ПРЕДИСЛОВИЕ

В основу настоящего учебного пособия положен трехсеместровый курс лекций, который читался автором в течение ряда лет на механико-математическом факультете Московского государственного университета и был частично издан ротапринтным способом под названием «Вероятность, статистика, случайные процессы, I, II», изд-во МГУ.

В соответствии с традицией первая часть курса (примерно один семестр) отводится на элементарную теорию вероятностей (глава I). Изложение начинается с построения вероятностных моделей с конечным числом исходов и введения основных вероятностных понятий таких, как элементарные события, события, вероятность, независимость, случайные величины, математические ожидания, корреляция, условные вероятности и др.

Многие вероятностно-статистические закономерности хорошо прослеживаются уже на примере простейшего случайного блуждания, порожденного схемой Бернулли. В связи с этим для этого случая излагаются как классические результаты (закон больших чисел, локальная и интегральная теоремы Муавра и Лапласа), так и более современные результаты (например, закон арксинуса).

Завершается первая глава рассмотрением зависимых случайных величин, образующих мартингал и марковскую цепь.

Главы II—IV являются расширенным изложением второй части курса (второй семестр). Здесь излагается (глава II) ставшая общепринятой аксиоматика теории вероятностей А. Н. Колмогорова и дается математический аппарат, составляющий арсенал средств современной теории вероятностей (σ-алгебры, меры и способы их задания, интеграл Лебега, случайные величины и случайные элементы, характеристические функции, условные математические ожидания относительно σ-алгебр, гауссовские системы и др.). Следует отметить, что два результата теории меры — теорема Каратеодори о продолжении меры и теорема Радона-Никодима — принимаются без доказательства.

Третья глава посвящается вопросам слабой сходимости вероятностных распределений и методу характеристических функций в доказательстве предельных теорем. Вводятся понятия относительной компактности и плотности семейства вероятностных распределений и доказывается (для случая числовой прямой) теорема Ю. В. Прохорова об эквивалентности этих понятий.

К этой же части курса отнесено рассмотрение свойств «с вероятностью единица» для последовательностей и сумм независимых случайных величин (глава IV). Приводятся доказательства законов «нуля или единицы» (Колмогоров, Хьюитт и Сэвидж), критерии сходимости рядов и даются условия справедливости усиленного закона больших чисел. Закон повторного логарифма формулируется для произвольных последовательностей независимых одинаково распределенных случайных величин с конечным вторым моментом и доказывается в предположении, что эти величины имеют гауссовское распределение.

Наконец, третья часть курса (главы V—VIII) отводится случайным процессам с дискретным временем (случайным последовательностям). Главы V и VI посвящены теории стационарных случайных последовательностей, где стационарность понимается как в узком, так и широком смыслах. Изложение теории стационарных в узком смысле случайных последовательностей ведется с привлечением понятий эргодической теории: сохраняющее меру преобразование, эргодичность, перемешивание. ... Приводится простое доказательство (данное А. Гарсиа) максимальной эргодической теоремы, что позволяет дать и простое доказательство эргодической теоремы Биркгофа — Хинчина.

Рассмотрение стационарных в широком смысле случайных последовательностей начинается с доказательства спектрального представления для ковариационной функции. Затем вводятся ортогональные стохастические меры, интегралы по ним и доказывается спектральное представление для самих последовательностей. Рассмотрен также ряд статистических задач: оценивание ковариационной функции и спектральной плотности, экстраполяция, интерполяция и фильтрация. В эту же главу включен также материал, относящийся к фильтру Калмана — Бьюси и его обобщениям.

В седьмой главе рассматриваются основные результаты теории мартингалов и родственных понятий. Излагаемый здесь материал стал включаться в традиционные курсы теории вероятностей лишь сравнительно недавно. В последней главе, посвященной марковским цепям, основное внимание уделяется вопросам асимптотического поведения цепей Маркова со счетным множеством состояний.

В конце каждого параграфа приводятся задачи, значимость которых может быть различной: в одних из них предлагается доказать утверждения, сформулированные, но не доказанные в основном тексте, другие содержат утверждения, используемые в последующем изложении, третьи преследуют цель дать дополнительные сведения к рассматриваемому кругу вопросов и, наконец, некоторые носят характер простых упражнений.

При составлении курса и настоящего пособия автор использовал разнообразную литературу по теории вероятностей. В историко-библиографической справке указываются как источники приводимых результатов, так и дополнительная литература, относящаяся к рассматриваемому материалу.

В книге применяется следующая нумерация и система ссылок. Каждый параграф содержит свою нумерацию теорем, лемм и формул (без указания номера главы и параграфа). При ссылке на соответствующий результат из другого параграфа той же главы применяется двойная нумерация, где первая цифра указывает номер параграфа (так, ссылка на формулу (2.10) означает формулу (10) из § 2). При ссылке на результаты из другой главы используется тройная нумерация (так, формула (II.4.3) означает формулу (3) из § 4 главы II).

Автор пользуется здесь случаем поблагодарить А. Н. Колмогорова, Б. В. Гнеденко, Ю. В. Прохорова, которые его учили и у которых он учился теории вероятностей и советами которых он имел возможность пользоваться. Автор приносит также свою признательность сотрудникам кафедр теории вероятностей и математической статистики механико-математического факультета МГУ и сотрудникам отдела теории вероятностей Математического института им. В. А. Стеклова АН СССР за обсуждения и советы.

Москва,
декабрь 1979

А. Ширяев

ВВЕДЕНИЕ

Предметом теории вероятностей является математический анализ случайных явлений, т. е. таких эмпирических феноменов, которые — при заданном комплексе условий — могут быть охарактеризованы тем, что

> Для них отсутствует *детерминистическая регулярность* (наблюдения над ними не всегда приводят к одним и тем же исходам)

и в то же самое время

> Они обладают некоторой *статистической регулярностью* (проявляющейся в статистической устойчивости частот).

Поясним сказанное на классическом примере «честного» подбрасывания «правильной» монеты. Ясно, что заранее невозможно с определенностью предсказать исход каждого подбрасывания. Результаты отдельных экспериментов носят крайне нерегулярный характер (то «герб», то «решетка») и кажется, что это лишает нас возможности познать какие-либо закономерности, связанные с этими экспериментами. Однако, если провести большое число «независимых» подбрасываний, то можно заметить, что для «правильной» монеты будет наблюдаться вполне определенная статистическая регулярность, проявляющаяся в том, что частота выпадания «герба» будет «близка» к $^1/_2$.

Статистическая устойчивость частот делает весьма правдоподобной гипотезу о возможности количественной оценки «случайности» того или иного события A, осуществляемого в результате экспериментов. Исходя из этого, теория вероятностей постулирует существование у события A определенной числовой характеристики $\mathsf{P}(A)$, называемой вероятностью этого события, естественное свойство которой должно состоять в том, что с ростом числа «независимых» испытаний (экспериментов) частота появления события A должна приближаться к $\mathsf{P}(A)$.

Применительно к рассмотренному примеру это означает, что вероятность события A, состоящего в выпадании «герба» при бросании «правильной» монеты, естественно считать равной $^1/_2$.

Число подобных примеров, в которых интуитивное представление о численном значении вероятности того или иного события складывается весьма легко, можно без труда приумножить. Однако все они будут носить сходный характер и сопровождаться неопределенными (пока) понятиями типа «честное» подбрасывание, «правильная» монета, «независимость» и т. п.

Призванная изучать количественные характеристики «случайности», теория вероятностей, как и всякая точная наука, стала таковой лишь тогда, когда было четко сформулировано понятие вероятностной модели, когда была создана ее аксиоматика. В этой связи естественно хотя бы кратко остановиться на основных этапах становления теории вероятностей.

Возникновение теории вероятностей как науки относится к середине XVII века и связано с именами Паскаля (1623—1662), Ферма (1601—1665), Гюйгенса (1629—1695). Хотя отдельные задачи, касающиеся подсчета шансов в азартных играх, рассматривались ранее — в XV—XVI вв. итальянскими математиками (Кардано, Пачоли, Тарталья и др.), первые общие методы решения таких задач были, по-видимому, даны в знаменитой переписке между Паскалем и Ферма, начавшейся в 1654 г., и в первой книге по теории вероятностей «De Ratiociniis in Aleae Ludo» («О расчетах в азартной игре»), опубликованной Гюйгенсом в 1657 г. Именно в этот период вырабатывается важное понятие «математическое ожидание», устанавливаются теоремы сложения и умножения вероятностей.

Истинная история теории вероятностей начинается с работы Я. Бернулли (1654—1705) «Ars Conjectandi» («Искусство предположения»), опубликованной в 1713 г., в которой была доказана (и вполне строго) первая предельная теорема теории вероятностей — закон больших чисел, и работы Муавра (1667—1754) «Miscellanea Analytica Supplementum» (примерный перевод может быть таков: «Аналитические методы» или «Аналитическая смесь»), 1730 г., в которой впервые была сформулирована и доказана (в симметричной схеме Бернулли) так называемая центральная предельная теорема.

Я. Бернулли был, вероятно, первым, кто осознал важность рассмотрения бесконечных последовательностей повторных испытаний и кто делал четкое различие между понятием вероятности события и частоты его появления. Муавру принадлежит заслуга в определении таких понятий, как независимость, математическое ожидание, условная вероятность.

В 1812 г. выходит большой трактат Лапласа (1749—1827) «Theorie Analytique des Probabilités» («Аналитическая теория вероятностей»), в которой он излагает свои собственные результаты в области теории вероятностей, а также результаты своих предшественников. В частности, он обобщил теорему Муавра на общий (несимметричный) случай схемы Бернулли, раскрыв тем самым более полным образом значение результата Муавра.

Весьма значителен вклад Лапласа, состоящий в применении вероятностных методов к теории ошибок наблюдений. Именно им была высказана плодотворная идея, что ошибка наблюдений должна рассматриваться как суммарный эффект сложения большого числа независимых элементарных ошибок. Отсюда следовало, что при достаточно общих условиях распределение ошибок наблюдений по крайней мере приближенно должно быть нормальным.

К этому же периоду в развитии теории вероятностей, когда центральное место в исследованиях занимали предельные теоремы, относятся работы Пуассона (1781—1840) и Гаусса (1777—1855).

С именем Пуассона в современной теории вероятностей связано понятие распределения и процесса, носящих его имя. Гауссу принадлежит заслуга создания теории ошибок и, в частности, обоснование одного из ее основных принципов — метода наименьших квадратов.

Следующий важный период в развитии теории вероятностей связан с именами П. Л. Чебышева (1821—1894), А. А. Маркова (1856—1922), А. М. Ляпунова (1857—1918), создавших эффективные методы доказательства предельных теорем для сумм независимых произвольно распределенных случайных величин.

Число публикаций Чебышева по теории вероятностей невелико — их всего четыре, но их роль в теории вероятностей и в создании классической русской школы теории вероятностей трудно переоценить.

«С методологической стороны основной переворот, совершенный Чебышевым, заключается не только в том, что он впервые с полной настойчивостью выдвинул требование абсолютной строгости доказательства предельных теорем... но главным образом в том, что Чебышев всюду стремился получить точные оценки отклонений от предельных закономерностей, возможных при хотя бы и большом, но конечном числе испытаний, в виде безусловно правильных при любом числе испытаний неравенств» (Колмогоров А. Н. [30]). До Чебышева основной интерес в теории вероятностей был связан с подсчетом вероятностей случайных событий. Им же впервые было ясно осознана и использована вся сила понятия случайной величины и математического ожидания случайной величины.

Лучшим выразителем идей Чебышева был его ближайший ученик Марков, которому принадлежит несомненная заслуга доведения до полной ясности результатов своего учителя. Значительным вкладом Маркова в теорию вероятностей явилось начатое им исследование предельных теорем для сумм зависимых случайных величин и создание одного из новых разделов теории вероятностей — теории зависимых случайных величин, связанных, как теперь принято говорить, в цепь Маркова.

«...классический курс исчисления вероятностей А. А. Маркова и его оригинальные мемуары, являющиеся образцом точности и ясности изложения, в наибольшей степени содействовали превращению теории вероятностей в одну из самых совершенных областей математики и широкому распространению направления и методов Чебышева» (Бернштейн С. Н. [3]).

Для доказательства центральной предельной теоремы теории вероятностей (о сходимости к нормальному закону) Чебышев и Марков применили так называемый метод моментов. При более общих условиях и более простым методом — методом характеристических функций эта теорема была получена Ляпуновым. Последующее развитие теории показало, что метод характеристических функций является мощным аналитическим средством доказательства самых разнообразных предельных теорем.

Современный период в развитии теории вероятностей начинается с установления аксиоматики. Первые работы в этом направлении принадлежат С. Н. Бернштейну (1880—1968), Р. Мизесу (1883—1953), Э. Борелю (1871—1956). В 1933 г. вышла книга А. Н. Колмогорова «Основные понятия теории вероятностей», в которой была предложена аксиоматика, получившая всеобщее признание и позволившая охватить не только все классические разделы теории вероятностей, но и дать строгую основу для развития ее новых разделов, вызванных запросами естествознания и связанных с бесконечномерными распределениями.

Изложение в настоящей книге основано на аксиоматическом подходе Колмогорова. При этом, чтобы формально-логическая сторона дела не заслоняла интуитивных представлений, наше изложение начинается с элементарной теории вероятностей, «элементарность» которой состоит в том, что в соответствующих вероятностных моделях рассматриваются эксперименты лишь с конечным числом исходов. После этого мы даем изложение основ теории вероятностей в ее наиболее общем виде.

Начиная с 20—30 годов в теории вероятностей бурно развивается один из ее новых разделов — теория случайных процессов, занимающаяся изучением семейств случайных величин, эволюционирующих во времени. Была создана теория марковских процессов, теория стационарных процессов, теория мартингалов, теория предельных теорем для случайных процессов. К недавнему времени относится возникновение теории информации.

Основное внимание в настоящей книге уделяется случайным процессам с дискретным временем — случайным последовательностям. Однако тот материал, который излагается во второй

ГЛАВА I

ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

§ 1. Вероятностная модель эксперимента с конечным числом исходов

1. Рассмотрим некоторый эксперимент, все мыслимые исходы которого описываются конечным числом различных исходов $\omega_1, \ldots, \omega_N$. Для нас несущественна реальная природа этих исходов, важно лишь то, что их число N конечно.

Исходы $\omega_1, \ldots, \omega_N$ будем называть *элементарными событиями*, а их совокупность

$$\Omega = \{\omega_1, \ldots, \omega_N\}$$

(конечным) *пространством элементарных событий* или *пространством исходов*.

Выделение пространства элементарных событий представляет собой первый шаг в формулировании понятия вероятностной модели того или иного эксперимента. Рассмотрим несколько примеров описания структуры пространства элементарных событий.

Пример 1. При однократном подбрасывании монеты пространство исходов Ω состоит из двух точек:

$$\Omega = \{\text{Г}, \text{Р}\},$$

где Г — «герб», Р — «решетка». (Мы исключаем возможности типа «монета стала на ребро», «монета исчезла» и т. п.)

Пример 2. При n-кратном подбрасывании монеты пространство элементарных событий

$$\Omega = \{\omega\colon \omega = (a_1, \ldots, a_n),\ a_i = \text{Г или Р}\}$$

и общее число $N(\Omega)$ исходов равно 2^n.

Пример 3. Пусть сначала подбрасывается монета. Если выпадет «герб», то бросается шестигранная кость (с цифрами 1, 2, 3, 4, 5, 6), если же выпадает «решетка», то снова подбрасывается монета. Пространство элементарных событий данного эксперимента будет таким:

$$\Omega = \{\text{Г1, Г2, Г3, Г4, Г5, Г6, РГ, РР}\}.$$

Рассмотрим теперь более сложные примеры, связанные с разными способами выбора n шаров из урны, содержащей M различных шаров.

2. Пример 4. *Выбор с возвращением.* Так называют эксперимент, в котором на каждом шаге извлеченный шар возвращается обратно. В этом случае каждая выборка из n шаров может быть записана в виде $(a_1, \ldots, a_n)$, где a_i — номер шара, извлеченного на i-м шаге. Понятно, что в случае выбора с возвращением каждое a_i может принимать любое из M значений $1, 2, \ldots, M$. Описание пространства элементарных событий существенно зависит от того, считаем ли мы выборки тождественного состава такие, как, скажем, (4, 1, 2, 1) и (1, 4, 2, 1), различными или одинаковыми. В связи с этим принято различать два случая: *упорядоченные* выборки и *неупорядоченные* выборки. В первом случае выборки, состоящие из одних и тех же элементов, но отличающиеся порядком следования этих элементов, объявляются различными. Во втором случае порядок следования элементов не принимается во внимание и такие выборки объявляются тождественными. Чтобы подчеркнуть, какие конкретно выборки мы рассматриваем, будем для упорядоченных выборок использовать обозначение $(a_1, \ldots, a_n)$, а для неупорядоченных — $[a_1, \ldots, a_n]$.

Итак, в случае упорядоченных выборок пространство элементарных событий Ω имеет следующую структуру:

$$\Omega = \{\omega\colon\ \omega = (a_1, \ldots, a_n),\quad a_i = 1, \ldots, M\}$$

и число (различных) исходов

$$N(\Omega) = M^n. \tag{1}$$

Если же рассматриваются неупорядоченные выборки, то

$$\Omega = \{\omega\colon\ \omega = [a_1, \ldots, a_n],\quad a_i = 1, \ldots, M\}.$$

Понятно, что $N(\Omega)$ (различных) неупорядоченных выборок меньше, чем число упорядоченных. Покажем, что для этого случая

$$N(\Omega) = C^n_{M+n-1}, \tag{2}$$

где $C^l_k \equiv \dfrac{k!}{l!\,(k-l)!}$ — число сочетаний из k элементов по l.

Будем вести доказательство по индукции. Обозначим $N(M, n)$ число интересующих нас исходов. Ясно, что для всех $k \leqslant M$

$$N(k, 1) = k = C^1_k.$$

Предположим теперь, что $N(k, n) = C^k_{k+n-1}$, $k \leqslant M$, и покажем, что эта формула остается справедливой при замене n на $n+1$.

