью единица сходится к пределу.

Доказательство. Введем обозначение

$$h_{ab} = \frac{\xi_a + \xi_{a+1} + \ldots + \xi_b}{b-a}.$$

Нам требуется показать, что с вероятностью единица величины h_{0b} при $b\to\infty$ стремятся к пределу. Обозначим случайное событие, состоящее в существовании этого предела, буквой $\overline{K}$. Нам нужно доказать, что $\mathsf{P}(\overline{K}) = 1$ или, что то же самое, $\mathsf{P}(K) = 0$.

Предположим обратное, что событие K (т.е. что величины h_{0b} при $b\to\infty$ не сходятся к пределу) имеет положительную вероятность и покажем, что это предположение приводит к противоречию.

С этой целью рассмотрим все сегменты (α_n, β_n) с рациональными концами $\alpha_n < \beta_n$. Множество всех таких сегментов счетно. Если $\lim\limits_{b\to\infty} h_{0b}$ не существует, то найдется такой сегмент (α_n, β_n), для которого $\limsup\limits_{b\to\infty} h_{0b} > \beta_n$ и $\limsup\limits_{b\to\infty} h_{0b} < \alpha_n$ (событие K_n). Таким образом, событие K распадается на счетное множество несовместимых случаев K_n. Так как по предположению $\mathsf{P}(K) > 0$, то найдется такое n, что $\mathsf{P}(K_n) > 0$.

Таким образом доказано, что если $\mathsf{P}(K) > 0$, то существуют два числа α и β ($\alpha < \beta$), для которых одновременно выполняются неравенства

$$\left.\begin{array}{l}\limsup h_{0b} > \beta, \\ \liminf h_{0b} < \alpha.\end{array}\right\} \qquad (1)$$

Предположим теперь, что все ξ_j приняли какие-то определенные значения. Если сегмент (a, b) таков, что $h_{ab} > \beta$, но при всех b', для которых $\alpha < b' < b$, $h_{ab'} \leqslant \beta$, то этот сегмент назовем *особым* (относительно β).

Легко обнаружить, что два особых сегмента не перекрываются. Действительно, если два особых сегмента (a, b) и (a_1, b_1) таковы, что $a < a_1 < < b < b_1$, то из равенства

$$h_{ab} = \frac{(a_1 - a)\,h_{aa_1} + (b - a_1)\,h_{a_1 b}}{b - a}$$

и неравенства $h_{ab} > \beta$ вытекает, что или $h_{aa_1} > \beta$ или $h_{a_1 b} > \beta$. Однако первое из этих неравенств невозможно, так как сегмент (a, b) особый, а второе неравенство так же невозможно, поскольку сегмент (a_1, b_1) особый.

Разность $b - a$ назовем *рангом сегмента* (a, b). Если сегмент (a, b) является особым, имеет ранг на превышающий s и не заключен ни в одном сегменте ранга, не превышающего s, то такой сегмент назовем s*-особым*.

Так как среди особых сегментов, заключающих в себе произвольный сегмент (α, β) ранга, не превышающего s, и имеющих также ранг не больший s, должен найтись хотя бы один наибольшей длины. Если бы таких сегментов нашлось два, то они перекрывались бы, что по ранее доказанному невозможно. Таким образом, каждый особый сегмент ранга не большего s, может находиться внутри только одного s-особого (или же совпадать с ним). Из определения следует, что два s-особых сегмента могут лежать только один вне другого.

Обозначим через K_s событие, состоящее в том, что выполнены неравенства (1) и, кроме того, существует такое $t \leqslant s$, что $h_{0t} > \beta$. Так как K является пределом для событий K_s, то

$$\mathsf{P}(K) = \lim_{s \to \infty} \mathsf{P}(K_s).$$

Отсюда следует, что для всех достаточно больших s имеет место неравенство $\mathsf{P}(K_s) > 0$. Далее мы ограничиваемся рассмотрением только таких значений s.

Пусть событие K_s имеет место. Тогда среди тех $t \leqslant s$, для которых $h_{0t} > \beta$, существует наименьшее t'. Сегмент $(0, t')$ — особый. Следовательно он заключен в некотором s-особом сегменте (a, b) (или же сам является таковым), для которого $a \leqslant 0 < b$. Верно и обратное: если существует s-особый сегмент (a, b), для которого $a \leqslant 0 < b$, то существует такое $t \leqslant s$, что $h_{0t} > \beta$. Для $a = 0$ это очевидно: достаточно положить $t = b$. Если же $a < 0$, то из равенства

$$h_{ab} = \frac{-ah_{a0} + bh_{0b}}{b - a}$$

и неравенств $h_{ab} > \beta$, $h_{a0} \leqslant \beta$ вытекает $h_{0b} > \beta$. Таким образом, и в этом случае можно положить $t = b$.

Обозначим $-a$ через p, $b - a$ через q. Так как s-особый сегмент $(-p, -p + q)$ может существовать только один, то событие K_s разбивается на несовместимые случаи K_{pq}, соответствующие наличию s-особых сегментов $(-p, -p + q)$:

$$K_s = \sum_{p,q} K_{pq} \quad (q = 1, 2, \ldots, s,\ p = 0, 1, \ldots, q - 1).$$

Замена нумерации последовательности $i' = i + p$ переводит случай K_{0q} в случай K_{pq}. Поэтому в силу стационарности*) $\mathbf{P}(K_{pq}) = \mathbf{P}(K_{0q})$ и $\mathbf{M}(\xi_0 \mid K_{pq}) = \mathbf{M}(\xi_0 \mid K_{0q})$. Так как

$$\mathbf{P}(K_s)\mathbf{M}(\xi_0 \mid K_s) = \sum_{p,q} \mathbf{P}(K_{pq})\mathbf{M}(\xi_0 \mid K_{pq}) =$$

$$= \sum_q \mathbf{P}(K_{0q}) \sum_p \mathbf{M}(\xi_p \mid K_{0q}) = \sum_q \mathbf{P}(K_{0q})\mathbf{M}(h_{0q} \mid K_{0q}),$$

то, приняв во внимание, что в случае K_{0q} имеет место неравенство $h_{0q} > \beta$, находим

$$\mathbf{P}(K_s)\mathbf{M}(\xi_0 \mid K_s) > \sum_q \mathbf{P}(K_{0q})\beta = \beta \sum \mathbf{P}(K_{pq}) = \beta \mathbf{P}(K_s).$$

Отсюда, так как по предположению $\mathbf{P}(K_s) \neq 0$, то

$$\mathbf{M}(\xi_0 \mid K_s) > \beta.$$

---

*) Обратим внимание, что только в этом пункте мы использовали предположение о стационарности.

Так как $K_s \to K$, то

$$\mathbf{M}(\xi_0 \mid K) \geqslant \beta.$$

Подобным же способом (если бы рассматривали особые сегменты относительно α) можно доказать, что

$$\mathbf{M}(\xi_0 \mid K) \leqslant \alpha.$$

Мы пришли к противоречию. Отсюда вытекает, что $\mathbf{P}(K) = 0$, что и требовалось доказать.

Исследование того, чему равен предел, к которому стремятся величины h_{0n} при $n \to \infty$, требует предварительных рассуждений. Мы ограничимся здесь доказательством одного предложения на эту тему.

Т е о р е м а. *Если случайные величины ξ_k стационарны, имеют конечную дисперсию и корреляционная функция $R(k) \to 0$ при $k \to \infty$, то*

$$\lim_{n \to \infty} \mathbf{P}\{h_{0n} \to a\} = 1 \quad (a = \mathbf{M}\xi_k).$$

Д о к а з а т е л ь с т в о. Рассмотрим дисперсию величины h_{0n}. В силу стационарности имеем

$$\mathbf{D}h_{0n} = \mathbf{M}\left[\frac{1}{n} \sum_{k=1}^{n} (\xi_k - a)\right]^2 = \frac{\mathbf{D}\xi_n}{n^2} \; [n + 2 \sum_{1 \leqslant i < j \leqslant n} R(j - i)].$$

Очевидно, что

$$\sum_{1 \leqslant i < j \leqslant n} R(j - i) = \sum_{k=1}^{n-1} (n - k) R(k).$$

Рассмотрим столь большое m, что при $k > m$ имеет место неравенство

$$|R(k)| \leqslant \epsilon \quad (\epsilon > 0).$$

Отсюда следует, что

$$\mathbf{D}h_{0n} \leqslant \frac{\mathbf{D}\xi_n}{n^2} \; [n + 2 \sum_{k=1}^{m} (n - k) R(k) + 2\epsilon \sum_{m+1}^{n-1} (n - k)].$$

Это неравенство, очевидно, усиливается следующим образом:

$$\mathbf{D}h_{0n} \leqslant \frac{\mathbf{D}\xi_n}{n^2} \; [n + 2m(n - 1) + \epsilon(n - m - 1)(n - m)].$$

Отсюда ясно, что если n достаточно велико, то правая часть этого неравенства может быть сделана меньше, чем 3ϵ. Таким образом, при $n \to \infty$ величины h_{0n} по вероятности сходятся к a, а так как h_{0n} сходятся при $n \to \infty$ с вероятностью единица, то отсюда следует утверждение теоремы.

Доказанная теорема представляет не только значительный теоретический интерес, но и находит широкие применения в статистической физике и непосредственной технической практике. Причина этого состоит в том, что для определения таких важных характеристик явления, какими являются $\mathsf{M}\xi(t)$, $\mathsf{D}\xi(t), R(u)$ в случае стационарных процессов не нужно знать распределения вероятностей возможных значений и вычислять эти величины по соответствующим формулам. Определение этих, как говорят в физике, пространственных средних требует от исследователя сведений, которых у него зачастую нет. И во всяком случае практическая оценка этих величин посредством эксперимента требует многократного осуществления испытаний для процесса $\xi(t)$. Эргодическая теорема Биркгофа – Хинчина показывает, что с вероятностью единица можно при этом ограничиться единственной реализацией процесса $\xi(t)$.

ГЛАВА 11

ЭЛЕМЕНТЫ СТАТИСТИКИ

§ 60. Основные задачи математической статистики

В теории вероятностей выводятся правила, которые позволяют по вероятностям одних случайных событий вычислять вероятности других, с ними связанных; по числовым характеристикам и функциям распределения одних случайных величин подсчитывать функции распределения и числовые характеристики других. Но естественно возникает вопрос: как найти эти исходные вероятности, функции распределения и числовые характеристики? Как оценить хотя бы приближенные их значения? Это является предметом исследования другой науки о массовых случайных явлениях, которая получила наименование математической статистики. Как наука с оформившейся тематикой и методами исследования математическая статистика возникла, в сущности, только в нашем двадцатом веке. Однако отдельные задачи возникали и рассматривались задолго до нашего времени — и в девятнадцатом, и в восемнадцатом и даже в семнадцатом веках.

Термин статистика происходит от латинского слова "статус" (status) — состояние. Первоначально, в XVIII веке, когда статистика начала оформляться в научную дисциплину, термин статистика связывался с системой описания фактов, характеризующих состояние государства. При этом даже не предполагалось, что ведению статистики подлежат только явления массового порядка. В настоящее время статистика включает в себя и большее и в то же время более определенное содержание. А именно, можно сказать, что статистика состоит из следующих трех разделов:

1) сбор статистических сведений, т.е. сведений, характеризующих отдельные единицы каких-либо массовых совокупностей;

2) статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе данных массового наблюдения;

3) разработка приемов статистического наблюдения и анализа статистических данных. Последний раздел, собственно, и составляет содержание *математической статистики.*

Сбор статистических сведений, касающихся главным образом населения, производился уже давно: имеются сведения, что в 2238 г. до нашей эры в Китае при императоре Яо была произведена перепись населения; производились переписи населения и в древнем Египте, древнем Иране, Римской империи; известны переписи населения в России в 1245, 1259, 1273, 1287 гг. и более поздние. Нужно, правда, отметить, что эти переписи были чрезвычайно примитивны и в Китае, например, в течение 200 лет население учитывалось путем копировки списков предыдущих переписей. Однако даже такие неполные и несовершенные переписи давали возможность намечать важные государственные мероприятия. Практическое значение статистики в наше время велико.

Большим мастером использования статистических методов был В.И. Ленин. Желая иметь дело не с отдельными фактами, не с отдельными событиями, а со всей совокупностью фактов, желая путем анализа массовых явлений вскрыть качественное их своеобразие, В.И. Ленин систематически обращался к статистике. В качестве образца конкретного статистического исследования можно привести работу В.И. Ленина "Развитие капитализма в России". Имеются несколько страниц, к сожалению незаконченной, работы В.И. Ленина "Статистика и социология" ("Ленинский сборник", т. XXX, с. 302), в которой, по-видимому, предполагалось развить общие принципы статистического анализа социальных явлений. В.И. Ленин высоко ценил статистику как орудие познания реального мира. Его слова, что социально-экономическая статистика – одно из самых могущественных орудий социального познания, в неменьшей мере относятся и к математической статистике, широко используемой не только в естествознании, но и в экономике, социологии, инженерном деле. Роль математической статистики не ограничивается вопросами обработки экспериментальных данных, а распространяется и на управление технологическими процессами, а также на большую проблему проверки соответствия теории того или иного явления экспериментальным данным.

Исходным материалом для статистического исследования реального явления служит набор результатов наблюдений над ним или же результатов специально поставленных испытаний. Вопросов, которые при этом возникают, очень много. Укажем теперь некоторые из них.

1. Оценка значения неизвестной вероятности случайного события.

2. Определение неизвестной функции распределения. Задача ставится так: в результате n независимых испытаний над случайной величиной ξ получены следующие ее значения:

$$x_1, x_2, \ldots, x_n.$$

Требуется определить, хотя бы и приближенно, неизвестную функцию распределения $F(x)$ величины ξ.

3. Определение неизвестных параметров распределения. Часто общетеоретические соображения позволяют сделать достаточно определенные заключения о типе функции распределения интересующей нас случайной величины. Так, например, теорема Ляпунова дает воз-

можность считать, что в определенных случаях функция распределения должна быть нормальной. При этом определение неизвестной функции распределения сводится к определению по результатам наблюдений только неизвестных значений параметров a и σ.

Общая задача ставится так: случайная величина ξ имеет функцию распределения определенного вида, зависящую от k параметров, значения которых неизвестны. На основании последовательных наблюдений величины ξ нужно найти значения этих параметров.

Очевидно, что определение неизвестной вероятности p события A является частным случаем только что сформулированной задачи, так как мы можем рассматривать случайную величину ξ, принимающую значение 1, если событие A появляется, и значение 0, если событие A не появляется. Функция распределения ξ зависит от единственного параметра p.

В § 50 мы рассмотрели результаты наблюдений над числом частиц золота, взвешенных в воде. Предполагалось, что это число должно подчиняться закону Пуассона и требовалось оценить параметр распределения λ. Рассматриваемая нами задача как раз включает приведенный пример.

Решение только что поставленной задачи будет нами дано лишь для нормального распределения

$$f(x\,|\,a,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

В этом случае вторая задача, очевидно, может быть разбита на 3 частных вопроса:

1) величина σ предполагается известной, требуется оценить неизвестное значение a;

2) величина a предполагается известной, требуется оценить неизвестное значение σ;

3) оба параметра a и σ неизвестны, требуется оценить их значения.

Более точно эти вопросы могут быть поставлены следующим образом: в результате n независимых испытаний величина ξ приняла следующие значения:

$$x_1, x_2, \ldots, x_n.$$

Требуется указать такие функции $\overline{a}=a(x_1,\ldots,x_n)$ и $\overline{\sigma}=\sigma(x_1, \ldots, x_n)$ (в первой задаче a может быть также функцией σ, а во второй задаче σ может быть функцией a), которые было бы рационально принять за приближенные значения оцениваемых величин a и σ. Помимо этого нужно также оценить среднюю точность этих приближенных формул.

Иногда предпочтительнее искать не приближенные значения неизвестных параметров a и σ в виде функций $\overline{a}$ и $\overline{\sigma}$, а такие функции a', a'' (σ' и σ'') от результатов испытаний и известных величин, чтобы с достаточной практической надежностью можно было утверждать, что

$$a' < a < a''$$

и, соответственно,

$$\sigma' < \sigma < \sigma''.$$

Функции $a', a''(\sigma', \sigma'')$ называются *доверительными границами* для $a(\sigma)$. Впоследствии мы изложим два подхода к решению этих задач.

4. Проверка статистических гипотез. Задача, которую мы здесь рассмотрим, ставится так: на основании некоторых соображений можно считать, что функция распределения случайной величины ξ есть $F(x)$; спрашивается, совместимы ли наблюденные значения с гипотезой, что ξ действительно имеет распределение $F(x)$?

В частности, если вид функции распределения не вызывает сомнений и в проверке нуждаются только значения некоторых параметров, характеризующих это распределение, то в задаче спрашивается: не опровергают ли результаты наблюдений ту гипотезу, что параметры распределения имеют предположенные значения? Это – задача проверки простой гипотезы. Если проверяемая гипотеза состоит в том, что параметры принимают не точно определенные значения, а какие-то из некоторых определенных множеств (например, в случае биномиального распределения, гипотеза $p < p_0$), то гипотеза называется сложной.

В качестве второго примера статистической гипотезы приведем проверку однородности статистического материала. Частым случаем этой задачи является следующий: имеются две последовательности независимых наблюдений над случайной величиной ξ с функцией распределения $F_1(x)$

$$x_1, x_2, \ldots, x_n$$

и над случайной величиной η с функцией распределения $F_2(x)$

$$y_1, y_2, \ldots, y_m.$$

Функции распределения $F_1(x)$ и $F_2(x)$ неизвестны; требуется оценить правдоподобность гипотезы $F_1(x) = F_2(x)$.

5. Оценка зависимости. Производится последовательность наблюдений сразу двух случайных величин ξ и η. Результаты наблюдений даны следующими парами значений: $x_1, y_1, x_2, y_2, \ldots, x_n, y_n$. Выяснить наличие функциональной или корреляционной связи между ξ и η.

6. Управление процесами. Пусть имеется случайный процесс от дискретного или непрерывного времени $\xi(t)$. Процесс под влиянием тех или иных причин может нарушить свое нормальное протекание и стать иным, скажем $\xi_1(t)$. Это нарушение нормального течения может привести к нежелательным последствиям и нам нужно своевременно заметить

момент "разладки" и оказать управляющее воздействие с целью восстановления нормального хода процесса.

В качестве примера мы можем указать на действие технологической линии, которая вырабатывает определенную продукцию. Время от времени в силу различных причин процесс выходит из нормального состояния. Этими причинами могут быть затупление инструмента, нарушение теплового или электромагнитного режима. Они приводят к ухудшению качества продукции. По наблюдениям нужно уловить момент разладки и восстановить ход процесса.

Заметим, что перечисленными задачами далеко не исчерпываются основные проблемы математической статистики. Совершенно новые задачи перед математической статистикой ставит промышленная и научная практика. В частности, само планирование испытаний является одной из основных задач математической статистики.

§ 61. Классический метод определения параметров распределения

Классический метод определения неизвестных параметров функции распределения случайной величины ξ состоит в том, что до наблюдения эти подлежащие оценке величины считаются случайными величинами, подчиненными некоторому "априорному" (доопытному) закону распределения вероятностей. Предполагая этот априорный закон распределения известным, можно вычислить, пользуясь теоремой Байеса, "апостериорный" (послеопытный) закон распределения параметров при условии, что результаты наблюдений над ξ оказались равными $x_1, x_2, \ldots, x_n$.

Как мы уже говорили раньше, все последующее изложение будет относиться к определению неизвестных параметров a и σ нормального закона распределения

$$p(x\,|\,a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}},$$

которому подчинена наблюдаемая случайная величина ξ.

Плотность распределения вероятностей того, что в результате n независимых наблюдений над величиной ξ будут получены значения $x_1, x_2, \ldots$ $\ldots, x_n$ при условии, что неизвестные параметры имеют значения a и σ, равна

$$f(x_1, x_2, \ldots, x_n\,|\,a, \sigma) = \frac{1}{(\sigma\sqrt{2\pi}\,)^n}\, e^{-\frac{s^2}{2\sigma^2}},$$

где

$$s^2 = \sum_{k=1}^{n} (x_k - a)^2.$$

Если ввести обозначения

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k,$$

$$s_1^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2,$$

то простой подсчет показывает, что

$$f(x_1, \ldots, x_n | a, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{n}{2\sigma^2}[s_1^2 + (\bar{x} - a)^2]} \tag{1}$$

Напомним, что в § 60 были поставлены следующие три задачи:

1) σ известно, требуется определить a;

2) a известно, требуется определить σ;

3) a и σ неизвестны, требуется их определить.

Если предположить, что σ известно и $\varphi_1(a)$ означает априорную плотность распределения величины a, то для условной плотности распределения вероятностей величины a при заданном σ и найденных значениях $x_1, x_2, \ldots, x_n$ получим такое выражение:

$$\varphi_1(a | x_1, x_2, \ldots, x_n; \sigma) = \frac{f(x_1, x_2, \ldots, x_n | a, \sigma)\varphi_1(a)}{\int f(x_1, x_2, \ldots, x_n | a, \sigma)\varphi_1(a)\,da}.$$

После подстановки вместо функции f ее значения по формуле (1) и последующих очевидных сокращений находим, что

$$\varphi_1(a | x_1, x_2, \ldots, x_n; \sigma) = \frac{e^{-\frac{n(a-\bar{x})^2}{2\sigma^2}} \cdot \varphi_1(a)}{\int e^{-\frac{n(a-\bar{x})^2}{2\sigma^2}} \cdot \varphi_1(a)\,da}. \tag{2}$$

Во второй и третьей задачах соответствующие формулы имеют вид:

$$\varphi_2(\sigma | x_1, x_2, \ldots, x_n; \sigma) = \frac{f(x_1, x_2, \ldots, x_n | a, \sigma)\varphi_2(\sigma)}{\int f(x_1, x_2, \ldots, x_n | a, \sigma)\varphi_2(\sigma)\,d\sigma}$$

и

$$\varphi_3(a, \sigma | x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n | a, \sigma)\varphi_3(a, \sigma)}{\iint f(x_1, x_2, \ldots, x_n | a, \sigma)\varphi_3(a, \sigma)\,da\,d\sigma},$$

где функции $\varphi_2(\sigma)$ и $\varphi_3(a, \sigma)$ обозначают априорные плотности распределения вероятностей величины σ и пары (a, σ).

После подстановки в эти формулы значения f по формуле (1) и последующих простых сокращений находим, что

$$\varphi_2(\sigma|x_1, x_2, \ldots, x_n; a) = \frac{\sigma^{-n} e^{-\frac{s^2}{2\sigma^2}} \varphi_2(\sigma)}{\int\limits_0^\infty \sigma^{-n} e^{-\frac{s^2}{2\sigma^2}} \varphi_2(\sigma)\, d\sigma} \tag{3}$$

и

$$\varphi_3(a, \sigma|x_1, x_2, \ldots, x_n) = \frac{\sigma^{-n} e^{-\frac{n}{2\sigma^2}(s_1^2 + (a-\bar{x})^2)} \varphi_3(a, \sigma)}{\int\limits_0^\infty \int \sigma^{-n} e^{-\frac{n}{2\sigma^2}(s_1^2 + (a-\bar{x})^2)} \varphi_3(a, \sigma)\, da\, d\sigma} \cdot \tag{4}$$

Полученные формулы непригодны для практического использования не только в силу их сложности, но главным образом потому, что входящие в них априорные вероятности, как правило, нам бывают неизвестны. Часто, не зная априорных плотностей, делают о них более или менее произвольные допущения и на их основе получают обозримые для практического применения формулы. Мы пойдем по иному пути: сделаем совершенно общие допущения о характере априорных распределений и из этих допущений выведем предельные закономерности (при $n \to \infty$) для апостериорных вероятностей.

Теорема 1. *Если априорная плотность распределения $\varphi_1(a)$ имеет ограниченную первую производную и*

$$\varphi_1(\bar{x}) \neq 0,$$

то равномерно относительно α

$$\psi_1(\alpha|x_1, x_2, \ldots, x_n; \sigma) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\alpha^2} \left[1 + O\left(\frac{\sigma}{\sqrt{n}}\right)(1 + |\alpha|)\right], \tag{5}$$

где

$$\alpha = \frac{\sqrt{n}}{\sigma}(a - \bar{x}), \tag{6}$$

а $\psi_1(\alpha| x_1, x_2, \ldots, x_n;\ \sigma)$ обозначает апостериорную плотность распределения величины α.

Доказательство. Действительно, из (6) находим, что

$$a = \bar{x} + \frac{\alpha\sigma}{\sqrt{n}} \tag{7}$$

и, значит *),

$$\frac{\sigma}{\sqrt{n}}\ \varphi_1(a\,|x_1, x_2, \ldots, x_n; \sigma) = \frac{e^{-\alpha^2/2}\varphi_1\left(\bar{x} + \dfrac{\alpha\sigma}{\sqrt{n}}\right)}{\int e^{-\alpha^2/2}\varphi_1\left(\bar{x} + \dfrac{\alpha\sigma}{\sqrt{n}}\right)d\alpha}.$$

По формуле конечных приращений

$$\varphi_1\left(\bar{x} + \frac{\alpha\sigma}{\sqrt{n}}\right) = \varphi_1(\bar{x}) + \frac{\alpha\sigma}{\sqrt{n}}\ \varphi_1'(z),$$

где $z = \bar{x} + \theta\dfrac{\alpha\sigma}{\sqrt{n}}$ и $0 < \theta < 1$.

По условию теоремы

$$|\varphi_1'(z)| \leqslant C < +\infty,$$

поэтому

$$\frac{\sigma}{\sqrt{n}}\ \varphi'(a\,|x_1, x_2, \ldots, x_n; \sigma) = \frac{e^{-\alpha^2/2}\left[\varphi_1(\bar{x}) + \dfrac{\alpha\sigma}{\sqrt{n}}\ \varphi_1'(z)\right]}{\sqrt{2\pi}\,[\varphi_1(\bar{x}) + r_n]},$$

где

$$r_n = \frac{\sigma}{\sqrt{2\pi\,n}} \int \alpha e^{-\alpha^2/2}\varphi_1'(z)\,d\alpha.$$

Легко сообразить, что

$$|r_n| < \frac{2C\sigma}{\sqrt{2\pi\,n}}. \tag{8}$$

---

*) Заметим, что плотность распределения величины α равна

$$\psi_1(\alpha\,|x_1, x_2, \ldots, x_n; \sigma) = \frac{\sigma}{\sqrt{n}}\ \varphi_1(a\,|x_1, x_2, \ldots, x_n; \sigma).$$

Несложные преобразования приводят нас к равенству

$$\psi_1(\alpha|x_1, x_2, \ldots, x_n; \sigma) = \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2}\left[1 + \frac{\sigma}{\sqrt{n}} \frac{\alpha\varphi'(z) - \frac{\sqrt{n}}{\sigma} r_n}{\varphi(\overline{x}) + r_n}\right].$$

Полученное равенство вместе с (8) доказывает теорему.

Т е о р е м а 2. *В условиях предыдущей теоремы*

$$\left.\begin{aligned} &\mathbf{M}(a|x_1, x_2, \ldots, x_n; \sigma) = \overline{x} + O\left(\frac{\sigma^2}{n}\right), \\ &\mathbf{M}[(a - \overline{x})^2 |x_1, x_2, \ldots, x_n; \sigma] = \frac{\sigma^2}{n}\left[1 + O\left(\frac{\sigma}{\sqrt{n}}\right)\right]. \end{aligned}\right\} \qquad (5')$$

Д о к а з а т е л ь с т в о. Действительно, из (7) находим, что

$$\mathbf{M}(a|B) = \overline{x} + \frac{\alpha}{\sqrt{n}} M(\alpha|B)$$

и

$$\mathbf{M}[(a - \overline{x})^2 |B] = \frac{\sigma^2}{n} M(\alpha^2 |B),$$

где B означает некоторое событие. Следовательно,

$$\mathbf{M}(a|x_1, x_2, \ldots, x_n; \sigma) = \overline{x} + \frac{\sigma}{\sqrt{n}} \int \alpha\psi_1(\alpha|x_1, \ldots, x_n; \sigma)\, d\alpha$$

и

$$\mathbf{M}[(a - \overline{x})^2 |x_1, x_2, \ldots, x_n; \sigma] = \frac{\sigma^2}{n} \int \alpha^2 \psi_1(\alpha|x_1, \ldots, x_n; \sigma)\, d\alpha.$$

Подстановка в эти равенства вместо функции ψ_1 ее значения по (5) и последующие несложные подсчеты доказывают теорему.

Доказанная теорема позволяет написать следующее приближенное равенство:

$$a \sim \overline{x},$$

средняя квадратическая ошибка которого приближенно равна σ^2/n.

Теорема 1 позволяет получить вероятность того, что a заключается в определенных границах при условии, что величины $\sigma, x_1, x_2, \ldots, x_n$ приняли определенные значения. Действительно,

$$\mathbf{P}\left\{|a - \overline{x}| < \frac{\sigma z}{\sqrt{n}} \,|x_1, \ldots, x_n; \sigma\right\} = \mathbf{P}\{|\alpha| < z|x_1, \ldots, x_n; \sigma\}$$

и, следовательно, в силу (5)

$$\mathsf{P}\left\{|a-\bar{x}| < \frac{\sigma z}{\sqrt{n}}\ \Big|\, x_1, \ldots, x_n; \sigma\right\} = \frac{2}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt + O\left(\frac{\sigma}{\sqrt{n}}\right).$$

Пренебрегая величиной $O(\sigma/\sqrt{n})$ (что можно сделать, вообще говоря, только тогда, когда или σ мало или n достаточно велико), мы можем считать, что

$$\mathsf{P}\left\{|a-\bar{x}| < \frac{\sigma}{\sqrt{n}}\, z \,|\, x_1, x_2, \ldots, x_n; \sigma\right\} = \frac{2}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt.$$

Теорема 3. *Если априорная плотность распределения $\varphi_2(\sigma)$ имеет ограниченную первую производную и $\varphi_2(\bar{s}) \neq 0$, то равномерно относительно z*

$$\psi_2(z|x_1, x_2, \ldots, x_n; a) = \frac{1}{\sqrt{\pi}}\, e^{-z^2}\left[1 + O\left(\frac{1}{\sqrt{n}}\right)(1+|z|)\right],$$

где ψ_2 – апостериорная плотность распределения величины

$$\beta = \frac{\sigma - \bar{s}}{\bar{s}} \sqrt{n} \tag{9}$$

и $\bar{s} = s/\sqrt{n}$.

Доказательство. Действительно, из (9) находим, что

$$\sigma = \bar{s}\left(1 + \frac{\beta}{\sqrt{n}}\right)$$

и что

$$\psi_2(z|x_1, x_2, \ldots, x_n; a) = \frac{\bar{s}}{\sqrt{n}}\, \varphi_2\!\left(\bar{s}\left(1 + \frac{z}{\sqrt{n}}\right)\Big|\, x_1, x_2, \ldots, x_n, a\right) =$$

$$= \frac{\left(1+\dfrac{z}{\sqrt{n}}\right)^{-n} e^{-\frac{n}{2\left(1+\frac{z}{\sqrt{n}}\right)^2}}\, \varphi_2\!\left(\bar{s} + \dfrac{z\bar{s}}{\sqrt{n}}\right)}{\displaystyle\int_{-\sqrt{n}}^{\infty}\left(1+\frac{\beta}{\sqrt{n}}\right)^{-n} e^{-\frac{s^2}{2\sigma^2}}\, \varphi_2(\sigma)\, d\sigma}.$$

По формуле конечных приращений

$$\varphi_2\left(\bar{s} + \frac{z\bar{s}}{\sqrt{n}}\right) = \varphi_2(\bar{s}) + \frac{z\bar{s}}{\sqrt{n}}\, \varphi_2'(u),$$

где $u = \overline{s} + \theta \dfrac{z\overline{s}}{\sqrt{n}}$, а $0 < \theta < 1$.

Согласно условию теоремы

$$|\varphi_2'(u)| \leqslant C < +\infty,$$

поэтому

$$\psi_2(z|x_1, x_2, \ldots, x_n; a) =$$

$$= \frac{\left(1 + \dfrac{z}{\sqrt{n}}\right)^{-n} e^{-\frac{n}{2\left(1+\frac{z}{\sqrt{n}}\right)^2}} \left[\varphi_2(\overline{s}) + \dfrac{z\overline{s}}{\sqrt{n}} O(1)\right]}{\varphi_2(\overline{s}) \displaystyle\int_{-\sqrt{n}}^{\infty} \left(1 + \dfrac{\beta}{\sqrt{n}}\right)^{-n} e^{-\frac{n}{2\left(1+\frac{\beta}{\sqrt{n}}\right)^2}} d\beta \left[1 + O\left(\dfrac{1}{\sqrt{n}}\right)\right]} =$$

$$= \frac{\left(1 + \dfrac{z}{\sqrt{n}}\right)^{-n} e^{-\frac{n}{2\left(1+\frac{z}{\sqrt{n}}\right)^2}} \left[1 + O\left(\dfrac{1}{\sqrt{n}}\right)(1 + |z|)\right]}{\displaystyle\int_{-\sqrt{n}}^{\infty} \left(1 + \dfrac{\beta}{\sqrt{n}}\right)^{-n} e^{-\frac{n}{2\left(1+\frac{\beta}{\sqrt{n}}\right)^2}} d\beta}. \tag{10}$$

Но

$$-n\ln\left(1 + \frac{z}{\sqrt{n}}\right) - \frac{n}{2\left(1 + \dfrac{z}{\sqrt{n}}\right)^2} = -n\left(\frac{z}{\sqrt{n}} - \frac{1}{2}\frac{z^2}{n} + \ldots\right) -$$

$$- \frac{n}{2}\left(1 - 2\frac{z}{\sqrt{n}} + 3\frac{z^2}{n} - \ldots\right) = -\frac{n}{2} - z^2 + O\left(\frac{1}{\sqrt{n}}\right) \tag{11}$$

и

$$J = \int_{-\sqrt{n}}^{\infty} \left(1 + \frac{\beta}{\sqrt{n}}\right)^{-n} e^{-\frac{n}{2\left(1+\frac{\beta}{\sqrt{n}}\right)^2}} d\beta = n^{-\frac{n}{2}+1} 2^{\frac{n-3}{2}} \int_0^{\infty} z^{\frac{n-3}{2}} e^{-z} dz =$$

$$= n^{-\frac{n}{2}+1} 2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right).$$

По формуле Стирлинга

$$\Gamma\left(\frac{n-1}{2}\right)=\sqrt{2\pi\,\frac{n-3}{2}}\left(\frac{n-3}{2}\right)^{\frac{n-3}{2}} e^{-\frac{n-3}{2}}[1+o(1)],$$

следовательно,

$$J=\sqrt{2\pi}\left(\frac{n-3}{2}\right)^{\frac{n}{2}-1} 2^{\frac{n-3}{2}}\, n^{-\frac{n}{2}+1}\, e^{-\frac{n}{2}+\frac{3}{2}}[1+o(1)]=$$

$$=\sqrt{\pi}\left(1-\frac{3}{n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}+\frac{3}{2}}[1+o(1)]=\sqrt{\pi}\, e^{-\frac{n}{2}}[1+o(1)]. \qquad (12)$$

Равенства (10), (11) и (12) доказывают теорему.