При рассмотрении неупорядоченных выборок $[a_1, \ldots, a_{n+1}]$ можно считать, что их элементы расположены в порядке неубывания: $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_{n+1}$. Очевидно, что число неупорядоченных выборок с $a_1 = 1$ равно $N(M, n)$, с $a_1 = 2$ равно $N(M-1, n)$ и т. д. Следовательно,

$$
\begin{aligned}
N(M, n+1) &= N(M, n) + N(M-1, n) + \ldots + N(1, n) = \\
&= C^n_{M+n-1} + C^n_{M-1+n-1} + \ldots + C^n_n = \\
&= \left(C^{n+1}_{M+n} - C^{n+1}_{M+n-1}\right) + \left(C^{n+1}_{M-1+n} - C^{n+1}_{M-1+n-1}\right) + \ldots \\
&\qquad \ldots + \left(C^{n+1}_{n+1} - C^n_n\right) + C^n_n = C^{n+1}_{M+n},
\end{aligned}
$$

где мы воспользовались следующим легко проверяемым свойством биномиальных коэффициентов:

$$C^{l-1}_k + C^l_k = C^l_{k+1}.$$

Пример 5. *Выбор без возвращения.* Будем предполагать, что $n \leqslant M$ и что извлеченные шары обратно не возвращаются. В этом случае также рассматриваются две возможности, связанные с различением упорядоченных и неупорядоченных выборок.

В случае упорядоченных выборок без возвращения пространство исходов

$$\Omega = \{\omega\colon \omega = (a_1, \ldots, a_n),\ a_1 \neq a_2 \neq \ldots \neq a_n,\quad a_i = 1, \ldots, M\},$$

а число элементов этого множества (называемых *размещениями*) равно $M(M-1)\ldots(M-n+1)$. Для этого числа используется обозначение $(M)_n$, или A^n_M, называемое «числом размещений из M по n».

В случае неупорядоченных выборок (называемых *сочетаниями*) пространство исходов

$$\Omega = \{\omega\colon \omega = [a_1, \ldots, a_n],\ a_1 \neq a_2 \neq \ldots \neq a_n,\ a_i = 1, \ldots, M\},$$

состоит из

$$N(\Omega) = C^n_M \tag{3}$$

элементов. Действительно, из каждого неупорядоченного набора $[a_1, \ldots, a_n]$, состоящего из различных элементов, можно получить ровно $n!$ упорядоченных наборов. Следовательно,

$$N(\Omega) \cdot n! = (M)_n$$

и, значит,

$$N(\Omega) = \frac{(M)_n}{n!} = C^n_M.$$

Результаты о числе исходов в случае n извлечений из урны с M шарами сведем в табл. 1.

Таблица 1

M^n	C^n_{M+n-1}	С возвращением
$(M)_n$	C^n_M	Без возвращения
Упорядоченный	Неупорядоченный	Выбор / Набор

Для случая $M=3$ и $n=2$ структура соответствующих пространств элементарных событий приводится в табл. 2.

Таблица 2

(1,1) (1,2) (1,3) (2,1) (2,2) (2,3) (3,1) (3,2) (3,3)	[1,1] [2,2] [3,3] [1,2] [1,3] [2,3]	С возвращением
(1,2) (1,3) (2,1) (2,3) (3,1) (3,2)	[1,2] [1,3] [2,3]	Без возвращения
Упорядоченный	Неупорядоченный	Выбор / Набор

Пример 6. *Размещение дробинок по ячейкам.* Рассмотрим вопрос о структуре пространства элементарных событий в задаче размещения n дробинок (шаров и т. п.) по M ячейкам (ящикам и т. п.). В статистической физике подобная задача возникает, например, при изучении распределения n частиц (это могут быть протоны, электроны, ...) по M состояниям (это могут быть энергетические уровни).

Пусть ячейкам присвоены номера $1, 2, \ldots, M$, и предположим сначала, что дробинки различимы (имеют номера $1, 2, \ldots, n$). Тогда распределение n дробинок по M ячейкам полностью описывается (упорядоченным) набором $(a_1, \ldots, a_n)$, где a_i — номер ячейки, куда попала дробинка с номером i. Если же рассматриваемые дробинки неразличимы, то их распределение по M ячейкам полностью описывается (неупорядоченным) набором $[a_1, \ldots, a_n]$, где a_i — номер ячейки, в которую попала дробинка на i-м шаге.

Сравнивая рассматриваемую ситуацию с примерами 4 и 5, видим, что имеют место следующие соответствия:

(*упорядоченные выборки*) $\leftrightarrow$ (*различимые дробинки*),
(*неупорядоченные выборки*) $\leftrightarrow$ (*неразличимые дробинки*),

означающие, что случаю упорядоченных (неупорядоченных) выборок в задаче выбора n шаров из урны с M шарами соответствует (один и только один) случай расположения различимых (неразличимых) дробинок в задаче размещения n дробинок по M ячейкам.

Аналогичный смысл имеют также следующие соответствия:

(*выбор с возвращением*) $\leftrightarrow$ (*в ячейке может находиться любое число дробинок*),

(*выбор без возвращения*) $\leftrightarrow$ (*в ячейке может находиться не более одной дробинки*).

Из этих соответствий можно сконструировать соответствия типа:

(*неупорядоченные выборки в задаче выбора без возвращения*) $\leftrightarrow$ (*неразличимые дробинки в задаче их размещения по ячейкам, когда в каждой из них не может находиться более одной дробинки*)

и т. д., что дает возможность использовать примеры 4 и 5 для описания структуры пространства элементарных событий в задаче распределения различимых и неразличимых дробинок по ячейкам с запретом (в ячейке может находиться не более одной дробинки) или без запрета (в ячейке может находиться любое число дробинок).

Табл. 3 показывает структуру расположения двух дробинок по трем ячейкам. В случае различимых дробинок они обозначаются Б (белая) и Ч (черная). В случае неразличимых дробинок их наличие в ячейке обозначается знаком $+$.

Указанная выше двойственность между рассматриваемыми задачами позволяет очевидным образом найти число исходов в задаче размещения дробинок по ячейкам. Соответствующие результаты, включающие в себя также и результаты табл. 1, сведены в табл. 4.

В статистической физике говорят, что различимые (неразличимые) частицы, не подчиняющиеся принципу запрета Паули, удовлетворяют (физической) статистике Максвелла — Больцмана (соответственно — статистике Бозе — Эйнштейна). Если же частицы неразличимы и подчиняются принципу запрета, то — статистике Ферми — Дирака (см. табл. 4). Известно, например, что электроны, протоны и нейтроны подчиняются статистике Ферми — Дирака. Фотоны и пи-мезоны — статистике Бозе — Эйнштейна. Известно также, что случай различимых частиц, подчиняющихся принципу запрета, в физике не встречается.

Таблица 3

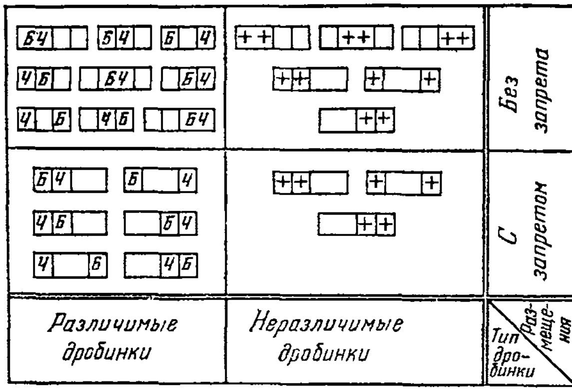


Таблица 4

$N(\Omega)$ в задаче размещения n дробинок по M ячейкам			
Размещение \ Тип дробинок	Различимые дробинки	Неразличимые дробинки	
Без запрета	M^n (статистика Максвелла — Больцмана)	C^n_{M+n-1} (статистика Бозе — Эйнштейна)	С возвращением
С запретом	$(M)_n$	C^n_M (статистика Ферми — Дирака)	Без возвращения
	Упорядоченные выборки	Неупорядоченные выборки	Выбор \ Набор
	$N(\Omega)$ в задаче выбора n шаров из урны с M шарами		

3. Наряду с понятием пространства элементарных событий введем теперь важное понятие *события*.

Экспериментаторы обычно интересуются не тем, какой конкретно исход имеет место в результате испытания, а тем, принадлежит ли исход тому или иному подмножеству всех исходов. Все те подмножества $A \subseteq \Omega$, для которых по условиям эксперимента возможен ответ одного из двух типов: «исход $\omega \in A$» или «исход $\omega \notin A$», — будем называть *событиями*.

Например, пусть осуществляется трехкратное подбрасывание монеты. Пространство всех исходов Ω состоит из восьми точек

$$\Omega = \{ГГГ,\ ГГР,\ \ldots,\ РРР\}$$

и если мы в состоянии записать (зафиксировать, «измерить» и т. п.) результаты всех трех подбрасываний, то, скажем, множество

$$A = \{ГГГ,\ ГГР,\ ГРГ,\ РГГ\}$$

является событием, состоящим в том, что выпадет по крайней мере два «герба». Однако если мы можем зафиксировать лишь результат первого подбрасывания, то рассматриваемое множество A нельзя будет называть событием, поскольку нельзя дать ни утвердительного, ни отрицательного ответа на вопрос о том, принадлежит ли конкретный исход ω множеству A.

Отправляясь от некоторой заданной системы множеств, являющихся событиями, можно образовывать новые события, отвечающие конструкциям высказываний с логическими связками «или», «и» и «не», чему на языке теории множеств соответствуют операции «объединения», «пересечения», «дополнения».

Если A и B — два множества, то под их *объединением*, обозначаемым $A \cup B$, понимается множество, состоящее из точек, входящих или в A, или в B:

$$A \cup B = \{\omega \in \Omega:\ \omega \in A \quad \text{или} \quad \omega \in B\}.$$

На языке теории вероятностей $A \cup B$ — событие, состоящее в том, что произошло или событие A, или событие B.

Пересечение двух множеств A и B, обозначаемое $A \cap B$, или AB, есть множество, состоящее из точек, входящих и в A и в B:

$$A \cap B = \{\omega \in \Omega:\ \omega \in A \ \text{и}\ \omega \in B\}.$$

Событие $A \cap B$ состоит в том, что одновременно произошло и событие A, и событие B.

Так, если $A = \{ГГ,\ ГР,\ РГ\}$ и $B = \{РР,\ РГ,\ ГР\}$, то

$$A \cup B = \{ГГ,\ ГР,\ РГ,\ РР\} \quad (= \Omega),$$
$$A \cap B = \{РГ,\ ГР\}.$$

Если A — некоторое подмножество Ω, то под его *дополнением*, обозначаемым в дальнейшем $\bar{A}$, понимается множество точек из Ω, не входящих в A.

Если через $B \setminus A$ обозначать *разность* множеств B и A (т. е. множество точек, входящих в B и не входящих в A), то $\bar{A} = \Omega \setminus A$. На языке теории вероятностей $\bar{A}$ — это событие, состоящее в ненаступлении события A. Так, если событие $A = \{$ГГ, ГР, РГ$\}$, то $\bar{A} = \{$РР$\}$ — событие, состоящее в том, что подряд выпадут две «решетки».

Множества A и $\bar{A}$ не имеют общих точек и, следовательно, множество $A \cap \bar{A}$ является пустым. Для пустого множества будем использовать обозначение $\varnothing$. В теории вероятностей множество $\varnothing$ называется *невозможным* событием. Множество Ω естественно назвать необходимым, или *достоверным*, событием.

Объединение $A \cup B$ множеств A и B в том случае, когда они не пересекаются ($AB = \varnothing$), называется *суммой* множеств A и B и обозначается $A + B$.

Если рассматривается некоторая система $\mathscr{A}_0$ множеств $A \subseteq \Omega$, то с помощью теоретико-множественных операций $\cup$, $\cap$ и $\setminus$ можно из элементов $\mathscr{A}_0$ построить новую систему множеств, которые также являются событиями. Присоединяя к этим событиям достоверное и невозможное события Ω и $\varnothing$, получаем систему множеств $\mathscr{A}$, которая является *алгеброй*, т. е. такой системой подмножеств множества Ω, что

1) $\Omega \in \mathscr{A}$,

2) если $A \in \mathscr{A}$, $B \in \mathscr{A}$, то множества $A \cup B$, $A \cap B$, $A \setminus B$ также принадлежит $\mathscr{A}$.

Из сказанного следует, что в качестве систем событий целесообразно рассматривать такие системы множеств, которые являются алгебрами. Именно такие системы событий мы и будем рассматривать далее.

Остановимся на некоторых примерах алгебр событий:

a) $\{\Omega, \varnothing\}$ — система, состоящая из Ω и пустого множества (так называемая *тривиальная* алгебра);

b) $\{A, \bar{A}, \Omega, \varnothing\}$ — система, порожденная событием A;

c) $\mathscr{A} = \{A\colon\ A \subseteq \Omega\}$ — совокупность *всех* (включая и пустое множество $\varnothing$) подмножеств Ω.

Нетрудно заметить, что все эти алгебры событий получены по следующему принципу.

Будем говорить, что система множеств

$$\mathscr{D} = \{D_1, \ldots, D_n\}$$

образует *разбиение* множества Ω, а D_i являются *атомами* этого разбиения, если множества D_i непусты, попарно не пересекаются и их сумма равна Ω:

Если, например, множество Ω состоит из трех точек, $\Omega = \{1, 2, 3\}$, то существует пять различных разбиений:

$$\begin{array}{lll}
\mathscr{D}_1 = \{D_1\} & \text{с } D_1 = \{1, 2, 3\}; & \\
\mathscr{D}_2 = \{D_1, D_2\} & \text{с } D_1 = \{1, 2\}, & D_2 = \{3\}; \\
\mathscr{D}_3 = \{D_1, D_2\} & \text{с } D_1 = \{1, 3\}, & D_2 = \{2\}; \\
\mathscr{D}_4 = \{D_1, D_2\} & \text{с } D_1 = \{2, 3\}, & D_2 = \{1\}; \\
\mathscr{D}_5 = \{D_1, D_2, D_3\} & \text{с } D_1 = \{1\}, & D_2 = \{2\}, \ D_3 = \{3\}.
\end{array}$$

(По поводу общего числа разбиений конечного множества см. задачу 2.)

Если рассматривать всевозможные объединения множеств из $\mathscr{D}$, то вместе с пустым множеством $\varnothing$ полученная система множеств будет алгеброй, которая называется *алгеброй порожденной разбиением* $\mathscr{D}$ и обозначается $\alpha(\mathscr{D})$. Таким образом, элементы алгебры $\alpha(\mathscr{D})$ составляются из пустого множества и сумм множеств, являющихся атомами разбиения $\mathscr{D}$.

Итак, если $\mathscr{D}$ — некоторое разбиение, то ему однозначным образом ставится в соответствие алгебра $\mathscr{B} = \alpha(\mathscr{D})$.

Справедливо и обратное утверждение. Пусть $\mathscr{B}$ — некоторая алгебра подмножеств конечного пространства Ω. Тогда найдется и притом единственное разбиение $\mathscr{D}$, атомы которого являются элементами алгебры $\mathscr{B}$, и такое, что $\mathscr{B} = \alpha(\mathscr{D})$. В самом деле, пусть множество $D \in \mathscr{B}$ и обладает тем свойством, что для всякого $B \in \mathscr{B}$ множество $D \cap B$ или совпадает с D, или является пустым множеством. Тогда совокупность таких множеств D образует разбиение $\mathscr{D}$ с требуемым свойством $\alpha(\mathscr{D}) = \mathscr{B}$. В случае примера а) в качестве $\mathscr{D}$ берется тривиальное разбиение, состоящее лишь из одного множества $D_1 = \Omega$; в случае b) $\mathscr{D} = \{A, \bar{A}\}$. Самое мелкое разбиение $\mathscr{D}$, составленное из одноточечных множеств $\{\omega_i\}$, $\omega_i \in \Omega$, порождает алгебру в примере с), т. е. алгебру всех подмножеств Ω.

Покажем, что если пространство Ω состоит, как было предположено выше, из конечного числа точек $\omega_1, \ldots, \omega_N$, то общее число $N(\mathscr{A})$ множеств, составляющих систему $\mathscr{A}$, равно 2^N. Действительно, каждое непустое множество $A \in \mathscr{A}$ может быть представлено в виде $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$, где $\omega_{i_j} \in \Omega$, $1 \leqslant k \leqslant N$. Поставим в соответствие этому множеству последовательность, состоящую из нулей и единиц:

$$(0, \ldots, 0, 1, 0, \ldots, 0, 1, \ldots),$$

где на $i_1, \ldots, i_k$ местах стоят единицы, а на остальных — нули. Тогда при фиксированном k число различных множеств A вида $\{\omega_{i_1}, \ldots, \omega_{i_k}\}$ совпадает с числом способов, которыми можно k единиц (k неразличимых дробинок) разместить по N местам (по N

ячейкам). Согласно табл. 4 (см. правую нижнюю клетку) число таких способов равно C_N^k. Отсюда (с учетом пустого множества) находим, что

$$N(\mathscr{A}) = 1 + C_N^1 + \ldots + C_N^N = (1+1)^N = 2^N.$$

4. Пока мы сделали два первых шага к определению вероятностной модели эксперимента с конечным числом исходов: выделили пространство элементарных событий и некоторую систему $\mathscr{A}$ его подмножеств, образующих алгебру и называемых событиями. Сделаем теперь последний шаг, а именно припишем каждому элементарному событию (исходу) $\omega_i \in \Omega$, $i = 1, \ldots, N$, некоторый «вес», обозначаемый $p(\omega_i)$, и называемый *вероятностью* исхода ω_i, который будем считать удовлетворяющим следующим условиям:

a) $0 \leqslant p(\omega_i) \leqslant 1$ (*неотрицательность*),

b) $p(\omega_1) + \ldots + p(\omega_N) = 1$ (*нормированность*).

Отправляясь от заданных вероятностей $p(\omega_i)$ исходов ω_i, определим *вероятность* $\mathsf{P}(A)$ любого события $A \in \mathscr{A}$ по формуле

$$\mathsf{P}(A) = \sum_{\{i:\ \omega_i \in A\}} p(\omega_i). \tag{4}$$

Наконец, скажем, что тройка

$$(\Omega, \mathscr{A}, \mathsf{P}),$$

где $\Omega = \{\omega_1, \ldots, \omega_N\}$, $\mathscr{A}$ — некоторая алгебра подмножеств Ω, $\mathsf{P} = \{\mathsf{P}(A);\ A \in \mathscr{A}\}$, определяет (задает) *вероятностную модель*, или *вероятностное пространство*, эксперимента с (конечным) пространством исходов Ω и алгеброй событий $\mathscr{A}$.

Из определения (4) вытекают следующие свойства вероятностей

$$\mathsf{P}(\varnothing) = 0, \tag{5}$$

$$\mathsf{P}(\Omega) = 1, \tag{6}$$

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B). \tag{7}$$

В частности, если $A \cap B = \varnothing$, то

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) \tag{8}$$

и

$$\mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A). \tag{9}$$

5. При построении вероятностных моделей в конкретных ситуациях выделение пространства элементарных событий Ω и алгебры событий $\mathscr{A}$, как правило, не является сложной задачей. При этом в элементарной теории вероятностей в качестве алгебры $\mathscr{A}$ обычно берется алгебра *всех* подмножеств Ω. Труднее обстоит дело с вопросом о том, как задавать вероятности элементарных событий.

В сущности, ответ на этот вопрос лежит вне рамок теории вероятностей, и мы его подробно не рассматриваем, считая, что основной нашей задачей является не вопрос о том, как приписывать исходам те или иные вероятности, а вычисление вероятностей сложных событий (событий из $\mathscr{A}$) по вероятностям элементарных событий.

С математической точки зрения ясно, что в случае конечного пространства элементарных событий с помощью приписывания исходам $\omega_1, \ldots, \omega_N$ неотрицательных чисел $p_1, \ldots, p_N$, удовлетворяющих условию $p_1+\ldots+p_N=1$, мы получаем все мыслимые (конечные) вероятностные пространства.

Правильность же назначенных для конкретной ситуации значений $p_1, \ldots, p_N$ может быть до известной степени проверена с помощью рассматриваемого далее закона больших чисел, согласно которому в длинных сериях «независимых» экспериментов, происходящих при одинаковых условиях, частоты появления элементарных событий «близки» к их вероятностям.

В связи с трудностью, связанной с вопросом о приписывании исходам значений их вероятностей, отметим, что существует много практических ситуаций, в которых из соображений симметрии представляется разумным рассматривать все мыслимые исходы как равновозможные. Поэтому, если пространство элементарных исходов Ω состоит из точек $\omega_1, \ldots, \omega_N$, где $N<\infty$, то полагают

$$p(\omega_1)=\ldots=p(\omega_N)=1/N,$$

и, следовательно, для любого события $A \in \mathscr{A}$

$$\mathsf{P}(A)=N(A)/N, \tag{10}$$

где $N(A)$ — число элементарных событий, составляющих A.

Такой способ задания вероятностей носит название *классического*. Ясно, что в этом случае подсчет вероятностей $\mathsf{P}(A)$ сводится к подсчету числа исходов, приводящих к событию A. Делают это обычно комбинаторными методами, в связи с чем *комбинаторика*, имеющая дело с конечными множествами, занимает значительное место в вероятностном исчислении.

Пример 6. *Задача о совпадениях.* Пусть урна содержит M шаров, занумерованных числами $1, 2, \ldots, M$. Производится выбор объема n с возвращением, при этом рассматриваемые выборки считаются упорядоченными. Ясно, что в этом случае

$$\Omega=\{\omega\colon\ \omega=(a_1, \ldots, a_n),\ a_i=1, \ldots, M\}$$

и $N(\Omega)=M^n$. В соответствии с классическим способом задания вероятностей будем считать все M^n исходов равновероятными и поставим вопрос о том, какова вероятность события

$$A=\{\omega\colon\ a_1 \neq a_2 \neq \ldots \neq a_n\},$$

т. е. события, заключающегося в отсутствии повторений. Понятно, что $N(A)=M(M-1)\ldots(M-n+1)$ и, значит,

$$\mathsf{P}(A)=\frac{(M)_n}{M^n}=\left(1-\frac{1}{M}\right)\left(1-\frac{2}{M}\right)\ldots\left(1-\frac{n-1}{M}\right). \qquad (11)$$

Эта задача допускает следующую интересную интерпретацию. Пусть в классе находится n учеников. Будем считать, что день рождения каждого ученика приходится на один из 365 дней и любой день равновозможен. Спрашивается, какова вероятность P_n того, что в классе найдутся по крайней мере два ученика, дни рождения которых совпадают? Если рассматривать выбор дня рождения как выбор шара из урны с $M=365$ шарами, то согласно (11)

$$P_n=1-\frac{(365)_n}{365^n}.$$

Следующая таблица дает значения вероятностей P_n для некоторых n:

n	4	16	22	23	40	64
P_n	0,016	0,284	0,476	0,507	0,891	0,997

Интересно отметить, что (вопреки ожидаемому!) размер класса, где с вероятностью 1/2 найдутся по крайней мере два ученика с совпадающими днями рождения, не столь уж велик: он равен всего лишь 23.

Пример 7. *Выигрыш в лотерею.* Рассмотрим лотерею, устроенную по следующему принципу. Имеется M билетов, занумерованных числами $1, 2, \ldots, M$, из которых n билетов с номерами $1, 2, \ldots, n$ являются выигрышными ($M \geqslant 2n$). Вы покупаете n билетов, и спрашивается, какова вероятность (обозначим ее P) того, что по крайней мере один билет будет выигрышным?

Поскольку порядок, в котором извлекаются билеты, не играет роли с точки зрения наличия или отсутствия в купленном наборе выигрышных билетов, то следует считать, что пространство элементарных событий имеет следующую структуру:

$$\Omega=\{\omega\colon\ \omega=[a_1, \ldots, a_n],\ a_1\neq a_2\neq\ldots\neq a_n,\ a_i=1, \ldots, M\}.$$

Согласно табл. 1 $N(\Omega)=C_M^n$. Пусть теперь

$$A_0=\{\omega\colon\ \omega=[a_1, \ldots, a_n],\ a_1\neq a_2\neq\ldots\neq a_n,\ a_i=n+1, \ldots, M\}$$

— событие, состоящее в том, что среди купленных билетов нет выигрышных. Опять-таки, согласно табл. 1, $N(A_0) = C^n_{M-n}$. Поэтому

$$\mathsf{P}(A_0) = \frac{C^n_{M-n}}{C^n_M} = \frac{(M-n)_n}{M_{(n)}} = \left(1 - \frac{n}{M}\right)\left(1 - \frac{n}{M-1}\right)\ldots\left(1 - \frac{n}{M-n+1}\right)$$

и, значит,

$$P = 1 - \mathsf{P}(A_0) = 1 - \left(1 - \frac{n}{M}\right)\left(1 - \frac{n}{M-1}\right)\ldots\left(1 - \frac{n}{M-n+1}\right).$$

Если $M = n^2$ и $n \to \infty$, то $\mathsf{P}(A_0) \to e^{-1}$ и

$$P \to 1 - e^{-1} \approx 0{,}632,$$

где сходимость довольно быстрая: уже при $n = 10$ вероятность $P = 0{,}670$.

6. Задачи.

1. Установите справедливость следующих свойств операций $\cup$ и $\cap$:

$A \cup B = B \cup A$, $AB = BA$ (коммутативность),
$A \cup (B \cup C) = (A \cup B) \cup C$, $A(BC) = (AB)C$ (ассоциативность),
$A(B \cup C) = AB \cup AC$, $A \cup (BC) = (A \cup B)(A \cup C)$ (дистрибутивность),
$A \cup A = A$, $AA = A$ (идемпотентность).

Показать также, что

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{AB} = \bar{A} \cup \bar{B}.$$

2. Пусть множество Ω состоит из N элементов. Показать, что общее число $d(N)$ различных разбиений множества Ω определяется формулой

$$d(N) = e^{-1} \sum_{k=0}^{\infty} \frac{k^N}{k!}. \tag{12}$$

(Указание. Доказать, что

$$d(N) = \sum_{k=0}^{N-1} C^k_{N-1}\, d(k), \quad \text{где } d(0) = 1,$$

и затем проверить, что ряды в (12) удовлетворяют этим рекуррентным соотношениям.)

3. Для любой конечной системы множеств $A_1, \ldots, A_n$

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

4. Пусть A и B — два события. Показать, что $A\bar{B} \cup B\bar{A}$ есть событие, состоящее в том, что произойдет в точности одно из

событий A или B. При этом

$$\mathsf{P}(A\bar{B} \cup B\bar{A}) = \mathsf{P}(A) + \mathsf{P}(B) - 2\mathsf{P}(AB).$$

5. Пусть $A_1, \ldots, A_n$ — события и величины $S_0, S_1, \ldots, S_n$ определены следующим образом: $S_0 = 1$,

$$S_r = \sum_{J_r} \mathsf{P}(A_{k_1} \cap \ldots \cap A_{k_r}), \qquad 1 \leqslant r \leqslant n,$$

где суммирование распространяется по неупорядоченным подмножествам $J_r = [k_1, \ldots, k_r]$ множества $\{1, \ldots, n\}$.

Пусть B_m — событие, состоящее в том, что одновременно произойдет в точности m событий из $A_1, \ldots, A_n$. Показать, что

$$\mathsf{P}(B_m) = \sum_{r=m}^{n} (-1)^{r-m} C_r^m S_r.$$

В частности, для $m = 0$

$$\mathsf{P}(B_0) = 1 - S_1 + S_2 - \ldots \pm S_n.$$

Показать также, что вероятность того, что одновременно произойдет по крайней мере m событий из $A_1, \ldots, A_n$, равна

$$\mathsf{P}(B_m) + \ldots + \mathsf{P}(B_n) = \sum_{r=m}^{n} (-1)^{r-m} C_{r-1}^{m-1} S_r.$$

В частности, вероятность того, что произойдет по крайней мере одно из событий $A_1, \ldots, A_n$ равна

$$\mathsf{P}(B_1) + \ldots + \mathsf{P}(B_n) = S_1 - S_2 + \ldots \pm S_n.$$

§ 2. Некоторые классические модели и распределения

1. Биномиальное распределение. Предположим, что монета подбрасывается n раз и результат наблюдений записывается в виде упорядоченного набора $(a_1, \ldots, a_n)$, где $a_i = 1$ в случае появления «герба» («успех») и $a_i = 0$ в случае появления «решетки» («неуспех»). Пространство всех исходов имеет следующую структуру:

$$\Omega = \{\omega: \omega = (a_1, \ldots, a_n),\ a_i = 0, 1\}.$$

Припишем каждому элементарному событию $\omega = (a_1, \ldots, a_n)$ вероятность

$$p(\omega) = p^{\sum a_i} q^{n - \sum a_i},$$

где неотрицательные числа p и q таковы, что $p + q = 1$. Прежде всего покажем, что этот способ задания «весов» $p(\omega)$ действительно является корректным. Для этого нам достаточно проверить, что $\sum_{\omega\in\Omega} p(\omega)=1$.

Рассмотрим все те исходы $\omega=(a_1, \ldots, a_n)$, для которых $\sum_i a_i=k$, где $k=0, 1, \ldots, n$. Согласно табл. 4 (размещение k неразличимых «единиц» по n местам) число таких исходов равно C_n^k. Поэтому

$$\sum_{\omega\in\Omega} p(\omega)=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=(p+q)^n=1.$$

Итак, пространство Ω вместе с системой $\mathscr{A}$ всех его подмножеств и вероятностями $\mathsf{P}(A)=\sum_{\omega\in A} p(\omega)$, $A\in\mathscr{A}$, определяет некоторую вероятностную модель. Естественно ее назвать вероятностной моделью, описывающей n-кратное подбрасывание монеты.

В случае $n=1$, когда пространство элементарных исходов состоит лишь из двух точек $\omega=1$ («успех») и $\omega=0$ («неуспех») вероятность $p(1)=p$ естественно назвать вероятностью «успеха». Далее мы увидим, что рассматриваемая нами вероятностная модель, описывающая n-кратное подбрасывание монеты, может быть получена как результат n «независимых» испытаний с вероятностью «успеха», на каждом шаге равной p.

Введем в рассмотрение события

$$A_k=\{\omega\colon \omega=(a_1, \ldots, a_n),\ a_1+\ldots+a_n=k\},$$
$$k=0, 1, \ldots, n,$$

означающие, что произойдет в точности k «успехов». Из сказанного выше следует, что

$$\mathsf{P}(A_k)=C_n^k p^k q^{n-k}, \tag{1}$$

причем $\sum_{k=0}^{n} \mathsf{P}(A_k)=1$.

Набор вероятностей $(\mathsf{P}(A_0), \ldots, \mathsf{P}(A_n))$ называется *биномиальным распределением* (числа «успехов» в выборке объема n). Это распределение играет исключительно важную роль в теории вероятностей, возникая в самых разнообразных вероятностных моделях. Обозначим $P_n(k)=\mathsf{P}(A_k)$, $k=0, 1, \ldots, n$. На рис. 1 воспроизведены биномиальные распределения для случая $p=\frac{1}{2}$ («симметричная» монета) и $n=5, 10, 20$.

Приведем еще одну модель (в сущности эквивалентную предшествующей), описывающую случайное блуждание некоторой «частицы».

Пусть частица выходит из нуля и через единицу времени делает шаг на единицу вверх или вниз (рис. 2).

Таким образом, за n шагов частица может переместиться максимум на n единиц вверх или n единиц вниз. Понятно, что каждая «траектория» ω движения частицы может быть полностью

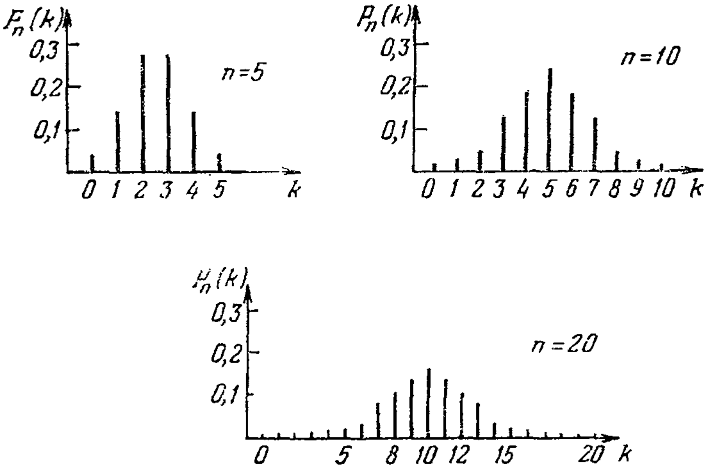


Рис. 1. Графики биномиальных вероятностей $P_n(k)$ для $n=5$, 10, 20.

описана набором $(a_1, \ldots, a_n)$, где $a_i=+1$, если на i-м шаге частица сдвигается вверх и $a_i=-1$, если сдвигается вниз. Припишем каждой траектории ω «вес» $p(\omega)=p^{\nu(\omega)}q^{n-\nu(\omega)}$, где $\nu(\omega)$ — число $+1$ в последовательности $\omega=(a_1, \ldots, a_n)$, т. е.

$$\nu(\omega)=\frac{(a_1+\ldots+a_n)+n}{2},$$

а неотрицательные числа p и q таковы, что $p+q=1$.

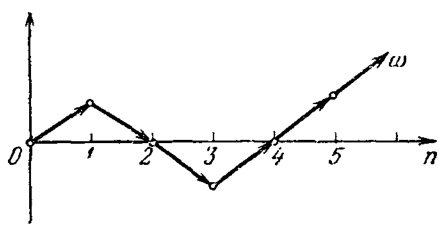


Рис. 2.

Поскольку $\sum_{\omega\in\Omega} p(\omega)=1$, то набор вероятностей $p(\omega)$ вместе с пространством Ω траекторий $\omega=(a_1, \ldots, a_n)$ и его подмножествами действительно определяет некоторую вероятностную модель движения частицы за n шагов.

Поставим следующий вопрос: какова вероятность события A_k, что за n шагов частица окажется в точке с ординатой, равной

k? Этому условию удовлетворяют все те траектории ω, для которых $\nu(\omega)-(n-\nu(\omega))=k$, т. е.

$$\nu(\omega)=\frac{n+k}{2}.$$

Число же таких траекторий (см. табл. 4) равно $C_n^{\frac{n+k}{2}}$ и, значит,

$$\mathsf{P}(A_k)=C_n^{\frac{n+k}{2}}p^{\frac{n+k}{2}}q^{\frac{n-k}{2}}.$$

Таким образом, биномиальное распределение $(\mathsf{P}(A_{-n}), \ldots, \ldots, \mathsf{P}(A_0), \ldots, \mathsf{P}(A_n))$ описывает, как говорят, распределение вероятностей положения частицы за n шагов.

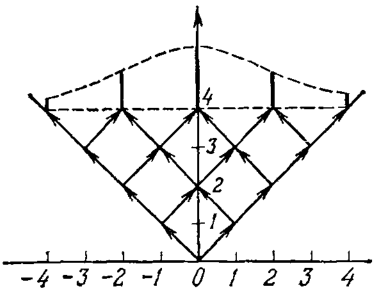


Рис. 3. Возникновение биномиального распределения.

Заметим, что в «симметричном» случае ($p=q=1/2$), когда вероятность отдельной траектории равна 2^{-n},

$$\mathsf{P}(A_k)=C_n^{\frac{n+k}{2}}\cdot 2^{-n}.$$

Рассмотрим асимптотику этих вероятностей при больших n.

Если число шагов равно $2n$, то из свойств биномиальных коэффициентов следует, что среди вероятностей $\mathsf{P}(A_k)$, $|k|\leqslant 2n$, максимальной является вероятность

$$\mathsf{P}(A_0)=C_{2n}^n\cdot 2^{-2n}.$$

Из формулы Стирлинга (см. формулу (6) в п. 4)

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

Поэтому

$$C_{2n}^n=\frac{(2n)!}{(n!)^2}\sim 2^{2n}\cdot\frac{1}{\sqrt{\pi n}}$$

и, значит, при больших n

$$\mathsf{P}(A_0)\sim\frac{1}{\sqrt{\pi n}}.$$

Рис. 3 дает представление о возникновении биномиального распределения при движении частицы за $2n$ шагов (в отличие от рис. 2 временна́я ось здесь направлена вверх).

---

*) Соотношение $f(n)\sim g(n)$ означает, что $\frac{f(n)}{g(n)}\to 1$ при $n\to\infty$.

2. Мультиномиальное распределение. В обобщение предшествующей модели предположим, что пространство исходов имеет следующую структуру:

$$\Omega=\{\omega:\ \omega=(a_1, \ldots, a_n),\ a_i=b_1, \ldots, b_r\},$$

где $b_1, \ldots, b_r$ — заданные числа. Пусть $\nu_i(\omega)$ — число элементов в последовательности $\omega=(a_1, \ldots, a_n)$, равных b_i, $i=1, \ldots, r$, и вероятность исхода ω определяется формулой

$$p(\omega)=p_1^{\nu_1(\omega)} \ldots p_r^{\nu_r(\omega)},$$

где $p_i \geqslant 0$ и $p_1+\ldots+p_r=1$. Заметим, что

$$\sum_{\omega\in\Omega} p(\omega)=\sum_{\substack{\{n_1\geqslant 0, \ldots, n_r\geqslant 0,\\ n_1+\ldots+n_r=n\}}} C_n(n_1, \ldots, n_r)\, p_1^{n_1} \ldots p_r^{n_r},$$

где $C_n(n_1, \ldots, n_r)$ — число (упорядоченных) последовательностей $(a_1, \ldots, a_n)$, у которых элемент b_1 встречается n_1 раз, ..., элемент b_r встречается n_r раз. Поскольку число способов, которым n_1 элементов b_1 можно расположить на n местах, равно $C_n^{n_1}$, n_2 элементов b_2 — на $n-n_1$ местах — $C_{n-n_1}^{n_2}$ и т. д., то

$$\begin{aligned} C_n(n_1, \ldots, n_r) &= C_n^{n_1}\cdot C_{n-n_1}^{n_2} \ldots C_{n-(n_1+\ldots+n_{r-1})}^{n_r}= \\ &=\frac{n!}{n_1!\,(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!\,(n-n_1-n_2)!}\ldots 1= \\ &=\frac{n!}{n_1!\ldots n_r!}. \end{aligned}$$

Поэтому

$$\sum_{\omega\in\Omega} p(\omega)=\sum_{\substack{\{n_1\geqslant 0, \ldots, n_r\geqslant 0,\\ n_1+\ldots+n_r=n\}}} \frac{n!}{n_1!\ldots n_r!}\, p_1^{n_1} \ldots p_r^{n_r}=(p_1+\ldots+p_r)^n=1,$$

и, следовательно, рассматриваемый способ задания вероятностей является корректным.