Из теоремы 3, очевидно, следует такой результат:

Теорема 4. *В условиях теоремы* 3 *имеют место следующие соотношения*:

$$\mathbf{M}(\sigma\,|\,x_1, x_2, \ldots, x_n; a)=\bar{s}+o(1/n)$$

и

$$\mathbf{M}[(\sigma-\bar{s})^2\,|\,x_1, x_2, \ldots, x_n; a]=\frac{\bar{s}^2}{2n}[1+o(1/\sqrt{n})].$$

Эта теорема позволяет нам заключить, что при больших n имеет место приближенное равенство

$$\sigma\sim\sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k-a)^2},$$

средняя квадратическая ошибка которого приближенно равна $\bar{s}^2/(2n)$.

Теорема 3 может быть использована также для определения вероятности того, что σ будет находиться в заданных границах. Так, пренебрегая величинами порядка $1/\sqrt{n}$, мы можем утверждать, что при заданных $x_1, x_2, \ldots, x_n$ и a с вероятностью

$$\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$$

σ находится в границах

$$\bar{s}\left(1-\frac{z}{\sqrt{n}}\right)<\sigma<\bar{s}\left(1+\frac{z}{\sqrt{n}}\right).$$

В отношении третьей поставленной нами задачи мы ограничимся только формулировкой результатов, так как их получение ничем не отличается от доказательства теоремы 3. Введем обозначения

$$\alpha_1 = \frac{a - \bar{x}}{\bar{s}_1}\sqrt{n}, \qquad \beta_1 = \frac{\sigma - \bar{s}_1}{\bar{s}_1}\sqrt{n},$$

где $\bar{s}_1 = s_1\sqrt{\dfrac{n}{n-1}}$.

Теорема 5. *Если априорная плотность распределения $\varphi_3(a, \sigma)$ имеет ограниченные первые производные по a и σ и $\varphi_3(\bar{x}, \bar{s}_1) \neq 0$, то равномерно относительно α_1 и β_1*

$$\psi_3(\alpha_1, \beta_1 | x_1, x_2, \ldots, x_n) =$$
$$= \frac{1}{\pi\sqrt{2}} e^{-\frac{\alpha_1^2}{2} - \beta_1^2}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)[1 + |\alpha_1| + |\beta_1|]\right),$$

где ψ_3 означает апостериорную плотность распределения пары (α_1, β_1).

Из теоремы 5 вытекает такой результат.

Теорема 6. *В условиях теоремы* 5

$$\mathsf{M}(a | x_1, x_2, \ldots, x_n) = \bar{x} + O\left(\frac{1}{n}\right),$$

$$\mathsf{M}[(a - \bar{x})^2 | x_1, x_2, \ldots, x_n] = \frac{\bar{s}_1^2}{n}\left[1 + O\left(\frac{1}{\sqrt{n}}\right)\right],$$

$$\mathsf{M}[\sigma | x_1, x_2, \ldots, x_n] = \bar{s}_1\left[1 + O\left(\frac{1}{n}\right)\right],$$

$$\mathsf{M}[(\sigma - \bar{s}_1)^2 | x_1, x_2, \ldots, x_n) = \frac{\bar{s}_1^2}{2n}\left[1 + O\left(\frac{1}{\sqrt{n}}\right)\right].$$

Как и теоремы 1 и 3, теорема 5 может быть использована для определения вероятностей того, что a и σ будут находиться в заданных границах при условии, что наблюденные значения оказались равными $x_1, x_2, \ldots, x_n$.

Практическое значение теорем 1, 3, 5 неодинаково. По теореме 1 точность приближенных формул (5) и (5′) увеличивается не только с увеличением n, но и с уменьшением σ. Поэтому для определения a при известном σ имеется основание пользоваться формулами (5) и (5′) даже при малых n, если только мало σ. В случае же теорем 3 и 5 остаточные члены полученных формул убывают только с возрастанием n, и поэтому при малых значениях n они не дают ничего.

Только что доказанные теоремы являются в некотором смысле обращениями следующих элементарных предложений. Если случайная величина ξ

нормально распределена, параметры a и σ известны, $x_1, x_2, \ldots, x_n$ являются результатами независимых наблюдений ξ, то

1. Плотность распределения величины

$$\alpha = \frac{\sqrt{n}}{\sigma}(\bar{x} - a)$$

равна

$$\psi(x|a, \sigma) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

2. $\mathbf{M}(\bar{x}|a, \sigma) = a$ и $\mathbf{D}(\bar{x}|a, \sigma) = \sigma^2/n$.

3. Плотность распределения величины

$$\beta = \frac{\bar{s} - \sigma}{\sigma}\sqrt{2n}$$

асимптотически равна

$$\psi_2(x|a, \sigma) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

4. $\mathbf{M}(\bar{s}|a, \sigma) = \sigma\{1 + O(1/n)\}$; $\mathbf{D}(\bar{s}|a, \sigma) = \dfrac{\sigma^2}{2n}[1 + O(1/n)]$.

5. Величины α и β независимы, и плотность распределения величины (α, β) асимптотически равна

$$\psi_3(x, y|a, \sigma) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

6. $\mathbf{M}(\bar{x}|a, \sigma) = a$; $\mathbf{D}(\bar{x}|a, \sigma) = \dfrac{\sigma^2}{n}$;

$$\mathbf{M}(\bar{s}|a, \sigma) \approx \sigma; \quad \mathbf{D}(\bar{s}|a, \sigma) \approx \frac{\sigma^2}{2n}.$$

Предложения 1 и 2 не требуют доказательства.

Докажем 3. В § 20 мы нашли, что плотность распределения величины $\bar{s}$ равна

$$\varphi(y) = \frac{\sqrt{2n}}{\sigma\Gamma(n/2)}\left(\frac{y\sqrt{n}}{\sigma\sqrt{2}}\right)^{n-1} e^{-\frac{ny^2}{2\sigma^2}}.$$

Легко проверить, что плотность распределения β есть

$$\psi_2(x|a,\sigma)=\frac{\sigma}{\sqrt{n}}\,\varphi\left(\frac{\sigma x}{\sqrt{n}}+\sigma\right).$$

Несложные преобразования приводят нас к равенству

$$\psi_2(x|a,\sigma)=\frac{1}{\sqrt{2n}}\,e^{-x^2/2}.$$

Для доказательства 4 заметим, что элементарные подсчеты приводят нас к равенствам

$$\mathbf{M}\bar{s}=\sqrt{\frac{2}{n}}\;\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)}\,\sigma\sim e^{-\frac{1}{4n}}\sigma;\quad \mathbf{M}\bar{s}^2=\frac{2}{n}\;\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma(n/2)}\,\sigma^2=\sigma^2.$$

Отсюда

$$\mathbf{M}\bar{s}\approx\sigma\left(1-\frac{1}{4n}\right)$$

и

$$\mathbf{D}\bar{s}=\frac{\sigma^2}{2n}\left(1-\frac{1}{8n}\right).$$

Независимость $\bar{x}$ и $\bar{s}$ будет нами доказана позднее. После того, как это будет сделано, остальные утверждения, содержащиеся в 5 и 6, становятся очевидными.

§ 62. Исчерпывающие статистики

Английским статистиком Фишером было введено весьма важное понятие, которое мы поясним сначала на частном примере. Предположим, что нами решается задача определения параметра a при известном σ по n наблюдениям над нормально распределенной случайной величиной. Если априорная плотность распределения параметра a существует и равна $\varphi_1(a)$, то полученная нами в предыдущем параграфе формула (2) показывает, что условная плотность распределения $\varphi_1(a|x_1,x_2,\ldots,x_n;\ \sigma)$ полностью определяется знанием $\varphi_1(a)$, σ и средним арифметическим результатов наблюдений $x_1,x_2,\ldots,x_n$. Таким образом, каково бы ни было априорное распределение вероятностей параметра a, все то новое (в случае известной дисперсии), что вносят в оценку a наблюдения, заключено в одной единственной величине $\bar{x}$. Говорят поэтому, что $\bar{x}$ является *исчерпывающей статистикой для параметра* a.

Точно так же при известных a и $\varphi_2(\sigma)$ все то новое, что вносят результаты наблюдений в определение параметра σ, заключено в одной величине $\bar{s} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2}$ [см. (3) § 61]. В задаче определения σ при известном a, таким образом, исчерпывающей статистикой будет величина $\bar{s}$.

Общее определение исчерпывающей статистики мы дадим, следуя А.Н. Колмогорову.

Пусть наблюдаемая случайная величина имеет функцию распределения, зависящую от k параметров $\theta_1, \theta_2, \ldots, \theta_k$, значения которых нам неизвестны. Любую функцию $\chi(x_1, x_2, \ldots, x_n)$ от результатов наблюдений и от параметров, значения которых известны, называют статистикой.

Определение исчерпывающей статистики получает следующее естественное обобщение: система функций

$$\chi_i(x_1, x_2, \ldots, x_n) \qquad (i = 1, 2, \ldots, s)$$

называется *исчерпывающей системой статистик для системы параметров* $\theta_1, \theta_2, \ldots, \theta_k$, если условное k-мерное распределение для этих параметров при известных $x_1, x_2, \ldots, x_n$ полностью определяется априорным распределением параметров $\theta_1, \theta_2, \ldots, \theta_k$ и значениями статистик $\chi_1, \chi_2, \ldots, \chi_s$.

Из формулы (4) § 61 мы заключаем, что для параметров a и σ исчерпывающей системой статистик являются функции $\chi_1 = \bar{x}$ и $\chi_2 = s_1$. Понятно, что для каждого параметра a и σ в отдельности система статистик χ_1 и χ_2 также является исчерпывающей.

Без большого труда читатель может самостоятельно убедиться в том, что если случайная величина ξ подчинена закону Пуассона

$$\mathsf{P}\{\xi = k\} = \frac{a^k e^{-a}}{k!} \qquad (k = 0, 1, 2, \ldots)$$

с неизвестным параметром a, то исчерпывающей статистикой для a будет $\bar{x}$ – среднее арифметическое результатов наблюдений.

Точно так же, если двумерная случайная величина (ξ, η) распределена нормально, но параметры a, b, σ_1, σ_2 и r неизвестны, то исчерпывающей системой статистик для указанной системы параметров будут следующие пять функций:

$$\chi_1(x_1, x_2, \ldots, x_n) = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x},$$

$$\chi_2(y_1, y_2, \ldots, y_n) = \frac{1}{n}\sum_{k=1}^{n} y_k = \bar{y},$$

$$\chi_3(x_1, x_2, \ldots, x_n) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2} = s_1,$$

$$\chi_4(y_1, y_2, \ldots, y_n) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \bar{y})^2} = s_2,$$

$$\chi_5(x_1, \ldots, x_n, y_1, \ldots, y_n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y}) = \bar{r}.$$

Здесь (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ – результаты наблюдений.

В качестве упражнения рекомендуем читателю самостоятельно определить исчерпывающие системы статистик для параметра 1) a, 2) b, 3) σ_1, 4) σ_2, 5) r.

§ 63. Доверительные границы и доверительные вероятности

Во вводном параграфе к настоящей главе мы указали, что задача определения неизвестных параметров иногда ставится следующим образом: требуется определить такие две функции $\theta'(x_1, x_2, \ldots, x_n)$ и $\theta''(x_1, x_2, \ldots$ $\ldots, x_n)$ от результатов наблюдений, чтобы была практическая уверенность в том, что неизвестный параметр θ находится в пределах между θ' и θ''. Функции θ' и θ'' называются *доверительными границами для параметра* θ. Для того чтобы доверительные границы для θ были удовлетворительны, нужно, очевидно, потребовать, чтобы условная вероятность

$$\mathsf{P}\{\theta' \leqslant \theta < \theta'' | x_1, x_2, \ldots, x_n\}$$

параметру θ находиться в промежутке от θ' и θ'' при заданных $x_1, x_2, \ldots$ $\ldots, x_n$ была достаточно близка к единице. Степень близости при этом определяется той практической задачей, с которой связано определение неизвестного параметра θ. Если известна априорная плотность распределения для параметра θ, то для определения доверительных границ $\theta'(x_1, x_2, \ldots$ $\ldots, x_n)$ и $\theta''(x_1, x_2, \ldots, x_n)$ естественно выбрать те θ' и θ'', при которых для заданного ω, близкого к единице, выполняется равенство

$$\omega = \mathsf{P}\{\theta' \leqslant \theta < \theta'' | x_1, x_2, \ldots, x_n\} = \frac{\int\limits_{\theta'}^{\theta''} f(x_1, x_2, \ldots, x_n | \theta)\varphi(\theta)\,d\theta}{\int\limits_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n | \theta)\varphi(\theta)\,d\theta},$$

и при этом разность $\theta'' - \theta'$ будет минимальна.

Задача определения доверительных границ в такой постановке сложна не только потому, что она приводит к сложным аналитическим операциям, но в первую очередь потому, что априорная плотность $\varphi(\theta)$ для параметра θ нам обычно бывает неизвестна. Мы видели, что задача получает осмысленное и простое решение не зависящее от априорного распределения для параметров, если число наблюдений n настолько велико, что имеется возможность пользоваться предельными теоремами.

Можно, правда, идти по другому пути, а именно искать правила такого рода: каковы бы ни были результаты наблюдений $x_1, x_2, \ldots, x_n$, требуется указать такие доверительные границы $\theta'(x_1, x_2, \ldots, x_n)$ и $\theta''(x_1, x_2, \ldots \ldots, x_n)$, чтобы с заданной уверенностью (вероятностью) можно было считать, что

$$\theta'(x_1, x_2, \ldots, x_n) \leqslant \theta < \theta''(x_1, x_2, \ldots, x_n).$$

Так как заранее неизвестно, каковы будут результаты наблюдений, то при решении вопроса о том, следует рекомендовать это правило или нет, нужно обращаться не к рассмотрению условных вероятностей

$$\mathsf{P}\{\theta' \leqslant \theta < \theta'' \mid x_1, x_2, \ldots, x_n\},$$

а к рассмотрению безусловной вероятности

$$\mathsf{P}\{\theta' \leqslant \theta < \theta''\} \tag{1}$$

того, что при применении правила не произойдет ошибки.

При заданном виде функций $\theta'(x_1, x_2, \ldots, x_n)$ и $\theta''(x_1, x_2, \ldots, x_n)$ вероятность (1) зависит, конечно, от функции распределения величин $x_1, x_2, \ldots, x_n$. Если это последнее распределение зависит от k параметров $\theta_1, \theta_2, \ldots, \theta_k$ и безусловная плотность распределения этих параметров дается функцией $\varphi(\theta_1, \ldots, \theta_k)$, то

$$\mathsf{P}\{\theta' \leqslant \theta \leqslant \theta''\} =$$
$$= \int \ldots \int \mathsf{P}\{\theta' \leqslant \theta < \theta'' \mid \theta_1, \theta_2, \ldots, \theta_k\} \varphi(\theta_1, \ldots, \theta_k) d\theta_1, \ldots, d\theta_k.$$

Особенно важным на практике является тот случай, когда условная вероятность

$$\mathsf{P}\{\theta' \leqslant \theta \leqslant \theta'' \mid \theta_1, \theta_2, \ldots, \theta_k\} \tag{2}$$

при любых значениях $\theta_1, \theta_2, \ldots, \theta_k$ остается неизменной, равной некоторому числу ω. В этом случае также

$$\mathsf{P}\{\theta' \leqslant \theta \leqslant \theta''\} = \omega,$$

т.е. безусловная вероятность (1) не зависит от априорного безусловного распределения параметров $\theta_1, \theta_2, \ldots, \theta_k$.

Сама гипотеза о существовании априорного распределения параметров не всегда осмысленна. В самом деле, как можно говорить о распределении параметра a в законе Пуассона для любой задачи, в которой характеризующая эту задачу случайная величина распределена по закону Пуассона? Однако, если условная вероятность (2) не зависит от значений параметров и равна одному и тому же значению ω, то естественно считать безусловную вероятность (1) существующей и равной ω даже в тех случаях, когда существование априорного распределения параметров не предполагается.

Условимся говорить, что предлагаемое *правило имеет доверительную вероятность* ω, если при всех возможных значениях параметров условная вероятность (2) равна ω.

Обратимся теперь к рассмотренным в § 60 задачам.

В случае первой из рассмотренных там задач положим

$$a' = \bar{x} + \frac{z_1}{\sqrt{n}}\sigma; \qquad a'' = \bar{x} + \frac{z_2}{\sqrt{n}}\sigma.$$

Каковы бы ни были значения параметров a и σ, мы имеем, очевидно:

$$\mathsf{P}\{a' \leqslant a < a'' \mid a, \sigma\} = \mathsf{P}\left\{a - \frac{z_2\sigma}{\sqrt{n}} \leqslant \bar{x} \leqslant a - \frac{z_1\sigma}{\sqrt{n}} \,\middle|\, a, \sigma\right\} =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-z_2}^{-z_1} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-t^2/2} dt.$$

Отсюда мы заключаем, что доверительная вероятность правила

$$\bar{x} + \frac{z_1\sigma}{\sqrt{n}} \leqslant a < \bar{x} + \frac{z_2\sigma}{\sqrt{n}}$$

равна

$$\omega = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-t^2/2} dt.$$

В частности, имеем, следовательно, равенство

$$\mathsf{P}\left\{|a - \bar{x}| \leqslant \frac{z}{\sqrt{n}}\sigma\right\} = \frac{2}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt.$$

Во второй задаче мы считаем известным параметр a, тогда как параметр σ подлежит оценке. Положим

$$\sigma' = \bar{s}\sqrt{n}/t_1, \qquad \sigma'' = \bar{s}\sqrt{n}/t_2,$$

где

$$s = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - a)^2}.$$

Легко видеть, что

$$\mathsf{P}\{\sigma' \leqslant \sigma < \sigma'' \mid a, \sigma\} = \mathsf{P}\{\sigma t_2/\sqrt{n} < \bar{s} < t_1\sigma/\sqrt{n} \mid a, \sigma\}.$$

В § 21 (распределение χ^2) мы нашли плотность распределения величины $\bar{s}$ при условии, что a и σ заданы. Именно

$$\varphi(y \mid a, \sigma) = \frac{\sqrt{2n}}{\sigma\Gamma(n/2)}\left(\frac{y\sqrt{n}}{\sigma\sqrt{2}}\right)^{n-1} e^{-\frac{y^2 n}{2\sigma^2}}.$$

Отсюда мы находим, что

$$\mathsf{P}\{\sigma' \leqslant \sigma < \sigma'' \mid a, \sigma\} = \frac{\sqrt{2n}}{\sigma\Gamma(n/2)} \int\limits_{\sigma t_2/\sqrt{n}}^{\sigma t_2/\sqrt{n}} \left(\frac{y\sqrt{n}}{\sigma\sqrt{2}}\right)^{n-1} e^{-\frac{y^2 n}{2\sigma^2}} dy =$$

$$= \frac{1}{2^{\frac{n-2}{2}}\Gamma(n/2)} \int\limits_{t_2}^{t_1} z^{n-1} e^{-z^2/2} dz.$$

Мы видим, что условная вероятность неравенств

$$\sigma' \leqslant \sigma < \sigma''$$

при условии, что параметры a и σ известны, не зависит от значений этих параметров. Следовательно, по предыдущему, доверительная вероятность правила

$$\frac{1}{t_1}\sqrt{\sum_{k=1}^{n}(x_k - a)^2} \leqslant \sigma < \frac{1}{t_2}\sqrt{\sum_{k=1}^{n}(x_k - a)^2}$$

равна

$$\omega = \frac{1}{2^{\frac{n-2}{2}}\Gamma(n/2)} \int\limits_{t_2}^{t_1} z^{n-1} e^{-z^2/2} dz.$$

Перейдем, наконец, к рассмотрению последней задачи, когда оба параметра a и σ неизвестны. Положим

$$a_1' = \bar{x} + c_1\sqrt{n}\,s_1 \quad \text{и} \quad a_1'' = \bar{x} + c_2\sqrt{n}\,s_1,$$

$$\sigma_1' = \sqrt{n}\,s_1/t_1, \quad \sigma_1'' = \sqrt{n}\,s_1/t_2,$$

где

$$s_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2}\,.$$

При условии, что a и σ заданы, мы имеем:

$$\mathsf{P}\{a_1' \leqslant a < a_1'' \mid a, \sigma\} = \mathsf{P}\left\{c_1 \leqslant \frac{a - \bar{x}}{s_1\sqrt{n}} < c_2 \mid a, \sigma\right\}$$

и

$$\mathsf{P}\{\sigma_1' \leqslant \sigma < \sigma_2' \mid a, \sigma\} = \mathsf{P}\left\{t_2 \leqslant \frac{s_1\sqrt{n}}{\sigma} \leqslant t_1 \mid a, \sigma\right\}.$$

Нам нужно найти теперь условные плотности распределения для $\dfrac{a - \bar{x}}{s_1\sqrt{n}}$ и $\dfrac{s_1\sqrt{n}}{\sigma}$ при условии, что a и σ известны. Так как

$$\frac{\bar{x} - a}{s_1\sqrt{n}} = \frac{\dfrac{1}{n}\sum\limits_{k=1}^{n}(x_k - a)}{\sqrt{\sum\limits_{k=1}^{n}[(x_k - a) - (\bar{x} - a)]^2}} = \frac{\bar{x}'}{\sqrt{\sum\limits_{k=1}^{n}(x_k' - \bar{x}')^2}}\,,$$

где $x_k' = x_k - a$ и $\bar{x}' = \dfrac{1}{n}\sum\limits_{k=1}^{n} x_k'$ (величины x_k' независимы и распределены нормально с математическим ожиданием 0 и дисперсией σ^2).

Введем теперь в n-мерном пространстве $(x_1', x_2', \dots, x_n')$ новую ортогональную систему координат $(y_1, y_2, \dots, y_n)$ так, чтобы $y_1 = \sqrt{n}\,\bar{x}_1'$.

При этом

$$n s_1^2 = \sum_{k=1}^{n} (x'_k - \bar{x}')^2 = \sum_{k=1}^{n} x_k'^2 - n\bar{x}'^2 = \sum_{k=1}^{n} y_k^2 - y_1^2 = \sum_{k=2}^{n} y_k^2$$

и, следовательно,

$$\frac{\bar{x} - a}{s_1 \sqrt{n}} = \frac{y_1}{\sqrt{n \sum\limits_{k=2}^{n} y_k^2}}.$$

Так как

$$y_k = \sum_{i=1}^{n} \alpha_{ki} x'_i,$$

где величины α_{ki} удовлетворяют соотношениям

$$\sum_{j=1}^{n} \alpha_{ij}\alpha_{jk} = \begin{cases} 1 & \text{при } i = k, \\ 0 & \text{при } i \neq k \end{cases}$$

и величины x'_i нормально распределены, то величины y_k также нормально распределены. Далее $\mathbf{M} y_k = 0$ $(k = 1, 2, \ldots, n)$. Наконец, из того, что

$$\mathbf{M} y_i y_k = \sum_{j=1}^{n} \alpha_{ij}\alpha_{jk} \mathbf{M} x_j'^2 = \sigma^2 \sum_{j=1}^{n} \alpha_{ij}\alpha_{jk} = \begin{cases} \sigma^2 & \text{при } i = k, \\ 0 & \text{при } i \neq k, \end{cases}$$

мы заключаем о независимости величин y_k $(k = 1, 2, \ldots, n)$ и о том, что $\mathbf{D} y_k = \sigma^2$ $(k = 1, 2, \ldots, n)$.

Так как далее

$$\frac{y_1}{\sqrt{\sum\limits_{k=2}^{n} y_k^2}} = \frac{y_1 / \sqrt{n-1}}{\sqrt{\left(\sum\limits_{k=2}^{n} y_k^2\right) / (n-1)}}$$

и в этой дроби числитель и знаменатель независимы, причем плотность распределения числителя равна

$$\frac{\sqrt{n-1}}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}(n-1)},$$

а плотность распределения знаменателя (согласно § 21 распределение χ^2) равна

$$\frac{\sqrt{2(n-1)}}{\sigma\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{y\sqrt{n-1}}{\sigma\sqrt{2}}\right)^{n-2} e^{-\frac{y^2}{2\sigma^2}(n-1)},$$

то согласно § 21 (распределение Стьюдента) плотность распределения частного $\dfrac{y_1}{\sqrt{\sum\limits_{k=2}^{n} y_k^2}}$ равна

$$\frac{\Gamma(n/2)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}(1+x^2)^{-n/2}.$$

Плотность распределения величины $\dfrac{y_1}{\sqrt{n\sum\limits_{k=2}^{n} y_k^2}}$, как в этом легко убедиться, равна

$$\varphi(x \mid a, \sigma) = \frac{\sqrt{n}\,\Gamma(n/2)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}(1+nx^2)^{-n/2},$$

следовательно,

$$\mathsf{P}\left\{c_1 \leqslant \frac{a-\bar{x}}{s_1\sqrt{n}} < c_2\right\} = \frac{\sqrt{n}\,\Gamma(n/2)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\int\limits_{c_1}^{c_2}(1+nx^2)^{-n/2}dx.$$

Эта вероятность не зависит от значений, которые принимают параметры a и σ. Мы можем поэтому сказать, что в третьей задаче доверительная вероятность правила

$$c_1 s_1 \sqrt{n} \leqslant a - \bar{x} < c_2 s_1 \sqrt{n}$$

равна

$$\omega = \frac{\sqrt{n}\,\Gamma(n/2)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{c_1}^{c_2} (1 + nx^2)^{-n/2} dx.$$

Нам остается найти доверительную вероятность правила, устанавливающего границы для σ. Использовав произведенные нами преобразования, находим, что

$$\mathsf{P}\{\sigma_1' \leqslant \sigma < \sigma_2' \mid a, \sigma\} = \mathsf{P}\left\{t_2 < \frac{s_1\sqrt{n}}{\sigma} \leqslant t_1 \mid a, \sigma\right\} =$$

$$= \mathsf{P}\left\{t_2\sigma < \sqrt{\sum_{k=2}^{n} y_k^2} \leqslant t_1\sigma \mid a, \sigma\right\}.$$

Отсюда, в силу результатов § 21 (распределение χ^2)

$$\mathsf{P}\{\sigma_1' \leqslant \sigma < \sigma_2' \mid a, \sigma\} = \frac{\sqrt{2(n-1)}}{\sigma\Gamma\left(\frac{n-1}{2}\right)} \int_{\frac{t_2\sigma}{\sqrt{n-1}}}^{\frac{t_1\sigma}{\sqrt{n-1}}} \left(\frac{y\sqrt{n-1}}{\sigma\sqrt{2}}\right)^{n-2} e^{-\frac{y^2(n-1)}{2\sigma^2}} dy =$$

$$= \frac{1}{2^{\frac{n-3}{2}}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{t_2}^{t_1} z^{n-2} e^{-\frac{z^2}{2}} dz.$$

Снова эта вероятность не зависит от значений параметров a и σ. Следовательно, правило

$$\frac{\sqrt{n}\,s_1}{t_1} \leqslant \sigma < \frac{\sqrt{n}\,s_1}{t_2}$$

имеет доверительную вероятность, равную

$$\omega = \frac{1}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)} \int_{t_2}^{t_1} z^{n-2} e^{-\frac{t^2}{2}} dt.$$

Заметим в заключение, что из того, что при любых $\theta_1, \theta_2, \ldots, \theta_k$ имеет место равенство

$$\mathsf{P}\{\theta' \leqslant \theta < \theta'' \mid \theta_1, \theta_2, \ldots, \theta_k\} = \omega$$

и из него вытекает равенство

$$\mathsf{P}\{\theta' \leqslant \theta < \theta''\} = \omega,$$

еще не следует, что

$$\omega = \mathsf{P}\{\theta'(x_1, x_2, \ldots, x_n) \leqslant$$
$$\leqslant \theta < \theta''(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n\}.$$

§ 64. Проверка статистических гипотез

Предположим, что нам известна функциональная форма распределения случайной величины ξ, но неизвестны значения параметров $\theta_1, \theta_2, \ldots, \theta_k$, от которых оно зависит. Имеются основания считать, что параметры имеют некоторые определенные значения $\theta_1 = \theta_1^0, \theta_2 = \theta_2^0, \ldots, \theta_k = \theta_k^0$ (простая гипотеза) или же принадлежат некоторому множеству (сложная гипотеза). Требуется выяснить, подтверждают или не подтверждают эту гипотезу результаты наблюдений над величиной ξ.

Для того чтобы подчеркнуть практическую важность задачи, рассмотрим примеры.

Пример 1. Имеется большая партия продукции некоторого производства. Каждая единица этого продукта относится к одной из двух категорий: годная, бракованная. Вся партия считается пригодной к сдаче, если относительное число p бракованных единиц продукта невелико, скажем, не больше, чем некоторое число p_0 $(0 < p_0 < 1)$. Число p нам неизвестно; его нужно найти путем исследования сравнительно небольшого (по отношению ко всему объему партии) числа изделий. Рассмотрим случайную величину ξ, равную 0, если взятое наудачу изделие окажется пригодным, и равную 1, если взятое наудачу изделие окажется бракованным. Функция распределения ξ равна

$$F(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ 1-p & \text{при} \quad 0 < x \leqslant 1, \\ 1 & \text{при} \quad x > 1. \end{cases}$$

Параметр p, от которого зависит распределение, неизвестен. Наша задача состоит в том, чтобы проверить гипотезу $p \leqslant p_0$.

П р и м е р 2. Случайная величина ξ распределена нормально

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

параметры a и σ неизвестны. Требуется проверить гипотезу, что

$$|a - a_0| \leqslant \alpha \quad \text{и} \quad \sigma < \sigma_0,$$

где a_0, σ_0 и α – некоторые известные числа. Эта и аналогичные задачи постоянно возникают в теории измерений, а также в естественно научных и производственных задачах.

Обозначим через n число наблюдений, на основании которых необходимо сделать заключение о подтверждении или опровержении сделанной гипотезы. Пусть

$$x_1, x_2, \ldots, x_n \tag{1}$$

результаты этих наблюдений. Процесс проверки, ведущий к подтверждению или опровержению гипотезы, есть некоторое правило, согласно которому множество всех возможных результатов n наблюдений разбивается на две непересекающиеся части R_{n1} и R_{n2}. При этом принадлежность чисел (1) к множеству R_{n1} будем считать подтверждением проверяемой гипотезы, а принадлежность их к множеству R_{n2} – отрицанием проверяемой гипотезы. Если мы станем изображать числа (1) как координаты n-мерного евклидова пространства R_n, то, очевидно, каждый процесс проверки означает разбиение пространства R_n на части R_{n1} и R_{n2}. При этом, если точка $(x_1, x_2, \ldots, x_n)$ оказывается в части R_{n1}, то гипотеза принимается, а если $(x_1, x_2, \ldots, x_n)$ оказывается в части R_{n2}, то гипотеза отбрасывается. Множество R_{n2} носит название к р и т и ч е с к о й о б л а с т и. Выбор правила проверки, таким образом, эквивалентен выбору критической области.

Для иллюстрации вернемся к примеру 1. Множество R_n в этом случае состоит из всевозможных совокупностей n чисел, каждое из которых может принимать лишь значения 0 и 1. Критическая область R_{n2} состоит из тех элементов R_n, для которых

$$\frac{1}{n}(x_1 + x_2 + \ldots + x_n) > p_0 .$$

Мы перейдем теперь к следующей частной задаче проверки гипотез, для которой имеется исчерпывающее решение: имеются две простые гипотезы H_1 и H_2. Гипотеза H_1 состоит в том, что $\theta_i = \theta_i'$ $(i = 1, 2, \ldots, k)$, гипотеза H_2 – в том, что $\theta_i = \theta_i''$ $(i = 1, 2, \ldots, k)$. Эти гипотезы конкурируют друг с другом и на основании произведенных наблюдений требуется одной из них отдать предпочтение.

Заметим, что при подтверждении или отрицании гипотезы H_1 мы можем совершить ошибки двух видов. Ошибку первого рода мы совершаем тогда, когда отвергаем H_1 в то время, когда в действительности она верна. Иными словами, ошибки первого рода имеют место тогда, когда точка $(x_1, x_2, \ldots, x_n)$ попадает в область R_{n2} в то время, когда верна гипотеза H_1. Ошибку второго рода мы совершаем, если принимаем H_1 в то время, когда она неверна. Если критическая область выбрана, то вероятности ошибок первого и второго рода можно рассчитать; обозначим их для данных n и R_{n2} соответственно буквами α_1 и α_2.

Понятно, что чем меньше для данной критической области числа α_1 и α_2, тем удачнее выбрана критическая область. Однако при данном числе испытаний n невозможно ни при каком выборе критической области одновременно сделать как угодно малыми оба числа α_1 и α_2. В то же время изменением критической области мы можем добиться произвольной малости ошибок первого или второго рода в отдельности. Так, если положить $R_{n2} = R_n$, то ясно, что в этом случае $\alpha_2 = 0$; если же положить $R_{n1} = R_n$, то $\alpha_1 = 0$. Отсюда вытекает следующий рациональный принцип выбора критической области: при заданных значениях α_1 и n нужно выбирать ту область R_{n2}, для которой α_2 достигает минимума. При этом, конечно, чем меньшее значение α_1 мы выбираем, тем большее значение получается для минимума α_2. Заранее нельзя сказать, какое α_1 нужно выбрать, чтобы метод проверки гипотезы был самым выгодным, так как основную роль при этом играет практическая сторона дела.

Пусть для примера отбрасывание или прием гипотезы H_1 связаны с материальными затратами. Если прием гипотезы H_1, в то время как она неверна, приводит к большим затратам (скажем, к необходимости ручной подгонки некоторых деталей, поступающих для сборки на некоторое предприятие), тогда как отбрасывание гипотезы H_1, в то время как она верна, приводит к сравнительно небольшим потерям, то ясно, что необходимо выбрать возможно меньшее α_2 и при этом можно помириться со сравнительно большими значениями α_1.

Предположим, что практические соображения приняты в расчет и величина α_1 выбрана; тогда имеет место следующее предложение, которое мы сформулируем лишь для того случая, когда величина ξ имеет конечную плотность распределения вероятностей, как при гипотезе H_1, так и при гипотезе H_2.

Т е о р е м а. *Среди всевозможных критических областей, для которых вероятность ошибок первого рода равна α_1, вероятность ошибок второго рода принимает наименьшее значение для критической области R^*_{n2}, состоящей из всех тех точек $(x_1, x_2, \ldots, x_n)$, для которых*

$$\prod_{k=1}^{n} f(x_k \mid H_2) \geqslant c \prod_{k=1}^{n} f(x_k \mid H_1)^{*)} . \tag{2}$$

Число c определяется из условия

$$\psi(c) = \mathsf{P}\{(x_1, x_2, \ldots, x_n) \subset R^*_{n2} \mid H_1\} = \alpha_1 . \tag{3}$$

Д о к а з а т е л ь с т в о. Так как (в случае независимых испытаний) вероятность точке $(x_1, x_2, \ldots, x_n)$ находиться в какой-нибудь области S равна

$$\mathsf{P}\{S \mid H_1\} = \int_S \ldots \int \prod_{k=1}^{n} f(x_k \mid H_1) dx_1 dx_2 \ldots dx_n$$

при условии, что верна гипотеза H_1 и

$$\mathsf{P}\{S \mid H_2\} = \int_S \ldots \int \prod_{k=1}^{n} f(x_k \mid H_2) dx_1 dx_2 \ldots dx_n$$

при условии, что верна гипотеза H_2, то согласно предположению,

$$\mathsf{P}\{R^*_{n2} \mid H_1\} = \alpha_1$$

и для любой другой рассматриваемой нами области R_{n2} также

$$\mathsf{P}\{R_{n2} \mid H_1\} = \alpha_1 .$$

Согласно аксиоме сложения вероятностей

$$\mathsf{P}\{R_{n2} - R_{n2}R^*_{n2} \mid H_1\} = \mathsf{P}\{R_{n2} \mid H_1\} - \mathsf{P}\{R_{n2}R^*_{n2} \mid H_1\} =$$
$$= \alpha_1 - P\{R_{n2}R^*_{n2} \mid H_1\}$$

---

*) Таким образом, R^*_{n2} является наивыгоднейшей критической областью.

и

$$\mathsf{P}\{R_{n2}^* - R_{n2}R_{n2}^* \mid H_1\} = \alpha_1 - \mathsf{P}\{R_{n2}R_{n2}^* \mid H_1\},$$

т. е.

$$\mathsf{P}\{R_{n2} - R_{n2}^*R_{n2} \mid H_1\} = \mathsf{P}\{R_{n2}^* - R_{n2}R_{n2}^* \mid H_1\}.$$

Согласно определению R_{n2}^* и последнему равенству

$$\mathsf{P}\{R_{n2}^* - R_{n2}R_{n2}^* \mid H_2\} \geqslant c\mathsf{P}\{R_{n2}^* - R_{n2}R_{n2}^* \mid H_1\} =$$
$$= c\mathsf{P}\{R_{n2} - R_{n2}R_{n2}^* \mid H_1\}. \qquad (4)$$

Но для любой точки $(x_1, x_2, \ldots, x_n)$, не принадлежащей R_{n2}^*,

$$\prod_{k=1}^{n} f(x_k \mid H_2) < c \prod_{k=1}^{n} f(x_k \mid H_1)$$

и, следовательно, поскольку область $R_{n2} - R_{n2}R_{n2}^*$ находится целиком вне R_{n2}^*, должно быть

$$c\mathsf{P}\{R_{n2} - R_{n2}R_{n2}^* \mid H_1\} > \mathsf{P}\{R_{n2} - R_{n2}R_{n2}^* \mid H_2\}.$$

Это неравенство вместе с (4) приводит нас к неравенству

$$\mathsf{P}\{R_{n2}^* - R_{n2}R_{n2}^* \mid H_2\} > \mathsf{P}\{R_{n2} - R_{n2}R_{n2}^* \mid H_2\}.$$

Прибавив к обеим частям последнего неравенства $\mathsf{P}\{R_{n2}R_{n2}^* \mid H_2\}$, находим, что

$$\mathsf{P}\{R_{n2}^* \mid H_2\} > \mathsf{P}\{R_{n2} \mid H_2\}.$$

А так как

$$\mathsf{P}\{R_n \mid H_2\} = 1$$

и $R_{n1}^* = R_n - R_{n2}^*$, $R_{n1} = R_n - R_{n2}$, то

$$\mathsf{P}\{R_{n1}^* \mid H_2\} < \mathsf{P}\{R_{n1} \mid H_2\}.$$

Так как $\mathsf{P}\{R_{n1} \mid H_2\}$ и $\mathsf{P}\{R_{n1}^* \mid H_2\}$ представляют собой ошибки второго рода для критических областей R_{n2} и соответственно R_{n2}^*, то теорема доказана.

Нам остается подтвердить, что выбор постоянной c действительно можно произвести по правилу (3). С этой целью заметим, что функция

$$\psi(c) = \mathsf{P}\{R_{n2}^* \mid H_1\}$$

с ростом c может только убывать (так как неравенству (2) будет удовлетворять все более и более "тощее" множество точек $(x_1, x_2, ..., x_n)$). Кроме того, ясно, что $\psi(0) = 1$ (так как для каждой точки $(x_1, x_2, ..., x_n)$

$$\prod_{k=1}^{n} f(x_k \mid H_2) \geqslant 0.$$

Далее из (2) следует, что

$$\mathsf{P}\{R_{n2}^* \mid H_2\} \geqslant c\, \mathsf{P}\{R_{n2}^* \mid H_1\}.$$

Заменив левую часть неравенства единицей и вспомнив определение $\psi(c)$, находим неравенство

$$1 \geqslant c\,\psi(c).$$

Итак,

$$0 \leqslant \psi(c) \leqslant 1/c.$$

Таким образом, $\psi(c) \to 0$ при $c \to \infty$. Так как функция $\psi(c)$ не возрастает, то при любом α_1 $(0 < \alpha_1 < 1)$ найдется такое c, что

$$\psi(c-0) \geqslant \alpha_1 \geqslant \psi(c+0).$$

Если в точке c функция $\psi(c)$ непрерывна, то выбор постоянной c согласно правилу (3) оправдан; если же в точке c функция $\psi(c)$ имеет разрыв, то положение несколько усложняется и требуется незначительно изменить определение множества R_{n2}^*, исключив из него часть точек $(x_1, x_2, ..., x_n)$, для которых

$$\prod_{k=1}^{n} f(x_k \mid H_2) = c \prod_{k=1}^{n} f(x_k \mid H_1),$$

и присоединив их к множеству R_{n1}^*, так чтобы вероятность ошибок первого рода была равна α_1.

Рассмотрим пример. Пусть известно, что ξ распределено нормально с известной дисперсией σ^2. Относительно математического ожидания a имеются две гипотезы, состоящие в том, что $a = a_1$ (гипотеза H_1) и $a = a_2$ (гипотеза H_2). Требуется найти выгоднейшую критическую область.

В нашем примере соотношение (2) может быть записано в следующем виде:

$$e^{-\frac{1}{2\sigma^2}\sum\limits_{k=1}^{n}[(x_k-a_2)^2-(x_k-a_1)^2]} \geqslant c.$$

Это неравенство, как легко подсчитать, эквивалентно следующему (в предположении, что $a_2 > a_1$):

$$\sum_{k=1}^{n} x_k \geqslant \frac{\sigma^2 \ln c}{a_2 - a_1} + \frac{n}{2}(a_1 + a_2)$$

или, что то же самое, неравенству

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (x_k - a_1) \geqslant \frac{\sigma \ln c}{(a_2 - a_1)\sqrt{n}} + \frac{\sqrt{n}}{2\sigma}(a_2 - a_1) = k_1.$$

Полученное неравенство определяет выгоднейшую критическую область R^*_{n2}.

Так как величина

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (x_k - a_1)$$

распределена нормально, с математическим ожиданием 0 и дисперсией 1, если только гипотеза H_1 имеет место, то по таблицам нормального распределения и заданному α_1 легко определить k_1 (и тем самым c). Пусть для определенности $\alpha_1 = 0{,}05$. Тогда $k_1 = 1{,}645$ и, следовательно, наивыгоднейшая критическая область при $\alpha_1 = 0{,}05$ определяется неравенством

$$\sum_{k=1}^{n} (x_k - a_1) \geqslant 1{,}645\, \sigma\sqrt{n}.$$

Интересно отметить, что критическая область в нашем примере не зависит от конкурирующего значения a_2.

Область R^*_{n1} определяется неравенством

$$\sum_{k=1}^{n} x_k < \frac{\sigma^2 \ln c}{a_2 - a_1} + \frac{n}{2}(a_2 + a_1),$$

которое, очевидно, может быть записано в таком виде:

$$\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(x_k - a_2) < k_1 - \frac{\sqrt{n}}{\sigma}(a_2 - a_1).$$

Величина, стоящая в левой части неравенства, в предположении, что имеет место гипотеза H_2, распределена нормально с математическим ожиданием 0 и дисперсией 1. Отсюда следует, что вероятность ошибки второго рода равна

$$\Phi\left(k_1 - \frac{\sqrt{n}}{\sigma}(a_2 - a_1)\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{k_1 - \frac{\sqrt{n}}{\sigma}(a_2 - a_1)} e^{-\frac{z^2}{2}}dz.$$

Если заданы величины α_1 и α_2, то возникает задача определения минимального числа $n = n(\alpha_1, \alpha_2)$ необходимых испытаний для того, чтобы ошибочные заключения могли быть сделаны с вероятностями, не большими, чем α_1 и α_2.

С ростом n величина $\alpha_{2n} = \alpha_2(\alpha_1, n)$ не возрастает и, вообще говоря, стремится к нулю. Очевидно, что $n(\alpha_1, \alpha_2)$ есть наименьшее из тех n, для которых $\alpha_2(\alpha_1, n) \leqslant \alpha_2$.

В только что рассмотренном примере число n по заданным значениям α_1 и α_2 найти очень просто. В самом деле, из того, что

$$1 - \Phi(k_1) = \alpha_1$$

и

$$\Phi\left(k_1 - \frac{\sqrt{n}}{\sigma}(a_2 - a_1)\right) = \alpha_2$$

мы получаем два уравнения:

$$k_1 = \psi(1 - \alpha_1)$$

и

$$k_1 - \frac{\sqrt{n}}{\sigma}(a_2 - a_1) = \psi(\alpha_2),$$

где ψ – функция, обратная для $\Phi(x)$. Отсюда

$$n = \frac{\sigma^2}{(a_2 - a_1)^2}[\psi(1 - \alpha_1) - \psi(\alpha_2)]^2.$$

Приведем небольшой числовой пример. Пусть

$$a_1 = 135, \quad a_2 = 150, \quad \sigma = 25, \quad \alpha_1 = 0{,}01, \quad \alpha_2 = 0{,}03.$$

Поскольку

$$\psi(0{,}99) = 2{,}33, \quad \psi(0{,}03) = -1{,}88,$$

то

$$n = \frac{25^2}{15^2}[2{,}33 + 1{,}88]^2 = \frac{25}{9} \cdot 4{,}21^2 \sim 49.$$

Таким образом, минимальное число наблюдений, которое необходимо провести для выбора между гипотезами H_1 и H_2 в случае нормального распределения и только что указанных данных, должно быть 49. Только при таком числе испытаний мы можем быть уверены в том, что если верна гипотеза H_1, то мы отбросить ее можем с вероятностью, не большей одной сотой, а если верна гипотеза H_2, то ее отбросить мы можем с вероятностью, не превосходящей 0,03.

ДОПОЛНЕНИЕ

ОЧЕРК ИСТОРИИ ТЕОРИИ ВЕРОЯТНОСТЕЙ

ГЛАВА 1

ПРЕДЫСТОРИЯ ПОНЯТИЯ ВЕРОЯТНОСТИ И СЛУЧАЙНОГО СОБЫТИЯ

§ 1. Первые данные

Сейчас уже трудно установить, кто впервые поставил вопрос, пусть и в несовершенной форме, о возможности количественного измерения возможности появления случайного события. Ясно одно, что мало-мальски удовлетворительный ответ на этот вопрос потребовал длительного времени и значительных усилий ряда поколений выдающихся исследователей. В течение долгого периода исследователи ограничивались рассмотрением разного рода игр, особенно игр в кости, поскольку их изучение позволяет ограничиваться простыми и прозрачными математическими моделями. Однако следует заметить, что многие отлично понимали то, что позднее было прекрасно сформулировано Христианом Гюйгенсом: " ... я полагаю, что при внимательном изучении предмета читатель заметит, что имеет дело не только с игрой, но что здесь закладываются основы очень интересной и глубокой теории".

Мы увидим, что при дальнейшем прогрессе теории вероятностей глубокие соображения как естественнонаучного, так и общефилософского характера играли большую роль. Эта тенденция продолжается и в наши дни: мы постоянно наблюдаем, как вопросы практики – научной, производственной, оборонной – выдвигают перед теорией вероятностей новые проблемы и приводят к необходимости расширения арсенала ее идей, понятий и методов исследования.

На первом этапе изучения случайных явлений внимание ученых было сосредоточено на трех задачах: 1) подсчет числа различных возможных исходов при бросании нескольких костей; 2) раздел ставки между игроками, когда игра прекращена где-то посередине; 3) определение числа бросаний двух или нескольких костей, при котором число случаев, благоприятствующих выпадению на всех костях одинаковых граней (например, "шестерок") хотя бы при одном бросании было бóльшим, чем число случаев, когда это событие не появится ни разу.

Число различных исходов при бросании трех игральных костей было определено в 960 г. епископом Виболдом из города Камбрэ. Он считал, что таких исходов 56 (он не принимал во внимание то обстоятельство, что данное число очков может появиться на любой из трех костей). Бросанию трех костей Виболд придал религиозную трактовку – с появлением каждого набора трех чисел он связал одну из 56 добродетелей. Описание правильных подсчетов было дано в XI веке летописцем Балдерикусом, а появилось оно в печати лишь в 1615 г.

Попытка подсчитать число исходов при бросании трех игральных костей, включая и перестановки, имеется в поэме Ричарда де Форниваль (1200–1250) "De vetula", написанной в промежутке от 1220 до 1250 г. В части поэмы, посвященной играм и спорту, имеются следующие рассуждения: "Одинаковое число очков на трех костях можно получить шестью способами. Если число очков на двух костях совпадает, а на третьей от него отлично, то мы имеем 30 способов, поскольку одна пара может быть выбрана 6 способами, а третье число лишь пятью. Если очки на всех костях различны, то мы имеем 20 способов, потому что 30 раз по 4 равно 120, но каждая возможность появляется 6 способами. Таким образом существует всего 56 возможностей. Одинаковые числа очков на всех костях можно получить только единственным способом; одинаковые числа очков на двух костях, а третье отличное от них – тремя способами".

Хотя в тексте явно указано лишь число случаев по Виболду (56), но фактически Ричард де Форниваль полностью подготовил подсчет общего числа равновероятных случаев при бросании трех костей, а именно $6 \cdot 1 + 30 \cdot 3 + 20 \cdot 6 = 216$.

Далее Форниваль привел таблицу, в которой вычислены числа способов, которыми может быть получена данная сумма очков на всех трех костях. Мы приведем эту таблицу в укороченном виде. В первых двух столбцах приведены суммы очков на трех

Таблица 18

Сумма		Число способов	Сумма		Число способов	Сумма		Число способов
3	18	1	6	15	10	9	12	25
4	17	3	7	14	15	10	11	27
5	16	6	8	13	21			

костях, а в третьем столбце – число различных случаев, при которых реализуется эта сумма. Все подсчеты выполнены без ошибок, да и рассуждения, проведенные автором, вполне логичны и даже можно сказать современны в нашем смысле слова. Это обстоятельство заслуживает быть отмеченным, поскольку эти же самые подсчеты через двести с лишним лет были выполнены неправильно. А именно в 1477 г. Бенвенуто Д'Имола издал в Венеции "Божественную комедию" Данте, снабдив ее комментариями. В комментарии к VI части "Чистилища", в которой говорится об игроке в кости, Д'Имола произвел подсчеты шансов. Согласно его рассуждениям сумма очков при бросании трех костей, равная 3, 4, 17 и 18, может получаться одним единственным способом. Ошибка Б. Д'Имола очевидна и ее нет нужды комментировать.

Заслуживает специального упоминания одна из первых математических книг начала эпохи итальянского Возрождения, написанная Лукой Пачоли (ок. 1445 – ок. 1514) и носившая наименование "Сумма знаний по арифметике, геометрии, отношениям и пропорциональности". Написана эта книга была в 1487 г., но издана лишь через семь лет в Венеции. Поскольку задачи Луки Пачоли сыграли определенную роль в формировании интереса к теории вероятностей, мы приведем их формулировку. В разделе "необычных задач" в упомянутой книге были помещены две следующие:

1. Компания играет в мяч до 60 очков и делает ставку в 22 дуката. В связи с некоторыми обстоятельствами игра прекращена до ее окончания причем одна сторона в этот момент имеет 50, другая – 30 очков. Спрашивается, какую часть общей ставки должна получить каждая сторона?

2. Трое соревнуются в стрельбе из арбалета. Кто первым достигнет 6 лучших попаданий, тот выигрывает. Ставка 10 дукатов. Когда первый получил 4 лучших попадания, второй 3, а третий 2, они не хотят продолжать и решают разделить приз справедливо. Спрашивается, какой должна быть доля каждого?

Пачоли предложил решение, которое позднее многократно оспаривалось, поскольку оно было признано ошибочным. А именно, он предложил делить ставку пропорционально числу выигранных партий. Таким образом в первой задаче решение Пачоли таково: первый должен получить 5/8 ставки, т.е. 13,75 дуката, а второй 3/8 ставки, т.е. 8,25 дуката. Во второй же задаче, согласно Пачоли, первый должен получить 4 и 4/9 дуката, второй 3 и 3/9 дуката и третий 2 и 2/9 дуката.

§ 2. Исследования Дж. Кардано и Н. Тарталья

Несомненно, что существенное продвижение в решении первичных задач теории вероятностей связано с именами итальянских ученых Дж.Кардано (1501–1575) и Н.Тарталья (ок.1499–1557). В рукописи "Книга об игре в кости", датированной самим Кардано 1526 г., но изданной лишь в 1563 г., были решены многие задачи, связанные с бросанием игральных костей и выпадением на них того или иного числа очков. Он правильно подсчитал числа различных случаев, которые могут произойти при бросании двух и трех костей. Словесные формулировки при этом достаточно сложны. Вот, для примера, что он писал в главе XI "О бросании двух костей": "При бросании двух костей возможны 6 случаев по два одинаковых числа и 15 случаев выпадения разного числа очков, т.е., считая и двойные, 30. Следовательно, всего возможно 36 случаев". Под двойными выпадениями он понимает выпадение на двух костях очков, получаемых перестановкой. Например, двойным к случаю выпадения на первой кости 2 очков, а на второй 5 будет выпадение 5 очков на первой кости и 2 на второй.

Кардано указал далее число возможных случаев появления хотя бы на одной кости определенного числа очков. Таких случаев оказалось 11. Заслуживают упоминания слова Кардано: "число это меньше, чем число случаев отсутствия данного числа очков. По отношению к общему числу случаев при бросании двух костей оно составляет больше одной шестой и меньше одной четверти". Здесь у Кардано ошибка: нужно было сказать меньше одной трети, поскольку 11/36 не меньше, а больше 1/4.

Это место заслуживает пристального внимания, поскольку Кардано дважды предложил рассматривать отношение, которое теперь мы называем классическим определением вероятности. А именно, 1/6 – это вероятность появления заданного числа очков при бросании одной кости, а 11/36 – вероятность получить хотя бы на одной кости грань с заданным числом очков. Означает ли это, что Кардано решил рассматривать вместо чисел благоприятствующих шансов вероятности случайных событий, т.е. ввел в рассмотрение классическое определение вероятности? Судя по всему, это было озарение только для данного примера. Повидимому, Кардано хотел выяснить вопрос: что чаще происходит – при бросании одной кости выпадение заданного числа очков или же при бросании двух костей выпадение этого числа очков хотя бы на одной кости? Ответ был найден и на этом Кардано успокоился. Единичное наблюдение он не сделал основой для общего заключения. В результате он не заметил, что стоял на пороге введения важного понятия для всего дальнейшего развития большой главы математики, да и всего количественного естествознания.

Этот вывод подкрепляется тем, что в следующей главе, в которой рассматривается бросание трех костей, Кардано уже не обращается к отношению числа благоприятствующих шансов к числу всех возможных. Все усилия Кардано затратил на подсчет числа возможных случаев.

В тринадцатой главе "О сложных числах, как до шести, так и свыше и как для двух, так и для трех костей", Кардано вновь возвратился к рассмотрению отношений числа благоприятствующих случаев к числу всех возможных случаев. Однако и здесь Кардано не заметил, что он находился на грани введения важного для науки понятия. Вот его подлинные слова: "Десять очков (в сумме – *Б.Г.*) может получиться из двух пятерок и из шестерки и четверки. Последнее сочетание возможно при этом в двух видах. Таким же образом девять очков может получиться из пятерки и четверки и из шестерки и тройки, так что это составляет 1/9 всей серии *) и две девятых ее половины. Восемь же очков получается из двух четверок, из 3 и 5 и из 6 и 2. Всего же 5 возможных случаев составляет приблизительно 1/7 часть из всей серии ... 7 очков составляется из 6 и 1, из 5 и 2, из 4 и 3. Всего, стало быть, имеется 6 возможных случаев, составляющих 1/6 всей серии. А 6 получается по такому же расчету, как и 8; 5 – как 9; 4 – как 10; 3 – как 11 и 2 – как 12".

Вновь Кардано оперирует фактически классическим понятием вероятности, но не замечает его значения для изучаемых им задач. Рассматриваемые им отношения воспринимаются им скорее чисто арифметически, как доля случаев, чем как характеристика возможности появления случайного события при испытании.

В главе XI имеется одно предложение, которое рядом авторов трактуется весьма широко, хотя, как мы сейчас увидим, его формулировка достаточно неопределенная. Вот эти слова Кардано: "Целая серия игр не дает отклонения, хотя в одной игре это может случиться ... при большом числе игр оказывается, что действительность весьма приближается к этому предположению". Ссылаясь на это место, В.В. Бобынин **) сделал далеко идущий вывод: "этот закон (больших чисел – *Б.Г.*) уже с достаточной ясностью был выражен в XVI столетии Карданом в его статье "De ludo aleae". Позднее О. Оре***) в книге, посвященной Кардано, писал, что этот последний формулировал и использовал закон больших чисел в рудиментарной форме. Вполне возможно, что мнение Оре имеет некоторые основания, но следует заметить, что формулировка Кардано весьма неопределенна.

В той же книге Кардано приблизился к определению безобидной игры, что видно из следующего предложения, заимствованного из этой книги: "Итак, имеется одно общее правило для расчета: необходимо учесть общее число возможных выпадений и число способов, которыми могут появиться данные выпадения, а затем найти отношение последнего числа к числу оставшихся возможных выпадений, приблизительно в такой же пропорции определяются относительные размеры ставок для того, чтобы игра шла на равных условиях". Однако мнение ряда авторов относительно того, что в этом месте Кардано приблизился к классическому определению понятия вероятности, мне представляется ошибочным.

Задача Пачоли о разделе ставки до окончания игры интересовала также и Кардано. В книге "Практика общей арифметики", изданной в 1539 г., Кардано привел ряд критических замечаний в связи с решением Пачоли. Он указал на то, что Пачоли, предлагая делить ставку пропорционально числу уже выигранных партий, никак не учитывает, как много партий еще нужно выиграть каждому из игроков. Согласно мнению Кардано, если s – число партий, которое следует выиграть, а p и q – числа фактически выигранных партий первым и вторым игроками, то ставка должна делиться между игроками в отношении

$$[1 + 2 + \ldots + (s - q)] : [1 + 2 + \ldots + (s - p)].$$

Как мы увидим позднее, решение, предложенное Карданом, в общем случае ошибочно и лишь в некоторых весьма частных случаях оно приводит к правильному результату.

---

*) Под серией Кардано принимал все возможные исходы, т.е. 36 при бросании двух костей и 216 при бросании трех костей.

**) Бобынин В.В. Яков I Бернулли и теория вероятностей // Мат. образование. – 1914. № 4.

***) Оre О. Cardano. The gambling scholer. – Princeton, 1953.

К задаче о разделе ставки вновь вернулся Н. Тарталья в книге "Общий трактат о мере и числе", которая была опубликована в 1556 году. Его подход изложен в § 20 книги, озаглавленном "Ошибка брата Луки из Борго". Критические замечания Тарталья верны и имеют под собой серьезный здравый смысл. Вот его слова: "Это его правило мне не кажется ни красивым, ни хорошим, потому что если бы одна из этих сторон имела 10, а другая вообще не имела никакого очка, то действуя по такому правилу, получилось бы, что одна сторона, имеющая указанные 10 очков, должна была бы взять все, а другая не получила бы ничего, что было бы совершенно лишено смысла."

Для первой задачи Пачоли (с измененным условием) Тарталья предложил следующее решение: первый игрок, набравший 10 очков, должен получить, во-первых, половину всей ставки и, во-вторых, $(10-0)/60$ всей ставки, или $22/6$ дуката, т.е. всего 14 и $2/3$ дуката, а второй 7 и $1/3$ дуката.

Мы увидим, что решение, предложенное Тарталья, также ошибочно. Но следует согласиться с тем, что трудно было бы требовать от него самого и его предшественников правильного решения, поскольку в науке для этого еще не было выработано необходимых понятий – понятия вероятности и математического ожидания. Следующее замечание Тарталья убедительно показывает, что он и сам не доверял своему решению. Вот эти слова: "Разрешение такого вопроса является скорее делом юриспруденции, чем разума, так что при любом способе решения этой задачи найдутся поводы для споров, но тем не менее наименее спорным, как мне кажется, будет следующее ... ". Далее он предложил делить ставку по такому правилу: отклонение выигрыша от половины ставки должно быть пропорционально разности выигранных партий. В только что приведенном примере, в котором игра шла до шестидесяти очков и ставка равнялась 22 дукатам, первый игрок выиграл 10 партий, а второй – 0, доли игроков согласно предложению Тарталья таковы:

$$14 \text{ и } 2/3 = 11 + (10-0)/60 \cdot 22 \quad \text{и} \quad 11 + (0-10)/60 \cdot 22 = 7 \text{ и } 1/3.$$

В 1558 г. была опубликована книга Г.Ф. Певероне "Два коротких и легких трактата по началам арифметики и основам геометрии". В этой книге без указания предшественников была рассмотрена задача о разделе ставки. Формулировка задачи такова: два лица A и B играют в мяч до выигрыша одним из них 10 партий. В тот момент, когда игрок A выиграл 7 партий, а игрок B – 9, они решили прекратить игру. Как следует разделить ставку между игроками?

Певероне предложил разделить ставку в отношении 1 к 6, т.е. игроку A отдать $1/7$ ставки, а игроку B – $6/7$ ставки. Это решение неправильно. Легко подсчитать, что A должен получить $1/8$, а игрок B – $7/8$ ставки. В то же время он дал правильное решение в двух случаях, когда игроки A и B выиграли по 9 партий, а также в случае, когда игрок A выиграл 8 партий, а игрок B – 9 партий.

§ 3. Исследования Галилео Галилея

Мы видим, что уже в XVI веке возникли задачи чисто вероятностного характера и упорно разыскивались подходы к их решению. Это неизбежно приводило к необходимости развития, с одной стороны, комбинаторных методов, а с другой стороны – к поиску тех понятий, в терминах которых было бы можно описывать возникающие ситуации. Ошибки, допущенные одними исследователями, подмечались другими. Эти другие предлагали свои способы решения, которые в свою очередь подвергались критическому анализу. Постепенно вырабатывались подходы, которые позднее становились основой новой теории и, во всяком случае, позволяли решать отдельные задачи.

Заслуживает внимания вклад в этот прогресс известного естествоиспытателя, ученого широких интересов и взглядов – Галилео Галилея (1564–1642). Его работа "О выходе очков при игре в кости", увидевшая свет только в 1718 г., была посвящена подсчету числа возможных случаев при бросании трех костей. Число всех возможных случаев Галилей подсчитал самым простым и естественным путем – он возвел 6 (число различных возможностей при бросании одной кости) в третью степень и получил $6^3 = 216$, что неоднократно непосредственным подсчетом получалось и ранее.

Далее Галилей подсчитал число различных способов, которыми может быть получено то или другое значение суммы выпавших на костях очков. Ясно, что эта сумма может принимать любое значение от 3 до 18. При подсчете Галилей пользовался полезной идеей – кости нумеровались (первая, вторая, третья) и возможные исходы записывались в виде троек чисел, причем на соответствующем месте стояло число очков, выпавшее на кости с данным номером. Эта простая мысль для своего времени была весьма полезной. Приведем теперь подлинные слова Галилея: "... хотя 9 и 12 получаются в результате стольких же комбинаций, как 10 и 11, и вследствие этого должны были бы признаваться равноценными, мы видим, тем не менее, что в результате продолжительных наблюдений игроки все же считают более выигрышными 10 и 11, чем 9 и 12. Совершенно очевидно, что 9 и 10 (мы говорим о них, имея в виду также 12 и 11) получаются из того же числа комбинаций: 9 из 1, 2,6 – 1,3,5 – 1,4,4 – 2,2,5 – 2,3,4 – 3,3,3, т.е. из шести троек, а 10 из 1,3,6 – 1,4,5 – 2,2,6 – 2,3,5 – 2,4,4 – 3,3,4 и ни при каких других сочетаниях, кроме этих шести" (G.Galilei, Opera, t. XIV, p. 293, Fiorentina, 1855).

Возникает естественный вопрос, почему же все-таки сумма 10 оказывается более предпочтительной, чем 9? Ответ заключается в следующем: "1. Тройки, или другими словами, числа, получающиеся при выпадении трех костей с тремя одинаковыми очками, не могут получиться иначе, как одним способом; 2. Тройки, образующиеся из двух одинаковых и третьего отличного от них, могут получаться тремя способами; 3. Те же, которые получаются из трех различных очков, могут получаться шестью способами. Из этих положений мы легко выводим, какими способами или, лучше сказать, при каких выходах трех костей могут получаться все числа" (там же. с. 295).

В завершающей части работы Галилей привел следующую таблицу.

В верхней строке указаны значения суммы чисел выпавших очков. Первые три цифры в каждой клетке указывают как может получиться сумма в соответствующем

Таблица 19

10		9		8		7		6		5		4		3	
631	6	621	6	611	3	511	3	411	3	311	3	211	3	111	1
622	3	531	6	521	6	421	6	321	6	221	3				
541	6	522	3	431	6	331	3	222	1						
532	6	441	3	422	3	322	3								
442	3	432	6	332	3										
433	3	333	1												
27		25		21		15		10		6		3		1	

столбце, четвертая цифра – число возможных различных случаев. Например, против тройки 631 указано 6 случаев; вот они: 631 – 136 – 316 – 613 – 163 – 361. Комбинация 361, для примера, означает, что на первой кости выпали 3 очка, на второй – 6 и на третьей – 1.

В таблице приведены результаты лишь для половины всех возможных сумм. Вторая половина вычисляется в точности таким же образом. В результате оказывается, что сумме 11 благоприятствует 27 различных возможностей, 12 – 25, 13 – 21, 14 – 15, 15 – 10, 16 – 6, 17 – 3 и 18 – 1. С учетом этого сумма всех возможных вариантов выпадения трех костей равна 2 (1 + 3 + 6 + 10 + 15 + 21 + 25 + 27) = 216.

Заметим, что Галилей, в сущности, повторил результаты, полученные значительно раньше рядом предшественников – епископом Виболдом, Ричардом де Форнивалем и рядом других. Однако эта, теперь такая простая для студента второго курса университета задача, в ту пору была серьезным испытанием и для мыслителя столь высокого ранга как Галилей. Вот что он сам писал по этому поводу: "Чтобы выполнить данное мне поручение, стоившее мне таких трудов, изложу мои соображения в надежде не только разрешить указанное недоразумение, но и указать путь к точнейшему изложению основания, которые позволят осветить все особенности игры" (там же, с. 293).