Пусть

$$A_{n_1, \ldots, n_r}=\{\omega:\ \nu_1(\omega)=n_1, \ldots, \nu_r(\omega)=n_r\}.$$

Тогда

$$\mathsf{P}(A_{n_1, \ldots, n_r})=C_n(n_1, \ldots, n_r)\, p_1^{n_1} \ldots p_r^{n_r}. \tag{2}$$

Набор вероятностей

$$\{\mathsf{P}(A_{n_1, \ldots, n_r})\}$$

носит название *мультиномиального* (полиномиального) распределения.

Подчеркнем, что возникновение этого распределения и его частного случая — биномиального распределения — связано с выбором *с возвращением*.

3. Многомерное гипергеометрическое распределение появляется в задачах, где имеет место выбор *без возвращения*.

Для примера рассмотрим урну, содержащую M различных шаров, занумерованных, скажем, числами $1, 2, \ldots, M$, из которых M_1 шар имеет «цвет» $b_1, \ldots, M_r$ шаров имеют «цвет» b_r, $M_1+\ldots+M_r=M$. Предположим, что осуществляется выбор без возвращения объема $n<M$. Пространство элементарных событий

$$\Omega=\{\omega\colon\ \omega=(a_1, \ldots, a_n),\ a_1\neq a_2\neq\ldots\neq a_n,\ a_i=1, \ldots, M\}$$

и $N(\Omega)=(M)_n$. Будем считать элементарные события равновозможными и найдем вероятность события $B_{n_1, \ldots, n_r}$, состоящего в том, что n_1 шар имеет цвет $b_1, \ldots, n_r$ шаров имеют цвет b_r, $n_1+\ldots+n_r=n$. Нетрудно показать, что

$$N\left(B_{n_1, \ldots, n_r}\right)=C_n(n_1, \ldots, n_r)(M_1)_{n_1}\ldots(M_r)_{n_r},$$

и, значит,

$$\mathsf{P}\left(B_{n_1, \ldots, n_r}\right)=\frac{N\left(B_{n_1, \ldots, n_r}\right)}{N(\Omega)}=\frac{C_{M_1}^{n_1}\ldots C_{M_r}^{n_r}}{C_M^n}. \tag{3}$$

Набор вероятностей $\left\{\mathsf{P}\left(B_{n_1, \ldots, n_r}\right)\right\}$ носит название *многомерного гипергеометрического распределения*. В случае $r=2$ это распределение называют просто *гипергеометрическим* в связи с тем, что так называемая производящая функция этого распределения есть гипергеометрическая функция.

Структура многомерного гипергеометрического распределения довольно сложна. Так, вероятность

$$\mathsf{P}\left(B_{n_1, n_2}\right)=\frac{C_{M_1}^{n_1}C_{M_2}^{n_2}}{C_M^n},\quad n_1+n_2=n,\quad M_1+M_2=M, \tag{4}$$

содержит девять факториалов. Однако легко показать, что если $M\to\infty$, $M_1\to\infty$, но так, что $\frac{M_1}{M}\to p$ $\left(\text{и, следовательно, } \frac{M_2}{M}\to 1-p\right)$, то

$$\mathsf{P}\left(B_{n_1, n_2}\right)\to C_{n_1+n_2}^{n_1}p^{n_1}(1-p)^{n_2}. \tag{5}$$

Иначе говоря, при сделанных предположениях гипергеометрическое распределение аппроксимируется биномиальным, что интуитивно понятно, поскольку при больших M и M_1 (конечный) выбор без возвращения должен давать почти тот же результат, что и выбор с возвращением.

Пример. Используем формулу (4) для нахождения вероятности угадывания шести «счастливых» номеров в известной лотерее «спортлото», суть которой состоит в следующем.

Имеется 49 шаров, занумерованных числами 1, 2, ..., 49, из которых шесть шаров «счастливых» (скажем, красного цвета; остальные — белого). Производится выбор без возвращения шести шаров. Спрашивается, какова вероятность того, что все шесть вытащенных шаров являются «счастливыми». Полагая $M=49$, $M_1=6$, $n_1=6$, $n_2=0$, видим, что интересующее нас событие

$$B_{6,0}=\{6 \text{ шаров — «счастливые»}\}$$

имеет, согласно (4), вероятность

$$\mathrm{P}(B_{6,0})=\frac{1}{C_{49}^{6}}\approx 7,2\cdot 10^{-8}.$$

4. Числа $n!$ с ростом n растут чрезвычайно быстро. Так,

$$10!=3\,628\,800,$$
$$15!=1\,307\,674\,368\,000,$$

а 100! содержит 158 знаков. Поэтому как с теоретической, так и вычислительной точки зрения важна следующая формула Стирлинга:

$$n!=\sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}\exp\left(\frac{\theta_n}{12n}\right),\quad 0<\theta_n<1, \tag{6}$$

доказательство которой имеется в большинстве руководств по математическому анализу (см. также [69]).

5. Задачи.

1. Доказать формулу (5).

2. Показать, что для мультиномиального распределения $\{\mathrm{P}(A_{n_1,\ \ldots,\ n_r})\}$ максимальное значение вероятности достигается в точке $(k_1, \ldots, k_r)$, удовлетворяющей неравенствам: $np_i-1<k_i\leqslant(n+r-1)p_i$, $i=1, \ldots, r$.

3. *Одномерная модель Изинга.* Пусть имеется n частиц, расположенных в точках 1, 2, ..., n. Предположим, что каждая из частиц относится к одному из двух типов, причем частиц первого типа n_1 и второго — n_2 $(n_1+n_2=n)$. Будем считать все $n!$ расположений частиц равновозможными.

Построить соответствующую вероятностную модель и найти вероятность события $A_n(m_{11},\ m_{12},\ m_{21},\ m_{22})=\{\nu_{11}=m_{11},\ \ldots,\ \nu_{22}=m_{22}\}$, где ν_{ij} — число частиц типа i, следующих за частицами типа j $(i,\ j=1,\ 2)$.

4. Используя вероятностные соображения, доказать справедливость следующих тождеств:

$$\sum_{k=0}^{n} C_n^k = 2^n,$$

$$\sum_{k=0}^{n} (C_n^k)^2 = C_{2n}^n,$$

$$\sum_{k=0}^{n} (-1)^{n-k} C_m^k = C_{m-1}^n, \quad m \geqslant n+1,$$

$$\sum_{k=0}^{n} k(k-1) C_m^k = m(m-1)\,2^{m-2}, \quad m \geqslant 2.$$

§ 3. Условные вероятности. Независимость

1. Понятие вероятности событий дает нам возможность ответить на вопрос такого типа: если урна содержит M шаров, из которых M_1 шаров белого цвета и M_2 — черного, то какова вероятность $\mathsf{P}(A)$ события A, состоящего в том, что вытащенный шар имеет белый цвет. В случае классического подхода $\mathsf{P}(A) = M_1/M$.

Вводимое ниже понятие *условной вероятности* позволяет отвечать на вопрос следующего типа: какова вероятность того, что второй извлеченный шар белого цвета (событие B), при условии, что первый шар также имеет белый цвет (событие A)? (Рассматривается выбор без возвращения.)

Естественно здесь рассуждать так: если первый извлеченный шар имел белый цвет, то перед вторым извлечением мы имеем урну с $M-1$ шаром, из которых M_1-1 шаров имеют белый цвет, а M_2 — черный; поэтому представляется целесообразным считать, что интересующая нас (условная) вероятность равна $\frac{M_1-1}{M-1}$.

Дадим теперь определение условной вероятности, согласующееся с интуитивными представлениями о ней.

Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — (конечное) вероятностное пространство и A — некоторое событие (т. е. $A \in \mathscr{A}$).

Определение 1. *Условной вероятностью* события B при условии события A с $\mathsf{P}(A) > 0$ (обозначается $\mathsf{P}(B \mid A)$) называется величина

$$\frac{\mathsf{P}(AB)}{\mathsf{P}(A)}. \tag{1}$$

В случае классического способа задания вероятностей $\mathsf{P}(A) = \frac{N(A)}{N(\Omega)}$, $\mathsf{P}(AB) = \frac{N(AB)}{N(\Omega)}$ и, значит,

$$\mathsf{P}(B \mid A) = \frac{N(AB)}{N(A)}. \tag{2}$$

Следующие свойства условных вероятностей непосредственно вытекают из определения 1:

$$\mathsf{P}(A \mid A) = 1,$$
$$\mathsf{P}(\varnothing \mid A) = 0,$$
$$\mathsf{P}(B \mid A) = 1, \quad B \supseteq A,$$
$$\mathsf{P}(B_1 + B_2 \mid A) = \mathsf{P}(B_1 \mid A) + \mathsf{P}(B_2 \mid A).$$

Из этих свойств следует, что при «закрепленном» множестве A условная вероятность $\mathsf{P}(\cdot \mid A)$ обладает на пространстве $(\Omega \cap A, \mathscr{A} \cap A)$, где $\mathscr{A} \cap A = \{B \cap A\colon\ B \in \mathscr{A}\}$, теми же свойствами, что и исходная вероятность $\mathsf{P}(\cdot)$ на $(\Omega, \mathscr{A})$.

Отметим, что

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

однако, вообще говоря,

$$\mathsf{P}(B \mid A) + \mathsf{P}(B \mid \bar{A}) \neq 1,$$
$$\mathsf{P}(B \mid A) + \mathsf{P}(\bar{B} \mid A) \neq 1.$$

Пример 1. Рассмотрим семьи, имеющие двух детей. Спрашивается, какова вероятность того, что в семье оба ребенка мальчики, в предположении, что:

a) старший ребенок — мальчик?

b) по крайней мере один из детей — мальчик?

Пространство элементарных событий —

$$\Omega = \{\text{ММ, МД, ДМ, ДД}\},$$

где МД означает, что старший ребенок — мальчик, младший — девочка и т. д.

Будем считать, что каждый исход равновозможен:

$$\mathsf{P}(\text{ММ}) = \mathsf{P}(\text{МД}) = \mathsf{P}(\text{ДМ}) = \mathsf{P}(\text{ДД}) = \frac{1}{4}.$$

Пусть A — событие «старший ребенок — мальчик», B — «младший ребенок — мальчик». Тогда $A \cup B$ есть событие «по крайней мере один из детей — мальчик», AB «оба ребенка — мальчики» и интересующая нас в вопросе a) вероятность есть условная вероятность $\mathsf{P}(AB \mid A)$, а в вопросе b) — условная вероятность $\mathsf{P}(AB \mid A \cup B)$.

Легко находим, что

$$\mathsf{P}(AB \mid A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)} = \frac{1/4}{1/2} = \frac{1}{2},$$

$$\mathsf{P}(AB \mid A \cup B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A \cup B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

2. Следующая простая, но важная формула (3), носящая название формулы полной вероятности, является основным средством при подсчете вероятностей сложных событий с использованием условных вероятностей.

Рассмотрим некоторое разбиение $\mathscr{D} = \{A_1, \ldots, A_n\}$ с $\mathsf{P}(A_i) > 0$, $i = 1, \ldots, n$ (часто такое разбиение называют также полной группой несовместимых событий). Ясно, что

$$B = BA_1 + \ldots + BA_n$$

и, значит,

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

Но

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

Тем самым имеет место *формула полной вероятности*

$$\mathsf{P}(B) = \sum_{i=1}^{n} \mathsf{P}(B \mid A_i)\,\mathsf{P}(A_i). \tag{3}$$

В частности, если $0 < \mathsf{P}(A) < 1$, то

$$\mathsf{P}(B) = \mathsf{P}(B \mid A)\,\mathsf{P}(A) + \mathsf{P}(B \mid \bar{A})\,\mathsf{P}(\bar{A}). \tag{4}$$

Пример 2. В урне имеется M шаров, среди которых m «счастливых». Спрашивается, какова вероятность извлечь на втором шаге «счастливый» шар (предполагается, что качество первого извлеченного шара неизвестно, рассматривается случай выбора без возвращения объема $n = 2$ и все исходы равновозможны). Пусть A — событие «первый шар — счастливый», B — «второй шар — счастливый». Тогда

$$\mathsf{P}(B \mid A) = \frac{\mathsf{P}(BA)}{\mathsf{P}(A)} = \frac{\dfrac{m(m-1)}{M(M-1)}}{\dfrac{m}{M}} = \frac{m-1}{M-1},$$

$$\mathsf{P}(B \mid \bar{A}) = \frac{\mathsf{P}(B\bar{A})}{\mathsf{P}(\bar{A})} = \frac{\dfrac{m(M-m)}{M(M-1)}}{\dfrac{M-m}{M}} = \frac{m}{M-1}$$

и

$$\mathsf{P}(B)=\mathsf{P}(B\mid A)\,\mathsf{P}(A)+\mathsf{P}(B\mid\bar{A})\,\mathsf{P}(\bar{A})=$$
$$=\frac{m-1}{M-1}\cdot\frac{m}{M}+\frac{m}{M-1}\cdot\frac{M-m}{M}=\frac{m}{M}.$$

Интересно отметить, что вероятность $\mathsf{P}(A)$ также равна m/M. Таким образом, то обстоятельство, что качество первого шара осталось неизвестным, не изменило вероятности того, что извлеченный на втором шаге шар оказался «счастливым».

Из определения условной вероятности ($\mathsf{P}(A)>0$)

$$\mathsf{P}(AB)=\mathsf{P}(B\mid A)\,\mathsf{P}(A). \tag{5}$$

Эта формула, носящая название *формулы умножения вероятностей*, обобщается (по индукции) следующим образом: если события $A_1, \ldots, A_{n-1}$ таковы, что $\mathsf{P}(A_1\ldots A_{n-1})>0$, то

$$\mathsf{P}(A_1\ldots A_n)=\mathsf{P}(A_1)\,\mathsf{P}(A_2\mid A_1)\ldots\mathsf{P}(A_n\mid A_1\ldots A_{n-1}) \tag{6}$$

(здесь $A_1\ldots A_n=A_1\cap A_2\cap\ldots\cap A_n$).

3. Предположим, что события A и B таковы, что $\mathsf{P}(A)>0$ и $\mathsf{P}(B)>0$. Тогда наряду с (5) справедлива также формула

$$\mathsf{P}(AB)=\mathsf{P}(A\mid B)\,\mathsf{P}(B). \tag{7}$$

Из (5) и (7) получаем так называемую формулу Байеса

$$\mathsf{P}(A\mid B)=\frac{\mathsf{P}(A)\,\mathsf{P}(B\mid A)}{\mathsf{P}(B)}. \tag{8}$$

Если события $A_1, \ldots, A_n$ образуют разбиение Ω, то из (3) и (8) следует так называемая теорема Байеса

$$\mathsf{P}(A_i\mid B)=\frac{\mathsf{P}(A_i)\,\mathsf{P}(B\mid A_i)}{\sum_{j=1}^{n}\mathsf{P}(A_j)\,\mathsf{P}(B\mid A_j)}. \tag{9}$$

В статистических применениях события $A_1, \ldots, A_n$ ($A_1+\ldots+A_n=\Omega$) часто называют «гипотезами», а $\mathsf{P}(A_i)$ — *априорной* *) вероятностью гипотезы A_i. Условная вероятность $\mathsf{P}(A_i\mid B)$ трактуется как *апостериорная* вероятность гипотезы A_i после наступления события B.

Пример 3. Пусть в урне находятся две монеты: A_1 — симметричная монета с вероятностью «герба» Г, равной 1/2, и A_2 — несимметричная монета с вероятностью «герба» Г, равной 1/3. Наудачу вынимается и подбрасывается одна из монет. Предполо-

*) A priori — до опыта, a posteriori — после опыта.

жим, что выпал герб. Спрашивается, какова вероятность того, что выбранная монета симметрична.

Построим соответствующую вероятностную модель. В качестве пространства элементарных событий естественно здесь взять множество $\Omega=\{A_1\Gamma,\ A_1\mathrm{P},\ A_2\Gamma,\ A_2\mathrm{P}\}$, описывающее все исходы выбора и подбрасывания ($A_1\Gamma$ означает, что вынута монета A_1 и в результате подбрасывания выпал герб и т. д.). Вероятности $p(\omega)$ рассматриваемых исходов должны быть заданы так, чтобы, согласно условиям задачи,

$$\mathsf{P}(A_1)=\mathsf{P}(A_2)=1/2$$

и

$$\mathsf{P}(\Gamma\,|\,A_1)=1/2,\quad \mathsf{P}(\Gamma\,|\,A_2)=1/3.$$

Этими условиями вероятности исходов определяются однозначно:

$$\mathsf{P}(A_1\Gamma)=1/4,\quad \mathsf{P}(A_1\mathrm{P})=1/4,\quad \mathsf{P}(A_2\Gamma)=1/6,\quad \mathsf{P}(A_2\mathrm{P})=1/3.$$

Тогда, согласно формуле Байеса, интересующая нас вероятность

$$\mathsf{P}(A_1\,|\,\Gamma)=\frac{\mathsf{P}(A_1)\,\mathsf{P}(\Gamma\,|\,A_1)}{\mathsf{P}(A_1)\,\mathsf{P}(\Gamma\,|\,A_1)+\mathsf{P}(A_2)\,\mathsf{P}(\Gamma\,|\,A_2)}=\frac{3}{5}$$

и, значит,

$$\mathsf{P}(A_2\,|\,\Gamma)=2/5.$$

4. Вводимое в этом пункте понятие *независимости* играет в определенном смысле центральную роль в теории вероятностей: именно это понятие определило то своеобразие, которое выделяет теорию вероятностей в общей теории, занимающейся исследованием измеримых пространств с мерой.

Если A и B — два события, то естественно сказать, что событие B не зависит от A, если знание того обстоятельства, что совершилось событие A, никак не влияет на вероятность совершения события B. Иначе говоря, «B не зависит от A», если

$$\mathsf{P}(B\,|\,A)=\mathsf{P}(B) \qquad (10)$$

(здесь мы предполагаем, что $\mathsf{P}(A)>0$).