Заметим, что и у Галилея, как и у его предшественников, рассуждения ведутся не над вероятностями случайных событий, а над числами шансов, которые им благоприятствуют.

Для теории вероятностей и математической статистики бо́льшее значение, чем только что рассмотренная работа, имеют его соображения по поводу теории ошибок наблюдений. До него никто этим не занимался. Таким образом, все, что он написал на эту тему, ново для его времени и важно даже в наши дни. Свои мысли и выводы он достаточно подробно изложил в одном из основных своих произведений "Диалог о двух главнейших системах мира птоломеевой и коперниковой" (М. – Л., 1948).

Согласно Галилею, ошибки наблюдений являются неизбежными спутниками каждого измерения, каждого экспериментального исследования. "В каждой комбинации наблюдений будет какая-нибудь ошибка; я думаю, что это неизбежно . . . " ("Диалог . . .", с. 214). При этом ошибки могут быть двух типов: систематические, связанные прочно со способом измерений и с используемыми инструментами, и случайные, которые меняются непредсказуемым образом от одного измерения к другому. Эта классификация сохранилась до нашего времени и широко используется во всех руководствах по теории ошибок измерений.

Случайные ошибки измерений обладают некоторыми характерными особенностями. Их Галилей старательно выделил и проанализировал. Во-первых, малые ошибки встречаются чаще, чем большие, поэтому, как правило, в результаты измерений следует вносить лишь небольшие поправки. Далее, положительные ошибки встречаются так же часто, как и отрицательные. "Можно одинаково легко ошибаться как тем, так и другим образом" (там же, с. 125). Далее Галилей отметил, что около истинного результата должно группироваться наибольшее число измерений. "Среди возможных мест истинное местонахождение, надо думать, будет то, вокруг которого группируется наибольшее число расстояний" (там же, с. 216).

Эти исследования Галилея имеют принципиальное значение, поскольку они положили начало новой научной дисциплине – теории ошибок наблюдений. Эта теория, несомненно, сыграла важную роль в формировании теории вероятностей, но еще большее значение она имела для развития математической статистики. Это тем более так, что теория случайных ошибок наблюдений в настоящее время рассматривается в качестве естественной задачи математической статистики.

§ 4. Вклад Б. Паскаля и П. Ферма в развитие теории вероятностей

Обычно считают, что теория вероятностей зародилась в переписке двух великих ученых – Б. Паскаля (1623 – 1662) и П. Ферма (1601 – 1665). От этой переписки сохранилось лишь три письма Паскаля (от 29 июля, 24 августа и 27 октября 1654 г.) и четыре письма П.Ферма (одно письмо без даты и письма от 9 августа, 29 августа, 25 сентября 1654 г.) Самое первое письмо Б. Паскаля утрачено и о его содержании можно судить лишь по ответу Ферма.

В 1950 – 1951 г., в связи с приближавшимся тогда 150-летним юбилеем М.В. Остроградского (1801 – 1862), мне было поручено изучить архивы этого ученого, хранящиеся в Государственной публичной библиотеке УССР. Среди рукописей нашелся фрагмент (лист 904), явно относившийся к вводной лекции по теории вероятностей. Из литературных источников известно, что в 1858 г. Остроградский прочитал в Михайловском артиллерийском училище двадцать лекций по теории вероятностей с целью развития кругозора слушателей и их научной инициативы. Более того, три из них даже были изданы. Однако ни одной из них мне не удалось найти. Тем интереснее было познакомиться с обнаружённым фрагментом, который я считаю полезным привести здесь полностью.

"Теорию вероятностей должно отнести к наукам нового времени, ибо настоящее ее начало не восходит дальше половины XVII столетия. Правда, некоторые предметы, относящиеся к этой науке, были известны во времена весьма отдаленные и постоянно делались расчеты, основанные на продолжительности средней жизни, известны были морские страхования, знали число случайностей в азартных играх, но только в самых простых, найдены были величины ставок или закладов, безобидных для игроков, но подобные выводы не были подчинены никаким правилам. Однако же теорию вероятностей считают наукой нового времени и ее начало относят к первой половине XVII столетия, ибо прежде этой эпохи вопросы о вероятностях не были подчинены математическому анализу и не имелось никаких точных общих правил для решения их.

Паскаль, а за ним Ферма, геометры XVII столетия, по справедливости считаются основателями науки о вероятностях. Первый вопрос, относящийся к этой науке, и довольно сложный, решен Паскалем. Вопрос, о котором говорим, был предложен Паскалю кавалером де Мере и состоял в следующем условии. Два игрока начали игру, состоящую из данного числа партий, положим 30-ти, розыгрыш каждой партии непременно выигрывается одним из игроков, и тот из них, кто выиграл бы прежде другого тридцать партий, считался окончательно выигравшим и вязл бы обе ставки, внесенные в начале игры. Но игроки согласились прекратить игру, не окончив ее, т.е. одному не хватало до выигрыша тридцати партий некоторого числа, например, трех партий, а другому, положим, пятнадцати партий. Внесенные ставки для безобидности, конечно, должны быть разделены между игроками так, чтобы тот, кому недостает до выигрыша бо́льшего числа партий, получил бы меньшую сумму, а противник его большую, именно безобидный раздел требует, чтобы каждый игрок получил часть внесенной суммы, пропорциональную вероятности своего выигрыша. Итак, нужно найти эту вероятность. Паскаль нашел ее, а потом вопрос де Мере предложил Ферма. Последний немедленно нашел решение и даже для случая более сложного, когда игра происходит не между двумя только, а между произвольным числом игроков.

Замечательно, что имя кавалера де Мере, человека светского и не имевшего никакого преуспевания на поприще математических наук, остается навсегда в истории этих наук".

Мы видим теперь, что оценка, данная роли Паскаля и Ферма Остроградским, несколько завышена. Впрочем, такой же точки зрения придерживаются многочисленные историки науки. Однако в переписке Паскаля с Ферма еще отсутствует понятие вероятности, и оба они ограничиваются рассмотрением числа благоприятствующих событию шансов. Конечно, у этих авторов впервые в истории имеется правильное решение задачи о разделе ставки, которая, как мы знаем, отняла много усилий у исследователей в течение длительного времени. Оба они исходили из одной и той же идеи: раздела ставки в отношении, пропорциональном, как мы теперь сказали бы, вероятностям окончательного выигрыша каждого игрока. В предложенных ими решениях можно увидеть зачатки использования математического ожидания и в весьма несовершенной форме теорем о сложении и умножении вероятностей. Точнее сказать не вероятностей, а шансов, благоприятствующих тому или иному событию. Это был серьезный шаг в создании предпосылок и интересов к задачам теоретико-вероятностного характера. Второй шаг был сделан также Паскалем, когда он существенно продвинул развитие комбинаторики и указал на ее значение для зарождающейся теории вероятностей.

Толчком к появлению интересов Паскаля к задачам, приведшим к теории вероятностей, послужили встречи и беседы с одним из придворных французского королевского двора – шевалье де Мере (1607 – 1648). Де Мере интересовался философией, литературой и одновременно был страстным игроком. В этой страсти были истоки тех задач, которые он предложил Паскалю. Вот эти вопросы:

1. Сколько раз надо подбросить две кости, чтобы число случаев, благоприятствующих выпадению хотя бы раз сразу двух шестерок, было больше, чем число случаев, когда ни при одном бросании не появляются две шестерки одновременно?

2. Как нужно разделить ставку между игроками, когда они прекратили игру, не набрав необходимого для выигрыша числа очков?

Де Мере претендовал, что первую задачу он решил. Однако при ближайшем рассмотрении в его рассуждениях легко обнаружить ошибку. А именно, в одном из писем де Мере Паскалю содержится такая фраза: "Если в одном случае есть один шанс из N_0 в единственной попытке и в другом случае один шанс из N_1, то отношение соответствующих чисел есть $N_0 : N_1$. Таким образом, $n_0 : N_0 = n_1 : N_1$".

Обозначения и смысл этой фразы требуют пояснения. В приведенном письме речь идет о следующем: при бросании одной кости имеется $N_0 = 6$ различных исходов и выпадению шестерки благоприятствует один из них. При бросании двух костей сразу выпадению шестерки на двух костях благоприятствует лишь один исход из $N_1 = 36$ возможных. При бросании одной кости n_0 (= 4) раз число благоприятствующих исходов для выпадения шестерки первосходит число благоприятствующих случаев ее невыпадения. Символом n_1 обозначим число бросаний двух костей, при котором число благоприятствующих случаев выпадения одновременно двух шестерок превзойдет число благоприятствующих случаев для их невыпадения ни разу. Из правила де Мере вытекает, что уже при 24 бросаниях двух костей наступает интересующее нас событие.

В действительности правило де Мере ошибочно, поскольку вероятность того, что при четырех бросаниях одной кости ни разу не появится шестерка, равна $(5/6)^4 = 625/1296$ и, значит, искомая вероятность равна $1 - 625/1296 = 671/1296$. В этом пункте де Мере оказался прав, но при 24 бросаниях двух костей вероятность ни разу не выбросить сразу две шестерки равна $(35/36)^{24} = 0{,}509$, а искомая вероятность хотя бы раз выбросить две шестерки сразу есть $1 - (35/36)^{24} = 0{,}491$. Легко понять, что двадцати четырех бросаний еще недостаточно, а нужно по меньшей мере двадцать пять бросаний двух костей, чтобы вероятность выпадения сразу двух шестерок превосходила 0,5.

При изложении мы воспользовались современным языком и употребляли понятие вероятности. Подход де Мере был обычным для того времени и ограничивался лишь подсчетом числа благоприятствующих тому или иному событию шансов.

Основное содержание писем Паскаля и Ферма посвящено разделу ставки. Решение, предложенное Паскалем, в подробностях изложено в письме от 29 июля:

"Вот примерно, что я делаю для определения стоимости каждой партии, когда два игрока играют, например, на три партии и каждым вложено по 32 пистоля.

Предположим, что один выиграл две партии, а другой одну. Они играют еще одну партию, и если выигрывает первый, то он получает всю сумму в 64 пистоля, вложенную в игру; если же эту партию выигрывает второй, то каждый игрок будет иметь по 2 выигранных партии, и, следовательно, если они намерены произвести раздел, каждый должен получить обратно свой вклад в 32 пистоля.

Примите же во внимание, монсеньер, что если первый выиграет, то ему причитается 64; если он проиграет, то ему причитается 32. Если же игроки не намерены рисковать на эту партию и хотят произвести раздел, то первый должен сказать: "Я имею 32 пистоля верных, ибо в случае проигрыша я их также получил бы, но остальные 32 пистоля могут быть получены либо мной, либо Вами, случайности равны. Разделим же эти 32 пистоля пополам, и дайте мне, кроме того, бесспорную сумму в 32 пистоля".

Далее Паскаль рассмотрел другой случай, когда первый игрок выиграл две партии, а второй ни одной и третий, когда, первый игрок выиграл одну партию, а второй ни одной. В обоих случаях рассуждения при решении подобны тем, которые уже были проведены. Ответы же, предложенные Паскалем, таковы: в первом случае один игрок должен получить 56, а второй – 8 пистолей: во втором же 44 и 20.

Решение, которое для задачи Паскаля предложил Ферма, дошло до нас только по изложению, которое содержится в письме Паскаля от 24 августа. Письмо же Ферма с оригинальным текстом не сохранилось. Пусть до выигрыша игроку A недостает двух партий, а игроку B – трех партий. Тогда для завершения игры достаточно сыграть еще максимум четыре партии. Их возможные исходы представлены в виде следующей таблицы:

Таблица 20

Партии	Возможные исходы партий			
1	*A A A A*	*A B A A*	*A B B A*	*B B B A*
2	*A A A B*	*B A A B*	*B A B A*	*B B A B*
3	*A A B A*	*B A A A*	*B B A A*	*B A B B*
4	*A A B B*	*A B A B*		*A B B B*
				B B B B
игра выиграна игроком	*A* после двух партий	*A* после четырех партий	*A* после трех партий	*B* после трех или четырех партий

В этой таблице символом A обозначен выигрыш соответствующей партии игроком A, символом B – игроком B. Номера партий идут по строкам. В первых одиннадцати исходах выигрывает игрок A, в последних пяти – игрок B. Таким образом ставка между игроками A и B должна быть разделена в отношении 11 к 5. Иными словами игрок A получит 11/16, а игрок B – 5/16 ставки. Совершенно очевидно, что Ферма, также как и Паскаль, делит ставку пропорционально вероятностям выигрыша каждым из игроков всей игры. Но этого понятия в их руках еще нет и они вынуждены искать иные способы выражения своих идей. В результате они сами не замечают, что их исходные позиции одинаковы. Это отчетливо видно из письма Паскаля от 27 октября, в котором он писал: "Сударь, я очень доволен Вашим последним письмом, я любуюсь методом в отношении партий, тем более, что я его хорошо понимаю, он полностью Ваш, ничего общего не имеет с моим и легко приводит к той же самой цели".

В письме от 24 августа Паскаль высказал сомнение в том, что метод Ферма можно распространить на число игроков, бóльшее двух. Однако Ферма показал, что теми же рассуждениями можно решить задачу о разделении ставки и для случая трех игроков. Это решение им было использовано в задаче о трех игроках, когда до окончания игры игроку A недостает одной выигранной партии, а игрокам B и C – по две. Это решение вновь сопровождается таблицей, смысл которой пояснять уже нет необходимости:

Таблица 21

A	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*B*	*B*	*B*	*C*	*C*	*C*	*B*	*C*	*A*	*B*	*B*	*B*	*C*	*C*	*C*	*C*	*C*	*B*
A	*A*	*A*	*B*	*B*	*B*	*C*	*C*	*C*	*A*	*A*	*A*	*A*	*A*	*A*	*C*	*B*	*B*	*B*	*B*	*C*	*B*	*C*	*C*	*C*	*B*	*C*
A	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*A*	*B*	*A*	*C*	*B*	*B*	*C*	*A*	*B*	*C*	*C*
A	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*B*	*B*	*B*	*B*	*B*	*C*	*C*	*C*	*C*	*C*

В своем письме Паскаль отметил, что Роберваль (1604 – 1675) спросил его зачем рассматривать продолжение игры до четырех партий в тех случаях, когда уже ясно какой из игроков выигрывает игру? Паскаль явно понимал, что это необходимо для сохранения равновозможности всех перечисляемых случаев. Так в первых четырех исходах первой таблицы игрок A выигрывает всю игру уже после двух партий. Точно также в первых девяти исходах второй таблицы игрок A выигрывает игру после первой партии. Тем не менее Ферма доводит таблицу до конца и рассматривает все возможные случаи исхода четырех партий. Этим самым Паскаль и Ферма избежали ошибки, которую допустил в следующем столетии Даламбер, когда подсчитывал число равновероятных случаев при бросании двух монет.

При рассмотрении второй таблицы Паскаль допустил неточность в рассуждениях. А именно он считал, что из 27 возможных исходов бесспорно благоприятствуют игроку A лишь 13, а исходы 5, 11, 19 столбцов, также как 9, 15 и 24 благоприятствуют сразу и игроку A и угроку B (как A, так и C), поэтому их следует брать с половинным весом. В результате Паскаль предлагал делить ставку в отношении 16 :5, 5:5, 5. Ошибка Паскаля нам теперь очевидна.

Паскаль одновременно с размышлениями над проблемами, составившими содержание его переписки с Ферма, разрабатывал вопросы комбинаторики. Результатом этого явился "Трактат об арифметическом треугольнике", опубликованный в 1665 г. и внесший серьезный вклад в развитие комбинаторики. В этом трактате имеется параграф, в котором изложены правила использования комбинаторных результатов в задаче о разделе ставки. Правило. предложенное Паскалем, состоит в следующем: пусть игроку до выигрыша всей игры не хватает m партий, а игроку В – n партий, тогда ставка должна делиться между игроками в таком отношении:

$$\sum_{i=0}^{n-1} C_{m+n-1}^{i} : \sum_{i=0}^{m-1} C_{m+n-1}^{i}.$$

§ 5. Работа Х. Гюйгенса

Несомненно, что на развитие теории вероятностей значительное влияние оказала работа Х. Гюйгенса (1629 – 1695). Интерес Гюйгенса к этим вопросам был вызван его поездкой в Париж в 1655 г., где он познакомился с рядом видных ученых и услышал от них сведения относительно задач о разделе ставки в азартных играх, которые разрабатывались Паскалем и Ферма. Повидимому, ему стали известны и идеи, которыми они руководствовались при решении. Задачи Гюйгенса заинтересовали и он самостоятельно занялся размышлениями над подобными же вопросами. Поскольку, как он позднее писал в трактате "О расчетах в азартных играх", ни Паскаль, ни Ферма не опубликовали разработанных ими методов, ему пришлось самому искать пути решения. Результатом явилась работа Гюйгенса, опубликованная в 1656 г. в виде дополнения к книге его учителя Ф. ван Схоутена "Математические этюды". Схоутен настолько высоко оценил эту работу Гюйгенса, что сам перевел ее на латинский язык.

Работа Гюйгенса состоит из небольшого введения и 14 предложений. Эти предложения весьма различны по своему содержанию. Первые три являются теми принципами, на основе которых Гюйгенс основывал последующие решения. Предложения 4 – 9 посвящены решению задач, связанных с безобидным делением ставки. Предложения 10 – 14 содержат различные задачи, связанные с бросанием костей. В конце мемуара помещены 5 задач без решений, которые Гюйгенс предложил читателям для самостоятельных размышлений. Их решения были им даны лишь в 1665 г.

Несомненно, что первые три предложения составляют идейную основу всего сочинения Гюйгенса и поэтому приведем их полностью.

П р е д л о ж е н и е 1. Если я имею равные шансы получить a или b, то это мне стоит $(a + b)/2$.

П р е д л о ж е н и е 2. Если я имею равные шансы на получение a, b, или c, то это мне стоит столько же, как если бы я имел $(a + b + c)/3$.

П р е д л о ж е н и е 3. Если число случаев, в которых получается сумма a, равно p, а число случаев, в которых получается сумма b, равно q, то стоимость моего ожидания равна $(ap + bq)/(p + q)$.

Для нас ясно, что этими предложениями Гюйгенс ввел понятие математического ожидания для случайной величины, принимающей два или три значения. Если использовать современные представления, то в первых двух предложениях значения, принимаемые случайными величинами, равновероятны, а в третьем предложении вероятность значения a равна $p/(p + q)$ и вероятность значения b равна $q/(p + q)$. У Гюйгенса еще понятие вероятности не выделено, и он все время оперирует с числами шансов, благоприятствующих тому или другому событию. Гюйгенс предпочел, так сказать, коммерческую терминологию и говорил о стоимости, за которую он готов уступить свое право на получение выигрыша. Термин "ожидание" был введен в употребление учителем Гюйгенса – Схоутеном – при переводе.

Предложения 1 и 2 представляют собой ничто иное как версию задачи о разделе ставки. Мы приведем текст Гюйгенса с тем, чтобы читатели убедились насколько близки его рассуждения к рассуждениям Паскаля.

"Предположим, что я играю против другого лица на то, кто первым выиграет 3 партии, и что я уже выиграл 2 партии, а он – 1. Я хочу знать какая часть ставки причитается мне, когда мы хотим прервать игру и справедливо разделить ставки . . . Нужно заметить сначала, что достаточно принять во внимание число партий недостающих той и другой стороне. Так как верно, что если бы мы играли на то, кто выиграет 20 партий, и если бы я выиграл 19 партий, а мой противник 18, то я имел бы такое же самое преимущество, как и в изложенном случае, где при трех партиях я выиграл две, а он только одну, а это потому, что в обоих случаях мне недостает только одной партии, а ему двух*). Затем, чтобы вычислить часть причитающуюся каждому из нас, нужно обратить внимание на то, что произошло бы, если бы мы продолжали игру. Верно и то, что выиграв партию, я получил бы полностью сумму ставки, которую обозначу a. Но если первую партию выиграет мой противник, то наши шансы станут равными, принимаю во внимание, что каждому из нас будет недоставать по одной партии; значит, каждый из нас имел бы право на $a/2$, что согласно первому предложению, эквивалентно сумме половин, т.е. $(3/4)a$, так что моему сопернику остается $(1/4)a$".

Разделение ставки между тремя игроками Гюйгенс рассмотрел в предложении VIII, когда первому игроку недостает до выигрыша всей игры одной партии, а второму и третьему – по две партии. В предложении IX он рассмотрел вопрос о разделе ставки между тремя игроками, но при произвольном состоянии игроков. Общего выражения для решения этой задачи им дано не было и он изложил только принципы сведения общей задачи к частным случаям.

Формулировки предложений 10 – 14 следует признать недостаточно четкими. Их содержание полностью проясняется лишь при рассмотрении предложенных Гюйгенсом вопросов. Приведем некоторые из них.

П р е д л о ж е н и е 10. Определить при скольких бросаниях можно обязаться выбросить одной костью шесть очков?

Конечно, задача сформулирована весьма неопределенно. Нам ясно, что автору нужно понятие вероятности для точной формулировки его вопроса, а этого понятия еще нет. Речь же идет о вероятности того, что при n бросаниях ($n = 1, 2, \ldots$) хотя бы раз появится на кости шестерка.

Решение Гюйгенса состоит в следующем: при бросании имеется один шанс выкинуть 6 очков и 5 шансов получить другие грани. Если разыгрывается сумма a, то шанс получить эту сумму, согласно предложению 3, будет стоить $(1 \cdot a + 5 \cdot 0)/6 = a/6$.

"Тому, кто предложил ему бросить кости, остается $5a/6$. Значит, тот, кто играет партию в одно бросание, может ставить только 1 против 5."

При двух бросаниях кости вычисления стоимости игры Гюйгенс проводит следующим путем. "Если 6 очков получается при первом бросании, то бросающий получает a, но на это у него имеется 1 шанс, и имеется 5 шансов, что это не произойдет. Но тогда имеется еще второе бросание, которое стоит ему, согласно предшествующему вычислению $a/6$. Отсюда следует, что игра должна стоить играющему $\left(1 \cdot a + \frac{5}{6}a\right)/6 =$ $= 11a/36$. Аналогичным путем Гюйгенс получает для трехкратного бросания кости стоимость игры, состоящей в том, что хотя бы раз выпадет грань с числом очков 6, равную $91a/216$. Далее он вычислил стоимость подобной игры при четырехкратном, пяти и шестикратном бросании кости. Результаты получились такими: $671a/1296$, $4651a/7776$, $31031a/46656$.

В предложении 11 Гюйгенс рассматривает такую задачу: "Найти во сколько бросаний можно обязаться выбросить две шестерки?" Для нас эта формулировка неопределенна. Она должна быть сформулирована так: если при бросании двух костей игрок

---

*) Заметим, что это место выглядит убедительно лишь в предположении равносильности игроков. Однако если из двух партий обе выиграл игрок A, то это наводит на мысль о том, что он играет лучше, чем игрок B. Если же A выиграл 20 партий, а B – 18, то представление об их равносильности становится более убедительным.

Позднее неоднократно рассматривались многочисленные задачи с учетом неравносильности игроков. Такие задачи затем получали интерпретацию на языке физики и инженерного дела.

выигрывает сумму a. то какую ставку он должен внести для участия в игре (при безобидной игре)? Легко подсчитать. что цена игры при одном бросании должна стоить $a/36$, при двух бросаниях $71a/1296$ и так далее. Далее Гюйгенс сделал такое замечание: "Я нахожу. что тот, кто играет при 24 бросаниях, имеет еще легкую невыгоду и что можно принимать с выгодою партию, играя только минимально при 25 бросаниях."

Мы приведем теперь формулировки остальных трех предложений работы Гюйгенса.

Предложение 12. Найти такое число костей, при котором можно обязательно выбросить две шестерки при первом бросании?

Предложение 13. Найти причитающуюся каждому из нас часть общей суммы при предположении. что я бросил две кости один раз с тем условием, что если выпадет 7 очков, то выигрываю я, и что выигрывает мой противник, если выпадает 10 очков. А если выпадает другое число очков, то мы делим общую сумму поровну.

Предложение 14. Другой игрок и я поочередно бросаем две кости при условии, что я выигрываю, как только я выброшу 7 очков, и он выигрывает, как только выбросит 6 очков, и я предоставляю ему бросить первому. Требуется найти отношение моих шансов и его."

Интересно отметить, что в письма к Каркави от 6 июля 1656 г. Гюйгенс писал, что предложение 14 его трактата соответствует одной из шести задач Ферма, которые последний сообщил Каркави.

Для полноты картины мы сформулируем все пять задач, предложенных Гюйгенсом читателям для самостоятельного решения. Их решение он опубликовал лишь в 1665 г.

1. A и B играют двумя костями на следующих условиях: A выигрывает. если он выбросит 6 очков, B выигрывает, если выбросит 7 очков. Первым бросает A один раз, затем B бросает дважды, затем A бросает два раза и т.д., пока кто-нибудь не выиграет. В каком отношении шансы A относятся к шансам B? Ответ: как 10355 к 12276.

2. Трое игроков берут 12 фишек, из которых 4 белых и 8 черных, и играют на таких условиях: первый вытянувший белую фишку побеждает. A тянет первым, B – вторым, а затем C, потом опять A и т.д. В каком отношении находятся шансы одного против других?

3. A держит пари против B, что из 40 карт (по 10 одинаковой масти) он выберет такие, что каждая будет различной масти.

Здесь величина шансов A против B определяется как 1000 к 8139

4. Имеем, как во второй задаче, 12 фишек, из которых 4 белых и 8 черных; A держит пари против B, что в выборе 7 фишек вслепую он будет иметь 3 белых. Спрашивается, в каком отношении стоят шансы A против B? Если Гюйгенс имел в виду, что будут вынуты точно 3 белых фишки, то результат 35 : 64; если же по меньшей мере 3, то 42 : 57.

5. A и B, каждый имеющий по 12 монет, играют тремя костями на условиях: если A выбросит 11 очков, то он должен дать B одну монету, но если он выбросит 14, тогда B должен дать одну монету A. Тот игрок выигрывает, который первым получит все монеты. Здесь шансы A относятся к шансам B как 244 140 625 к 282 429 536 481.

Последняя задача является ничем иным как разновидностью задачи о разорении игрока.

Спустя десять лет после кончины известного философа Б. Спинозы (1632–1677), в Гааге была опубликована анонимная работа, состоящая из двух частей, далеких друг от друга по содержанию. "Исследование о радуге" и "Заметки о математической вероятности". Проведенные исследования подтверждают предположение о том, что эти сочинения были написаны Спинозой. Во второй части работы содержалось решение первой задачи Гюйгенса и были приведены формулировки остальных четырех. Нас должно заинтересовать то обстоятельство, что в названии работы уже говорится о математической вероятности, но хотя в самой работе вероятность не определяется и рассуждения ведутся над числом благоприятствующих событию случаев.

Для дальнейшего нам полезно сделать следующее замечание. В 1692 г. Д. Арбутнот (1667–1735) предпринял издание английского перевода книги Гюйгенса и к этому переводу он добавил ряд новых задач, в том числе задачу иной природы. Формулировка этой задачи такова: на плоскость наудачу бросается прямоугольный параллелепипед с ребрами, находящимися в отношении $a : b : c$. Найти отношение шансов выпадения параллелепипеда гранями ab, bc и ca.

К концу XVII века завершался длительный период накопления первичных сведений о случайных событиях, точно поставленных задач и подходов к их решению. Многие выдающиеся умы занимались этими вопросами и с разных позиций подходили к количественной оценке возможности наступления случайного события. Ферма фактически пользовался понятием математического ожидания, использование которого для решения разнообразных задач было широко развито Гюйгенсом; Паскаль, Ферма и Гюйгенс использовали представления о теоремах сложения и умножения вероятностей и подошли вплотную к понятию вероятности, однако его они не ввели. Казалось бы, что этот шаг – переход от рассмотрения числа возможных исходов, благоприятствующих наступлению события, к рассмотрению отношения этого числа к числу всех возможных исходов – был естественен. Однако никто этого шага не сделал. Рассуждения, благодаря этому, были сложны и формулировки задач не очень точны. И если бы исследователи того времени задали себе вопрос, что возможнее при четырехкратном бросании кости хотя бы раз выбросить шестерку или при двадцатипятикратном бросании двух костей хотя бы раз выбросить на обеих костях шестерки? – они были бы вынуждены ввести классическое понятие вероятности и далее его использовать. Однако этого в XVII веке не произошло и введение в науку классического понятия вероятностей принадлежит лишь XVIII столетию. Однако оно было исследованиями XVII века хорошо подготовлено. Период предыстории завершался и начинался период собственно истории теории вероятностей. Для этого уже был создан достаточно прочный фундамент.

§ 6. О первых исследованиях по демографии

В следующей главе мы узнаем, что одним из толчков для развития основных понятий теории вероятностей сыграли исследования Джона Граунта (1620–1675) и Вильяма Петти (1623–1687) по демографии или, как тогда говорили, по политической арифметике. Их работы наглядно продемонстрировали каким мощным орудием могут служить для изучения массовых явлений статистические наблюдения, если их соответствующим образом обработать. Их книги получили большое распространение, старательно изучались учеными самых разнообразных направлений деятельности, в том числе и математиками.

Первой работой, с которой начинается история статистики как области научного знания, следует назвать книгу Д. Граунта, опубликованную в 1662 г. под названием "Естественные и политические наблюдения, перечисленные в прилагаемом оглавлении и сделанные над бюллетенями смертности. По отношению к управлению, религии, торговле, росту, воздуху, болезням и разным изменениям означенного города". Нам нет нужды давать описание всего содержания этой книги, но оттенить отдельные моменты, необходимые для дальнейшего, следует.

Основная задача, которая заинтересовала Граунта, состояла в указании метода, который позволял бы установить с достаточной точностью возрастной состав населения города в результате наблюдений за возрастом умерших. С этой целью им были проанализированы результаты 229250 регистраций смертей в Лондоне происшедших за 20 лет. Среди этих смертей было отмечено 71124 смерти детей от 0 до 6 лет. Причины смерти были тщательно перечислены Граунтом. Он специально отметил, что отношение числа смертей детей от 0 до 6 лет к общему числу смертей за тот же период времени, равное 71124/229250, приблизительно равняется 1/3. Иными словами Граунт ввел представление о частоте события. Для развития теории вероятностей это обстоятельство сыграло огромную роль, как, впрочем, и его замечание: ". . . мы хотели бы отметить, что некоторые из случайностей имеют постоянное отношение к числу всех похорон" (цитированная книга, с. 32). Здесь Граунт вплотную подошел к представлению о статистической устойчивости средних.

Он установил, что для Лондона число рождений мальчиков к числу рождений девочек относится как 14 : 13, что в среднем на каждые 11 семейств ежегодно умирают 3 их члена, что одна из 2000 женщин умирает от родов, что в среднем на каждые 63 покойника приходится 52 новорожденных. Тем самым численность населения Лондона пополняется систематически за счет провинции. Он установил на основании таблиц смертности, что в Лондоне на каждые 100 мужчин 34 имеют возраст от 16 до 56 лет. Так что по его данным в ту пору из 199112 жителей мужского пола 67694 имели возраст от 16 до 56 лет.

Им была составлена первая таблица смертности, которую мы теперь приведем: из каждых 100 новорожденных доживает до

6 лет	64	36 лет	16	66 лет	3
16 лет	40	46 лет	10	76 лет	1
26 лет	25	56 лет	6	86 лет	0.

В этой таблице поражает огромная детская и юношеская смертность: только 64 % в ту пору доживали до 6 лет и только 40 % – до 16 лет.

Граунт прекрасно понимал, что точность его выводов тем больше, чем больше наблюдений имеется для обработки. Именно в связи с этим он отметил, что недостаточно ограничиваться обработкой бюллетеней смертности только за одну неделю для получения полноценных выводов о составе населения.

Понятие частоты оказалось полезным и его сразу подхватили другие авторы. Так в небольшой книге В. Петти "Два очерка по политической арифметике, относящиеся к людям, зданиям, больницам в Лондоне, Париже", вышедшей в 1682 г. в Лондоне, а через два года во французском переводе – в Париже, были даны сравнительные данные о смертности в госпиталях шарите *) Парижа и Лондона. Так в одном из госпиталей шарите Парижа в течение года из 2647 больных скончались 338, а в двух госпиталях Лондона из 3281 больных ушли из жизни 461. Частоты госпитальной смертности для Парижа и Лондона оказываются соответственно равными 0,136 и 0,140. Петти не использовал десятичных дробей и обе частоты считал приблизительно равными 1/7. Еще больший процент смертности оказался в парижском госпитале "Божий дом" (L'hotel dieu), а именно в нем из 21591 больных скончалось 5360. Таким образом для этого госпиталя частота окончательного исцеления от всех болезней и печалей оказалась равной $5360/21491 \approx 0{,}262$. Петти принимал ее за 1/4.

В этой же книге Петти установил, что в Лондоне в среднем умирает один житель из 30, а в сельской местности – один из 37. Среди же членов парламента одна смерть приходится на 50 человек. Он также утверждал, что о численности населения города можно судить по бюллетеням смертности. Так, для примера, в Лондоне было зарегистрировано 22331 смертей. Значит, поскольку коэффициент смертности для Лондона равен 1/30, число жителей в этом городе должно быть близко к 669 930.

---

*) La charité – милосердие. Так назывались больницы, организованные церковью для бедняков.

Несомненно, что работы Граунта, Петти и ряда их последователей представляют собой ничто иное как первые шаги в области математической статистики.

Непосредственным продолжателем исследований, начатых Граунтом и Петти, был знаменитый английский астроном Эдмунт Галлей (1656–1742). В 1693 г. Галлей опубликовал в изданиях Лондонского королевского общества две статьи "Оценка степеней смертности человечества, выведенная на основании любопытных таблиц рождений и погребений города Бреславля, с попыткой установить цену пожизненных рент" и "Несколько дальнейших замечаний по поводу Бреславльских бюллетеней смертности". В основу этих статей были положены данные о движении населения Бреславля за 1687–1691 гг., присланные по просьбе секретаря общества Генриха Жюстелля пастором Каспаром Нейманом. Более Галлей к этим вопросам не возвращался.

Одна из причин интереса Галлея к таблицам смертности состоит в том, что сами Граунт и Петти сознавали недостаточную обоснованность своих выводов, поскольку у них отсутствовали численность населения и возраст умерших (зачастую). Кроме того, в городах, которые они изучали – Лондон и Дублин – был большой приток населения извне. Это обстоятельство делает указанные города "неподходящими в качестве стандарта для этой цели, которая требует, если это возможно, чтобы население, с которым имеют дело, было совершенно закрытым, т.е. таким, где все умирают там, где они родились, где нет никаких эмигрантов и иммигрантов" (Галлей, первый мемуар). По словам Галлея, бреславльские материалы не имеют указанных дефектов.

На основании имевшихся у него данных Галлей составил таблицу смертности, которую он рассматривал одновременно и как таблицу доживающих по возрасту лиц, так и как распределение населения по возрасту. Он ввел в науку понятие о вероятной продолжительности жизни, как о возрасте, которого одинаково можно достигнуть и не достигнуть. На современном языке это медиана длительности жизни. Сам Галлей не вводил ни термина медиана, ни термина вероятная продолжительность жизни. В вычислениях Галлея можно заметить использование им принципов, лежащих в основе теорем сложения и умножения вероятностей, а также рассуждения, близкие к формулировке закона больших чисел.

Работы Галлея имели очень большое значение для развития науки и применений статистических исследований о народонаселении к вопросам страхования.

ГЛАВА 2

ПЕРИОД ФОРМИРОВАНИЯ ОСНОВ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 7. Возникновение классического определения вероятности

Образование основных математических понятий представляет важные этапы в процессе математического развития. Мы видели, что до конца XVII века наука так и не подошла к введению классического определения вероятности, а продолжала оперировать только с числом шансов, благоприятствующих тому или иному интересующему исследователей событию. Отдельные попытки, которые нами были отмечены у Кардано и у позднейших исследователей, не привели к ясному пониманию значения этого нововведения и остались инородным телом в завершенных работах. Однако, в тридцатых годах XVIII столетия классическое понятие вероятности стало общеупотребительным и никто из ученых этих лет не мог бы ограничиться только подсчетом числа благоприятствующих событию шансов. Кто же ввел это понятие и настолько ясно показал его необходимость, чтобы в дальнейшем уже не возникло сомнения в его целесообразности для развития науки? Мы должны заметить, что введение классического определения вероятности произошло не в результате однократного действия, а заняло длительный промежуток времени, на протяжении которого происходило непрерывное совершенствование формулировки, переход от частных задач к общему случаю.

Внимательное изучение, показывает, что еще в книге Х. Гюйгенса "О расчетах в азартных играх" (1657) нет понятия вероятности как числа, заключенного между 0 и 1 и равного отношению числа благоприятствующих событию шансов к числу всех возможных. А в трактате Я. Бернулли "Искусство предположений" *) (1713) понятие это введено, хотя и в далеко несовершенной форме, но, что особенно важно, широко используется. Что же произошло за те полстолетия, которое прошло между публикациями этих книг? Что заставило Я. Бернулли ввести в научный обиход классическое понятие вероятности?

Несомненно, что формулировка закона больших чисел, осуществленная Я. Бернулли, сама по себе является достаточным для этого основанием. Однако имеется и другое соображение, которое, несомненно, оказало сильное влияние на ход мыслей ряда исследователей, в том числе и Я. Бернулли. Речь идет о работах Граунта и Петти, о которых было сказано в предыдущем параграфе. Эти произведения решающим образом воздействовали на лучшие умы того времени и не было ни одного мало-мальски крупного математика, который не изучал бы их и не находился под их воздействием. Этого влияния не избежал и Я. Бернулли. Произведения Граунта и Петти убедительно показали преимущества понятия частоты перед понятием численности. Именно понятие частоты, т.е. отношение числа наблюдений, в которых появляется определенное свойство, к числу всех наблюдений, позволяет получить серьезные практические выводы, тогда как рассмотрение численностей оставляет исследователя в состоянии неопределенности. Отсюда оставался лишь один шаг до введения понятия классической вероятности. Заметим, что выводы Граунта и Петти относительно устойчивости частоты некоторых событий подготовили почву и к формулировке закона больших чисел.

В весьма несовершенной форме классическое определение вероятности у Я. Бернулли появилось в первой главе четвертой части "Искусства предположений". Там он сказал следующие слова: "Вероятность есть степень достоверности и отличается от нее, как часть от целого". Далее было пояснение сказанного на примере, который отчетливо показывает, что Я. Бернулли в данную им формулировку фактически вкладывал тот же самый смысл, какой мы вкладываем в классическое определение вероятности. Вот его пояснение: "Именно: если полная и безусловная достоверность, обозначаемая нами буквой α или единицей 1, будет, для примера, предположена состоящей из пяти вероятностей, как бы частей, из которых три благоприятствуют существованию или осуществлению какого-либо события, остальные же неблагоприятствуют, то будет сказано, что это событие имеет $\frac{3}{5}\alpha$ или $\frac{3}{5}$ достоверности".

При формулировке главного предложения в пятой главе четвертой части Я. Бернулли вновь писал об отношении числа благоприятствующих случаев к числу всех возможных. Но при этом он не оговаривал, а предполагал само собой разумеющимся, что эти случаи должны быть равновероятными. Наряду с этим отношением, которое вошло в науку, Бернулли предлагал и другое – число благоприятствующих к числу неблагоприятствующих. В науке привилось только первое из этих отношений, второе же не привилось, быть может по той причине, что оно изменяется от 0 до бесконечности, а может быть по причине неаддитивности этих отношений.

Интересны рассуждения четвертой главы четвертой части сочинения Я. Бернулли. Он задал вопрос: как определить вероятность случайного события, если у нас нет

---

*) Часть четвертая этой книги переведена на русский язык и с содержательными комментариями издана недавно: Я. Бернулли, "О законе больших чисел", изд. Наука, 1986, редактор Ю.В. Прохоров.

возможности подсчитать числа всех возможных и благоприятствующих ему шансов? Ответ им был сформулирован следующим образом: "Но здесь нам открывается другая дорога для достижения искомого. И то, что не дано вывести á priori, то, по крайней мере, можно получить á posteriori, т.е. из многократного наблюдения результатов в подобных примерах. . . Ибо, если, например, при наблюдениях, сделанных некогда над тремя сотнями людей того же возраста и сложения, в каких находится теперь Тит, было замечено, что из них двести до истечения десяти лет умерли, а остальные остались в живых и дальше, то можно заключить с достаточным основанием, что имеется вдвое больше случаев Титу умереть в течение ближайшего десятилетия, чем остаться в живых по истечении этого срока . . . Этот опытный способ определения числа случаев по наблюдениям не нов и не необычен."

Нам важно теперь подчеркнуть, что в высказанных отрывках достаточно четко прослеживается мысль о статистическом определении вероятности. Наверняка при этом Я. Бернулли основывался и на работах Граунта и Петти.

Таким образом в трактате Я. Бернулли присутствуют обе концепции вероятности – классическая и статистическая. Обе они изложены не очень четко, но существенно то, что они уже введены в рассмотрение и использованы. Этим был сделан принципиальный шаг в науке о случае – введено в рассмотрение понятие вероятности случайного события как числа, заключенного между 0 и 1. Достоверному событию при этом приписывается максимально возможное значение вероятности единица, а невозможному – минимальное – ноль. Кроме того было ясно сказано, что это число может быть определено двумя различными способами: путем подсчета числа равновозможных случаев, которые благоприятствуют событию, и всех возможных случаев и вычисления их отношения или же путем проведения большого числа независимых испытаний и вычисления частоты события. Можно считать, что теория вероятностей с этого момента начала свою историю. До этого же была предыстория, которая подготовляла почву для формирования основных понятий и задач теории вероятностей.

Я. Бернулли обдумывал свое "Искусство предположений" долгие годы, по его словам, по меньшей мере двадцать лет. Но свет оно увидело лишь в 1713 г., восемь лет спустя после смерти автора. Однако содержание этого произведения многие годы до его публикации уже было известно научной общественности по рукописи, которая стала доступна многим. Об этом говорится, в частности, в публикациях Фонтенеля и Сорена, посвященных заслугам покойного и вышедшим в свет соответственно в 1705 и 1706 годах. На эти публикации позднее ссылался П. Монмор (1687–1719) в своей книге "Обзор анализа азартных игр" (1-е изд. – 1706 г.; 2-е изд. – 1713 г.); а также сделал подробный анализ содержания "Искусства предположений". Таким образом трактат Я. Бернулли оказывал влияние на развитие теории вероятностей задолго до его опубликования. Об этом позднее мы скажем еще по другому поводу.

Монмор в упомянутой книге использовал понятие вероятности и применил его к решению достаточно сложных задач. В частности, Монмор рассмотрел и правильно решил следующую задачу: имеется n предметов, пронумерованных числами от 1 до n. Спрашивается, чему равна вероятность того, что при последовательном вынимании этих предметов наудачу (без возвращения) хотя бы один предмет будет вынут так, что номер вынимания совпадет с присвоенным ему номером. Эта вероятность оказалась равной $1 - 1/2! + 1/3! - \ldots + (-1)^{n-1}/n!$. Мы знаем, что этой задаче теперь придаются различные формулировки.

А. Муавр воспринял классическое определение вероятности, данное Бернулли, и вероятность события определил почти в точности так, как это делаем мы теперь. Он писал: "Следовательно мы строим дробь, числитель которой будет число случаев появления события, а знаменатель – число всех случаев, при которых оно может появиться или не появиться, такая дробь будет выражать действительную вероятность его появления." После этого определения Муавр привел в точности пример, о котором мы упоминали при рассказе о вкладе Бернулли, а именно: "если какое-то событие имеет 3 благоприятствующих шанса, 2 неблагоприятствующих, дробное выражение 3/5 будет точно говорить о вероятности его появления и может рассматриваться как ее мера" ("Доктрина шансов"). Обратим внимание на то, что Муавр, как и Я. Бернулли, не оттенял то обстоятельство, что шансы должны быть равновероятными. Это замечание впервые было введено в определение классической вероятности лишь П. Лапласом в его "Аналитической теории вероятностей". Лагранж об этом еще не задумывался и давал определение вероятности в точности по Муавру. По-видимому, на Лапласа повлияла дискуссия, начатая Д'Аламбером, который при решении задачи о вероятности выпадения (при бросании двух монет) герба на одной из монет и решки – на другой, определил ее равной 1/3. Это он мотивировал тем, что имеются лишь три возможности: 1) на обоих монетах выпадает герб; 2) на обоих монетах выпадает решка; 3) на одной монете выпадает герб, а на другой – решка.

§ 8. О формировании понятия геометрической вероятности

Уже в первой половине XVIII века выяснилось, что классическое понятие вероятности имеет ограниченную область применений и возникают ситуации, когда оно не действует, а потому необходимо какое-то естественное его расширение. Обычно считают, что таким толчком послужили работы французского естествоиспытателя Ж. Бюффона (1707–1788), в которых он сформулировал знаменитую задачу о бросании иглы на разграфленную плоскость и предложил ее решение. Это утверждение требует поправки, поскольку исторически оно неверно. Дело в том, что задолго до рождения Бюффона появилась работа, в которой фактически уже был поставлен вопрос о нахождении геометрической вероятности. Правда, в ту пору еще не было и определения вероятности.

В 1692 г. в Лондоне был опубликован английский перевод книги Х. Гюйгенса "О расчетах в азартных играх", выполненный Д. Арбутнотом (1667–1735). В конце первой части переводчик добавил несколько задач, среди которых была сформулирована задача совсем иной природы, по сравнению с теми, которые были рассмотрены великим автором. Он назвал эту задачу трудной и поместил ее в дополнении "для того, чтобы она была решена теми, кто считает такого рода проблемы достойными внимания". Задача, предложенная Арбутнотом, состоит в следующем: на плоскость наудачу бросается прямоугольный параллелепипед с ребрами, равными a, b, c. Спрашивается, как часто параллелепипед будет выпадать гранью ab?

Сам Арбутнот не сделал даже попытки решения придуманной им задачи. Это было осуществлено значительно позднее Т. Симпсоном (1710–1761) в книге "Природа и законы случая" (1740), где задача была приведена под номером XXVII.

Идея решения, предложенная Симпсоном, состоит в следующем: опишем около параллелепипеда сферу и спроектируем из центра на поверхность ее все ребра, боковые грани и основания. В результате поверхность сферы разобьется на шесть непересекающихся областей, соответствующих граням параллелепипеда. Далее Симпсон написал: "Нетрудно заметить, что определенная часть сферической поверхности, ограниченная траекторией, описанной таким образом радиусом, будет находиться в таком же соотношении к общей площади поверхности, как вероятность появления некоторой грани к единице".

В том, что было только что сказано, в полной мере заключены принципы разыскания геометрических вероятностей: вводится мера множества благоприятствующих событию случаев и берется ее отношение к мере множества всех возможных случаев. В нашем случае полная мера сводится к площади поверхности шара. Заметим, что Симпсон ни слова не сказал о физической интерпретации решения. Ведь для того, чтобы параллелепипед упал на плоскость определенной гранью, необходимо, чтобы его центр тяжести находился над ее проекцией на плоскость падения. Однако в решении Симпсона это требование соблюдено.

Введем для дальнейшего обозначения: $R^2 = a^2 + b^2 + c^2$, P_{ab}, P_{bc}, P_{ca} – вероятности выпадения на какую-то определенную грань соответственно ab, bc, ca. Вероятности выпадения на какую-то из граней ab (соответственно bc и ca) должны быть увеличены вдвое. Формулы, о которых идет речь, должны быть таковы:

$$P_{ab} = \frac{1}{\pi}\operatorname{arctg}\frac{ab}{cR}, \quad P_{bc} = \frac{1}{\pi}\operatorname{arctg}\frac{bc}{aR}, \quad P_{ca} = \frac{1}{\pi}\operatorname{arctg}\frac{ac}{bR}.$$

Бюффон дважды публиковал работы, посвященные геометрическим вероятностям. Первая его публикация на эту тему относится к 1733 г., когда он сделал в Парижской академии наук доклад, напечатанный под называнием "Мемуар об игре франк-карро" (Мемуар об игре прямо в клетку). Позднее, в 1777 г. этот мемуар был целиком включен в "Опыт нравственной арифметики", являвшейся дополнением к тому IV его "Естественной истории".

Цель, которую ставил перед собой Бюффон, состояла в том, чтобы показать, что "геометрия может быть использована в качестве аналитического инструмента в области теории вероятностей", в то время как до тех пор "геометрия казалось мало пригодной для этих целей", поскольку для них использовалась только арифметика.

Игра франк-карро состоит в следующем: пол разграфлен на одинаковые фигуры. На пол бросается монета, ее диаметр $2r$ меньше каждой из сторон и монета целиком укладывается внутрь фигуры. Чему равна вероятность того, что брошенная научачу монета пересечет одну или две стороны фигуры?

Для определенности рассмотрим покрытие плоскости прямоугольниками со сторонами a и b, $b > 2r$, $a > 2r$. Легко подсчитать, что площадь полосы между основным прямоугольником со сторонами, параллельными сторонам основного на расстоянии r от каждой из его сторон и целиком расположенного внутри основного, равна $2r(a + b - 2r)$. Легко понять, что центр монеты, попав внутрь малого прямоугольника, не только не пересечет, но даже не коснется сторон основного. Значит, вероятность того, что монета пересечет по меньшей мере одну из сторон основного прямоугольника равна $2r\dfrac{a + b - 2r}{ab}$.

Вторая задача, сформулированная и решенная Бюффоном, состоит в следующем: плоскость разграфлена равноотстоящими параллельными прямыми. На плоскость наудачу бросается игла. Один игрок утверждает, что игла пересечет одну из параллельных прямых, другой — что не пересечет. Определить вероятность выигрыша каждого из игроков. Решение этой задачи хорошо известно, и нет необходимости приводить его здесь. Менее известна задача Бюффона об игре, когда игла бросается на плоскость, разграфленную на квадраты. В решении этой задачи Бюффон допустил ошибку, позднее исправленную Лапласом. Именно, Бюффон считал, что искомая вероятность равна $2r\dfrac{a - r}{\pi a^2}$, тогда как в действительности она равна $4r\dfrac{2a - r}{\pi a^2}$.

После Бюффона задачи на геометрические вероятности стали систематически включаться в трактаты и учебники по теории вероятностей. Так, в знаменитую книгу Лапласа "Аналитическая теория вероятностей" были включены и подробно рассмотрены все задачи Бюффона. Но Лаплас не счел нужным отметить, откуда они были заимствованы и кто автор этих задач.

Следует отметить, что терминология Лапласа далека от совершенства. Так, для примера, он писал, что "$8r$ равняется сумме всех случаев, в которых игла пересекает одну или другую параллельные линии" и что $2a\pi$ равно "числу всех возможных комбинаций". Здесь $2r$ означает длину иглы, а a – расстояние между параллельными прямыми.

Во второй задаче, рассмотренной Лапласом, плоскость разграфлена двумя системами параллельных прямых, представляющих ничто иное как систему координатных линий на плоскости. Расстояние между линиями первой системы равно a, второй системы – b. На плоскость бросается игла длины $2r$ ($2r < a$, $2r < b$). Чему равна вероятность того, что игла пересечет хотя бы одну линию? Решение, предложенное Лапласом, предполагает, что дело идет о системах взаимно перпендикулярных прямых. Это Лапласом не оговорено. В результате вычисления – числа благоприятствующих и "числа всех возможных случаев" – Лаплас определил, что вероятность пересечения одной из линий брошенной иглой равна $4r\dfrac{a+b-r}{\pi ab}$.

В прекрасном для своего времени учебнике "Основания математической теории вероятностей" (1846) В.Я. Буняковского (1804–1889) имеется довольно большой раздел, посвященный геометрической вероятности. В него включена задача Бюффона о бросании иглы и частный случай игры франк-карро, когда плоскость разбита на равнобедренные треугольники. С современных позиций терминология Буняковского далека от совершенства. Пример такого словоупотребления мы сейчас и приведем: "... иногда встречаются такие ситуации, в которых число благоприятствующих статочностей, а равно и всех возможных бывает бесконечное. Искомая вероятность определится тогда отношением этих двух бесконечных чисел ... ". Она будет "числом конечным и совершенно определенным".

Серьезный шаг в развитии геометрических вероятностей связан с именами Ламе (1795–1870), Барбье, Д. Сильвестра (1814–1897), М. Крофтона, которые не только поставили новые задачи, но и привлекли к их решению понятие меры множества (пусть еще и на интуитивном уровне). На базе их рассмотрений позднее возникла новая ветвь геометрии, получившая наименование интегральной геометрии.

В 1860 г. Ламе на факультете наук Парижской нормальной школы прочитал курс лекций по геометрии. В этом курсе он рассмотрел задачу Бюффона о бросании иглы и применил ее к тому случаю, когда центр иглы бросается наудачу в центр эллипса или правильного многоугольника. Среди слушателей был Барбье, обобщивший рассуждения Ламе на случай любого выпуклого контура. В сущности Барбье не внес ничего нового в сам метод. Он только заметил, что рассуждения Ламе не связаны жестко ни с рассмотрением эллипса, ни с правильными многоугольниками, а легко обобщаются на любой выпуклый контур.

Сильвестр первым после Бюффона расширил тематику задач на геометрические вероятности. Им была предложена задача о четырех точках или задача Сильвестра. Ее формулировка такова: четыре точки взяты наудачу внутри выпуклой области. Чему равна вероятность того, что, взяв эти точки в качестве вершин, можно составить выпуклый четырехугольник?

Сильвестр предложил следующее решение: обозначим через A площадь выпуклой области. Бросим в нашу область сначала три точки и построим по этим точкам треугольник. Пусть его средняя площадь равна M. Бросим теперь наудачу четвертую точку. Если она попадет внутрь треугольника, то по этим четырем точкам выпуклого четырехугольника составить нельзя. Но четвертую точку мы можем выбирать четырьмя различными способами, следовательно, при бросании четырех точек вероятность получить невыпуклый четырехугольник равна $p = 4M/A$. Отсюда заключаем, что вероятность получения при этом выпуклого четырехугольника равна $1 - 4M/A$. Среднее значение M зависит от области, в которую бросают точки. Для некоторых выпуклых фигур значение M вычислено. М. Крофтон в статье "Вероятность", опубликованной в Британской энциклопедии (9-е издание, т. 19, стр. 786, Эдинбург, 1885).

привел таблицу, сославшись на работу Вольхауза, из которой легко получить значения M для соответствующих выпуклых областей.

Таблица 22

вероятность	треугольник	параллелограмм	правильный 6-угольник	окружность
p	$1/3 = 0{,}3333...$	$11/36 = 0{,}3056$	$289/972 \approx$ $\approx 0{,}2971$	$35/(12\pi^2) =$ $= 0{,}2955$

Сильвестр отчетливо понимал, что при вычислении геометрических вероятностей приходится брать отношение площадей или объемов (общее мер) тех областей, которые благоприятствуют событию и в которых помещаются всевозможные события. Фактически так поступали и раньше. Но при этом произносили другие слова, которые или не имели определенного смысла или же не соответствовали производимым действиям. Сравнив результаты вычислений для различных областей, Сильвестр предложил найти те области, для которых вероятность получения выпуклого четырехугольника достигает максимума и минимума. Первые результаты принадлежат Крофтону и опубликованы в ранее указанной статье. Он доказал, что минимум достигается для круга. Там же он высказал предположение, что минимум достигается и для эллипса. Это предложение было доказано лишь В. Блашке (Vorlesungen über differential Geometrie – Berlin, 1923). Дельтейль показал, что максимальная вероятность формирования выпуклого четырехугольника достигается для треугольной области.

В учебной литературе широко известна задача о встрече. Спрашивается, когда она появилась и кто был ее автором? При изучении многочисленной литературы моей ученице М.Т. Лориньо Перес удалось найти ответ на этот вопрос. В книге Уайтворта "Выбор и шанс" (Choice and chance – London, 1886, chap III, p. 242–243) была рассмотрена следующая задача. Лица A и B независимо один от другого отправляются на прием в парке. Лицо A прибывает на прием в наудачу выбранный момент между 3 и 5 часами пополудни, а B – между 4 и 7 часами пополудни. Каждый из них остается на приеме в течение часа. Чему равна вероятность того, что они окажутся на приеме одновременно хотя бы одно мгновение?

Задача была решена Уайтвортом обычным путем, какой используется и в настоящее время. Легко подсчитать, что искомая вероятность равна 1/3. Позднее эта задача перекочевывала из книги в книгу в качестве иллюстративного примера, а также находила применения в задачах организации производства.

Несомненно, что в XIX веке на развитие проблематики геометрических вероятностей особое влияние оказал Крофтон. Он начал изучать пересечение случайными прямыми заданных выпуклых контуров. Мы не станем излагать его результаты, поскольку они вошли в курсы интегральной геометрии и монографии по геометрическим вероятностям, а потому легко доступны.

На необходимость совершенствования понятия геометрической вероятности несомненное влияние оказала книга Ж. Бертрана (1822–1900) (Calcul de probabilité. – Paris, 1899), в которой на хорошо подобранных примерах было показано, что логически понятие геометрической вероятности не выдерживает критики. Играя на неопределенности терминологии, казалось бы для одной и той же задачи, ему удалось получить несколько различных ответов. В качестве основной мишени им была избрана известная задача о проведении наудачу хорды внутри круга. Само собой разумеется, что критика Бертрана привлекла внимание математиков к общим вопросам логического обоснования теории вероятностей.

В XX веке интерес к геометрическим вероятностям не ослабел, а вырос, поскольку, помимо чисто математического интереса, они приобрели и серьезное прикладное значение в физике, биологии, медицине, инженерном деле и др. Этот аспект геометрических вероятностей заслуживает специального рассмотрения.

§ 9. Основные теоремы теории вероятностей

Мы обратимся теперь к следующему естественному вопросу: когда и кто выделил в теории вероятностей основные ее теоремы – сложения и умножения и полной вероятности? В конечном счете на этих простых результатах покоится вся теория вероятностей и ее многочисленные применения. Именно поэтому представляет интерес выяснение процесса их формирования. В книге Л.Е. Майстрова "Теория вероятностей", исторический очерк, 1967 (с. 65) утверждается, что в XVII веке уже "были известны теоремы сложения и умножения вероятностей, которые широко применялись при решении задач". Однако, нам не удалось заметить ни в переписке Ферма с Паскалем, ни в трактате Гюйгенса ни формулировки, этих теорем, ни мало-мальски осознанного их использования. Однако зачатки этих теорем можно проследить буквально с первых шагов теории вероятностей как математической науки.

Так в работах Паскаля можно усмотреть, что он отчетливо понимал как следует подсчитывать число благоприятствующих шансов для события A, если нам известны шансы для несовместимых событий A_j, составляющих событие A. Это, конечно, еще не теорема сложения, но важный шаг на пути ее формулировки. При решении задачи о разделе ставки Паскаль рассуждал следующим образом: пусть игроку А для выигрыша игры недостает трех партий, а игроку В – четырех. Тогда для завершения игры достаточно шести партий. Игрок А выигрывает, если из этих шести партий он выиграет все шесть, пять или четыре, или три партии. Таким образом, число благоприятствующих шансов для выигрыша А игры оказывается равным

$$C_6^6 + C_6^5 + C_6^4 + C_6^3 = 1 + 6 + 15 + 20 = 42.$$

Это рассуждение Паскаль предложил и в общем случае, когда для окончания игры игроку А недостает m партий, а игроку В – n партий.

В работах Я. Бернулли и Н. Бернулли дается отчетливая формулировка правила вычисления вероятности противоположного события, если известна вероятность прямого события.

При выводе формул, получивших наименование формул Бернулли, Я. Бернулли сознательно использовал правила сложения и умножения вероятностей, но самих правил он не сформулировал. Они в его рассуждениях присутствуют как бы неявно.

Одно замечание Я. Бернулли показывает, что он отчетливо понимал особенности теоремы сложения для совместимых событий. Вот это замечание: "Если два человека, достойные смертной казни, принуждаются бросить кости при условии, что тот, кто выбросит меньшее число очков, понесет свое наказание, а другой, который выбросит бóльшее число очков, сохранит свою жизнь, и что оба они сохранят жизнь, если выбросят одинаковое число очков, то мы найдем для ожидания одного 7/12 . . . Но из этого нельзя заключить, что ожидание другого равно 5/12 жизни, так как очевидно, что обе участи одинаковы. Другой также будет ожидать 7/12, что дает для обоих 7/6 жизни, т.е. больше целой жизни. Причиной этого является то, что нет ни одного случая, в котором хотя бы один не останется живым, а имеется несколько случаев, когда они оба могут остаться в живых." Нет нужды добавлять к словам Я. Бернулли, что он находился рядом с предложением, которое мы теперь записываем в следующем виде:

$$\mathsf{P}\{A + B\} = \mathsf{P}\{A\} + \mathsf{P}\{B\} - \mathsf{P}\{AB\}.$$

Впрочем, если с современных позиций рассматривать работу Кардано "Книга об игре в кости", то в главе XIV "О соединении очков" можно в частном примере усмотреть этот же результат, но не для вероятностей, а для числа шансов. Он рассматривал число случаев выпадения при бросании двух костей хотя бы на одной кости одного очка. Это число равно 11, поскольку шестью различными способами может появиться 1 на первой кости: (1, 1), (1, 2), (1, 3) (1, 4), (1, 5), (1, 6) и столькими же на второй. Но случай (1, 1) при этом мы указываем дважды. Так что различных случаев будет не 12, а лишь 11.

Однако, как ни важны приведенные наблюдения, мы не должны приписывать ни Кардано, ни Паскалю и Ферма, ни Я. Бернулли формулировку теоремы сложения вероятностей, как важнейшего положения теории вероятностей.

Первая четкая и окончательная формулировка теоремы сложения вероятностей находится в работе Т. Байеса (1702–1761), носящей длинное название – "Опыт решения задач по теории вероятностей покойного достопочтеиного мистера Байеса, члена Королевского общества. Сообщено мистером Прайсом в письме Джону Кентону, магистру искусств, члену королевского общества". Работа Байеса была зачитана на заседании Лондонского королевского общества 27 декабря 1763 г., спустя два года после смерти автора. В определении 1 работы содержится определение несовместимых событий. Байес употребляет другой термин "неплотные события" (inconsistent). Согласно Байесу, "несколько событий являются неплотными, если наступление одного из них исключает наступление других". Формулировка же теоремы сложения дается Байесом в предложении 1, которое состоит в следующем: "Если несколько событий являются неплотными, то вероятность того, что наступит какое-то из них, равно сумме вероятностей каждого из них". В этом предложении мы видим четкую формулировку теоремы сложения вероятностей во вполне современной форме.

Точно так же теорема умножения вероятностей длительный период формировалась на рассмотрении частных примеров и на подсчете числа шансов, благоприятствующих наступлению произведения двух или нескольких событий. Такого рода подсчеты встречаются практически у всех предшественников Я. Бернулли. Я. Бернулли широко использует эти правила при выводе своих знаменитых формул. Широко использовал правила сложения и умножения вероятностей Монмор. Однако формулировки теоремы умножения ни у кого из них не встречается. Четкое выделение теоремы умножения было осуществлено лишь Муавром. Во введении к "доктрине шансов" в § 8 он определил важное понятие независимости случайных событий. А именно, он формулирует следующее положение: "Мы скажем, что два события независимы, когда каждое из них не имеет никакого отношения к другому, а появление одного из них не оказывает никакого влияния на появление другого." Еще более определенно им дано определение зависимых событий. А именно, "два события зависимы, когда они связаны друг с другом и когда вероятность появления одного из них изменяется при появлении другого".

Эти определения Муавр снабдил простеньким примером. Пусть имеются две кучки карт одной масти, в каждой кучке от двойки до туза. Тогда вероятность того, что из каждой кучки наудачу удается вынуть по тузу будет равна $1/13 \cdot 1/13 = 1/169$. Мы имеем дело с двумя независимыми событиями. Если же мы вынимаем две карты из одной кучки и спрашиваем о вероятности того, что при первом вынимании извлечем туз, а при втором – двойку, то здесь вероятность первого события равна $1/13$, а второго $1/12$. Таким образом вероятность интересующего нас события равна уже $1/13 \cdot 1/12 = 1/156$.

Нам особенно важно привести сейчас следующую формулировку Муавра: "... вероятность появления двух зависимых событий равна произведению вероятности появления одного из них на вероятность того, что другое должно появиться, если первое из них уже появилось. Это правило может быть обобщено на случай нескольких событий".

Мы видим, таким образом, что формулировку теоремы умножения вероятностей и введение понятия условной вероятности удалось осуществить только Муавру. Это им было сделано уже в 1718 г. в первом издании его "Доктрины шансов".

О вероятности совместного наступления нескольких событий Муавр писал следующее: "... надо обозначить одно из них как первое, другое как второе и т.д. Тогда вероятность появления первого должна рассматриваться как независимая от остальных; вторая – в предположении, что первое произошло, третье – в предположении наступления первого и второго и т.д. Следовательно, вероятность наступления всех событий равна произведению всех только что указанных вероятностей". Далее Муавр отметил, что разыскание условных вероятностей, как правило, представляет собой сложное занятие.

Общее положение Муавр продемонстрировал на решении ряда задач. Вот одна из них. Пусть события A, B и C независимы в совокупности и x, y, z означают вероятности их наступления. Тогда xyz есть вероятность наступления всех трех событий, а $1 - (1 - x) \cdot (1 - y) \cdot (1 - z)$ – вероятность наступления хотя бы одного из событий A, B, C.

В упомянутой ранее работе Байеса содержится формулировка теоремы умножения вероятностей (предложение 3): "Вероятность того, что наступят оба взаимосвязанные события, есть соотношение, получающееся от перемножения вероятности первого события на вероятность наступления второго в предположении, что первое наступило". Мы уже видели, что это предложение было четко сформулировано Муавром и, поскольку произведение Муавра было широко известно, Байес, несомненно, заимствовал его у своего знаменитого предшественника. Единственно, в чем Байес пошел дальше Муавра – это в формулировке следствия о вычислении вероятности $\mathsf{P}\{B|A\}$ по вероятностям $\mathsf{P}\{AB\}$ и $\mathsf{P}\{A\}$. Это предложение дало основание приписывать Байесу формулы, носящие его имя. В действительности у него их нет, поскольку он не знал формулы полной вероятности.