Поскольку

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

то из (10) находим, что

$$\mathsf{P}(AB)=\mathsf{P}(A)\,\mathsf{P}(B). \qquad (11)$$

Точно так же, если $\mathsf{P}(B)>0$, то естественно сказать, что «A не зависит от B», если

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

Отсюда снова получаем соотношение (11), которое симметрично относительно A и B и имеет смысл также и тогда, когда вероятность этих событий равна нулю.

Исходя из этого, примем следующее

Определение 2. События A и B называются *независимыми* или *статистически независимыми* (относительно вероятности P), если

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

В теории вероятностей часто приходится рассматривать независимость не только событий (множеств), но и систем событий (множеств).

Приведем соответствующее

Определение 3. Две алгебры событий (множеств) $\mathscr{A}_1$ и $\mathscr{A}_2$ называются *независимыми* или *статистически независимыми* (относительно вероятности P), если независимы любые два множества A_1 и A_2, принадлежащие соответственно $\mathscr{A}_1$ и $\mathscr{A}_2$.

Для примера рассмотрим две алгебры

$$\mathscr{A}_1=\{A_1,\ \bar{A}_1,\ \varnothing,\ \Omega\}\quad \text{и}\quad \mathscr{A}_2=\{A_2,\ \bar{A}_2,\ \varnothing,\ \Omega\},$$

где A_1 и A_2 — некоторые множества из Ω. Нетрудно показать, что $\mathscr{A}_1$ и $\mathscr{A}_2$ независимы тогда и только тогда, когда независимы события A_1 и A_2. Действительно, независимость $\mathscr{A}_1$ и $\mathscr{A}_2$ означает независимость шестнадцати событий: A_1 и A_2, A_1 и $\bar{A}_2$, ..., Ω и Ω. Следовательно, A_1 и A_2 независимы. Обратно, если A_1 и A_2 независимы, то надо показать, что независимы остальные пятнадцать пар событий. Проверим, например, независимость A_1 и $\bar{A}_2$. Имеем

$$\begin{aligned}\mathsf{P}(A_1\bar{A}_2)&=\mathsf{P}(A_1)-\mathsf{P}(A_1A_2)=\mathsf{P}(A_1)-\mathsf{P}(A_1)\,\mathsf{P}(A_2)=\\&=\mathsf{P}(A_1)\cdot(1-\mathsf{P}(A_2))=\mathsf{P}(A_1)\,\mathsf{P}(\bar{A}_2).\end{aligned}$$

Независимость остальных пар проверяется аналогичным образом.

5. Понятие независимости двух множеств и двух алгебр множеств распространяется на случай любого конечного числа множеств и алгебр множеств.

Именно, говорят, что множества $A_1, \ldots, A_n$ *независимы* или *статистически независимы* в совокупности (относительно вероятности P), если для любых $k=1, \ldots, n$ и $1\leqslant i_1<i_2<\ldots<i_k\leqslant n$

$$\mathsf{P}(A_{i_1}\ldots A_{i_k})=\mathsf{P}(A_{i_1})\ldots\mathsf{P}(A_{i_k}). \tag{12}$$

Алгебры множеств $\mathscr{A}_1, \ldots, \mathscr{A}_n$ называются *независимыми* или *статистически независимыми* в совокупности (относительно вероятности P), если независимы любые множества $A_1, \ldots, A_n$, принадлежащие соответственно $\mathscr{A}_1, \ldots, \mathscr{A}_n$.

Отметим, что из *попарной независимости* событий *не следует* их независимость. Действительно, если, например $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ и все исходы равновозможны, то нетрудно проверить, что события

$$A = \{\omega_1, \omega_2\}, \quad B = \{\omega_1, \omega_3\} \quad C = \{\omega_1, \omega_4\}$$

попарно независимы, но

$$\mathsf{P}(ABC) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3 = \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C).$$

Отметим также, что из того, что для некоторых событий A, B и C

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

вовсе не следует попарная независимость этих событий. В самом деле, пусть пространство Ω состоит из 36 упорядоченных пар (i, j), где $i, j = 1, 2, \ldots, 6$ и все эти пары равновозможны. Тогда для $A = \{(i, j)\colon j = 1, 2 \text{ или } 5\}$, $B = \{(i, j)\colon j = 4, 5 \text{ или } 6\}$, $C = \{(i, j)\colon i + j = 9\}$ имеем

$$\mathsf{P}(AB) = \frac{1}{6} \neq \frac{1}{4} = \mathsf{P}(A)\,\mathsf{P}(B),$$

$$\mathsf{P}(AC) = \frac{1}{36} \neq \frac{1}{18} = \mathsf{P}(A)\,\mathsf{P}(C),$$

$$\mathsf{P}(BC) = \frac{1}{12} \neq \frac{1}{18} = \mathsf{P}(B)\,\mathsf{P}(C),$$

но в то же время

$$\mathsf{P}(ABC) = \frac{1}{36} = \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C).$$

6. С точки зрения понятия независимости рассмотрим подробнее классическую модель $(\Omega, \mathscr{A}, \mathsf{P})$, введенную в § 2 и приведшую к возникновению биномиального распределения.

В этой модели

$$\Omega = \{\omega\colon \omega = (a_1, \ldots, a_n),\ a_i = 0, 1\},$$
$$\mathscr{A} = \{A\colon A \subseteq \Omega\}$$

и

$$p(\omega) = p^{\Sigma a_i} q^{n - \Sigma a_i}. \tag{13}$$

Пусть событие $A \subseteq \Omega$. Будем говорить, что это событие зависит от испытания в k-й момент времени, если оно определяется лишь значением a_k. Примером таких событий являются события

$$A_k = \{\omega\colon a_k = 1\}, \quad \bar{A}_k = \{\omega\colon a_k = 0\}.$$

Рассмотрим последовательность алгебр $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$, где $\mathscr{A}_k = \{A_k, \bar{A}_k, \varnothing, \Omega\}$, и покажем, что в случае (13) эти алгебры независимы.

Ясно, что

$$\mathsf{P}(A_k) = \sum_{\{\omega:\ a_k=1\}} p(\omega) = \sum_{\{\omega:\ a_k=1\}} p^{\Sigma a_i} q^{n-\Sigma a_i} =$$
$$= p \sum_{(a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n)} p^{a_1+\ldots+a_{k-1}+a_{k+1}+\ldots+a_n} \times$$
$$\times q^{(n-1)-(a_1+\ldots+a_{k-1}+a_{k+1}+\ldots+a_n)} = p \sum_{l=0}^{n-1} C_{n-1}^l p^l q^{(n-1)-l} = p,$$

и аналогичный подсчет показывает, что $\mathsf{P}(\bar{A}_k) = q$ и при $k \neq l$

$$\mathsf{P}(A_k A_l) = p^2, \quad \mathsf{P}(A_k \bar{A}_l) = pq, \quad \mathsf{P}(\bar{A}_k A_l) = q^2.$$

Отсюда легко выводится, что алгебры $\mathscr{A}_k$ и $\mathscr{A}_l$, $k \neq l$, независимы.

Точно так же показывается, что независимы и алгебры $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$. Это дает основание назвать рассматриваемую модель $(\Omega, \mathscr{A}, \mathsf{P})$ моделью, отвечающей «n независимым испытаниям с двумя исходами и вероятностью «успеха» p». Я. Бернулли был первый, кто систематически изучал эту модель и доказал для нее справедливость закона больших чисел (§ 5). В связи с этим эту модель называют также схемой Бернулли (с двумя исходами — «успехом» и «неуспехом» — и вероятностью «успеха» p).

Детальное рассмотрение вероятностного пространства в схеме Бернулли показывает, что оно имеет структуру «прямого произведения вероятностных пространств», состоящую в следующем.

Предположим, что задан набор $(\Omega_1, \mathscr{B}_1, \mathsf{P}_1), \ldots, (\Omega_n, \mathscr{B}_n, \mathsf{P}_n)$ конечных вероятностных пространств. Образуем пространство $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ точек $\omega = (a_1, \ldots, a_n)$, где $a_i \in \Omega_i$. Обозначим $\mathscr{A} = \mathscr{B}_1 \otimes \ldots \otimes \mathscr{B}_n$ — алгебру подмножеств Ω, состоящую из сумм множеств вида $A = B_1 \times B_2 \times \ldots \times B_n$ с $B_i \in \mathscr{B}_i$. Наконец, для $\omega = (a_1, \ldots, a_n)$ положим $p(\omega) = p_1(a_1) \ldots p_n(a_n)$ и определим $\mathsf{P}(A)$ для множеств $A = B_1 \times B_2 \times \ldots \times B_n$ формулой:

$$\mathsf{P}(A) = \sum_{\{a_1 \in B_1, \ldots, a_n \in B_n\}} p_1(a_1) \ldots p_n(a_n).$$

Нетрудно проверить, что $\mathsf{P}(\Omega) = 1$ и, следовательно, тройка $(\Omega, \mathscr{A}, \mathsf{P})$ определяет некоторое вероятностное пространство. Это пространство называют *прямым произведением вероятностных пространств* $(\Omega_1, \mathscr{B}_1, \mathsf{P}_1), \ldots, (\Omega_n, \mathscr{B}_n, \mathsf{P}_n)$.

Отметим одно легко проверяемое свойство прямого произведения вероятностных пространств: относительно вероятности P события

$$A_1 = \{\omega:\ a_1 \in B_1\}, \ \ldots, \ A_n = \{\omega:\ a_n \in B_n\},$$

где $B_i \in \mathscr{B}_i$, являются независимыми. Точно так же алгебры множеств пространства Ω

$$\mathscr{A}_1 = \{A_1 \colon A_1 = \{\omega \colon a_1 \in B_1\},\ B_1 \in \mathscr{B}_1\},$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\mathscr{A}_n = \{A_n \colon A_n = \{\omega \colon a_n \in B_n\},\ B_n \in \mathscr{B}_n\}$$

являются независимыми.

Из приведенных конструкций видно, что схема Бернулли

$$(\Omega,\ \mathscr{A},\ \mathsf{P}) \quad \text{с} \quad \Omega = \{\omega \colon \omega = (a_1, \dots, a_n),\ a_i = 0,\ 1\},$$
$$\mathscr{A} = \{A \colon A \subseteq \Omega\} \quad \text{и} \quad p(\omega) = p^{\Sigma a_i} q^{n - \Sigma a_i}$$

может быть получена как прямое произведение вероятностных пространств $(\Omega_i,\ \mathscr{B}_i,\ \mathsf{P}_i)$, $i = 1, 2, \dots, n$, где

$$\Omega_i = \{0,\ 1\}, \quad \mathscr{A}_i = \{\{0\},\ \{1\},\ \varnothing,\ \Omega_i\},$$
$$\mathsf{P}_i(\{1\}) = p, \quad \mathsf{P}_i(\{0\}) = q.$$

7. Задачи.

1. Привести примеры, показывающие, что, вообще говоря, равенства

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

неверны.

2. Урна содержит M шаров, из которых M_1 шаров белого цвета. Рассматривается выбор объема n. Пусть B_j — событие, состоящее в том, что извлеченный на j-м шаге шар имел белый цвет, а A_k — событие, состоящее в том, что в выборке объема n имеется в точности k белых шаров. Показать, что как для выбора с возвращением так и для выбора без возвращения

$$\mathsf{P}(B_j \mid A_k) = k/n.$$

3. Пусть $A_1, \dots, A_n$ — независимые события. Тогда

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} \mathsf{P}(\bar{A}_i).$$

4. Пусть $A_1, \dots, A_n$ — независимые события с $\mathsf{P}(A_i) = p_i$. Тогда вероятность P_0 того, что ни одно из этих событий не произойдет, определяется формулой

$$P_0 = \prod_{i=1}^{n} (1 - p_i).$$

5. Пусть A и B — независимые события. В терминах $\mathsf{P}(A)$ и $\mathsf{P}(B)$ выразить вероятности событий, состоящих в том, что про-

изойдет в точности k, по меньшей мере k и самое большее k из событий A и B ($k=0, 1, 2$).

6. Пусть событие A таково, что оно не зависит от самого себя, т. е. A и A независимы. Показать, что тогда $\mathsf{P}(A)$ равно 0 или 1.

7. Пусть событие A таково, что $\mathsf{P}(A)$ равно 0 или 1. Показать, что A и любое событие B независимы.

8. Рассматривается электрическая схема, изображенная на рис. 4:

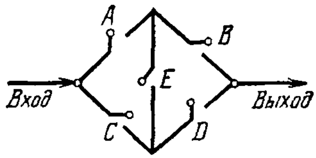


Рис. 4.

Каждое из реле A, B, C, D и E, работающих независимо, открывается и закрывается с вероятностями p и q соответственно. Спрашивается, какова вероятность того, что сигнал, поданный на «вход», будет получен на «выходе»? Какова условная вероятность того, что реле E было открыто, если на «выходе» был получен сигнал?

§ 4. Случайные величины и их характеристики

1. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — вероятностная модель некоторого эксперимента с конечным числом исходов, $N(\Omega) < \infty$, и алгеброй $\mathscr{A}$ всех подмножеств Ω. Можно заметить, что в рассмотренных выше примерах, связанных с подсчетом тех или иных вероятностей событий $A \in \mathscr{A}$, собственно природа пространства элементарных событий Ω не представляла интереса. Основной интерес представляли лишь некоторые числовые характеристики, значения которых зависели от элементарных событий. Так, мы интересовались вопросами о том, какова вероятность определенного числа успехов в серии из n испытаний, каково распределение вероятностей числа дробинок по ячейкам и т. п.

Вводимое сейчас (и далее — в более общем виде) понятие случайной величины призвано определить величины, подлежащие «измерению» в случайных экспериментах.

Определение 1. Всякая числовая функция $\xi = \xi(\omega)$, определенная на (конечном) пространстве элементарных событий Ω, будет называться (простой) *случайной величиной*. (Происхождение термина «простая» случайная величина станет понятным после введения общего понятия случайной величины в § 4 гл. II.)

Пример 1. В модели двукратного подбрасывания монеты с пространством исходов $\Omega=\{ГГ,\ ГР,\ РГ,\ РР\}$ определим случайную величину $\xi=\xi(\omega)$ с помощью таблицы

ω	ГГ	ГР	РГ	РР
$\xi(\omega)$	2	1	1	0

Здесь $\xi(\omega)$ по своему смыслу есть не что иное, как число «гербов», отвечающих исходу ω.

Другим простейшим примером случайной величины ξ является *индикатор* (иначе — *характеристическая функция*) некоторого *множества* $A\in\mathscr{A}$:

$$\xi=I_A(\omega),$$

где *)

$$I_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A.\end{cases}$$

Когда экспериментатор имеет дело со случайными величинами, описывающими те или иные показания, то основной вопрос, который его интересует, — это вопрос о том, с какими вероятностями эта случайная величина принимает те или иные значения. С этой точки зрения интерес представляет не распределение вероятностей P на $(\Omega,\ \mathscr{A})$, а распределение вероятностей на множестве значений случайной величины. Поскольку в рассматриваемом сейчас случае Ω состоит из конечного числа точек, то множество значений X случайной величины ξ также конечно. Пусть $X=\{x_1,\ \ldots,\ x_m\}$, где (различными) числами $x_1,\ \ldots,\ x_m$ исчерпываются все значения ξ.

Обозначим $\mathscr{X}$ — совокупность всех подмножеств множества X, и пусть $B\in\mathscr{X}$. Множество B можно также интерпретировать как некоторое событие, когда пространство исходов есть X — множество значений ξ.

Рассмотрим на $(X,\ \mathscr{X})$ вероятность $P_\xi(\cdot)$, индуцируемую случайной величиной ξ по формуле

$$P_\xi(B)=\mathsf{P}\{\omega\colon\ \xi(\omega)\in B\},\qquad B\in\mathscr{X}.$$

Ясно, что значения этих вероятностей полностью определяются вероятностями

$$P_\xi(x_i)=\mathsf{P}\{\omega\colon\ \xi(\omega)=x_i\},\qquad x_i\in X.$$

*) Для индикатора используется также обозначение $I(A)$. По поводу часто используемых далее свойств индикаторов см. задачу 1.

Набор чисел $\{P_\xi(x_1), \ldots, P_\xi(x_m)\}$ называется *распределением вероятностей случайной величины* ξ.

Пример 2. Случайная величина ξ, принимающая два значения 1 и 0 с вероятностями («успеха») p и («неуспеха») q, называется *бернуллиевской* *). Ясно, что для нее

$$P_\xi(x) = p^x q^{1-x}, \qquad x = 0, 1. \tag{1}$$

Биномиальной (или биномиально распределенной) *случайной величиной* ξ называется случайная величина, принимающая $n+1$ значение 0, 1, ..., n с вероятностями

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

Заметим, что в этих и во многих последующих примерах мы не конкретизируем структуру основного вероятностного пространства $(\Omega, \mathscr{A}, \mathsf{P})$, а интересуемся лишь значениями случайных величин и их распределениями вероятностей.

Вероятностная структура случайных величин ξ полностью описывается распределением вероятностей $\{P_\xi(x_i),\ i = 1, \ldots, m\}$. Вводимое ниже понятие функции распределения дает эквивалентное описание вероятностной структуры случайных величин.

Определение 2. Пусть $x \in R^1$. Функция

$$F_\xi(x) = \mathsf{P}\{\omega\colon \xi(\omega) \leqslant x\}$$

называется *функцией распределения* случайной величины ξ.

Ясно, что

$$F_\xi(x) = \sum_{\{i:\, x_i \leqslant x\}} P_\xi(x_i)$$

и

$$P_\xi(x_i) = F_\xi(x_i) - F_\xi(x_i -),$$

где $F_\xi(x-) = \lim\limits_{y \uparrow x} F_\xi(y)$.

Если считать, что $x_1 < x_2 < \ldots < x_m$ и положить $F_\xi(x_0) = 0$, то

$$P_\xi(x_i) = F_\xi(x_i) - F_\xi(x_{i-1}), \qquad i = 1, \ldots, m.$$

Следующие графики (рис. 5) дают представление о $P_\xi(x)$ и $F_\xi(x)$ для биномиальной случайной величины.

Непосредственно из определения 2 следует, что функция распределения $F_\xi = F_\xi(x)$ обладает такими свойствами:

(1) $F_\xi(-\infty) = 0$, $F_\xi(+\infty) = 1$;

(2) $F_\xi(x)$ непрерывна справа $(F_\xi(x+) = F_\xi(x))$ и кусочно-постоянна.

---

*) Обычно в литературе вместо выражений «бернуллиевская, биномиальная, пуссоновская, гауссовская случайная величина», используемых здесь, говорится о «случайных величинах, имеющих распределение Бернулли, биномиальное, Пуассона, Гаусса».