Результат, приписываемый Байесу, по-видимому, впервые получил современную формулировку у Лапласа в его "Опыте философии теории вероятностей". В главе "Общие принципы теории вероятностей" он сформулировал принцип VI, который относится к вероятности гипотез или, как писал Лаплас, вероятности причин. Пусть некоторое событие A может произойти с одним из n несовместимых событий B_1, $B_2, \ldots, B_n$ и только с ними; эти события Лаплас называет причинами. Спрашивается, если известно, что событие наступило, чему равна вероятность того, что осуществилась и причина B_i? Вот формулировка ответа, данного Лапласом: "вероятность существования какого-либо из этих причин равна, следовательно, дроби, числитель которой есть вероятность события, вытекающая из этой причины, а знаменатель есть сумма подобных вероятностей, относящихся ко всем причинам: если эти различные причины, рассматриваемые á priori, не одинаково вероятны, то вместо вероятности события, вытекающей из каждой причины, следует взять произведение этой вероятности на вероятность самой причины". Легко понять, что Лаплас словесно сформулировал известное "правило Байеса"

$$\mathsf{P}\{B_i|A\} = \mathsf{P}\{B_i\} \cdot \mathsf{P}\{A|B_i\} / \left(\sum_{j=1}^{n} \mathsf{P}\{B_j\} \cdot \mathsf{P}\{A|B_j\}\right).$$

Более того, этот принцип Лапласа содержит и формулу полной вероятности, которой с начала XVIII века широко пользовались в своих работах многочисленные математики, работавшие в области теории вероятностей. Они понимали как использовать принцип, заложенный в формуле полной вероятности, но его не формулировали.

Мы видим, таким образом, что основные принципы действия с вероятностями вычленялись длительным путем. Их многократно использовали при решении отдельных задач и использовали правильно, но не формулировали их в качестве особых предложений. И потребовалось почти целое столетие, чтобы после введения в науку понятия вероятности сформулировать для этого понятия систему правил действия с ним. Как постоянно происходит в истории науки, такие правила широко использовались фактически, но потребности в их формулировании не ощущали. Попутно при этом вводились и дополнительные понятия, которые позволяли глубже вникать в природу вещей. В нашем случае этими понятиям являются понятия несовместимости и независимости случайных событий.

§ 10. Задача о разорении игрока

Несомненно, что задача о разорении игрока в развитии теории вероятностей играла серьезную роль – она позволяла оттачивать методы решения сложных вопросов и в какой-то мере является исходным пунктом для развития теории случайных процессов. Действительно именно в этой задаче впервые начали изучать состояние системы в зависимости от времени. Точнее – положение игроков после заданного числа партий. Задача о разорении игрока была впервые сформулирована Гюйгенсом в книге "О расчетах в азартных играх" (см. § 5 первой главы настоящего очерка, задача 5). Этой задачей занимались многие выдающиеся математики прошлого – Я. Бернулли, Н. Бернулли, Муавр, Лаплас и др. Интересно отметить, что Я. Бернулли критиковал Гюйгенса за то, что тот решал и предлагал трудные задачи, но не в буквенной форме, а в числовом виде и тем самым ограничивал возможности выявления общих закономерностей.

Первые подходы к решению задачи о разорении игрока почти одновременно были предложены тремя математиками – П. Монмором (1687–1719), А. Муавром и Н. Бернулли (1687–1759). Их результаты относились к 1710–1711 г. Задачи Гюйгенса в их формулировке слегка преобразилась и приобрела привычный для нас вид: игроки А и В имеют соответственно a и b франков и при каждой партии некоторой игры один из них выигрывает у другого 1 франк. Вероятность выигрыша игрока А для каждой партии равна p, для игрока В вероятность выигрыша равна $q = 1 - p$. Спрашивается, чему равны вероятности p_a и p_b того, что игрок А выиграет (соответственно игрок В) игру (т.е. игрок А выиграет все деньги В раньше, чем В выиграет их у А).

Муавр опубликовал свои результаты в журнале Philosophicol Transactions за 1711 г. Он нашел, что

$$p_a = \frac{(q/p)^a - 1}{(q/p)^{a+b} - 1}, \quad p_b = \frac{(p/q)^b - 1}{(p/q)^{a+b} - 1}$$

и что математическое ожидание числа N необходимых для завершения игры партий равно

$$\mathbf{M}N = \frac{bp_a - ap_b}{p - q}.$$

Ему же удалось найти вероятности $p_{a,n}$ и $p_{b,n}$ того, что игрок А выиграет игру за n партий (соответственно выиграет игру за n партий игрок В). В современных обозначениях искомая формула имеет следующий вид:

$$p_{a,n} = \sum_t \{p^{ts-b} q^{ts} \sum_i c_n^i (p^{n-b-2ts-i} q^i - q^{n-s-2ts-i} p^i)\} -$$
$$- \sum_t \{p^{ts-i} q^{ts+i} \sum_i c_n^i (p^{n-b-2ts-2a-i} q^i - q^{n-b-2ts-2a-i} p^i)\}.$$

Здесь введено обозначение $s = a + b$; суммирование распространяется на те значения t, при которых все показатели неотрицательны.

Вдобавок им был подробно рассмотрен случай, когда $a = \infty$.

В 1710 г. формулы для $p_{a,n}, p_{b,n}$ в случае $p = q$ нашел Монмор. Свои соображения он переслал Иоганну Бернулли, который передал письмо своему племяннику Николаю. Ответное письмо Николая Бернулли от 26 февраля 1711 г. содержало решение и для случая $p \neq q$. Это письмо Монмор опубликовал в 1713 г. в трактате "Опыт анализа азартных игр" (P. Monmort, Essai d'analyse jeux asard).

Я. Бернулли также рассматривал задачу о разорении игрока, как в частных случаях (для $a = b = 2$), так и в общем случае. При ее решении он следовал методу Гюйгенса и получил довольно далеко идущие результаты (для вероятностей p_a и p_b).

Рассмотрение решений, предложенных Я. Бернулли, Н. Бернулли, Монмором и Муавром, ясно показывают, что все они владели приемами оперирования с вероятностями сложных событий. Практически они безукоризненно точно использовали теоремы сложения и умножения вероятностей, а также формулу полной вероятности, хотя в ту пору они еще не получили четкой формулировки. Происходило накопление опыта и выделение тех правил, которые постоянно необходимы при подсчете вероятностей сложных событий.

§ 11. Возникновение предельных теорем теории вероятностей

На последующее развитие теории вероятностей огромное воздействие оказала идея, впервые высказанная и осуществленная Я. Бернулли – рассматривать не только точные решения задач теории вероятностей, но и их асимптотические постановки при неограниченном увеличении некоторого параметра. Конечно, в первую очередь следует указать в этом плане на закон больших чисел в форме Я. Бернулли. Именно он послужил источником для различного рода уточнений как в XVIII, так и в последующие столетия.

Сам Я. Бернулли дал формулировку своей теоремы в отличном от принятого теперь виде. Мы приведем его формулировку несколько позднее. Сейчас же отметим, что и принятая им терминология отлична от современной и связана с демографией. Так, Я. Бернулли использовал для обозначения испытаний, при которых интересующее нас событие происходит, слова "плодовитый", "фертильный", а для противоположных исходов – слово "стерильный". Теперь мы можем перейти к оригинальной формулировке теоремы Я. Бернулли, которую он ценил и вынашивал по его словам свыше двадцати лет.

"Пусть число фертильных случаев к числу стерильных случаев относится точно или приближенно как $\frac{r}{s}$ или же это число относится к числу всех случаев как $\frac{r}{r+s}$ или же как $\frac{r}{t}$ *). Последнее отношение находится, следовательно, между $\frac{r-1}{t}$ и $\frac{r+1}{t}$. Нужно доказать, что можно произвести столь большое число опытов, что число поя-

---

*) Я. Бернулли счел излишним говорить, что $t = r + s$. Заметим также, что r, s и t не фиксированы, а могут принимать любые значения, лишь бы отношение $\frac{r}{t}$ имело заданное значение. Отсюда следует, что $\frac{1}{t}$ может быть сделано как угодно малым.

вившихся фертильных наблюдений к числу всех опытов будет больше, чем $\frac{r-1}{t}$, и меньше, чем $\frac{r+1}{t}$."

Ясно, что эта формулировка лишь словесно отличается от принятой теперь.

Мы уже говорили, что книга "Искусство предположений" Я. Бернулли была широко известна многим математикам задолго до ее публикации. В частности, она была тщательно изучена его племянником Н. Бернулли, который в 1709 г. защитил диссертацию для получения ученой степени лиценциата прав под названием "О применении искусства предположений в вопросах прав". Во второй главе "О способе установления вероятностей человеческой жизни", исходя из таблиц Граунта, он изучал вопрос о вероятности дожития до определенного возраста. Нам сейчас интересно отметить, что он отметил факт, подмеченный из изучения долголетних регистраций рождений, что мальчиков рождается больше, чем девочек. При этом отношение числа рождений мальчиков к числу рождения девочек оказывается, как он считал, равным 18 : 17. Подробное изучение содержания этой главы показывает, что Н. Бернулли принимал вероятность рождения мальчика равной $p_М = 18 : 35 \approx 0,514$ и соответственно вероятность рождения девочки равной $p_Д = 17 : 35 \approx 0,486$.

Далее Н. Бернулли рассмотрел пример, когда имеется 14000 рождений. Тогда, согласно формулам Я. Бернулли, имеет место равенство (μ означает фактическое число рождений мальчиков)

$$\mathbf{P}\{|\mu - 7200| < 163\} = \mathbf{P}\{7037 < \mu < 7363\} = \sum_{i=7038}^{7362} C_{14000}^{i} p^{i} q^{14000-i}.$$

Фактическое число рождений мальчиков зависит от случая. Приведенная формула позволяет вычислять вероятность того, что число рождений мальчиков будет заключено в указанных границах. Однако вычисления, которые при этом необходимо произвести, сложны.

Интересно, что в точности этот пример рассмотрен Лапласом в "Аналитической теории вероятностей" (1-е изд., с. 281). В качестве искомого значения вероятности неравенства $7037 < \mu < 7363$ Лаплас указал величину 0,994303.

В двух последних изданиях книги Муавра "Доктрина шансов" был помещен перевод его статьи 1733 г. "Approximation ad summum terminonum Binomii $(a + b)^n$ in serien expansis". Согласно словам самого автора "Я помещаю здесь перевод моей работы, написанной 12 ноября 1733 года и сообщенной некоторым друзьям, но никогда не публиковавшейся" ("Доктрина шансов", 1756, с. 242). В кратком введении Муавр отметил, что для решения ряда задач теории вероятностей необходимо подсчитывать суммы $\sum_{m=l}^{k} P_n(m)$ членов биномиального распределения и что вычисления становятся громоздкими при больших значениях числа испытаний n. В результате перед Муавром возник вопрос о разыскании асимптотической формулы. Эта задача им была благополучно решена. Основная трудность, которая при этом возникла, состояла в оценке факториала $m!$ при больших значениях m. Муавру удалось доказать, что имеет место асимптотическое равенство $m! \approx B\sqrt{m}\,e^{-m} m^m$, где B – постоянное. При этом оказалось, что

$$\ln B = 1 - 1/12 + 1/360 - 1/1260 + 1/1680 - \ldots$$

Муавр нашел, что приблизительно $B \approx 2,5074$, однако это его не удовлетворило и ему хотелось связать эту константу с ранее введенными в математику. Он обратился со своей проблемой к Д. Стирлингу (1692–1770). Стирлинг с успехом разрешил вопрос и показал, что $B = \sqrt{2\pi} \approx 2,506628\ldots$ В связи со сказанным хотелось бы отметить, что известную формулу Стирлинга для приближенного вычисления факториала в случае больших чисел следовало бы называть точнее формулой Муавра и, самое меньшее – формулой Муавра – Стирлинга. Заметим дополнительно, что Муавр впервые вычислил и опубликовал таблицу функции $\ln n!$ для значений n от 10 до 900.

Использовав найденную им формулу "Стирлинга", Муавр первоначально выяснил, что в случае $p = q = 0,5$ средний член бинома $(1/2 + 1/2)^n$ асимптотически равен $1/\sqrt{2n\pi pq}$, а затем доказал локальную теорему, носящую теперь его имя ("Доктрина шансов", с. 243–244). То, что Муавр начал со случая $p = q = 0,5$ вполне естественно, поскольку именно этот случай играет значительную роль в простейших задачах демографии. Далее Муавр получил локальную теорему для $p \neq 0,5$ фактически в принятом теперь виде.

Имея в руках локальную теорему, Муавр без затруднений сформулировал и интегральную теорему, правда, только для симметричных границ. Впрочем, интегральная теорема, доказанная для симметричных границ, без труда распространяется и на общий случай. Он оценил важность выражения $\sqrt{npq}$ для теории и предложил для него специальное наименование – модуль.

Использовав метод приближенного интегрирования Ньютона – Котса, Муавр вычислил для случая $p = q = 0,5$ вероятность

$$\mathbf{P}\left\{\frac{1}{2}n - \sqrt{n} < \mu < \frac{1}{2}n + \sqrt{n}\right\}.$$

Согласно его подсчетам, она оказалась равной 0,95428. Теперь, используя таблицы, несложно проверить его расчеты и убедиться в том, что допущенная им ошибка невелика, только в четвертой значащей цифре (табличное значение равно 0,95450). Точно так же он подсчитал вероятность

$$\mathbf{P}\left\{\frac{1}{2}n - \frac{3}{2}\sqrt{n} < \mu < \frac{1}{2}n + \frac{3}{2}\sqrt{n}\right\}.$$

Его результат – 0,99874. Табличное значение с таким же числом значащих цифр – 0,99731.

Муавр отметил, что интегральную теорему можно использовать и для оценки неизвестной вероятности p, т.е. для решения обратной задачи – задачи математической статистики.

§ 12. Контроль качества продукции

В связи с переходом промышленности на массовое изготовление изделий, за последние пятьдесят – шестьдесят лет резко увеличился интерес к вопросам проверки качества изделий, входящих в принимаемую партию. Появилась глубокая по содержанию и значительная по своим практическим применениям теория статистических методов приемочного контроля, основанная на широком использовании теории вероятностей.

Первым шагом, относящимся к этому кругу идей, по-видимому, следует считать одну из задач, рассмотренных Т. Симпсоном в книге "Природа и законы случая" (1740). Вот формулировка этой задачи: имеется данное число вещей различного сорта – n_1 вещей первого n_2 – второго, ... Наудачу берутся m вещей. Найти вероятность того, что при этом будет взято m_1 вещей первого сорта, m_2 вещей второго и т.д. В настоящее время эта задача не представляет труда для студентов, приступающих к изучению основ теории вероятностей. В ту пору она была предметом серьезного научного трактата.

Спустя сто с небольшим лет, к этой задаче вновь вернулся М.В. Остроградский (1801–1862) в работе "Об одном вопросе, касающемся вероятностей" (1846). В математическом отношении это произведение Остроградского не представляет большого интереса, но глубокое понимание самой практической задачи заслуживает нашего внимания. По-видимому, в этом отношении он имеет приоритет перед всеми исследователями. Во всяком случае Симпсон практических следствий из своих подсчетов не делал, а Остроградский вычислил и необходимые для практических применений таблицы. Приведем подлинные слова Остроградского.

"В сосуде имеются белые и черные шары, общее количество которых нам известно, но мы не знаем, сколько из них какого цвета. Мы извлекаем некоторое количество шаров, подсчитав, сколько из них белых и сколько черных, снова кладем в сосуд. Требуется определить вероятность того, что общее число белых не выходит из наперед заданных пределов. Или, лучше сказать, мы ищем зависимость между этой вероятностью и пределами, о которых идет речь.

Чтобы понять важность этого вопроса, представим себя на месте того, кто должен получить большое число предметов, причем должны выполняться некоторые условия, и кто, чтобы проверить эти условия, должен на каждый предмет потратить некоторое время. Перед армейскими поставщиками часто стоят такого рода задачи. Для них шары, содержащиеся в сосуде, представляют получаемые предметы, белые, например – предметы приемлемые, как удовлетворяющие определенным условиям, а черные – неприемлемые.

Таким образом, если бы вопрос, который мы перед собой поставили, был решен, поставщик мог бы воспользоваться этим, чтобы свести приблизительно к двадцатой доле часто очень утомительную механическую работу, как, например, проверку большого количества мешков муки или штук сукна".

Общее число шаров в урне известно, но неизвестен ее состав. Его и следует оценить по выборке, взятой из урны наудачу. Для этой цели Остроградский использует формулы Байеса. Однако его рассуждения базируются на одном предположении, которое вызывает серьезные возражения, поскольку в реальной практике не может встретиться. Именно, он предположил, что если n – общее число шаров в урне, то одинаково вероятны все гипотезы о распределении среди них белых и черных шаров, т.е. что одинаково вероятны все следующие $n + 1$ предположения о числе белых шаров: $0, 1, 2, \ldots, n$. В действительности ближе к истине предположение, которое используется теперь в задачах приемочного статистического контроля качества. Предполагается, что имеется причина, в силу которой каждое изделие может оказаться бракованным с вероятностью p. Для хорошо организованного производства p должно быть малым и практически неизменным. Если же технологический процесс налажен плохо, то вероятность p зависит от времени и может достигать большой величины. Но очевидно, что в этом случае статистический контроль не может принести пользы. Если же p мало и постоянно, то вероятность среди n изделий встретить m бракованных задается формулой Бернулли

$$P_n(m) = C_n^m p^m q^{n-m}.$$

Статистические методы приемочного контроля получили особенно бурное развитие в годы Второй мировой войны, поскольку было необходимо принимать огромные партии однородной продукции, а проверять ее сплошь не было возможностей по ряду причин, из которых укажем лишь на следующие: для некоторых видов продукции проверка равносильна уничтожению рабочих свойств изделия (фотобумага, взрыватели и т.д.); для сплошной проверки требовалось такое количество рабочей силы и рабочего времени, что ни то, ни другое не могло быть обеспечено в условиях военного времени. Нет возможности здесь перечислить даже основные этапы развития теории статистических методов приемочного контроля. Большое число исследователей работали над различными проблемами этой тематики и внесли в ее развитие крупный вклад. Из отечественных ученых заслуживают быть отмеченными А.Н. Колмогоров, В.И. Романовский (1879–1954), С.Х. Сираждинов (р. 1921), Ю.К. Беляев (р. 1932) и др.

В период Великой Отечественной войны 1941–1945 г. огромное внимание было уделено разработке методов управления качеством продукции в процессе производства. Это весьма важное направление работы, поскольку оно должно не просто разбраковывать изготовленную продукцию, но своевременно вмешиваться в производственный процесс и не допускать изготовления некачественной продукции. Именно к этому и должно стремиться любое производство. Итак, для управления качеством нужно разработать методы, которые позволяли бы поймать тот момент, когда бракованная продукция еще не производится, но возникает повышенная вероятность начала ее производства.

Идея статистического метода управления качеством в процессе производства состоит в том, чтобы время от времени проверять небольшие партии продукции (5–10 штук), только что сошедших со станка. По результатам таких проверок судят о качестве наладки. Эти проверки осуществляются не слишком часто, чтобы не лихорадить переналадками оборудования производственный процесс, и не слишком редко, чтобы не пропустить момент его разладки. Далее результаты наблюдений наносятся на так называемые контрольные карты, которые позволяют судить, что следует предпринимать после каждой серии таких наблюдений – прекращать работу для переналадки оборудования или же продолжить производственный процесс.

Если на ряде производств первичное произведение замеров параметров, определяющих качество продукции, допустимо и теперь оценивать вручную, то на других производствах оно уже требует заметного усовершенствования и перехода к автоматизации замеров и обработки результатов измерений. Дело в том, что во многих случаях приходится иметь дело с огромной скоростью технологических операций. Скорость настолько велика, что пока оператор производит измерение параметров отобранных изделий, автомат успевает изготовить сотни других изделий, а прокатный стан – прокатать сотни метров продукции. В результате при ручном измерении оказывается, что запаздывает информация о наладке процесса, а вместе с ней и управляющее воздействие. Вот почему теперь предложены автоматы, которые замеряют необходимые параметры и сами выполняют математические операции, необходимые для управления качеством.

Методы приемочного контроля и статистические методы управления качеством оказались весьма эффективным средством упорядочения производства и экономии станочного времени, материалов, рабочей силы. Экономический эффект от использования этих методов исчисляется миллиардами рублей (долларов, марок и т.д.). По-настоящему история статистических методов контроля и управления еще недостаточно изучена и нуждается в энтузиастах.

Широта применения теории вероятностей в начале двадцатого века настойчиво требовала глубокого осмысления самих основ этой науки. Наивные представления о случайном событии и его вероятности уже перестали удовлетворять запросы науки. Вот почему ряд ученых – Э. Борель, С.Н. Бернштейн, Р. Мизес и другие предприняли серьезные усилия в пересмотре основ теории вероятностей. В тексте книги мы познакомились с широко распространенным подходом к этой проблеме, предложенным А.Н. Колмогоровым в статье 1927 г. и нашедшем завершение в его монографии "Основные понятия теории вероятностей" (1933).

ГЛАВА 3

К ИСТОРИИ ФОРМИРОВАНИЯ ПОНЯТИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

§ 13. Развитие теории ошибок наблюдений

Мы уже упоминали в первой главе о том, что Галилео Галилей заложил основы теории ошибок измерений и ввел в рассмотрение ряд важных понятий, которые сохранили значение и в наши дни.

Позднее под влиянием в первую очередь астрономических и геодезических наблюдений интерес к ошибкам измерений заметно возрос. Знаменитый астроном-наблюдатель Тихо Браге (1546–1601) обратил внимание на то, что каждое отдельное измерение несет в себе возможную ошибку и точность измерений значительно повышается, если произвести несколько измерений и взять из них среднее арифметическое. Впрочем рекомендация пользоваться средиим арифметическим для уточнения размеров была дана задолго до Тихо Браге. Так, согласно литературным данным, в одном древнеиндийском математическом трактате рекомендовалось при подсчете объема земляных работ делать измерения в нескольких местах и затем оперировать со средними арифметическими (см. Л. Е. Майстров "Развитие понятия вероятности", с. 74).

Казалось бы от И. Кеплера (1571–1630), сделавшего так много для формирования законов движения планет, следовало ожидать повышенного внимания к методам обработки результатов наблюдений. Но эти вопросы фактически остались в стороне от его интересов и он заметил только то, что хороший наблюдатель производит измерения с ошибками ограниченной величины. В этом плане интересны его слова, относящиеся к Тихо Браге "благость божья дала нам в лице Тихо столь точного наблюдателя, что ошибка в восемь минут невозможна, поблагодарим бога и воспользуемся этой выгодой. Эти восемь минут, которыми пренебречь нельзя, дадут средство преобразовать всю астрономию".

Первые попытки построить математическую теорию ошибок измерений принадлежат Р. Котсу (1682–1716), Т. Симпсону (1710–1761) и Д. Бернулли (1700–1782).

Предположения, которые были высказаны указанными авторами о закономерностях распределения ошибок измерения, были весьма различны. Котс считал, что ошибки равномерно распределены в некотором отрезке $(-a, a)$. Симпсон исходил из предположения, что малые ошибки допускаются чаще, чем большие и также ограничены по абсолютной величине некоторым числом a. Симпсон считал, что ошибки подчинены треугольному распределению, плотность которого равна 0 в отрезках от $-\infty$ до $-a$ и от a до $+\infty$; в отрезке $(-a, 0)$ ее уравнение будет $x - 2a^2 y = -a$ и, наконец, в отрезке $(0, a)$ имеет уравнение $x + 2a^2 y = a$. Следует заметить, что как Котс, так и Симпсон не рассматривали в сущности плотности распределения, поскольку они считали, что ошибки укладываются в арифметическую прогрессию с очень малой разностью и неопределенным числом возможных значений.

Симпсон для избранного им распределения доказал, что среднее арифметическое дает лучшую точность, чем каждое отдельное измерение. Этому результату он придавал большое значение и опубликовал его в работе "О преимуществе выбора среднего из некоторого числа наблюдений в практической астрономии" (Philos, Trans., 1755).

Значительный интерес представляет работа Д. Бернулли (1700–1782) "Наиболее вероятное определение по нескольким расходящимся между собой наблюдениям и устанавливаемое отсюда наиболее правдоподобное заключение", опубликованная в 1778 г. в изданиях Петербургской Академии наук. Эта работа интересна тем, что в ней впервые был высказан и использован для оценки неизвестного параметра принцип максимального правдоподобия. Д. Бернулли начал свой мемуар с сомнений о целесообразности применения всеобще принятого принципа среднего арифметического. В качестве плотности распределения он принял функцию, определенную равенством $y=\sqrt{R^2-(x-\bar{x})^2}$, в котором параметр R известен, а $\bar{x}$ должно быть определено по результатам наблюдений. Заслуживает внимания то, что Д. Бернулли не обратил внимания на следующее обстоятельство: интеграл от принятой им плотности распределения равен $\frac{\pi}{2}R^2$ и, следовательно, только при одном значении R может быть плотностью распределения вероятностей.

К этой работе Д. Бернулли был написан комментарий Л. Эйлером (1707–1783), в котором, во-первых, критиковался метод максимального правдоподобия (конечно, тогда этого термина еще не было и в помине) и, во-вторых, предлагалось отбрасывать наблюдения, далекие от истинного значения параметра, поскольку они мало вероятны.

Следует отметить работы И. Ламберта (1728–1777), который в статьях 1760 и 1765 г. изложил цели теории ошибок измерений (кстати, ему принадлежит и сам этот термин), свойства ошибок, оценку точности наблюдений и правила подбора кривых по наблюденным точкам, содержащим случайные ошибки. Позднее появилась работа Ж. Лагранжа (1736–1813), посвященная выяснению роли среднего арифметического при оценке истинного значения измеряемой величины.

Для П. Лапласа (1749–1827) теория вероятностей была не столько математической, сколько естественно-научной дисциплиной. В связи с его занятиями астрономией, он неизбежно должен был прийти к вопросам теории ошибок наблюдений и вместе с тем заинтересоваться теорией вероятностей. Лаплас получил ряд важных результатов в теории ошибок наблюдений, которые вошли в практику обработки данных наблюдений. Мы не станем здесь вдаваться в подробности его исследований, поскольку для нас описание теории ошибок измерений не является самоцелью. Нас интересует ее связи с развитием теории вероятностей. В этом плане особый интерес представляют две идеи П. Лапласа. Первая из них вызвала к жизни значительное увеличение интереса к предельным теоремам для сумм независимых случайных величин. Именно, согласно Лапласу, наблюденные ошибки измерений являются результатом суммирования очень большого числа элементарных ошибок. Если эти ошибки равномерно малы, то Лаплас предположил, что распределение их суммы должно быть близко к нормальному.

Вторая идея касается оценки измеряемой величины по результатам $x_1, x_2, \ldots, x_n$ измерений. В качестве оценки неизвестного значения a измеряемой величины Лаплас предложил брать то значения $\hat{a}=a(x_1,\ldots,x_n)$, при котором обращается в минимум сумма $\sum_{k=1}^{n}|x_k-a|$. Оказывается, что $\hat{a}$ при этом равняется эмпирической медиане, т.е. тому значению x, слева и справа от которого расположено одинаковое число наблюденных значений. Этот прием не получил в ту пору распространения, поскольку вскоре был предложен другой метод, приводящий к более простым результатам. Разработка этого нового метода связана с именами Гаусса (1777–1855), Лежандра (1752–1833) и американского математика Р. Эдрейна (1775–1843). Их работы составили в теории ошибок наблюдений настоящую эпоху. Гауссом и Лежандром был предложен и разработан метод наименьших квадратов. Гаусс предложил его второй книге большого трактата "Теория движения небесных тел, вращающихся вокруг солнца по коническим сечениям" (1809). Лежандр же изложил свои идеи в работе "Новые методы для определения орбит комет" (1806), к которой было сделано специальное дополнение "О методе наименьших квадратов". Сам Гаусс неоднократно писал, что он пользовался этим методом, начиная с 1795 г. Гауссом и Эдрейном было показано, что при некоторых весьма широких условиях плотность ошибок измерений имеет вид $\varphi(\Delta) = \frac{h}{\sqrt{\pi}} \exp(-h^2 \Delta^2)$.

Необходимо сказать, что влияние Гаусса, Лежандра и Эдрейна на развитие науки оказалось весьма различным. Статья Эдрейна, опубликованная в мало распространенном американском журнале, прошла практически незамеченой. Работы же Гаусса и Лежандра почти мгновенно стали известны научному миру. Ученые восприняли предложенный ими метод и начали систематически использовать его в своей практической работе.

Большой вклад в дальнейшее развитие этой теории внес С. Пуассон (1781–1840). В частности, Пуассон задался вопросом: всегда ли среднее арифметическое дает лучший результат по сравнению с отдельным наблюдением? Ответ оказался отрицательным. Именно: ему удалось указать распределение, для которого этого правило ошибочно. Плотность этого распределения равна $p(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.

Пуассон обнаружил, что сумма двух независимых случайных величин с только что указанной плотностью распределения имеет с точностью до масштаба такое же распределение. Далее он обнаружил, что среднее арифметическое из независимых наблюдений над такой случайной величиной имеет в точности такое же распределение. Через двадцать лет (в 1853 г.) О. Коши (1789–1857) повторил эти результаты, после чего указанное распределение получило наименование распределения Коши. Пуассон же, первооткрыватель этих результатов, был забыт.

Позднее теория ошибок измерений привлекала внимание практически всех видных специалистов в области теории вероятностей. П.Л.Чебышев и А.А. Марков (1856–1922) и многие другие уделяли внимание как методу наименьших квадратов, так и другим вопросам теории ошибок. Теория ошибок оказала серьезное влияние на постановки задач и разработку методов математической статистики. Теперь теория ошибок включается в качестве естественной части в математическую статистику.

§ 14. Формирование понятия случайной величины

Мы неоднократно говорили о том, что формирование научных понятий проходит длительный и сложный путь, прежде чем войти во всеобщее употребление. Как правило, необходимое понятие еще не введено в научный обиход, а фактически им уже пользуются как при решении практических задач, так и при выводе общетеоретических закономерностей. Этот путь характерен и для случайной величины – основного понятия теории вероятностей и современного естествознания. Введение этого понятия связано с именами многих ученых, которые хотя и не использовали этого термина, но фактически исследовали отдельные его свойства.

Действительно, мы уже знаем, что,начиная с Котса, Симпсона и Д. Бернулли, в XVIII веке начала развиваться теория ошибок наблюдений, возникшая в первую очередь под влиянием астрономии. Ошибка измерения в зависимости от случая может принимать различные значения. Эта позиция была высказана Галилеем задолго до работ только что упомянутых ученых. Он же ввел в обиход термин "случайная" и "систематическая ошибка" измерения. Вторая тесно связана с качеством изготовления прибора, мастерством наблюдателя, условиями наблюдений. Первая же зависит от многочисленных причин, влияние которых невозможно учесть и которые изменяются от наблюдения к наблюдению, от измерения к измерению. Теперь мы ясно видим, что ошибка измерения представляет собой случайную величину с каким-то неизвестным нам распределением вероятностей.

Но с понятием случайной величины встречались уже Я. Бернулли, Н. Бернулли, Монмор, Муавр. В самом деле, Я. Бернулли рассмотрел число появлений интересующего его события A в n независимых испытаниях. Для нас теперь это случайная величина, способная принимать значения 0, 1, 2, . . . , n с вероятностями, задаваемыми формулами Бернулли. Н. Бернулли, Монмор и Муавр, исследуя задачу о разорении игрока, также имели дело со случайной величиной – числом партий, которые необходимы для разорения. Муавр пошел еще дальше – он ввел в рассмотрение нормальное распределение вероятностей. Однако никто из перечисленных ученых не заметил, что в науку властно постучалась необходимость введения нового понятия – случайной величины. Первый из них оставался на уровне схемы последовательности случайных событий, остальные же ограничились той частной задачей, которая перед ними стояла. Для Муавра нормальное распределение было лишь аппроксимирующей функцией, дающей хорошее приближение к точному значению искомых вероятностей.

Мы говорили, что первоначально считалось, что возможные значения ошибок измерений составляют арифметическую прогрессию с неопределенной, но очень малой разностью. Затем постепенно от этого предположения отказались и стали представлять себе, что возможные значения, принимаемые ошибками наблюдений, заполняют целый отрезок, вероятности возможных значений определялись посредством плотности распределения. И если Д. Бернулли в отношении плотности распределения вероятностей допускал еще определенные вольности, то у Лапласа, Гаусса, Лежандра с плотностью распределения было уже все в порядке. Это была неотрицательная функция, интеграл от которой по всей прямой равен единице, а вероятность попадания в тот или иной отрезок равнялся интегралу от плотности, взятому по этому отрезку. Лапласу уже была известна формула для разыскания плотности распределения суммы по плотностям распределения слагаемых. В знаменитой книге "Аналитическая теория вероятностей" Лаплас умело оперирует с плотностями распределения, ставит и решает ряд интересных задач, но нигде не вводит понятия случайной величины. Он либо использует язык теории ошибок измерений, либо язык математического анализа и не ощущает потребности в новом понятии теории вероятностей.

Первая половина XIX века принесла новые задачи, которые нуждаются в понятии случайной величины. Прежде всего – это исследования бельгийского естествоиспытателя А. Кетле (1796 – 1874), заметившего, что размеры органов животных определенного возраста подчиняются нормальному распределению. Изучение уклонений снаряда от цели явилось предметом исследования многих ученых; они также привели к выводу о нормальном распределении этой величины. С середины XIX века начались замечательные работы Д.К. Максвелла (1831 – 1879) и ряда других ученых по математической теории молекулярной физики газов. И здесь снова нормальное распределение завоевало почетное место.

Заслуживает внимания постановка еще одной задачи Гауссом. Он сформулировал ее 25 октября 1800 г. (именно за этот день в его дневнеке под № 113 сделана соответствующая запись). Через двенадцать лет он сформулировал ее в письме к Лапласу от 30 января 1812 г. Эта задача относится к интересному, начавшему развиваться лишь в нашем веке разделу математики – метрической теории чисел; одновременно она имеет самое непосредственное отношение к изучению равномерно распределенных случайных величин. В постановке задачи предоставим слово самому Гауссу. В упомянутом письме к Лапласу он писал: ". . . я вспоминаю любопытную задачу, которой я занимался уже 12 лет назад, но для которой я не нашел тогда удовлетворяющего меня решения.

Быть может, Вы соблаговолите заняться ею несколько минут: в этом случае, я убежден, Вы найдете более полное решение. Вот она. Пусть M – неизвестная величина, заключенная между пределами 0 и 1, для которой все значения или одинаково вероятны или же более или менее следуют данному закону: предположено, что она разложена в непрерывную дробь $M = 1/a^{(1)} + 1/a^{(2)} + \ldots$. Чему равна вероятность того, что, отбросив в разложении конечное число членов до $a^{(n)}$, следующая дробь $1/a^{(n)} + 1/a^{(n+1)} + \ldots$ будет заключена в пределах от 0 до x? Я обозначаю ее через $P(n, x)$ и предполагаю, что для M все значения одинаково вероятны: $P(0, x) = x$."

Гипотеза Гаусса состояла в том, что

$$\lim_{n \to \infty} P(n, x) = \ln(1 + x)/\ln 2.$$

Он писал далее, что при решении этой задачи "усилия, которые я делал, . . . оказались бесплодными". Решение этой задачи появилось лишь в 1928 г., его дал Р.О. Кузьмин (1891 – 1949). Через год П. Леви (1886 – 1971) дал для этой задачи чисто вероятностное решение, получив для быстроты сходимости к пределу лучшую, чем у Кузьмина, оценку. Позднее было доказано, что результат сохраняется для любой случайной величины M, для которой $P(0, x)$ имеет ограниченную производную. Это замечание делает более ясными неопределенные слова Гаусса о том, что для величины M "все значения или одинаково вероятны или же более или менее следуют данному закону".

Заслуживает упоминания то обстоятельство, что функция $P(0, x)$, также как и $P(n, x)$, представляет собой функцию распределения.

Мы видим, что многочисленные исследования многих крупных математиков подготовили почву для введения понятия случайной величины. По-видимому, первый шаг был сделан Пуассоном в мемуаре 1832 г. "О вероятности средних результатов наблюдений". Этот факт мне сообщил О.Б. Шейнин. Термина случайная величина у Пуассона еще нет, но он пишет о "некоторой вещи", которая способна принять значения $a_1, a_2, \ldots, a_\lambda$ соответственно с вероятностями $p_1, p_2, \ldots, p_\lambda$. Он рассмотрел также непрерывные случайные величины и их плотности распределения.

Итак, Пуассоном был сделан важный шаг в науке – он ввел в научный обиход новое понятие – случайную величину. Его первоначальный термин "вещь" не привился и скоро перестал употребляться. Чебышев в своих мемуарах по теории вероятностей уже использует термин "величина" и автоматически считает все случайные величины, с которыми имеет дело, независимыми. В работах же Ляпунова по теории вероятностей систематически используется термин "случайная величина" и всюду, где это необходимо, оговаривается, что автор имеет дело с независимыми случайными величинами.

Отметим еще одно обстоятельство. В самом начале § 4 работы "Об одном предложении теории вероятностей" Ляпунов определил функцию распределения точно также, как мы делаем это теперь. Он привел в этой работе широко используемую формулу

$$\mathbf{P}\{a \leqslant \xi \leqslant b\} = F(b) - F(a).$$

Полезно заметить, что в трактатах по теории вероятностей А. Пуанкаре (1854 – 1912) "Исчисление вероятностей", Э. Бертрана "Исчисление вероятностей", Чубера "Теория вероятностей и математическая статистика" понятие функции распределения не вводилось (книги, изданные до 1912 г.)

Определение случайной величины, данное Пуассоном, теперь уже не может считаться математическим. Это скорее описание реального объекта изучения, обращение к интуиции, полученной в результате житейского и научного опыта. Это описание широко используется и в наши дни на начальных ступенях педагогического процесса, связанного с изложением основ теории вероятностей. Даже несложный логический анализ этого определения показывает, что из него совсем не следует правила для действий над случайными величинами – сложения, вычитания, умножения и пр. Для того, чтобы случайная величина приобрела статус полноценного математического понятия, ей необходимо дать строго формализованное определение. Это было сделано в конце двадцатых годов А.Н. Колмогоровым в небольшой статье, посвященной аксиоматике теории вероятностей, а затем в подробностях изложено в его знаменитой книге "Основные понятия теории вероятностей". Подход Колмогорова стал теперь общепринятым, поскольку он полноценно включил теорию вероятностей в общий стиль современного изложения, принятый в математике.

§ 15. Закон больших чисел

Знаменитая теорема Я. Бернулли о сближении при увеличении числа наблюдений вероятности события A с частотой его появления получила первое обобщение лишь в 1837 г. в работе С. Пуассона "Исследования о вероятностях в решении судебных дел уголовных и гражданских". Именно в этом мемуаре он ввел сам термин "закон больших чисел".

Пуассон рассмотрел последовательность n независимых испытаний, в каждом из которых может появиться событие A, но с вероятностью p_k, зависящей от номера испытания. Если через μ_n обозначить число появлений события A в n последовательных испытаниях, то при любом $\epsilon > 0$ имеет место соотношение: при $n \to \infty$

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \ldots + p_n}{n}\right| < \epsilon\right\} \to 0.$$

По поводу этой теоремы Пуассона в небольшой заметке 1843 г. Чебышев писал: ". . . как ни остроумен способ, употребленный знаменитым геометром, он не доставляет предела погрешности, которую пускает этот приближенный анализ, и вследствие такой неизвестности величины погрешности доказательство не имеет надлежащей строгости"(Чебышев П.Л., Собр. соч. – АН СССР, 1947 – Т. II. – С. 14). Оценку числа n, для которого при заданных ϵ и η имеет место неравенство

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \ldots + p_n}{n}\right| < \epsilon\right\} > 1 - \eta,$$

Чебышев указал в этой заметке.

Как ни интересны эти результаты, они не внесли в теорию вероятностей существенного прогресса, поскольку в идейном плане они не выходили за пределы концепции Я. Бернулли. Существенный сдвиг в этом направлении связан с работой П.Л. Чебышева "О средних величинах" (1867), опубликованной одновременно на русском и французском языках. В этой работе он перешел от рассмотрения случайных событий к случайным величинам и тем самым перенес центр тяжести интересов теории вероятностей к изучению случайных величин. Нужно заметить, что Чебышев не упоминал, что он интересуется только независимыми случайными величинами, а, согласно традициям того времени, считал, что других величин не рассматривают. Теорема Чебышева теперь излагается во всех учебниках теории вероятностей. Она неоднократно позднее служила источником обобщений.

В 1909 г. Э. Борель для $p = 0{,}5$ показал, что в случае схемы Бернулли имеет более сильное предложение, чем закон больших чисел. Именно, он доказал, а в 1917 г.

это предложение на произвольное p распространил итальянский математик Кантелли, что

$$\mathbf{P}\left\{\lim_{n\to\infty} \frac{\mu_n}{n} = p\right\} = 1.$$

Это предложение получило наименование усиленного закона больших чисел.

Широкое обобщение усиленного закона больших чисел было дано А.Н. Колмогоровым в работе 1930 г., а также в 1934 г. в его монографии "Основные понятия теории вероятностей" (1932 г.).

Необходимые и достаточные условия для усиленного закона больших чисел были найдены в ряде работ Ю.В. Прохорова 1958 – 1959 г. (см. "Об усиленном законе больших чисел", Изв. АН СССР, сер. матем. 14, 6. 1958; "Несколько замечаний к усиленному закону больших чисел", Теория вероятностей и ее применения, т. IV, вып. 2, 215–220, 1959).

В 1935 г. А.Я. Хинчин ввел новое понятие относительной устойчивости сумм, которое должно было дать максимально общую форму закона больших чисел для положительных случайных величин. Пусть $\xi_1, \xi_2, \ldots$ – последовательность неотрицательных случайных величин. Про суммы $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ говорят, что они *относительно устойчивы*, если можно найти такие положительные константы $A_n > 0$, что при любом $\epsilon > 0$ и $n \to \infty$ выполняется соотношение

$$\mathbf{P}\left\{\left|\frac{S_n}{A_n} - 1\right| > \epsilon\right\} \to 0.$$

В случае одинаково распределенных величин ξ_n Хинчину удалось найти необходимое и достаточное условие для относительной устойчивости сумм S_n. Ученик А.Я. Хинчина А.А. Бобров (р. 1912) распространил этот результат на случай разнораспределенных слагаемых.

Закон больших чисел рассматривался вплоть до 1939 г. как особая предельная теорема и рассматривался обособленно от остальных предельных теорем для сумм независимых случайных величин. В работе Б.В. Гнеденко, о которой речь пойдет в § 17, закон больших чисел был включен в общую теорию предельных теорем, когда предельное распределение имеет единственную точку роста в нуле. Точно также теоремы об относительной устойчивости сумм являются предельными для того случая, когда предельное распределение имеет единственную точку роста при $x = 1$.

Существенное расширение проблематики, связанной с законом больших чисел, было осуществлено В.И. Гливенко в работах, относящихся к 1929 – 1933 гг., когда он начал рассматривать предельные теоремы для случайных величин со значениями в функциональных пространствах. Пожалуй, вершиной его результатов является замечательная теорема о сходимости эмпирических распределений к истинной функции распределения наблюдаемой случайной величины. Теорема Гливенко, сразу же после ее опубликования, была названа Кантелли основной теоремой математической статистики.

Теорема Гливенко неоднократно обобщалась. В этом направлении работало большое число исследователей, среди которых отметим лишь фразцузских ученых Р. Форте (р. 1912) и Э. Мурье.

§ 16. Центральная предельная теорема

Теорема Муавра о сходимости распределений центрированного и нормированного числа появлений события A в n независимых испытаниях, в каждом из которых событие A может наступить с одной и той же вероятностью p, к нормальному распределению долгое время служила образцом для последующих обобщений. Пожалуй, первое обобщение принадлежит Лапласу и уже формулируется как предельная теорема для сумм независимых случайных величин ξ_k, каждая из которых равномерно распределена на отрезке $(-h, h)$. Это было сделано в работе "Mémoire sur les approximation des formules qui sont fonctions de trés grands nombres et sur leur application aux probabilités" (1809). Лаплас в нем рассматривал дискретные случайные величины с увеличивающимся числом возможных значений. Этим самым давалась аппроксимация непрерывного распределения дискретным. Лаплас доказал, что для каждого s имеет место такой результат:

$$\lim_{n\to\infty} \mathbf{P}\left\{-s \leqslant \frac{S_n}{\sqrt{n}} \leqslant s\right\} = \frac{2\sqrt{3}}{h\sqrt{2\pi}} \int_0^s e^{-\frac{x^2}{2\sigma^2}}\, dx,$$

или

$$S_n = \sum_{k=1}^{n} \xi_k, \qquad \sigma^2 = h^2/3.$$

Заслуживает внимания тот факт, что Лаплас при доказательстве этого результата использовал метод характеристических функций, который, естественно, так тогда еще не назывался.

Существенное продвижение исследований по предельной теореме связано с именем Пуассона. В знаменитом мемуаре 1837 г. "Recherches sur la probabilité des jujements . . . "он рассмотрел схему последовательности независимых испытаний с разными вероятностями появления события A в каждом из испытаний. Пусть вероятность наступления A при k-м испытании равна p_k. Пуассон доказал для этого случая локальную теорему: Если ряд $\sum_{k=1}^{\infty} p_k(1-p_k)$ расходится, то вероятность того, что в n испытаниях событие появится m раз равна

$$\mathbf{P}\{m = np - \theta c\sqrt{n}\} = \frac{1}{c\sqrt{\pi n}}\, e^{-\theta^2} - \frac{h\theta}{2c^4 n\sqrt{\pi}}\,(3 + 2\theta^2)\, e^{-\theta^2},$$

где

$$p = \frac{1}{n}\sum_{k=1}^{n} p_k, \quad c^2 = 2\sum p_k(1-p_k), \qquad h = \frac{4}{3n}\sum_{k=1}^{n}(2p_k - 1)\,p_k(1-p_k).$$

В той же книге Пуассон дал ошибочное обобщение этой теоремы на суммы произвольных независимых случайных величин, имеющих конечные дисперсии, при условии их центрирования суммами математических ожиданий и нормирования корнем квадратным из суммы дисперсий слагаемых.

Справедливости ради следует сказать, что в частном случае одинаковой распределенности слагаемых эта теорема верна, однако строгое ее доказательство пришло значительно позднее и связано с именами наших современников – Линдеберга, Феллера и Хинчина.

Заметим, что как работы Лапласа, так и работы Пуассона и всех последующих исследователей, занимавшихся центральной предельной теоремой, были непосредственно связаны с теорией ошибок измерений. И во всех работах говорилось не о сложении абстрактных случайных величин, а о сложении ошибок. По-видимому, впервые от этой традиции отошел Чебышев.

Интерес к нормальному распределению в начале XIX века возрос в связи с появлением знаменитых исследований Лежандра и Гаусса по формулировке и обоснованию метода наименьших квадратов. Ф.В. Бессель еще в 1818 г. в работе "Fundamenta Astronomiae pro anno 1755 deducta ex observationibus viri incomportabilis James Bradley in specula Grénovicensi pro annis 1750–1762 institutis" заметил, что наблюдения гринвического астронома Брэдли прекрасно укладываются в схему нормального распределения. Объяснение, которое он предложил этому факту, совпадает с идеей, которую перед этим в течение тридцати лет вынашивал Лаплас: результирующая большого числа случайных воздействий, каждое из которых оказывается малым по сравнению с суммой остальных, подчиняется общему закону и этот общий закон должен быть нормальным. Эту мысль в совершенно отчетливой форме повторил Бессель также в работе 1838 г. (Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler//Astr. Nachr. – 1838. – С. 231). Справедливости ради следует сказать, что Бессель обратил внимание на то, что это правило не является всеобщим и могут у ошибок наблюдений встречаться другие, отличные от нормального распределения. Так, если при измерении углов один источник ошибок превалирует над всеми остальными, то плотность распределения результирующей ошибки может быть $f(t) = \dfrac{1}{\pi\sqrt{a^2 - x^2}}$.

Та же концепция обоснования нормального закона, как закона распределения ошибок, дважды встречается в известном учебнике А. Пуанкаре "Calcul des probabilités" (Paris, 1912). Первый раз в конце § 140 он писал, что "ошибка, связанная с инструментом, есть результирующая очень большого числа независимых одна от другой ошибок, таких, что каждая из них приносит лишь слабую долю в результат; результирующая ошибка следует закону Гаусса". Затем в самом начале § 144 был подведен следующий итог:

"Резюмируя, предположим, что окончательная ошибка будет результирующей очень большого числа частых погрешностей, независимых друг от друга, и что нет ошибок систематических; предположим также, что эти ошибки будут иметь приблизительно один и тот же порядок величины, внося каждая в общий результат лишь незначительную долю. В этом случае результирующая ошибка следует приблизительно закону Гаусса.

Такой, мне кажется, лучший довод, который можно дать в пользу закона Гаусса."

Следует сказать, что все эти идеи носят лишь качественный характер и нуждаются в математическом оформлении и последующем доказательстве строгих теорем.

Второй толчок, который вызвал дополнительный интерес к предельным теоремам теории вероятностей, была статистическая физика, начала которой были построены в середине XIX века. Первый общий результат в этом направлении был сформулирован в 1887 г. в работе "О двух теоремах относительно вероятностей" П.Л. Чебышева. Сформулированная Чебышевым теорема звучит следующим образом:

Если математические ожидания величин $u_1, u_2, \ldots$ равны 0, а математические ожидания всех их степеней имеют числовую величину ниже какого-либо предела, вероятность того, что сумма $u_1 + u_2 + \ldots + u_n$, деленная на квадратный корень из удвоенной суммы математических ожиданий их квадратов, заключается между двумя какими-нибудь пределами t и t' с возрастанием n до ∞ имеет пределом интеграл $\frac{1}{\sqrt{\pi}} \int\limits_t^{t'} e^{-z^2} dz$.

Для доказательства этого предложения Чебышевым был разработан весьма сильный метод, получивший наименование метода моментов и являющийся одним из крупнейших достижений науки того времени. Однако в формулировке теоремы и в ее доказательстве был допущен ряд промахов, которые сразу же взялся исправить ученик П.Л. Чебышева А.А. Марков. Критика мемуара Чебышева содержится в письмах Маркова к профессору Казанского университета А.В. Васильеву, который счел необходимым опубликовать ее в 1898 г. ("Закон больших чисел и метод наименьших квадратов"). В этих письмах Марковым была совершенно строго доказана несколько исправленная теорема Чебышева:

Если S_n – последовательность сумм $u_1 + u_2 + \ldots + u_n$ и $\Phi_n(x)$ – их функции распределения, то из предположения, что при любом целом положительном k имеют место соотношения

$$\int\limits_{-\infty}^{\infty} x^k d\Phi_n(x) \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x^k e^{-x^2/2} dx$$

вытекает, что при любых a и b имеет место равенство

$$\mathsf{P}\left\{a < \frac{S_n}{\mathsf{D}S_n} < b\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} dz.$$

Метод моментов, которым работал Чебышев, торжествовал победу. Была доказана сильная и, казалось бы, окончательная теорема. Некоторую неудовлетворенность приносило только то, что для простого результата требовалось выполнение счетного множества условий. Неожиданно в нескольких публикациях А.М. Ляпунова на протяжении 1900 и 1901 гг. было обнаружено, что окончательный результат имеет место при выполнении только одного очень простого условия, которое, вдобавок, выясняло смысл тех предположений, которые должны приводить к сходимости распределений нормированных и центрированных сумм к нормальному распределению.

Сначала Ляпунов показал, что если величины имеют конечные третьи моменты $c_k = \mathsf{M}|\xi_k - a_k|^3$, $C_n = \sum\limits_{k=1}^{n} c_k$, $B_n^2 = \sum\limits_{k=1}^{n} \mathsf{D}\xi_k$ и соотношение C_n/B_n^3 при $n \to \infty$ стремится к нулю, то имеет место сходимость функций распределения сумм S_n к нормальному распределению.

На следующий год Ляпунов же обнаружил, что для окончательного результата не обязательно требовать существования третьих моментов слагаемых. Достаточно, если существуют моменты некоторого порядка $2 + \delta$, где $\delta > 0$. Ляпунов показал, что для сходимости нормированных корнем из дисперсии сумм независимых слагаемых к нормальному распределению достаточно выполнения следующего условия: пусть $c_k = \mathsf{M}|\xi_k - a_k|^{2+\delta}$, $C_n = \sum\limits_{k=1}^{n} c_k$; отношение $C_n/B_n^{2+\delta}$ должно с ростом n стремиться к 0.

Ляпунов сделал несколько большее: он оценил скорость сходимости к предельному распределению функций распределения сумм. Порядок этой оценки оказался равным $n^{-1/2}\ln n$.

Точно также в упомянутой статье Чебышева, помимо предложения о сходимости к нормальному распределению, было дано асимптотическое разложение по степеням $\sqrt{n^{-1}}$.

Общность результатов Ляпунова произвела огромное впечатление на современников. По-видимому, именно в ту пору и появился термин "центральная предельная теорема" для обозначения условий сходимости функций распределения нормированных и центрированных математическими ожиданиями сумм к нормальному распределению. Марков подошел к результатам Ляпунова с далеко иных позиций. В связи с этим полезно привести подлинные слова Маркова: "Общность выводов в последней работе Ляпунова далеко превзошла ту, которая была достигнута методом математических ожиданий. Достигнуть столь общих выводов методом математических ожиданий казалось даже невозможным, ибо он основан на рассмотрении таких математических ожиданий в неограниченном числе, существование которых в случаях Ляпунова не предполагается.

Для восстановления поколебленного таким образом значения метода математических ожиданий необходимо было выяснить, что вышеупомянутыми работами он не исчерпан до конца". Марков в 1908 г. выступил с замечательной идеей – урезания случайных величин. Этот прием дал возможность доказать предельную теорему в условиях Ляпунова методом моментов или, как говорил Марков, методом математических ожиданий. Идея урезания прочно вошла в жизнь теории вероятностей.

Дальнейшая судьба центральной предельной теоремы такова: в 1922 г. финскому математику Линдебергу удалось пойти дальше Ляпунова и отказаться от предположения существования даже каких-либо моментов, кроме вторых. А именно, он доказал, что если при любом $\tau > 0$ имеет место соотношение

$$\lim_{n\to\infty} \frac{1}{B_n} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2\,dF_k(x) = 0,$$

то функция распределения сумм центрированных математическими ожиданиями и нормированных корнем квадратным из суммы дисперсий слагаемых сходятся к стандартному нормальному распределению.

Через 12 лет В. Феллер показал, что условие Линдеберга является и необходимым в предположении, что слагаемые равномерно малы.

Ясно, что из теоремы Линдеберга в качестве следствия получается давно ожидавшийся результат: если случайные величины независимы, одинаково распределены и имеют конечную дисперсию, отличную от 0, то к суммам таких величин применима центральная предельная теорема теории вероятностей.

В работе 1927 г. С.Н. Бернштейн рассмотрел несколько более общую задачу: имеется последовательность независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, относительно которых не предполагается ни существования дисперсий, ни существования математических ожиданий. Спрашивается: когда можно подыскать такие постоянные $B_n > 0$ и A_n, что функции распределения сумм $\dfrac{S_n - A_n}{B_n}$ сходятся к нормальному распределению?

Достаточные условия для этой задачи были найдены Бернштейном в той же работе 1927 г.; через восемь лет Феллер показал, что эти условия не только достаточны, но и необходимы в предположении, что слагаемые равномерно малы в смысле теории вероятностей.

В том же 1935 г. независимо один от другого А.Я. Хинчин и П. Леви в постановке С.Н. Бернштейна нашли необходимое и достаточное условие сходимости к нормальному распределению функций распределения сумм независимых, одинаково распределенных случайных величин.

Еще в 1926 г. в специальном курсе по предельным теоремам А.Я. Хинчин задал следующий вопрос: имеется ли связь между законом больших чисел и центральной предельной теоремой? Ответ был найден Д.А. Райковым и А.А. Бобровым, которые доказали следующую теорему: чтобы функции распределения сумм

$$\frac{\xi_1 + \ldots + \xi_n - A_n}{B_n}$$

при надлежащем выборе действительных постоянных $B_n > 0$ и A_n сходились к нормальному распределению, необходимо и достаточно чтобы суммы

$$(\xi_1 - a_1)^2 + (\xi_2 - a_2)^2 + \ldots + (\xi_n - a_n)^2$$

были относительно устойчивы, $a_n = \int_{-\epsilon}^{\epsilon} x\,dF_n(x)$, $\epsilon > 0$ – произвольно.

Исследование вопросов сходимости функций распределения к нормальному закону не окончились и в наши дни, но теперь исследуются другие вопросы: быстрота сходимости к предельному распределению, сходимость случайного числа случайных слагаемых, суммирование неравномерно малых случайных величин.

§ 17. Общие предельные распределения для сумм

Естественный вопрос о том, какие распределения вообще возможны в качестве предельных для сумм независимых случайных величин при условии, что они примерно одинаковы по величине, возник только в двадцатые–тридцатые годы нашего столетия. Раньше во всей общности этот вопрос не возникал, хотя частные результаты по этому поводу и появлялись. В этом отношении заслуживает упоминания мемуар С. Пуассона "О вероятности средних результатов наблюдений", в котором, пользуясь аппаратом характеристических функций, он вывел распределение суммы большого числа независимых ошибок наблюдений и рассмотрел распределение, которое получило впоследствии название распределения Коши. Для этого распределения Пуассон нашел плотность

$$f(t) = \frac{1}{\pi(1 + x^2)}$$

и доказал, что оно обладает двумя следующими свойствами:

1) среднее арифметическое ошибок наблюдений, распределенных по закону Коши, имеет то же распределение, что и каждое слагаемое;

2) для этого распределения точность не повышается от того, что берется среднее арифметическое результатов нескольких наблюдений.

Этот мемуар был опубликован в 1832 г.

На тридцать лет позднее, в 1853 г. в мемуаре "О средних результатах наблюдений той же природы и о результатах наиболее вероятных" О. Коши получил характеристическую функцию для всех тех распределений, для которых функция распределения суммы только на множитель при аргументе (коэффициент растяжения) отличается от распределения отдельных слагаемых. Коши нашел, что все такие функции имеют вид $f(x) = \exp(-t^{\mu})$, где μ – положительное число. Позднее выяснилось, что $f(t)$ тогда и только тогда является характеристической функцией, когда $0 < \mu \leqslant 2$.

П. Леви в книге "Calcul des probabilité" (1925) в главе VI "Экспоненциальные распределения" построил первую теорию устойчивых распределений. Эта теория, естественно продолжала исследования О. Коши, уйдя от них далеко вперед. Пусть $F(x)$ – функция распределения и $f(t)$ – ее характеристическая функция. Распределение $F(x)$ называется *устойчивым*, если при любых положительных постоянных a_1 и a_2 найдется такое положительное постоянное a, что выполняется равенство

$$f(a_1 t) \cdot f(a_2 t) = f(at).$$

В терминах случайных величин рассмотренный класс распределений обладает следующим характеристическим свойством: если ξ_1 и ξ_2 – независимые случайные величины с одним и тем же распределением вероятностей, a_1 и a_2 – произвольные положительные числа, то для каждой пары a_1 и a_2 найдется такое положительное число a, что сумма $a_1\xi_1 + a_2\xi_2$ имеет такое же распределение как $a\xi_1$.

П. Леви указал, что для устойчивых распределений функция $f(t)$ имеет вид $\exp[-c(1 + i\beta\frac{t}{|t|})|t|^\alpha]$, где $0 < \alpha \leqslant 2$. П. Леви также ввел понятие *области притяжения* устойчивого закона: множество всех тех распределений $F(x)$, для которых функции распределения независимых и распределенных по этому закону случайных величин при соответствующем нормировании сходятся к данному устойчивому распределению.

В 1935 г. А.Я. Хинчин пополнил понятие устойчивого распределения, введенного П. Леви, а именно: он предложил называть *устойчивыми* те распределения, для которых линейная форма $a_1\xi_1 + a_2\xi_2$ при произвольных положительных постоянных a_1 и a_2 имеет такое же распределение, как $a\xi_1 + b$, где a – некоторое положительное, а b – вещественное постоянное. Класс устойчивых в смысле Хинчина распределений оказался несколько шире класса П. Леви.

В 1939 г. независимо друг от друга Б.В. Гнеденко и В. Деблин нашли области притяжения устойчивых распределений. Условия принадлежности области притяжения устойчивого закона очень просты и сводятся к поведению "хвостов" распределений – поведению исходного распределения при больших значениях аргумента.

Основной результат, принадлежащий П. Леви и А.Я. Хинчину, можно сформулировать так: если $\xi_1, \xi_2, \ldots$ – последовательность одинаково распределенных независимых случайных величин, то суммы

$$S_n = \frac{\xi_1 + \ldots + \xi_n - A_n}{B_n} \qquad (1)$$

при надлежащем выборе постоянных $B_n > 0$ и вещественных A_n могут сходиться только к устойчивым законам распределения. Каждый устойчивый закон является предельным для функций распределения сумм (1).

В заметке 1936 г. П. Леви и А.Я. Хинчин дали окончательное представление устойчивых распределений через логарифмы характеристической функции. Чтобы функция $\varphi(t)$ была характеристической функцией устойчивого распределения, необходимо и дотаточно следующее ее представление:

$$\ln \varphi(t) = i\gamma t - c|t|^\alpha \left\{1 + i\beta\frac{t}{|t|}\omega(t, \alpha)\right\},$$

где α, β, γ, c – вещественные постоянные $(-1 \leqslant \beta \leqslant 1, c > 0, 0 < \alpha \leqslant 2)$ и

$$\omega(t, \alpha) = \begin{cases} \operatorname{tg} \dfrac{\pi}{2}\alpha, & \text{если} \quad \alpha \neq 1, \\ \dfrac{2}{\pi} \ln |t|, & \text{если} \quad \alpha = 1. \end{cases}$$

Этот результат полностью завершил исследования, которые были начаты Пуассоном и Коши.

Естественный вопрос о классе предельных распределений для сумм (1), когда слагаемые могут быть распределены не одинаково, был поставлен А.Я. Хинчиным в письме к П. Леви. Вскоре ответ был найден П. Леви. По предложению А.Я. Хинчина этот класс распределений получил наименование класса L. На слагаемые суммы ξ_k/B_n при этом естественно наложить требование: каждое из слагаемых оказывает на сумму незначительное влияние. Это требование можно представить так: величины ξ_k/B_n предельно постоянны, т.е. для них можно найти такую последовательность постоянных m_{nk}, что равномерно относительно k $(1 \leqslant k \leqslant n)$ для любого $\epsilon > 0$ выполняется соотношение

$$\mathsf{P}\left\{\left|\frac{\xi_k}{B_n} - m_{nk}\right| > \epsilon\right\} \to 0 \quad \text{при} \quad n \to \infty.$$

Характеристическое свойство законов класса L, найденное П. Леви, состоит в следующем: чтобы функция $\varphi(t)$ была характеристической функцией закона класса L, необходимо и достаточно выполнение следующего условия: при каждом α $(0 < \alpha \leqslant 1)$ имеет место равенство $\varphi(t) = \varphi(\alpha t) f_\alpha(t)$, где $f_\alpha(t)$ – некоторая характеристическая функция.

На этом история вопроса не завершилась, поскольку для полного завершения проблематики оставалось ответить еще на один вопрос, поставленный Б.В. Гнеденко. Каковы классы возможных предельных распределений, если случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ могут быть распределены только по k $(1 \leqslant k < \infty)$ различным законам распределения $F_1(x), F_2(x), \ldots, F_k(x)$? Полное решение этой задачи было дано в 1971 г. А.А. Зингером.

В рассмотренном круге вопросов была изучена еще одна задача: а что будет, если рассматривать суммы (1) одинаково распределенных независимых слагаемых не по всем значениям n, а только по некоторой подпоследовательности? Какие предельные распределения при этом могут встретиться? Этот вопрос был поставлен А.Я. Хинчиным; он же дал на него ответ: класс возможных предельных распределений в только что указанном смысле совпадает с классом безгранично делимых распределений, в 1930 г. введенным итальянским математиком Бруно де Финетти и подробно исследованным А.Н. Колмогоровым, П. Леви и А.Я. Хинчиным. Случайная величина называется безгранично делимой, если для любого целого числа n ее можно представить в виде суммы n независимых одинаково распределенных слагаемых. Отсюда и название этих величин.

В 1933 г. А.Н. Колмогоров высказал гипотезу, что если суммируются примерно равноправные независимые случайные величины, то при увеличении числа слагаемых из распределения будут приближаться к безгранично делимым законам и, следовательно, если распределения последовательных сумм будут сходиться к предельному, то этот предельный закон обязательно должен быть безгранично делимым. В предположении, что слагаемые имеют конечные дисперсии, а дисперсии последовательных сумм ограничены эту гипотезу доказал ученик А.Н. Колмогорова Г.М. Бавли (1908–1941) в 1934 г. В полном объеме эта гипотеза была доказана А.Я. Хинчиным с привлечением довольно громоздких аналитических средств через три года. Отправляясь от этой работы, Б.В. Гнеденко построил теорию суммирования независимых случайных величин, основанную на сравнительно легко доказываемом факте: если суммируются предельно постоянные независимые слагаемые и функции распределения соответствующих центрированных сумм сходятся к некоторому предельному, то можно построить последовательность безгранично делимых случайных величин, функции распределения которых сближаются с функциями распределения сумм. Эти безгранично делимые величины получили название сопровождающих. Из этого предложения в качестве частных случаев получались теоремы Бавли и Хинчина. Кроме того, этот подход давал возможность совершенно прозрачно найти условия существования предельных распределений и условия сходимости функций распределения сумм к любому возможному предельному распределению. В частности, были найдены необходимые и достаточные условия для закона больших чисел, для сходимости к нормальному распределению, распределению Пуассона, устойчивым распределениям. Весь круг этих вопросов нашел отражение в монографии Б.В. Гнеденко и А.Н. Колмогорова "Предельные распределения для сумм независимых случайных величин" (1949).

В последние годы большое число исследователей приступило к изучению предельного поведения сумм независимых случайных слагаемых в случайном числе. Первоначально усилия были сосредоточены только на условиях сходимости к нормальному распределению и выполнимости закона больших чисел. Позднее были поставлены вопросы о классе предельных распределений и об условиях существования предельного распределения. Эту задачу удалось решить в условиях одинаковой распределенности и независимости слагаемых, а также независимости индекса суммирования от слагаемых. Заметим также, что к самой постановке этих задач привели вопросы теории надежности и физики. Основная теорема, относящаяся к названной проблематике, получила наименование теоремы переноса.

§ 18. Закон повторного логарифма

От закона больших чисел взяла начало новая предельная закономерность, получившая наименование закона повторного логарифма. Эта теорема не ставит перед собой цели разыскания предельного распределения, но зато переводит задачу рассмотрения последовательных сумм совсем в новую область, а именно изучает поведение этих сумм всех вместе. Мы вначале рассмотрим эту задачу для простейшего случая – для схемы Бернулли. Это вполне естественно; тем более, что это соответствует историческому ходу исследований.

Обозначим через μ_n число появлений события A в n независимых испытаниях и рассмотрим разности $S_n = \mu_n - np$. В 1909 г. Э. Борель дал обобщенную формулировку закона больших чисел, показав, что имеет более сильное утверждение, а именно

$$\mathbf{P}\ \{S_n/n \to 0\} = 1.$$

Через четыре года Ф. Хаусдорф (1868 – 1942) доказал, что имеет место еще более сильное утверждение, а именно, что при любом $\epsilon > 0$

$$\mathbf{P}\{S_n/\sqrt{n^{1+\epsilon}} \to 0\} = 1.$$

Год спустя, Г. Харди (1877 – 1947) и Дж. Литвуд (1885 – 1977) обнаружили еще более сильное предложение, согласно которому с вероятностью единица отношение $|S_n|/\sqrt{n \ln n}$ остается ограниченным. В 1922 г. А.Я. Хинчин дал для роста сумм S_n оценку $S_n = O(\sqrt{n \ln \ln n})$. Через два года А.Я. Хинчин нашел окончательный результат.