Наряду со случайными величинами часто приходится рассматривать *случайные векторы* $\xi=(\xi_1, \ldots, \xi_r)$, компоненты которых являются случайными величинами. Например, при рассмотрении мультиномиального распределения мы имели дело со случайным вектором $\nu=(\nu_1, \ldots, \nu_r)$, где $\nu_i=\nu_i(\omega)$ — число элементов в последовательности $\omega=(a_1, \ldots, a_n)$, равных b_i, $i=1, \ldots, r$.

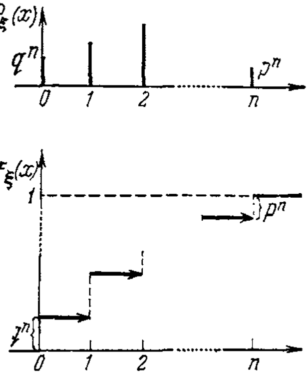


Рис. 5.

Набор вероятностей

$$P_\xi(x_1, \ldots, x_r)=\mathrm{P}\{\omega:\ \xi_1(\omega)=x_1, \ldots, \xi_r(\omega)=x_r\},$$

где $x_i \in X_i$ — области допустимых значений ξ_i, называется *распределением вероятностей случайного вектора* ξ, а функция

$$F_\xi(x_1, \ldots, x_r)=\mathrm{P}\{\omega:\ \xi_1(\omega)\leqslant x_1, \ldots, \xi_r(\omega)\leqslant x_r\},$$

где $x_i \in R^1$, называется *функцией распределения случайного вектора* $\xi=(\xi_1, \ldots, \xi_r)$.

Так, для упомянутого выше вектора $\nu=(\nu_1, \ldots, \nu_r)$

$$P_\nu(n_1, \ldots, n_r)=C_n(n_1, \ldots, n_r)\,p_1^{n_1}\ldots p_r^{n_r}$$

(см. (2.2)).

2. Пусть $\xi_1, \ldots, \xi_r$ — некоторый набор случайных величин, принимающих значения в (конечном) множестве $X\subseteq R^1$. Обозначим через $\mathscr{X}$ алгебру всех подмножеств X.

Определение 3. Случайные величины $\xi_1, \ldots, \xi_r$ называются *независимыми* (*независимыми в совокупности*), если для любых $x_1, \ldots, x_r \in X$

$$\mathrm{P}\{\xi_1=x_1, \ldots, \xi_r=x_r\}=\mathrm{P}\{\xi_1=x_1\}\ldots\mathrm{P}\{\xi_r=x_r\},$$

или, что эквивалентно, для любых $B_1, \ldots, B_r \in \mathscr{X}$

Простейший пример независимых случайных величин можно получить, рассматривая схему Бернулли. Именно, пусть

$$\Omega=\{\omega:\ \omega=(a_1,\ \ldots,\ a_n),\ a_i=0,\ 1\},\ p(\omega)=p^{\Sigma a_i}q^{n-\Sigma a_i}$$

и $\xi_i(\omega)=a_i$ для $\omega=(a_1,\ \ldots,\ a_n)$, $i=1,\ \ldots,\ n$. Тогда случайные величины $\xi_1,\ \xi_2,\ \ldots,\ \xi_n$ являются независимыми, что вытекает из установленной в § 3 независимости событий

$$A_1=\{\omega:\ a_1=1\},\ \ldots,\ A_n=\{\omega:\ a_n=1\}.$$

3. В дальнейшем нам не раз придется сталкиваться с вопросом о распределении вероятностей случайных величин, являющихся функциями $f(\xi_1,\ \ldots,\ \xi_r)$ от случайных величин $\xi_1,\ \ldots,\ \xi_r$. Рассмотрим сейчас лишь вопрос об отыскании распределения суммы случайных величин $\zeta=\xi+\eta$.

Если ξ принимает значения в множестве $X=\{x_1,\ \ldots,\ x_k\}$, а η — в множестве $Y=\{y_1,\ \ldots,\ y_l\}$, то случайная величина $\zeta=\xi+\eta$ принимает значения в множестве $Z=\{z:\ z=x_i+y_j,\ i=1,\ \ldots,\ k;\ j=1,\ \ldots,\ l\}$, и ясно, что

$$P_\zeta(z)=\mathsf{P}\{\zeta=z\}=\mathsf{P}\{\xi+\eta=z\}=\sum_{\{(i,\ j):\ x_i+y_j=z\}}\mathsf{P}\{\xi=x_i,\ \eta=y_j\}.$$

Особо важен случай независимых случайных величин ξ и η. Тогда

$$\mathsf{P}\{\xi=x_i,\ \eta=y_j\}=\mathsf{P}\{\xi=x_i\}\,\mathsf{P}\{\eta=y_j\},$$

и, значит, для любого $z\in Z$

$$P_\zeta(z)=\sum_{\{(i;\ j):\ x_i+y_j=z\}}P_\xi(x_i)\,P_\eta(y_j)=\sum_{i=1}^{k}P_\xi(x_i)\,P_\eta(z-x_i),\qquad(3)$$

где в последней сумме $P_\eta(z-x_i)$ полагается равным нулю, если $z-x_i\notin Y$.

Если, например, ξ и η — независимые бернуллиевские случайные величины, принимающие каждая значения 1 и 0 с вероятностями p и q соответственно, то $Z=\{0,\ 1,\ 2\}$ и

$$\begin{aligned}
&P_\zeta(0)=P_\xi(0)\,P_\eta(0)=q^2,\\
&P_\zeta(1)=P_\xi(0)\,P_\eta(1)+P_\xi(1)\,P_\eta(0)=2pq,\\
&P_\zeta(2)=P_\xi(1)\,P_\eta(1)=p^2.
\end{aligned}$$

По индукции легко устанавливается, что если $\xi_1,\ \xi_2,\ \ldots,\ \xi_n$ — независимые бернуллиевские случайные величины с $\mathsf{P}\{\xi_i=1\}=p$, $\mathsf{P}\{\xi_i=0\}=q$, то случайная величина $\zeta=\xi_1+\ldots+\xi_n$ имеет биномиальное распределение

$$P_\zeta(k)=C_n^kp^kq^{n-k},\qquad k=0,\ 1,\ \ldots,\ n.\qquad(4)$$

4. Перейдем теперь к важному понятию математического ожидания, или среднего значения, случайных величин.

Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — (конечное) вероятностное пространство и $\xi = \xi(\omega)$ — некоторая случайная величина, принимающая значения в множестве $X = \{x_1, \ldots, x_k\}$. Если положить $A_i = \{\omega: \xi = x_i\}$, $i = 1, \ldots, k$, то, очевидно, ξ можно представить в таком виде:

$$\xi(\omega) = \sum_{i=1}^{k} x_i I(A_i), \tag{5}$$

где множества $A_1, \ldots, A_k$ образуют разбиение пространства Ω (т. е. они попарно не пересекаются и их сумма равна Ω; см. п. 3 § 1).

Обозначим $p_i = \mathsf{P}\{\xi = x_i\}$. Интуитивно ясно, что если наблюдать за значениями случайной величины ξ в «n повторных независимых экспериментах», то значение x_i должно встретиться примерно $p_i n$ раз, $i = 1, \ldots, k$. Таким образом, среднее значение, подсчитанное по результатам n экспериментов, есть примерно

$$\frac{1}{n}[np_1x_1 + \ldots + np_kx_k] = \sum_{i=1}^{k} p_i x_i.$$

Это замечание делает понятным следующее

Определение 4. *Математическим ожиданием* или *средним значением* случайной величины $\xi = \sum_{i=1}^{k} x_i I(A_i)$ называется число

$$\mathsf{M}\xi = \sum_{i=1}^{k} x_i P(A_i). \tag{6}$$

Поскольку $A_i = \{\omega: \xi(\omega) = x_i\}$ и $P_\xi(x_i) = \mathsf{P}(A_i)$, то

$$\mathsf{M}\xi = \sum_{i=1}^{k} x_i P_\xi(x_i). \tag{7}$$

Вспомнив определение функции распределения $F_\xi = F_\xi(x)$ и обозначив

$$\Delta F_\xi(x) = F_\xi(x) - F_\xi(x-),$$

находим, что $P_\xi(x_i) = \Delta F_\xi(x_i)$ и, следовательно,

$$\mathsf{M}\xi = \sum_{i=1}^{k} x_i \Delta F_\xi(x_i). \tag{8}$$

Прежде чем переходить к рассмотрению свойств математических ожиданий, заметим, что часто приходится иметь дело с различными представлениями случайной величины ξ в виде

$$\xi(\omega)=\sum_{j=1}^{l} x_j' I(B_j),$$

где $B_1+\ldots+B_l=\Omega$, но среди x_j' могут быть, вообще говоря, одинаковые значения. В этом случае $\mathsf{M}\xi$ можно подсчитывать по формуле $\sum_{j=1}^{l} x_j'\mathsf{P}(B_j)$, не переходя к представлению (5), где все x_i различны. Действительно,

$$\sum_{\{j:\, x_j'=x_i\}} x_j'\mathsf{P}(B_j)=x_i \sum_{\{j:\, x_j'=x_i\}} \mathsf{P}(B_j)=x_i\mathsf{P}(A_i)$$

и, значит,

$$\sum_{j=1}^{l} x_j'\mathsf{P}(B_j)=\sum_{i=1}^{k} x_i\mathsf{P}(A_i).$$

5. Сформулируем основные свойства математических ожиданий:

1) *Если* $\xi\geqslant 0$, *то* $\mathsf{M}\xi\geqslant 0$.
2) $\mathsf{M}(a\xi+b\eta)=a\mathsf{M}\xi+b\mathsf{M}\eta$, a, b — постоянные.
3) *Если* $\xi\geqslant\eta$, *то* $\mathsf{M}\xi\geqslant\mathsf{M}\eta$.
4) $|\mathsf{M}\xi|\leqslant\mathsf{M}|\xi|$.
5) *Если* ξ *и* η *независимы, то* $\mathsf{M}\xi\eta=\mathsf{M}\xi\cdot\mathsf{M}\eta$.
6) $(\mathsf{M}|\xi\eta|)^2\leqslant\mathsf{M}\xi^2\cdot\mathsf{M}\eta^2$ (*неравенство Коши — Буняковского*).
7) *Если* $\xi=I(A)$, *то* $\mathsf{M}\xi=\mathsf{P}(A)$.

Свойства 1) и 7) очевидны. Для доказательства 2) пусть

$$\xi=\sum_i x_i I(A_i),\quad \eta=\sum_j y_j I(B_j).$$

Тогда

$$a\xi+b\eta=a\sum_{i,j} x_i I(A_i\cap B_j)+b\sum_{i,j} y_j I(A_i\cap B_j)=$$
$$=\sum_{i,j}(ax_i+by_j)\, I(A_i\cap B_j)$$

и

$$\mathsf{M}(a\xi+b\eta)=\sum_{i,j}(ax_i+by_j)\,\mathsf{P}(A_i\cap B_j)=$$
$$=\sum_i ax_i\mathsf{P}(A_i)+\sum_j by_j\mathsf{P}(B_j)=$$
$$=a\sum_i x_i\mathsf{P}(A_i)+b\sum_j y_j\mathsf{P}(B_j)=a\mathsf{M}\xi+b\mathsf{M}\eta.$$

Свойство 3) следует из 1) и 2). Свойство 4) очевидно, поскольку

$$|\mathsf{M}\xi| = \Big|\sum_i x_i \mathsf{P}(A_i)\Big| \leqslant \sum_i |x_i| \mathsf{P}(A_i) = \mathsf{M}|\xi|.$$

Для доказательства свойства 5) заметим, что

$$\mathsf{M}\xi\eta = \mathsf{M}\Big(\sum_i x_i I(A_i)\Big)\Big(\sum_j y_j I(B_j)\Big) =$$
$$= \mathsf{M}\sum_{i,j} x_i y_j I(A_i \cap B_j) = \sum_{i,j} x_i y_j \mathsf{P}(A_i \cap B_j) =$$
$$= \sum_{i,j} x_i y_j \mathsf{P}(A_i)\mathsf{P}(B_j) =$$
$$= \Big(\sum_i x_i \mathsf{P}(A_i)\Big)\cdot\Big(\sum_j y_j \mathsf{P}(B_j)\Big) = \mathsf{M}\xi\cdot\mathsf{M}\eta,$$

где мы воспользовались тем, что для независимых случайных величин события

$$A_i = \{\omega:\ \xi(\omega) = x_i\} \quad \text{и} \quad B_j = \{\omega:\ \eta(\omega) = y_j\}$$

являются независимыми: $\mathsf{P}(A_i \cap B_j) = \mathsf{P}(A_i)\mathsf{P}(B_j)$.

Чтобы доказать свойство 6), заметим, что

$$\xi^2 = \sum_i x_i^2 I(A_i), \quad \eta^2 = \sum_j y_j^2 I(B_j)$$

и

$$\mathsf{M}\xi^2 = \sum_i x_i^2 \mathsf{P}(A_i), \quad \mathsf{M}\eta^2 = \sum_j y_j^2 \mathsf{P}(B_j).$$

Пусть $\mathsf{M}\xi^2 > 0$, $\mathsf{M}\eta^2 > 0$. Положим

$$\tilde{\xi} = \frac{\xi}{\sqrt{\mathsf{M}\xi^2}}, \quad \tilde{\eta} = \frac{\eta}{\sqrt{\mathsf{M}\eta^2}}.$$

Поскольку $2|\tilde{\xi}\tilde{\eta}| \leqslant \tilde{\xi}^2 + \tilde{\eta}^2$, то $2\mathsf{M}|\tilde{\xi}\tilde{\eta}| \leqslant \mathsf{M}\tilde{\xi}^2 + \mathsf{M}\tilde{\eta}^2 = 2$. Значит, $\mathsf{M}|\tilde{\xi}\tilde{\eta}| \leqslant 1$ и $(\mathsf{M}|\xi\eta|)^2 \leqslant \mathsf{M}\xi^2\cdot\mathsf{M}\eta^2$.

Если же, скажем, $\mathsf{M}\xi^2 = 0$, то это означает, что $\sum_i x_i^2 \mathsf{P}(A_i) = 0$ и, следовательно, среди значений, принимаемых случайной величиной ξ, есть значение 0, причем $\mathsf{P}\{\omega:\ \xi(\omega) = 0\} = 1$. Поэтому, если по крайней мере одно из значений $\mathsf{M}\xi^2$ или $\mathsf{M}\eta^2$ равно нулю, то, очевидно, $\mathsf{M}|\xi\eta| = 0$ и, следовательно, неравенство Коши — Буняковского также выполняется.

Замечание. Свойство 5) обобщается очевидным образом на любое конечное число случайных величин: если $\xi_1, \ldots, \xi_r$ независимы, то

$$\mathsf{M}\xi_1 \ldots \xi_r = \mathsf{M}\xi_1 \ldots \mathsf{M}\xi_r.$$

Доказательство здесь то же, что и для случая $r=2$ или по индукции.

Пример 3. Пусть ξ — бернуллиевская случайная величина, принимающая значения 1 и 0 с вероятностями p и q. Тогда

$$\mathsf{M}\xi = 1\cdot \mathsf{P}\{\xi=1\}+0\cdot\mathsf{P}\{\xi=0\}=p.$$

Пример 4. Пусть $\xi_1, \ldots, \xi_n$ — n бернуллиевских случайных величин с $\mathsf{P}\{\xi_i=1\}=p$, $\mathsf{P}\{\xi_i=0\}=q$, $p+q=1$. Тогда для

$$S_n=\xi_1+\ldots+\xi_n$$

находим, что

$$\mathsf{M}S_n=np.$$

К этому результату можно прийти и другим путем. Нетрудно понять, что $\mathsf{M}S_n$ не изменится, если предположить, что бернуллиевские случайные величины $\xi_1, \ldots, \xi_n$ независимы. При этом предположении, согласно (4),

$$\mathsf{P}(S_n=k)=C_n^k p^k q^{n-k}, \qquad k=0,\ 1,\ \ldots,\ n.$$

Поэтому

$$\mathsf{M}S_n=\sum_{k=0}^{n} k\mathsf{P}(S_n=k)=\sum_{k=0}^{n} kC_n^k p^k q^{n-k}=$$

$$=\sum_{k=0}^{n} k\cdot\frac{n!}{k!\,(n-k)!}\,p^k q^{n-k}=$$

$$=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!\,((n-1)-(k-1))!}\,p^{k-1}q^{(n-1)-(k-1)}=$$

$$=np\sum_{l=0}\frac{(n-1)!}{l!\,((n-1)-l)!}\,p^l q^{(n-1)-l}=np.$$

Впрочем, первый способ приводит к результату быстрее, нежели последний.

6. Пусть $\xi=\sum_i x_i I(A_i)$, где $A_i=\{\omega\colon \xi(\omega)=x_i\}$, и $\varphi=\varphi(\xi(\omega))$ — некоторая функция от $\xi(\omega)$. Если $B_j=\{\omega\colon \varphi(\xi(\omega))=y_j\}$, то

$$\varphi(\xi(\omega))=\sum_j y_j I(B_j),$$

и, следовательно,

$$\mathsf{M}\varphi=\sum_j y_j\mathsf{P}(B_j)=\sum_j y_j\mathsf{P}_\varphi(y_j). \tag{9}$$

Но ясно также, что

$$\varphi(\xi(\omega))=\sum_i \varphi(x_i)\,I(A_i).$$

Поэтому наряду с (9) для подсчета математического ожидания случайной величины $\varphi = \varphi(\xi)$ можно пользоваться формулой

$$M\varphi(\xi) = \sum_i \varphi(x_i) P_\xi(x_i).$$

7. Следующее важное понятие дисперсии случайной величины ξ характеризует степень разброса значений ξ относительно ее математического ожидания.

Определение 5. *Дисперсией* случайной величины ξ (обозначается $\mathsf{D}\xi$) называется величина

$$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2.$$

Величина $\sigma = +\sqrt{\mathsf{D}\xi}$ называется стандартным отклонением.

Поскольку

$$\mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{M}(\xi^2 - 2\xi \cdot \mathsf{M}\xi + (\mathsf{M}\xi)^2) = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2,$$

то

$$\mathsf{D}\xi = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2.$$

Ясно, что $\mathsf{D}\xi \geqslant 0$. Из определения дисперсии также следует, что

$$\mathsf{D}(a + b\xi) = b^2\mathsf{D}\xi, \quad a, \ b \text{ — постоянные.}$$

В частности, $\mathsf{D}a = 0$, $\mathsf{D}(b\xi) = b^2\mathsf{D}\xi$.

Пусть ξ и η — две случайные величины. Тогда

$$\begin{aligned}\mathsf{D}(\xi + \eta) &= \mathsf{M}((\xi - \mathsf{M}\xi) + (\eta - \mathsf{M}\eta))^2 = \\ &= \mathsf{D}\xi + \mathsf{D}\eta + 2\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta).\end{aligned}$$

Обозначим

$$\operatorname{cov}(\xi, \eta) = \mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta).$$

Эта величина называется *ковариацией* случайных величин ξ и η. Если $\mathsf{D}\xi > 0$, $\mathsf{D}\eta > 0$, то величина

$$\rho(\xi, \eta) = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\mathsf{D}\xi \cdot \mathsf{D}\eta}}$$

называется *коэффициентом корреляции* случайных величин ξ и η. Нетрудно показать (см. далее задачу 7), что если $\rho(\xi, \eta) = \pm 1$, то величины ξ и η линейно зависимы:

$$\eta = a\xi + b,$$

где $a > 0$, если $\rho(\xi, \eta) = 1$, и $a < 0$, если $\rho(\xi, \eta) = -1$.

Сразу отметим, что если ξ и η независимы, то независимы $\xi - \mathsf{M}\xi$ и $\eta - \mathsf{M}\eta$, а значит, по свойству 5) математических ожиданий

$$\operatorname{cov}(\xi, \eta) = \mathsf{M}(\xi - \mathsf{M}\xi) \cdot \mathsf{M}(\eta - \mathsf{M}\eta) = 0.$$

С учетом введенного обозначения для ковариации

$$\mathsf{D}(\xi + \eta) = \mathsf{D}\xi + \mathsf{D}\eta + 2\operatorname{cov}(\xi, \eta), \tag{10}$$

если же ξ и η независимы, то дисперсия суммы $\xi+\eta$ равна сумме дисперсий

$$\mathsf{D}(\xi+\eta)=\mathsf{D}\xi+\mathsf{D}\eta. \tag{11}$$

Как следует из (10), свойство (11) остается выполненным, и при меньшем предположении, нежели независимость ξ и η. Именно, достаточно предположить, что величины ξ и η некоррелированы, т. е. $\operatorname{cov}(\xi, \eta)=0$.

Замечание. Из некоррелированности ξ и η, вообще говоря, не следует их независимость. Вот простой пример. Пусть случайная величина α принимает значения 0, $\pi/2$ и π с вероятностями $1/3$. Тогда $\xi=\sin\alpha$ и $\eta=\cos\alpha$ некоррелированы; в то же время они не только стохастически зависимы (т. е. не независимы относительно вероятности P):

$$\mathsf{P}\{\xi=1,\ \eta=1\}=0\neq 1/9=\mathsf{P}\{\xi=1\}\,\mathsf{P}\{\eta=1\},$$

но и функционально зависимы: $\xi^2+\eta^2=1$.

Свойства (10), (11) очевидным образом распространяются на произвольное число случайных величин $\xi_1, \ldots, \xi_n$:

$$\mathsf{D}\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}\mathsf{D}\xi_i+2\sum_{i>j}\operatorname{cov}(\xi_i, \xi_j). \tag{12}$$

В частности, если величины $\xi_1, \ldots, \xi_n$ попарно независимы (достаточно их попарной некоррелированности), то

$$\mathsf{D}\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}\mathsf{D}\xi_i. \tag{13}$$

Пример 5. Если ξ — бернуллиевская случайная величина, принимающая два значения 1 и 0 с вероятностями p и q, то

$$\mathsf{D}\xi=\mathsf{M}(\xi-\mathsf{M}\xi)^2=\mathsf{M}(\xi-p)^2=(1-p)^2p+p^2q=pq.$$

Отсюда следует, что если $\xi_1, \ldots, \xi_n$ — последовательность независимых (одинаково распределенных) бернуллиевских случайных величин и $S_n=\xi_1+\ldots+\xi_n$, то

$$\mathsf{D}S_n=npq. \tag{14}$$

8. Рассмотрим две случайные величины ξ и η. Предположим, что наблюдению подлежит лишь случайная величина ξ. Если величины ξ и η коррелированы, то можно ожидать, что знание значений ξ позволит вынести некоторые суждения и о значениях ненаблюдаемой величины η.

Всякую функцию $f=f(\xi)$ от ξ будем называть *оценкой* для η. Будем говорить также, что *оценка* $f^*=f^*(\xi)$ *оптимальна в среднеквадратическом смысле*, если

$$\mathsf{M}(\eta-f^*(\xi))^2=\inf_f \mathsf{M}(\eta-f(\xi))^2.$$

Покажем, как найти оптимальную оценку в классе линейных оценок $\lambda(\xi)=a+b\xi$. Для этого рассмотрим функцию $g(a, b)=\mathsf{M}(\eta-(a+b\xi))^2$. Дифференцируя $g(a, b)$ по a и b, получаем

$$\frac{\partial g(a, b)}{\partial a}=-2\mathsf{M}[\eta-(a+b\xi)],$$

$$\frac{\partial g(a, b)}{\partial b}=-2\mathsf{M}[(\eta-(a+b\xi))\xi],$$

откуда, приравнивая производные к нулю, находим, что оптимальная в среднеквадратическом смысле линейная оценка есть $\lambda^*(\xi)=a^*+b^*\xi$, где

$$a^*=\mathsf{M}\eta-b^*\mathsf{M}\xi, \qquad b^*=\frac{\operatorname{cov}(\xi, \eta)}{\mathsf{D}\xi}. \tag{15}$$

Иначе говоря,

$$\lambda^*(\xi)=\mathsf{M}\eta+\frac{\operatorname{cov}(\xi, \eta)}{\mathsf{D}\xi}(\xi-\mathsf{M}\xi). \tag{16}$$

Величина $M(\eta-\lambda^*(\xi))^2$ называется среднеквадратической ошибкой оценивания. Простой подсчет показывает, что эта ошибка равна

$$\Delta^*=\mathsf{M}(\eta-\lambda^*(\xi))^2=\mathsf{D}\eta-\frac{\operatorname{cov}^2(\xi, \eta)}{\mathsf{D}\xi}=\mathsf{D}\eta[1-\rho^2(\xi, \eta)]. \tag{17}$$

Таким образом, чем больше (по модулю) коэффициент корреляции $\rho(\xi, \eta)$ между ξ и η, тем меньше среднеквадратическая ошибка оценивания Δ^*. В частности, если $|\rho(\xi, \eta)|=1$, то $\Delta^*=0$ (ср. с результатом задачи 7). Если же случайные величины ξ и η не коррелированы ($\rho(\xi, \eta)=0$), то $\lambda^*(\xi)=\mathsf{M}\eta$, т. е. в случае отсутствия корреляции между ξ и η лучшей оценкой η по ξ является просто $\mathsf{M}\eta$ (ср. с задачей 4).

9. Задачи.

1. Проверить следующие свойства индикаторов $I_A=I_A(\omega)$:

$$I_\varnothing=0, \quad I_\Omega=1, \quad I_A+I_{\bar{A}}=1,$$

$$I_{AB}=I_A\cdot I_B,$$

$$I_{A\cup B}=I_A+I_B-I_{AB},$$

$$I_{\bigcup\limits_{i=1}^n A_i}=1-\prod_{i=1}^n(1-I_{A_i}), \qquad I_{\overline{\bigcup\limits_{i=1}^n A_i}}=\prod_{i=1}^n(1-I_{A_i}),$$

$$I_{\sum\limits_{i=1}^n A_i}=\sum_{i=1}^n I_{A_i},$$

$$I_{A\triangle B}=(I_A-I_B)^2,$$

где $A \triangle B$ — *симметрическая разность* множеств A и B, т. е. множество $(A \setminus B) \cup (B \setminus A)$.

2. Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины и

$$\xi_{\min} = \min(\xi_1, \ldots, \xi_n), \quad \xi_{\max} = \max(\xi_1, \ldots, \xi_n).$$

Показать, что

$$\mathsf{P}\{\xi_{\min} \geqslant x\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i \geqslant x\},$$
$$\mathsf{P}\{\xi_{\max} < x\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i < x\}.$$

3. Пусть $\xi_1, \ldots, \xi_n$ — независимые бернуллиевские случайные величины с

$$\mathsf{P}\{\xi_i = 0\} = 1 - \lambda_i \Delta,$$
$$\mathsf{P}\{\xi_i = 1\} = \lambda_i \Delta,$$

где Δ — малое число, $\Delta > 0$, $\lambda_i > 0$.

Показать, что

$$\mathsf{P}\{\xi_1 + \ldots + \xi_n = 1\} = \left(\sum_{i=1}^{n} \lambda_i\right) \Delta + O(\Delta^2),$$
$$\mathsf{P}\{\xi_1 + \ldots + \xi_n > 1\} = O(\Delta^2).$$

4. Показать, что $\inf\limits_{-\infty < a < \infty} \mathsf{M}(\xi - a)^2$ достигается при $a = \mathsf{M}\xi$ и, следовательно,

$$\inf_{-\infty < a < \infty} \mathsf{M}(\xi - a)^2 = \mathsf{D}\xi.$$

5. Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$ и m_e — медиана $F_\xi(x)$, т. е. такая точка, что

$$F_\xi(m_e -) \leqslant \frac{1}{2} \leqslant F_\xi(m_e).$$

Показать, что

$$\inf_{-\infty < a < \infty} \mathsf{M}|\xi - a| = \mathsf{M}|\xi - m_e|.$$

6. Пусть $P_\xi(x) = \mathsf{P}\{\xi = x\}$ и $F_\xi(x) = \mathsf{P}\{\xi \leqslant x\}$. Показать, что для $a > 0$ и $-\infty < b < \infty$

$$P_{a\xi+b}(x) = P_\xi\left(\frac{x-b}{a}\right),$$
$$F_{a\xi+b}(x) = F_\xi\left(\frac{x-b}{a}\right).$$

Если $y \geqslant 0$, то

Пусть $\xi^+ = \max(\xi, 0)$. Тогда

$$F_{\xi^+}(x) = \begin{cases} 0, & x < 0, \\ F_\xi(0), & x = 0, \\ F_\xi(x), & x > 0. \end{cases}$$

7. Пусть ξ и η — две случайные величины с $\mathsf{D}\xi > 0$, $\mathsf{D}\eta > 0$, и $\rho = \rho(\xi, \eta)$ — их коэффициент корреляции. Показать, что $|\rho| \leqslant 1$. При этом, если $|\rho| = 1$, то найдутся такие константы a и b, что $\eta = a\xi + b$. Более того, если $\rho = 1$, то

$$\frac{\eta - \mathsf{M}\eta}{\sqrt{\mathsf{D}\eta}} = \frac{\xi - \mathsf{M}\xi}{\sqrt{\mathsf{D}\xi}}$$

(и, значит, $a > 0$), если же $\rho = -1$, то

$$\frac{\eta - \mathsf{M}\eta}{\sqrt{\mathsf{D}\eta}} = -\frac{\xi - \mathsf{M}\xi}{\sqrt{\mathsf{D}\xi}}$$

(и, значит, $a < 0$).

8. Пусть ξ и η — две случайные величины с $\mathsf{M}\xi = \mathsf{M}\eta = 0$, $\mathsf{D}\xi = \mathsf{D}\eta = 1$ и коэффициентом корреляции $\rho = \rho(\xi, \eta)$. Показать, что

$$\mathsf{M}\max(\xi^2, \eta^2) \leqslant 1 + \sqrt{1 - \rho^2}.$$

9. Используя равенство $I_{\bigcup\limits_{i=1}^{n} A_i} = \prod\limits_{i=1}^{n}(1 - I_{A_i})$, доказать формулу $\mathsf{P}(B_0) = 1 - S_1 + S_2 + \ldots \pm S_n$ из задачи 4 § 1.

10. Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины, $\varphi_1 = \varphi_1(\xi_1, \ldots, \xi_k)$ и $\varphi_2 = \varphi_2(\xi_{k+1}, \ldots, \xi_n)$ — две функции от $\xi_1, \ldots, \xi_k$ и $\xi_{k+1}, \ldots, \xi_n$ соответственно. Показать, что случайные величины φ_1 и φ_2 независимы.

11. Показать, что случайные величины $\xi_1, \ldots, \xi_n$ независимы тогда и только тогда, когда для всех $x_1, \ldots, x_n$

$$F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \ldots F_{\xi_n}(x_n),$$

где $F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \mathsf{P}\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\}$.

12. Показать, что случайная величина ξ не зависит от самой себя (т. е. ξ и ξ независимы) в том и только том случае, когда $\xi \equiv \mathrm{const}$.

13. При каких условиях на ξ случайные величины ξ и $\sin\xi$ независимы?

14. Пусть ξ и η — независимые случайные величины и $\eta \neq 0$. Выразить вероятности событий $\mathsf{P}\{\xi\eta \leqslant z\}$ и $\mathsf{P}\left\{\frac{\xi}{\eta} \leqslant z\right\}$ через вероятности $P_\xi(x)$ и $P_\eta(y)$.

§ 5. Схема Бернулли. I. Закон больших чисел

1. В соответствии с данными выше определениями тройка

$$(\Omega,\ \mathscr{A},\ \mathsf{P})\ \text{с}\ \Omega=\{\omega\colon \omega=(a_1,\ \ldots,\ a_n),\ a_i=0,\ 1\},$$
$$\mathscr{A}=\{A\colon A\subseteq\Omega\},\quad p(\omega)=p^{\sum a_i}q^{n-\sum a_i}$$

была названа вероятностной моделью, отвечающей n независимым испытаниям с двумя исходами, или схемой Бернулли.

В этом и следующем параграфе мы изучим некоторые предельные (в указываемом ниже смысле) свойства схем Бернулли, которые оказывается удобным вести в терминах случайных величин и вероятностей событий, связанных с ними.

Введем случайные величины $\xi_1, \ldots, \xi_n$, полагая для $\omega=(a_1, \ldots, a_n)$, что $\xi_i(\omega)=a_i$, $i=1, \ldots, n$. Как мы уже видели, бернуллиевские величины $\xi_i(\omega)$ независимы и одинаково распределены:

$$\mathsf{P}\{\xi_i=1\}=p,\quad \mathsf{P}\{\xi_i=0\}=q,\qquad i=1,\ \ldots,\ n.$$

Понятно, что случайная величина ξ_i характеризует результат испытания на i-м шаге (в i-й момент времени).

Положим $S_0(\omega)\equiv 0$ и

$$S_k=\xi_1+\ldots+\xi_k,\qquad k=1,\ \ldots,\ n.$$

Как было найдено выше, $\mathsf{M}S_n=np$ и, следовательно,

$$\mathsf{M}\,\frac{S_n}{n}=p. \tag{1}$$

Иначе говоря, среднее значение частоты появления «успеха», т. е. S_n/n, совпадает с вероятностью «успеха» p. Отсюда естественно возникает вопрос о том, как велики отклонения частоты S_n/n появления «успеха» от его вероятности p.

Прежде всего отметим, что не приходится рассчитывать на то, что при достаточно малых $\varepsilon>0$ и даже при больших значениях n отклонения частоты S_n/n от вероятности p будут меньше ε для всех ω, т. е. что будет выполнено неравенство

$$\left|\frac{S_n(\omega)}{n}-p\right|\leqslant\varepsilon,\qquad \omega\in\Omega. \tag{2}$$

Действительно, при $0<p<1$

$$\mathsf{P}\left\{\frac{S_n}{n}=1\right\}=\mathsf{P}\{\xi_1=1,\ \ldots,\ \xi_n=1\}=p^n,$$
$$\mathsf{P}\left\{\frac{S_n}{n}=0\right\}=\mathsf{P}\{\xi_1=0,\ \ldots,\ \xi_n=0\}=q^n,$$

откуда следует, что неравенство (2) не выполняется при достаточно малых $\varepsilon>0$.

Однако мы замечаем, что при больших n вероятности событий $\left\{\frac{S_n}{n}=1\right\}$ и $\left\{\frac{S_n}{n}=0\right\}$ малы. Естественна поэтому мысль, что суммарная вероятность исходов ω, для которых $\left|\frac{S_n(\omega)}{n}-p\right|>\varepsilon$, будет при достаточно больших n также мала.

В связи с этим постараемся оценить вероятность события $\left\{\omega\colon \left|\frac{S_n(\omega)}{n}-p\right|>\varepsilon\right\}$, для чего воспользуемся следующим неравенством, открытым П. Л. Чебышевым.

Неравенство Чебышева. *Пусть* $(\Omega, \mathscr{A}, \mathsf{P})$ — *некоторое вероятностное пространство и* $\xi=\xi(\omega)$ — *неотрицательная случайная величина. Тогда для всякого* $\varepsilon>0$

$$\mathsf{P}\{\xi\geqslant\varepsilon\}\leqslant\frac{\mathsf{M}\xi}{\varepsilon}. \tag{3}$$

Доказательство. Заметим, что

$$\xi=\xi I(\xi\geqslant\varepsilon)+\xi I(\xi<\varepsilon)\geqslant\xi I(\xi\geqslant\varepsilon)\geqslant\varepsilon I(\xi\geqslant\varepsilon),$$

где $I(A)$ — индикатор множества A.

Поэтому по свойствам математических ожиданий

$$\mathsf{M}\xi\geqslant\varepsilon\mathsf{M}I(\xi\geqslant\varepsilon)=\varepsilon\mathsf{P}(\xi\geqslant\varepsilon),$$

что и доказывает (3).

Следствия. Если ξ — произвольная случайная величина, то для $\varepsilon>0$

$$\begin{gathered}\mathsf{P}\{|\xi|\geqslant\varepsilon\}\leqslant\frac{\mathsf{M}|\xi|}{\varepsilon},\\ \mathsf{P}\{|\xi|\geqslant\varepsilon\}=\mathsf{P}\{\xi^2\geqslant\varepsilon^2\}\leqslant\frac{\mathsf{M}\xi^2}{\varepsilon^2},\\ \mathsf{P}\{|\xi-\mathsf{M}\xi|\geqslant\varepsilon\}\leqslant\frac{\mathsf{D}\xi}{\varepsilon^2}.\end{gathered} \tag{4}$$

Воспользуемся последним неравенством, взяв $\xi=S_n/n$. Тогда с учетом (4.14) получим

$$\mathsf{P}\left\{\left|\frac{S_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{\mathsf{D}\left(\frac{S_n}{n}\right)}{\varepsilon^2}=\frac{\mathsf{D}S_n}{n^2\varepsilon^2}=\frac{npq}{n^2\varepsilon^2}=\frac{pq}{n\varepsilon^2}.$$

Итак,

$$\mathsf{P}\left\{\left|\frac{S_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{pq}{n\varepsilon^2}\leqslant\frac{1}{4n\varepsilon^2}, \tag{5}$$

откуда видно, что при больших n вероятность отклонения частоты «успеха» S_n/n от его вероятности p больше чем на ε достаточно мала.