Оказалось, что

$$\mathbf{P}\left\{\limsup_{n\to\infty}\frac{|S_n|}{\sqrt{2npq\ln\ln n}}=1\right\}=1.$$

В 1926 г. А.Я. Хинчину удалось распространить этот результат на случай схемы Пуассона, т.е. на случай последовательных испытаний с переменной вероятностью появления события A.

Работа А.Н. Колмогорова 1929 г. значительно перекрывала результаты А.Я. Хинчина, которые являлись для нее простыми следствиями. Этими словами мы не хотим преуменьшить значения работ А.Я. Хинчина, поскольку открыть новую закономерность даже на простом случае заслуживает самой высокой оценки.

Пусть имеется последовательность $\xi_1, \xi_2, \ldots$, взаимно независимых случайных величин, имеющих математические ожидания $a_k = \mathbf{M}\xi_k$ и дисперсии $b_k = \mathbf{D}\xi_k$; $B_n = \sum_{k=1}^{n} b_k$, $S_n = \sum_{k=1}^{n}(\xi_k - a_k)$. Если последовательность ξ_k удовлетворяет еще двум условиям: при $n\to\infty$ 1) $B_n\to\infty$, 2) $|\xi_n| < m_n = o\left(\sqrt{\frac{B_n}{\ln\ln B_n}}\right)$, то она удовлетворяет закону повторного логарифма, т.е. для нее выполняется соотношение

$$\mathbf{P}\left\{\limsup_{n\to\infty}\frac{|S_n|}{\sqrt{2B_n\ln\ln B_n}}=1\right\}=1.$$

Иными словами было высказано следующее утверждение: в высказанных предположениях при любых положительных ϵ и δ можно указать столь большое целое число N, что

1) вероятность того, что хотя бы при одном $n > N$ выполнится неравенство

$$|S_n| > (1+\delta)\sqrt{2B_n\ln\ln B_n}$$

меньше ϵ и

2) вероятность того, что хотя бы для одного $n > N$ будет выполнено неравенство

$$|S_n| > (1-\delta)\sqrt{2B_n\ln\ln B_n}$$

больше $1-\epsilon$.

Позднее задачей повторного логарифма занимались многочисленные исследователи – П. Леви, В. Феллер, Зигмунд и Марцинкевич, Хартман, Т.А. Сарымсаков, В.В. Петров, Б.В. Гнеденко и др. Среди многих прекрасных результатов мы выделим лишь один: если случайные величины ξ_k одинаково распределены и имеют конечную дисперсию (конечно, отличную от нуля), то это условие достаточно для выполнения закона повторного логарифма. Как показал А.И. Мартикайнен, этот результат допускает обращение *)

---

*) Обращение закона повторного логарифма для случайного блуждания //Теория вероятностей и ее применения. – 1980. – Т. 25, вып. 2. – С. 364 – 366.

Аналогичная задача была поставлена для устойчивых распределений, отличных от нормального. При этом выяснилось (Б.В. Гнеденко), что для любой неубывающей функции $u(n)$ и для любого устойчивого закона с показателем α $(0 < \alpha < 2)$ с вероятностью единица отношение $\limsup\limits_{n \to \infty} \dfrac{|S_n|}{u(n)}$ равно 0 или бесконечности.

§ 19. Формирование понятий математического ожидания и дисперсии

Понятие математического ожидания в самых начальных его элементах было введено в теорию вероятностей очень рано: впервые оно появилось в известной переписке Паскаля с Ферма. В более явной форме оно было введено Гюйгенсом. Именно, первые три предложения являются ничем иным как определением математического ожидания для случайных величин, способных принимать два или три значения. Как мы уже говорили в первой главе сам термин ожидание был предложен Схоутеном – учителем Гюйгенса. Этот термин прижился и сохранился до нашего времени. Но в ту пору этому термину придавался смысл ожидания той средней цены, которую можно дать за приобретение случайной величины, дающей выигрыш x_1 с вероятностью p_1, выигрыш x_2 с вероятностью $p_2, \ldots$, выигрыш x_n с вероятностью p_n.

Эта мысль красной строкой проходит и в книге Н. Бернулли "О применении искусства предположений в вопросах права". Он писал там, что "правило это (вычисления ожидания – *Б.Г.*) тождественно с тем, с помощью которого обыкновенно отыскиваются среднее арифметическое нескольких данных величин, а также и с тем правилом смешения, на которое счел уместным сослаться мой дядя". Далее он рассмотрел пример, заимствованный из рукописи книги Я.Бернулли "Искусство предположений": "Если три кружки пива ценой по 13 смешиваются с 2 кружками ценой по 8, то после перемножения 3 на 13 и 2 на 8 получится общая цена всех кружек – 55, что дает путем деления на число всех кружек, т.е. 5, среднюю цену одной кружки смеси, равную 11. Такова же должна быть, согласно правилу, и оценка величины ожидания чего-либо, что будет иметь 3 случая по 13 и 2 случая по 8". Заметим, что сказанное является ничем иным как повторением правил Гюйгенса. Заслуживает внимания не только то, что Н. Бернулли рассмотрел ожидание для случайных величин, принимающих не только два или три значения, но и бóльшее число значений, но и нечто совсем новое, а именно сравнение формулы для вычисления математического ожидания с правилом вычисления координат центра тяжести системы материальных точек. Вот подлинные слова Н. Бернулли, из той же книги.

"Еще более заслуживает быть отмеченным особое и исключительное совпадение, наблюдающееся между этим правилом и тем, которое рекомендуется для нахождения центра тяжести нескольких грузов; действительно, ведь сумма моментов, т.е. сумма произведений весов на соответствующие расстояния от какой-либо данной точки, деленная на сумму весов, показывает расстояние от центра тяжести, т.е. той точки, по отношению к которой подвешенные грузы находятся в равновесии, точно также и та средняя, которая получается согласно настоящему правилу, является, так сказать, центром тяжести всех вероятностей, который их так уравновешивает, что ни та, ни другая из них, отклоняясь в ту или другую сторону от средней, не перевешивают друг друга. В целях соблюдения такого же равновесия в сомнительных и темных делах наши юристы придерживаются обычно середины."

Для XVIII века обращение к математическому ожиданию было не характерным. Все внимание привлекало понятие вероятности случайного события. В энциклопедии науки о вероятностях – знаменитой книге П. Лапласа "Аналитическая теория вероятностей" – нет определения математического ожидания и тем более правил действий с ним. Возможно, это связано с тем, что Лаплас не рассматривал и понятия случайной величины, вместо этого он изучал ошибки наблюдений, плотности их распределений и даже вывел и использовал формулу для плотности суммы двух независимых ошибок. Правда, при этом он не говорил о том, что рассматривались независимые, поскольку другие и не изучались.

Казалось бы создание и развитие теории ошибок наблюдений должно было стимулировать развитие числовых характеристик случайных величин (которые в ту пору еще назывались ошибками измерений). Однако этого не случилось. Впрочем, для нормального распределения были введены понятия истинного значения и точности наблюдений; было известно как их вычислять по плотности распределения. Таким образом для этого частного случая уже была известна формула для вычисления математического ожидания и дисперсии.

Обратим внимание на то, что в начале XIX века нормальное распределение затмило собой все остальные, поскольку с ним столкнулись в теории ошибок наблюдений и, казалось, доказали в работах Гаусса и Лежандра, что распределение ошибок наблюдений должно быть нормальным. С ним же столкнулись в теории стрельбы. Бельгийский биолог Кетле давал многочисленные свидетельства того, что и в биологии нормальное распределение играет центральную роль. Остальные распределения потеряли интерес, о них попросту не думали. Несомненно, в связи с этим никто и не помышлял о доказательстве теорем относительно математических ожиданий и дисперсий, поскольку для нормального распределения все уже было известно. В связи со сказанным интересно заметить, что в книге П.Л. Чебышева "Опыт элементарного анализа теории вероятностей" (М., 1845) понятия случайной величины, математического ожидания и дисперсии даже не упоминаются. Однако в курсе лекций по теории вероятностей, который систематически он читал в Петербургском университете, Чешев говорит о величинах (имея в виду случайные величины), их математическом ожидании и дисперсии. Более того, в этих лекциях (записанных А.М. Ляпуновым, переписанных у него А.Н. Крыловым и изданных в 1936 г. в издательстве АН СССР) было сказано, что "óно (понятие математического ожидания) имеет бóльшее значение на практике, чем сама вероятность, потому что на основании ее у нас составляется суждение о том, что мы можем ожидать перед появлением известного события" (с. 159). Само это утверждение не очень понятно, но, несомненно, Чебышев имел в виду какое-то определенное замечательное свойство математического ожидания. По-видимому, свою роль сыграла и формулировка закона больших чисел в форме Чебышева.

Заслуживает пристального внимания то обстоятельство, что в этих записках лекций имеется доказательство и формулировка теорем о математическом ожидании и дисперсии суммы случайных величин. Там же он привел и вывод своего знаменитого неравенства. При этом он предполагал как нечто самоочевидное, что речь идет о независимых величинах. Следует отметить, что сам факт о том, что дисперсия суммы равна сумме дисперсий, имеется и использован Чебышевым в статье "О средних величинах". Там же впервые встречается и неравенство Чебышева. Следует отметить, что в распространенных учебниках (А. Пуанкаре и Ж. Бертрана) конца прошлого века и начала текущего столетия вообще нет теорем о математическом ожидании и дисперсии.

Естественно спросить себя: когда же стал известен факт, что математическое ожидание суммы равно сумме математических ожиданий всегда, а не только при независимых слагаемых? Пока на это можно лишь ответить, что в учебнике Чубера (1908) и переиздании книги А. Пуанкаре (1912) такой теоремы нет, а в знаменитом для своего времени учебнике "Исчисление вероятностей" (1913, 1924) строго доказываются и теорема о математическом ожидании произведения и о математическом ожидании суммы со специальным упоминанием о том, что она верна не только для независимых величин.

В заключение мы должны сказать, что история понятий математического ожидания и дисперсии изучена совершенно недостаточно. Мы видим, что основы понятия математического ожидания возникли одновременно с понятием вероятности, но выделены основные его свойства были очень поздно – только во второй половине прошлого – начале нашего столетия. Неясно, в какой мере на понятие дисперсии влияло уже существовавшее понятие момента инерции. Впрочем, заслуживает внимания и исследование истории становления и развития теории случайных величин. То, что изложено в настоящей главе может считаться лишь первым приближением к истории этого важного раздела научных знаний.

ГЛАВА 4

К ИСТОРИИ ТЕОРИИ СЛУЧАЙНЫХ ПРОЦЕССОВ

§ 20. Общие представления

Понятие случайного процесса принадлежит нашему столетию и связано с именами А.Н. Колмогорова, А.Я. Хинчина, Е.Е. Слуцкого, Н. Винера (1894–1965). Это понятие в наши дни является одним из центральных не только в теории вероятностей, но также в естествознании, инженерном деле, экономике, организации производства, теории связи. Теория случайных процессов принадлежит к категории наиболее быстро развивающихся математических дисциплин. Несомненно, что это обстоятельство в значительной мере определяется ее глубокими связями с практикой.

XX век не мог удовлетвориться тем идейным наследием, которое было получено им от прошлого. Действительно, в то время как физика, биолога, инженера интересовал процесс, т.е. изменение изучаемого явления во времени, теория вероятностей предлагала им в качестве математического аппарата лишь средства, изучавшие стационарные состояния. Для исследования изменения во времени теория вероятностей конца прошлого – начала нашего века не имела ни разработанных частных схем, ни тем более общих приемов. А необходимость их создания буквально стучала в окна и двери математической науки. Изучение броуновского движения в физике подвело математику к порогу создания теории случайных процессов. В исследованиях датского ученого А.К. Эрланга была начата новая важная область поисков, связанных с изучением загрузки телефонных сетей. Число абонентов изменяется во времени случайно, а длительность каждого разговора обладает большой индивидуальностью. И вот в этих-то условиях двойной случайности следует производить расчет пропускной способности телефонных сетей, коммутационной аппаратуры и управляющих связью систем. Несомненно, что работы Эрланга оказали значительное влияние не только на решение чисто телефонных задач, но и на формирование элементов теории случайных процессов, в частности, процессов гибели и размножения.

Во втором десятилетии XX века начались исследования динамики биологических популяций. Итальянский математик Вито Вольтерра разработал математическую теорию этого процесса на базе чисто детерминистских соображений. Позднее ряд биологов и математиков развивали его идеи уже на основе стохастических представлений. Первоначально и в этой теории применялись исключительно идеи процессов гибели и размножения. Собственно именно от задач биологии и пошло наименование этого очень частного типа случайных процессов.

Представим себе, что мы задались целью проследить за движением какой-нибудь молекулы газа или жидкости. Эта молекула в случайные моменты сталкивается с другими молекулами и меняет при этом направление движения и скорость. Состояние молекулы, таким образом, подвержено случайным изменениям и представляет собой ничто иное, как случайный процесс. Это процесс определяется шестью параметрами – тремя координатами и тремя компонентами скорости. Многие физические явления для своего изучения требуют умения вычислять вероятности того, что определенная доля молекул успеет за заданный промежуток времени перейти из одной области пространства в другую. Например, если приведены в соприкосновение две жидкости, то начинается взаимное проникновение молекул одной жидкости в другую. Происходит диффузия. Как быстро происходит процесс диффузии, по каким законам и когда образующаяся смесь становится практически однородной? На эти и многие другие вопросы дает ответ статистическая теория диффузии, базирующаяся на использовании теории случайных процессов. Очевидно, что подобные же задачи возникают в химии, когда приступают к изучению процессов химических реакций. Какая часть молекул уже вступила в реакцию, какая особенность протекания реакции со временем, когда реакция практически уже закончилась?

Весьма важный круг явлений протекает по принципу радиоактивного распада. Суть его состоит в том, что атомы радиоактивного вещества распадаются, превращаясь в атомы другого элемента. Распад каждого атома происходит мгновенно, подобно взрыву, с выделением некоторого количества энергии. Многочисленные наблюдения показывают, что распад отдельных атомов происходит в случайно взятые моменты времени и расположение этих моментов, если количество распадающегося вещества не превосходит некоторого определенного критического предела, не зависят друг от друга. Для изучения процесса радиоактивного распада весьма важно определить вероятность того, что за определенный промежуток времени распадается то или иное число атомов. Формально, если задаться целью выяснеиия только математической стороны явления, аналогично происходят многие другие процессы: обрывы нитей в прядильной машине, число броуновских частиц, оказавшихся в данный момент в определенной области пространства, вызовы от абонентов, поступающие на телефонную станцию и т.д.

Теория броуновского движения, исходящая из теоретико-вероятностных предпосылок, была разработана в 1905 г. двумя известными физиками М. Смолуховским (1872–1917) и А. Эйнштейном (1879–1955). Позднее высказанные ими идеи использовались неоднократно как при изучении физических явлений, так и в различных инженерных задачах. В частности, именно с их работ, как, впрочем, и с работ Эрланга, начался широкий интерес к процессу Пуассона. Впрочем, сам Пуассон ввел в рассмотрение только распределение Пуассона и о процессе Пуассона даже не мечтал, но он заслужил того, чтобы его имя произносилось и при рассмотрении случайных процессов, связанных с его распределением. Это не единственный случай, когда в честь того или другого исследователя новым понятиям присваиваются их имена, хотя до этих понятий они и не доходили. Теперь широко распространены гауссовские случайные процессы, хотя сам Гаусс о них не имел никакого представления, да и само исходное распределение задолго до его рождения было получено Муавром, Лапласом и др. В теории же ошибок измерений одновременно с Гауссом к нему пришел также Лежандр.

Попытка изучения средствами теории вероятностей явления диффузии была предпринята в 1914 г. двумя известными физиками – М. Планком (1858–1947) и Фоккером.

Н. Винер в середине двадцатых годов при изучении броуновского движения ввел в рассмотрение процессы, удовлетворяющие следующим условиям:

1) функция распределения разности $\xi(t_0 + t) - \xi(t_0)$ не зависит от начального момента t_0 (однородность во времени);

2) приращения процесса $\xi(t)$ за непересекающиеся промежутки времени (s_i, t_i) в конечном числе взаимно независимы (независимость приращений);

3) величины $\xi(t_0 + t) - \xi(t_0)$ нормально распределены со средним значением равным 0 и дисперсией $\sigma^2 t$.

Мы должны упомянуть еще о двух важных группах исследований, начатых в разное время и по разным поводам. Во-первых, это работы А.А. Маркова (1856–1922) по изучению цепных зависимостей. Во-вторых, работах Е.Е. Слуцкого (1880–1948) по теории случайных функций. Оба эти направления играли очень существенную роль в формировании общей теории случайных процессов. Для этой цели уже был накоплен значительный исходный материал и необходимость построения теории как бы носилась в воздухе. Оставалось осуществить глубокий анализ имеющихся работ, высказанных в них идей и результатов и на его базе осуществить необходимый синтез.

В 1931 г. была опубликована большая статья А.Н. Колмогорова "Об аналитических методах в теории вероятностей", а через три года работа А.Я. Хинчина "Теория корреляции стационарных стохастических процессов", которые следует считать началом построения общей теории случайных процессов. В первой из этих работ были заложены основы теории марковских процессов, а во второй – основы стационарных процессов. Они были источником огромного числа последующих исследований, среди которых следует отметить статью В. Феллера "К теории стохастических процессов" (1936), давшую интегро-дифференциальные уравнения для скачкообразных марковских процессов.

Обе только что упомянутые основополагающие работы содержат не только математические результаты, но и глубокий философский анализ причин, послуживших исходным пунктом для построения основ теории случайных процессов. Приведем с целью ознакомления с этим аспектом исследований довольно большой отрывок из введения к работе А.Н. Колмогорова.

"Желая подвергнуть математической обработке явления природы или социальной жизни, необходимо предварительно эти явления схематизировать; дело в том, что к исследованию процесса изменения некоторой системы математический анализ применим лишь в том случае, если предположить, что каждое возможное состояние этой системы может быть вполне определено с помощью известного математического аппарата, например, при помощи значений, принимаемых известным числом параметров; такая математически определимая система есть не сама действительность, но лишь схема, пригодная для описания действительности.

Классическая механика пользуется лишь такими схемами, при которых состояние y системы для момента времени t однозначным образом определяется ее состоянием x в любой предшествующий момент t_0; математически это выражается формулой $y = f(x, t_0, t)$.

Если такая однозначная функция существует, как это всегда предполагается в классической механике, то мы говорим, что наша схема есть схема *вполне детерминированного* процесса. К числу вполне детерминированных процессов можно было отнести и те, в которых состояние y не вполне определяется заданием состояния x для единственного момента времени t, существенным образом зависит еще от характера изменения этого состояния перед моментом t. Однако обычно предпочитают избегать такой зависимости от предшествующего поведения системы, для чего расширяют само понятие состояния системы в момент времени t и соответственно этому вводят новые параметры *).

---

*) Хорошо известный пример применения этого метода мы имеем при описании состояния некоторой механической системы не только координатами ее точек, но также и компонентами их скоростей.

Вне области классической механики, наряду со схемами вполне детерминированных процессов, часто рассматриваются и такие схемы, где состояние x системы в некоторый момент времени t_0 обуславливает лишь известную вероятность для наступления возможного состояния y в некоторый последующий момент $t > t_0$. Если для любых заданных t_0, $t > t_0$ и x существует определенная функция распределения вероятностей для состояния y, мы говорим, что наша схема есть схема *стохастически определенного* процесса. В общем случае эта функция распределения представляется в виде $\mathbf{P}(t_0, x, t, A)$, причем A обозначает некоторое множество состояний, а $\mathbf{P}$ есть вероятность того, что в момент t окажется реализованным одно из состояний A, принадлежащих этому множеству".

Но не общефилософское содержание является основным достоинством этой работы А.Н. Колмогорова. В ней были заложены основы теории случайных процессов без последействия и получены дифференциальные уравнения (прямые и обратные), которые управляют вероятностями перехода. В этой же работе был дан набросок теории скачкообразных процессов без последействия, подробное развитие которой позднее было дано В. Феллером и В.М. Дубровским.

В настоящее время теория марковских процессов превратилась в большую и разветвленную главу математической науки, которая получила огромное число различных применений в физике, инженерном деле, геофизике, химии и ряде других областей знания.

Построение основ другого класса случайных процессов на базе физических задач было осуществлено А.Я. Хинчиным в упомянутой нами работе. Он ввел понятие стационарного процесса в широком и узком смысле и получил знаменитую формулу для коэффициента автокорреляций. Эта работа послужила основанием для последующих исследований Г. Крамера, Г. Вальда, А.Н. Колмогорова и многих других ученых.

В процессе развития теории случайных процессов произошло разделение близких понятий. Если случайная величина $\xi(t)$ или вектор $(\xi_1(t), \ldots, \xi_n(t))$ со значениями на числовой прямой зависит от одного вещественного параметра t, то принято говорить о случайном процессе $\xi(t)$. При этом, как правило, параметр t носит название времени. Если время принимает дискретную последовательность значений $t_1, t_2, \ldots$ то говорят не о случайном процессе, а о случайной последовательности. Если же случайная величина ξ (или вектор) зависит не от одного, а от нескольких параметров, то ее называют случайным полем.

Со случайными полями столкнулись раньше всего в биологии и геофизике, а затем оказалось, что практически все области знания приводят к необходимости рассмотрения наряду со случайными процессами и случайных полей. Рассмотрим примеры.

Обозначим через $\rho(t, x, y, z)$ плотность воды в океане. Эта величина изменяется от одной точки к другой и от одного момента времени к другому. Как показывают многочисленные наблюдения, ρ можно рассматривать как случайное поле.

Рассмотрим изменение силы и направления ветра. Для каждого момента времени и каждой точки пространства сила ветра $f(t, x, y, z)$ является скалярной величиной, а направление ветра $\xi(t, x, y, z), \eta(t, x, y, z), \zeta(t, x, y, z)$ – случайным вектором. Это типичный пример скалярного и векторного полей.

Число примеров случайных полей, относящихся к различным областям знания, можно продолжать практически неограниченно.

В истории каждой науки постоянно приходится сталкиваться с такими ситуациями, когда эта наука еще не создана, а исследователи рассматривают отдельные задачи, которые относятся к ее компетенции. Так было с арифметикой и геометрией, алгеброй и теорией чисел. С таким же положением мы сталкиваемся и в теории случайных процессов. Этой теории еще не было, не было и свойственных ей понятий, не было даже идеи рассмотрения изменения случайной величины во времени, а отдельные задачи в этом направлении уже изучались. Для примера еще Н. Бернулли, Монмор и

Муавр занимались задачей о разорении игрока и состоянии игроков после n партий. Это типичная задача теории случайных процессов, в которой число сыгранных партий играет роль времени. Такая же ситуация складывается и с задачей Лапласа перекладывания шаров из урны в урну и подсчета содержания урны после n перекладываний. Всегда новое рождается в недрах старого и со временем вырастает из становящихся тесными рамок уже установившихся представлений и понятий. В результате появляется необходимость выделения специальной области научных исследований. Первоначально же отдельные новые задачи решаются в рамках старых представлений, как правило специальными приемами, создаваемыми для каждой задачи. Но время еще не созрело для выделения соответствующей новой ветви научного знания. Требуется иногда длительный срок, чтобы первоначальные идеи и отдельные задачи сформировались и дали начало новой теории со своими постановками проблем и методами исследования, позволяющими продвинуться по пути познания явлений окружающего нас мира.

Теория вероятностей имеет богатую и поучительную историю. Она наглядно показывает как возникали ее основные понятия и развивались методы из задач, с которыми сталкивался общественный прогресс. История теории вероятностей еще далека от совершенства и требуется систематическая работа для того, чтобы восстановить пройденный путь и воздать должное ее создателям. При этом мы увидим как человечество переходило от первичных догадок к более полному и совершенному знанию, как создание теории вероятностей позволяло переходить от строгих детерминистических представлений к более широким стохастическим концепциям, тем самым открывая новые возможности для глубоких заключений о природе вещей.

Теория вероятностей продолжает бурно развиваться, в ней появляются новые направления исследований – оптимальное управление случайными процессами, теория мартингалов, теория просачивания, случайные операторы, вероятностные закономерности на алгебраических и топологических структурах. Эти направления представляют значительный общетеоретический и прикладной интерес. Практически исторический очерк ограничивается во времени сороковыми годами нашего столетия и только отдельные замечания относятся к более позднему времени. Я надеюсь на то, что вопросы истории теории вероятностей заинтересуют некоторых читателей и им удастся существенно дополнить настоящий очерк в ряде направлений.

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046

(*Окончание*)

x	0	1	2	3	4	5	6	7	8	9
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблица значений функции $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}}\, dz$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0159	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0909	0948	0987	1026	1103	1064	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2356	2389	2421	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2703	2734	2764	2793	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3437	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3728	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3906	3925	3943	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545

(*Окончание*)

x	0	1	2	3	4	5	6	7	8	9
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4648	4656	4664	4671	4678	4686	4692	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4874	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4924	4927	4929	4930	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4958	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

3,0	0,4986	3,1	4990	3,2	49931	3,3	49952	3,4	49966
3,5	4998	3,6	4998	3,7	49989	3,8	49993	3,9	49995
4,0	499968								
4,5	499997								
5,0	49999997								

Таблица значений функции $P_k(a) = \dfrac{a^k e^{-a}}{k!}$

k \ a	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5		0,000002	0,000015	0,000057	0,000158	0,000356
6			0,000001	0,000004	0,000013	0,000035
7					0,000001	0,000003

(*Продолжение*)

k \ a	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367879	0,135335	0,049787
1	0,347610	0,359463	0,365913	0,367879	0,270671	0,149361
2	0,121663	0,143785	0,164661	0,183940	0,270671	0,224042
3	0,028388	0,038343	0,049398	0,061313	0,180447	0,224042
4	0,004968	0,007669	0,011115	0,015328	0,090224	0,168031
5	0,000695	0,001227	0,002001	0,003066	0,036089	0,100819
6	0,000081	0,000164	0,000300	0,000511	0,012030	0,050409
7	0,000008	0,000019	0,000039	0,000073	0,003437	0,021604
8		0,000002	0,000004	0,000009	0,000859	0,008101
9				0,000001	0,000191	0,002701
10					0,000038	0,000810
11					0,000007	0,000221
12					0,000001	0,000055
13						0,000013
14						0,000003
15						0,000001

k \ a	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111
2	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998
3	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994
4	0,195367	0,175467	0,133853	0,091226	0,057252	0,033737
5	0,156293	0,175467	0,160623	0,127717	0,091604	0,060727
6	0,104194	0,146223	0,160623	0,149003	0,122138	0,091090
7	0,059540	0,104445	0,137677	0,149003	0,139587	0,117116
8	0,029770	0,065278	0,103258	0,130377	0,139587	0,131756
9	0,013231	0,036266	0,068838	0,101405	0,124077	0,131756
10	0,005292	0,018133	0,041303	0,070983	0,099262	0,118580
11	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020
12	0,000642	0,003434	0,011262	0,026350	0,048127	0,072765
13	0,000197	0,001321	0,005199	0,014188	0,029616	0,050376
14	0,000056	0,000472	0,002228	0,007094	0,016924	0,032384
15	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431
16	0,000004	0,000049	0,000334	0,001448	0,004513	0,010930

(*Окончание*)

k \ a	4,0	5,0	6,0	7,0	8,0	9,0
17	0,000001	0,000014	0,000118	0,000596	0,002124	0,005786
18		0,000004	0,000039	0,000232	0,000944	0,002893
19		0,000001	0,000012	0,000085	0,000397	0,001370
20			0,000004	0,000030	0,000159	0,000617
21			0,000001	0,000010	0,000061	0,000264
22				0,000003	0,000022	0,000108
23				0,000001	0,000008	0,000042
24					0,000003	0,000016
25					0,000001	0,000006
26						0,000002
27						0,000001

Таблица значений функции $\sum\limits_{m=0}^{k} \dfrac{a^m e^{-a}}{m!}$

k \ a	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,995321	0,982477	0,963063	0,938448	0,909796	0,878099
2	0,999845	0,998852	0,996390	0,992074	0,985612	0,977885
3	0,999996	0,999943	0,999724	0,999224	0,998248	0,997642
4	1,000000	0,999998	0,999974	0,999939	0,999828	0,999606
5	1,000000	1,000000	0,999999	0,999996	0,999986	0,999962
6	1,000000	1,000000	1,000000	1,000000	0,999999	0,999997
7	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000

k \ a	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367879	0,135335	0,049787
1	0,844195	0,808792	0,772483	0,735759	0,406006	0,199148
2	0,965858	0,952577	0,937144	0,919699	0,676677	0,423190

(*Окончание*)

k \ a	0,7	0,8	0,9	1,0	2,0	3,0
3	0,994246	0,990920	0,988542	0,981012	0,857124	0,647232
4	0,999214	0,998589	0,997657	0,996340	0,947348	0,815263
5	0,999909	0,999816	0,999658	0,999406	0,983437	0,916082
6	0,999990	0,999980	0,999958	0,999917	0,995467	0,966491
7	0,999998	0,999999	0,999997	0,999990	0,998904	0,988095
8	1,000000	1,000000	1,000000	0,999999	0,999763	0,996196
9				1,000000	0,999954	0,998897
10					0,999992	0,999707
11					0,999999	0,999928
12					1,000000	0,999983
13						0,999996
14						0,999999
15						0,100000

СПИСОК ЛИТЕРАТУРЫ

1. *Боровков А.А.* Теория вероятностей. – 2-е изд. – М.: Наука, 1986.
2. *Гихман И.И., Скороход А.В., Ядренко М.И.* Теория вероятностей и математическая статистика. – Киев: Вища школа, 1979.
3. *Гихман И.И., Скороход А.В.* Введение в теорию случайных процессов. – М.: Наука, 1977.
4. *Гнеденко Б.В., Колмогоров А.Н.* Предельные распределения для сумм независимых случайных величин. – М.–Л.: Гостехиздат, 1949.
5. *Вентцель А.Д.* Курс теории случайных процессов. – М.: Наука, 1975.
6. *Карлин С.* Основы теории случайных процессов. – М.: Мир, 1975.
7. *Кендалл М., Моран П.* Геометрические вероятности. – М.: Наука, 1972.
8. *Климов Г.П.* Теория вероятностей и математическая статистика. – М.: Изд-во МГУ, 1983.
9. *Климов Г.П., Кузьмин А.Д.* Вероятность, процессы, статистика: Задачи с решениями. – М.: Изд-во МГУ, 1985.
10. *Коваленко И.Н., Кузнецов Н.Ю., Шуренков В.М.* Случайные процессы; справочник. – Киев: Наукова думка, 1983.
11. *Коваленко И.Н., Филиппова А.А.* Теория вероятностей и математическая статистика. – М.: Высш. школа, 1973.
12. *Королюк В.С.* (редактор). Справочник по теории вероятностей и математической статистике. – Киев: Наукова думка, 1978.
13. *Лоэв М.* Теория вероятностей. – М.: ИЛ, 1962.
14. *Лукач Е.* Характеристические функции. – М.: Наука, 1979.
15. *Майстров Л.Е.* Развитие понятия вероятности. – М.: Наука, 1980.
16. *Мешалкин Л.Д.* Сборник задач по теории вероятностей. – М.: Изд-во МГУ, 1963.
17. *Невё Ж.* Математические основы теории вероятностей. – М.: Мир, 1969.
18. *Нейман Ю.* Вводный курс теории вероятностей и математической статистики. – М.: Наука, 1969.
19. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей. – 2-е изд. – М.: Наука, 1985.
20. *Розанов Ю.А.* Случайные процессы. – М.: Наука, 1971.
21. *Савельев Л.Я.* Комбинаторика и вероятность. – Новосибирск: Наука, 1975.
22. *Сантало Л.* Интегральная геометрия и геометрические вероятности. – М.: Наука, 1983.
23. *Севастьянов Б.А.* Курс теории вероятностей и математической статистики. – М.: Наука, 1982.
24. *Севастьянов Б.А., Чистяков В.П., Зубков А.М.* – Сборник задач по теории вероятностей. – М.: Наука, 1982.
25. *Стоянов Й.* и др. Ръководство за упражнения по теория на вероятностите. – София: Наука и изкуство, 1985.
26. *Феллер В.* Введение в теорию вероятностей и ее приложения. – М.: Мир, 1984.
27. *Чжун Кай-Лай.* Однородные цепи Маркова. – М.: Мир, 1964.
28. *Ширяев А.Н.* Вероятность. – М.: Наука, 1980.

Борис Владимирович Гнеденко

КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

Редактор *А.С. Чистопольский*
Художественный редактор *Т.Н. Кольченко*
Технические редакторы *С.Н. Баронина, О.Б. Черняк*
Корректоры *Н.П. Круглова, Т.В. Обод*

Набор осуществлен в издательстве
на наборно-печатающих автоматах

ИБ № 32527

Сдано в набор 25.08.87. Подписано к печати 16.12.87
Формат 60 х 88/16. Бумага книжно-журнальная
Гарнитура Пресс-Роман. Печать офсетная
Усл.печ.л. 27,44. Усл.кр.-отт. 27,44. Уч.-изд.л. 27,28
Тираж 53 000 экз. Тип. зак. 152. Цена 1 р. 20 к.

Ордена Трудового Красного Знамени
издательство "Наука"
Главная редакция физико-математической литературы
117071 Москва В-71, Ленинский проспект, 15

Типография им. Котлякова
издательства "Финансы и статистика"
Государственного комитета СССР
по делам издательств, полиграфии и книжной торговли
195273 Ленинград, ул. Руставели 13