Обозначим для всех $n \geqslant 1$ и $0 \leqslant k \leqslant n$

$$P_n(k) = C_n^k p^k q^{n-k}.$$

Тогда

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} = \sum_{\left\{k:\, \left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\}} P_n(k),$$

и, в сущности, мы установили, что

$$\sum_{\left\{k:\, \left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\}} P_n(k) \leqslant \frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}, \tag{6}$$

т. е. доказали некоторое неравенство, которое можно было бы получить аналитически, без использования вероятностной интерпретации.

Из (6) ясно, что

$$\sum_{\left\{k:\, \left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\}} P_n(k) \to 0, \qquad n \to \infty. \tag{7}$$

Графически это утверждение можно пояснить следующим образом. Изобразим биномиальное распределение $\{P_n(k),\ 0 \leqslant k \leqslant n\}$, как это сделано на рис. 6.

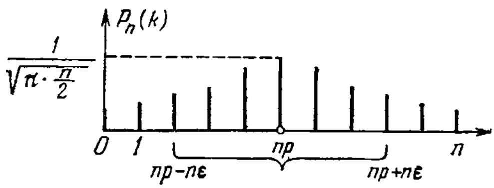


Рис. 6.

Тогда с ростом n вся картина «расплывается», в то же время «сжимаясь» по высоте. При этом сумма величин $P_n(k)$ по k таким, что $np - n\varepsilon \leqslant k \leqslant np + n\varepsilon$, стремится к единице.

Будем представлять последовательность случайных величин $S_0, S_1, \ldots, S_n$ как *траекторию* некоторой блуждающей частицы. Тогда результат (7) означает следующее.

Проведем прямые kp, $k(p+\varepsilon)$ и $k(p-\varepsilon)$. Тогда в среднем траектория движется вдоль прямой kp, и для любого $\varepsilon > 0$ можно утверждать, что для достаточно больших n с большой вероятностью точка S_n, характеризующая положение частицы в момент n, будет лежать в интервале $[n(p-\varepsilon), n(p+\varepsilon)]$; см. рис. 7.

Утверждение (7) хотелось бы записать в таком виде:

$$\mathsf{P}\left\{\left|\frac{S_n}{n}-p\right|\geqslant\varepsilon\right\}\to 0,\qquad n\to\infty. \tag{8}$$

Однако надо иметь в виду, что здесь существует определенная тонкость. Дело в том, что эта запись была бы вполне оправданной, если бы P была вероятностью на некотором пространстве $(\Omega, \mathscr{A})$, на котором определена *бесконечная* последовательность независимых бернуллиевских случайных величин ξ_1,

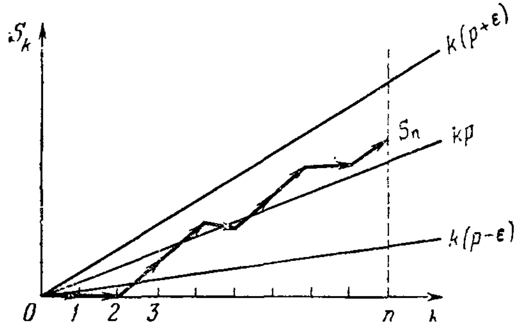


Рис. 7.

$\xi_2, \ldots$ Эти объекты действительно можно построить и тем самым придать утверждению (8) совершенно строгий вероятностный смысл (см. далее следствие 1 к теореме 1 § 9 в гл. II). Пока же, если желать придать смысл аналитическому утверждению (7), пользуясь языком теории вероятностей, мы доказали лишь следующее.

Пусть $(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathsf{P}^{(n)})$, $n\geqslant 1$, — последовательность схем Бернулли таких, что

$$\Omega^{(n)}=\left\{\omega^{(n)}\colon\ \omega^{(n)}=\left(a_1^{(n)},\ldots,a_n^{(n)}\right),\ a_i^{(n)}=0,\ 1\right\},$$

$$\mathscr{A}^{(n)}=\left\{A\colon\ A\subseteq\Omega^{(n)}\right\},$$

$$p^{(n)}\left(\omega^{(n)}\right)=p^{\sum a_i^{(n)}}q^{n-\sum a_i^{(n)}}$$

и

$$S_k^{(n)}\left(\omega^{(n)}\right)=\xi_1^{(n)}\left(\omega^{(n)}\right)+\ldots+\xi_k^{(n)}\left(\omega^{(n)}\right),$$

где для каждого $n\geqslant 1$ $\xi_1^{(n)},\ldots,\xi_n^{(n)}$ — последовательности независимых одинаково распределенных бернуллиевских случайных величин.

Тогда

$$\mathsf{P}^{(n)}\left\{\omega^{(n)}:\ \left|\frac{S_n^{(n)}(\omega^{(n)})}{n}-p\right|\geqslant\varepsilon\right\}=\sum_{\left\{k:\ \left|\frac{k}{n}-p\right|\geqslant\varepsilon\right\}} P_n(k)\to 0,\quad n\to\infty. \tag{9}$$

Утверждения типа (7) — (9) носят название **закона больших чисел Я. Бернулли.** Отметим, что доказательство Я. Бернулли именно и состояло в установлении утверждения (7), что было сделано им вполне строго с использованием оценок для «хвостов» биномиальных вероятностей $P_n(k)$ (при тех k, для которых $\left|\frac{k}{n}-p\right|\geqslant\varepsilon$). Непосредственное вычисление суммы вероятностей «хвостов» биномиального распределения $\sum_{\left\{k:\ \left|\frac{k}{n}-p\right|\geqslant\varepsilon\right\}} P_n(k)$ представляет для больших n довольно трудоемкую задачу, к тому же получаемые формулы мало пригодны для практической оценки того, с какой вероятностью частоты S_n/n отличаются от p меньше чем на ε. Именно поэтому большое значение имели открытые Муавром (в случае $p=1/2$) и затем Лапласом (для произвольного $0<p<1$) простые асимптотические формулы для вероятностей $P_n(k)$, что позволило не только заново доказать закон больших чисел, но и получить его уточнения — так называемые локальные и интегральные предельные теоремы, суть которых состоит в том, что при больших n и по крайней мере для $k\sim np$

$$P_n(k)\sim\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}},$$

а

$$\sum_{\left\{k:\ \left|\frac{k}{n}-p\right|\leqslant\varepsilon\right\}} P_n(k)\sim\frac{1}{\sqrt{2\pi}}\int_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} e^{-\frac{x^2}{2}}\,dx.$$

2. Следующий параграф посвящен точным формулировкам и доказательствам этих результатов. Сейчас же мы остановимся на вопросе о том, каков реальный смысл закона больших чисел, какова его эмпирическая интерпретация?

Пусть производится большое число, скажем, N, серий экспериментов, каждая из которых состоит из «n независимых испытаний с вероятностью интересующего нас события C, равной p». Пусть S_n^i/n — частота появления события C в i-й серии и N_ε — число серий, в которых частоты отклоняются от p меньше чем на ε:

$$N_\varepsilon \text{ равно числу тех } i, \text{ для которых } \left|\frac{S_n^i}{n}-p\right|\leqslant\varepsilon.$$

Тогда

$$\frac{N_\varepsilon}{N} \sim P_\varepsilon, \tag{10}$$

где $P_\varepsilon = \mathsf{P}\left\{\left|\frac{S_n^1}{n} - p\right| \leqslant \varepsilon\right\}$.

Важно подчеркнуть, что попытка уточнить соотношение (10) неминуемо приводит к необходимости использования некоторой вероятностной меры точно так же, как оценка отклонения частоты S_n/n от p оказывается возможной лишь после привлечения вероятностной меры P.

3. Рассмотрим полученную выше оценку

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} = \sum_{\left\{k:\ \left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\}} P_n(k) \leqslant \frac{1}{4n\varepsilon^2} \tag{11}$$

для ответа на следующий, типичный для математической статистики вопрос: каково наименьшее гарантированное число наблюдений n, при котором (для любого $0 < p < 1$)

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha, \tag{12}$$

где α — заданное (обычно малое) число?

Из (11) следует, что таким числом является наименьшее целое n, для которого

$$n \geqslant \frac{1}{4\varepsilon^2\alpha}. \tag{13}$$

Если, например, $\alpha = 0{,}05$ и $\varepsilon = 0{,}02$, то число наблюдений, равное 12 500, гарантирует выполнение неравенства (12) независимо от значения неизвестного параметра p.

Далее мы увидим (п. 5, § 6), что это число наблюдений сильно завышено; это объясняется тем, что неравенство Чебышева дает слишком грубую оценку сверху вероятности $P\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\}$.

4. Обозначим

$$C(n, \varepsilon) = \left\{\omega:\ \left|\frac{S_n(\omega)}{n} - p\right| \leqslant \varepsilon\right\}.$$

Из доказанного закона больших чисел следует, что для всякого $\varepsilon > 0$ при достаточно больших n вероятность множества $C(n, \varepsilon)$ близка к единице. В этом смысле траектории (реализации) ω из $C(n, \varepsilon)$ естественно назвать *типичными* (или (n, ε)-типичными).

Поставим следующий вопрос: каково число типичных реализаций и вес $p(\omega)$ каждой типичной реализации?

С этой целью заметим сначала, что общее число точек $N(\Omega) = 2^n$, и если $p = 0$ или 1, то множество типичных траекторий

$C(n, \varepsilon)$ состоит всего лишь из одной траектории $(0, 0, \ldots, 0)$ или $(1, 1, \ldots, 1)$. Но если $p = 1/2$, то интуитивно понятно, что «почти все» траектории (за исключением траектории типа $(0, 0, \ldots \ldots, 0)$ или $(1, 1, \ldots, 1)$) будут типичными и, следовательно, их число должно быть близко к 2^n.

Оказывается, что на поставленный вопрос можно дать исчерпывающий ответ для произвольных $0 < p < 1$; при этом выясняется, что как число типичных реализаций, так и их веса $p(\omega)$ определяются некоторой специальной функцией от p, называемой энтропией.

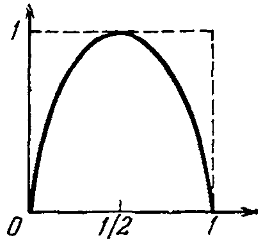


Рис. 8. Функция $H(p) = -p \ln p - (1-p) \times \ln (1-p)$.

Чтобы глубже раскрыть содержание соответствующего результата, полезно рассмотреть несколько более общую схему из п. 2 § 2, нежели схема Бернулли.

Пусть $(p_1, p_2, \ldots, p_r)$ — некоторое конечное распределение вероятностей, т. е. набор неотрицательных чисел, удовлетворяющих условию $p_1 + \ldots + p_r = 1$. *Энтропией* этого распределения называется величина

$$H = -\sum_{i=1}^{r} p_i \ln p_i, \tag{14}$$

где $0 \cdot \ln 0 = 0$. Ясно, что $H \geqslant 0$, причем $H = 0$ тогда и только тогда, когда все вероятности p_i, кроме одной, равны нулю. Функция $f(x) = -x \ln x$, $0 \leqslant x \leqslant 1$, выпукла кверху и, как хорошо известно из свойств выпуклых функций,

$$\frac{f(x_1) + \ldots + f(x_r)}{r} \leqslant f\left(\frac{x_1 + \ldots + x_r}{r}\right).$$

Следовательно,

$$H = -\sum_{i=1}^{r} p_i \ln p_i \leqslant -r \cdot \frac{p_1 + \ldots + p_r}{r} \cdot \ln\left(\frac{p_1 + \ldots + p_r}{r}\right) = \ln r.$$

Иначе говоря, энтропия достигает своего максимального значения при $p_1 = \ldots = p_r = 1/r$ (см. рис. 8 для функции $H = H(p)$ в случае $r = 2$).

Если рассматривать распределение вероятностей $(p_1, p_2, \ldots, p_r)$ как вероятности появления некоторых событий, скажем, $A_1, A_2, \ldots \ldots, A_r$, то совершенно понятно, что «степень неопределенности» в свершении того или иного события различна для различных распределений. Если, например, $p_1 = 1$, $p_2 = \ldots = p_r = 0$, то ясно, что такое распределение не обладает никакой неопределенностью: с полной уверенностью можно сказать, что в результате опыта произойдет событие A_1. Однако если $p_1 = \ldots = p_r = 1/r$, то такое распределение обладает максимальной неопределенностью в том смысле, что невозможно отдать предпочтение в свершении тому или иному событию.

Важно поэтому иметь количественную характеристику меры неопределенности различных распределений вероятностей, что позволяло бы их сравнивать с этой стороны. Такой удачной характеристикой меры неопределенности оказалась энтропия, играющая существенную роль в статистической механике и во многих важных задачах кодирования и теории связи.

Предположим теперь, что пространство исходов

$$\Omega=\{\omega:\ \omega=(a_1,\ \ldots,\ a_n),\ a_i=1,\ \ldots,\ r\}$$

и $p(\omega)=p_1^{\nu_1(\omega)}\ldots p_r^{\nu_r(\omega)}$, где $\nu_i(\omega)$ — число элементов i в последовательности ω, а $(p_1,\ \ldots,\ p_r)$ — некоторое распределение вероятностей.

Для $\varepsilon>0$ и $n=1,\ 2,\ \ldots$ положим

$$C(n,\ \varepsilon)=\left\{\omega:\left|\frac{\nu_i(\omega)}{n}-p_i\right|<\varepsilon,\quad i=1,\ \ldots,\ r\right\}.$$

Ясно, что

$$\mathsf{P}(C(n,\ \xi))\geqslant 1-\sum_{i=1}^{r}\mathsf{P}\left\{\left|\frac{\nu_i(\omega)}{n}-p_i\right|\geqslant\varepsilon\right\},$$

и для достаточно больших n в силу закона больших чисел, примененного к случайным величинам

$$\xi_k(\omega)=\begin{cases}1, & a_k=i,\\ 0, & a_k\neq i,\end{cases}\qquad k=1,\ \ldots,\ n,$$

вероятности $\mathsf{P}\left\{\left|\frac{\nu_i(\omega)}{n}-p_i\right|\geqslant\varepsilon\right\}$ достаточно малы. Тем самым при больших n вероятность события $C(n,\ \varepsilon)$ близка к единице. Поэтому, как и в случае $r=2$, траектории, входящие в $C(n,\ \varepsilon)$, будем называть типичными.

Если все $p_i>0$, то для любого $\omega\in\Omega$

$$p(\omega)=\exp\left\{-n\sum_{k=1}^{r}\left(-\frac{\nu_k(\omega)}{n}\ln p_k\right)\right\}.$$

Поэтому, если ω — типичная траектория, то

$$\left|\sum_{k=1}^{r}\left(-\frac{\nu_k(\omega)}{n}\ln p_k\right)-H\right|\leqslant-\sum_{k=1}^{r}\left|\frac{\nu_k(\omega)}{n}-p_k\right|\ln p_k\leqslant-\varepsilon\sum_{k=1}^{r}\ln p_k.$$

Отсюда следует, что для типичных траекторий вероятность $p(\omega)$ близка к e^{-nH} и — поскольку в силу закона больших чисел при больших n типичные траектории «почти» исчерпывают Ω — число таких траекторий должно быть порядка e^{nH}. Эти соображения приводят к следующему предложению.

Теорема (Макмиллан). *Пусть* $p_i > 0$, $i = 1, \ldots, r$ *и* $0 < \varepsilon < 1$. *Тогда существует* $n_0 = n_0(\varepsilon;\ p_1, \ldots, p_r)$ *такое, что для всех* $n > n_0$:

a) $e^{n(H-\varepsilon)} \leqslant N(C(n,\ \varepsilon_1)) \leqslant e^{n(H+\varepsilon)}$;

b) $e^{-n(H+\varepsilon)} \leqslant p(\omega) \leqslant e^{-n(H-\varepsilon)}, \quad \omega \in C(n,\ \varepsilon_1)$;

c) $\mathsf{P}(C(n,\ \varepsilon_1)) = \sum\limits_{\omega \in C(n,\ \varepsilon_1)} p(\omega) \to 1, \quad n \to \infty,$

где

$$\varepsilon_1 = \min\left(\varepsilon,\ \frac{\varepsilon}{-2\sum\limits_{k=1}^{r} \ln p_k}\right).$$

Доказательство. Утверждение c) следует из закона больших чисел. Для доказательства остальных утверждений заметим, что если $\omega \in C(n,\ \varepsilon_1)$, то

$$np_k - \varepsilon_1 n < \nu_k(\omega) < np_k + \varepsilon_1 n, \qquad k = 1, \ldots, r,$$

и, значит,

$$p(\omega) = \exp\left\{-\sum \nu_k \ln p_k\right\} < \exp\left\{-n\sum p_k \ln p_k - \varepsilon_1 n \sum \ln p_k\right\} \leqslant$$
$$\leqslant \exp\left\{-n\left(H - \frac{\varepsilon}{2}\right)\right\}.$$

Аналогично

$$p(\omega) > \exp\left\{-n\left(H + \frac{\varepsilon}{2}\right)\right\}.$$

Следовательно, b) и подавно выполнено.

Далее, поскольку

$$\mathsf{P}(C(n,\ \varepsilon_1)) \geqslant N(C(n,\ \varepsilon_1)) \cdot \min_{\omega \in C(n,\ \varepsilon_1)} p(\omega),$$

то

$$N(C(n,\ \varepsilon_1)) \leqslant \frac{\mathsf{P}(C(n,\ \varepsilon_1))}{\min\limits_{\omega \in C(n,\ \varepsilon_1)} p(\omega)} < \frac{1}{e^{-n\left(H + \frac{\varepsilon}{2}\right)}} = e^{n\left(H + \frac{\varepsilon}{2}\right)}$$

и аналогично

$$N(C(n,\ \varepsilon_1)) \geqslant \frac{\mathsf{P}(C(n,\ \varepsilon_1))}{\max\limits_{\omega \in C(n,\ \varepsilon_1)} p(\omega)} > \mathsf{P}(C(n,\ \varepsilon_1))\, e^{n\left(H - \frac{\varepsilon}{2}\right)}.$$

Поскольку $\mathsf{P}(C(n,\ \varepsilon_1)) \to 1$, $n \to \infty$, то найдется n_1 такое, что для $n > n_1$ $\mathsf{P}(C(n,\ \varepsilon_1)) > 1 - \varepsilon$ и, значит,

$$N(C(n,\ \varepsilon_1)) \geqslant (1 - \varepsilon)\exp\left\{n\left(H - \frac{\varepsilon}{2}\right)\right\} =$$
$$= \exp\left\{n(H - \varepsilon) + \left(\frac{n\varepsilon}{2} + \ln(1 - \varepsilon)\right)\right\}.$$

Пусть n_2 таково, что для $n > n_2$

$$\frac{n\varepsilon}{2} + \ln(1-\varepsilon) > 0.$$

Тогда для $n \geqslant n_0 = \max(n_1, n_2)$

$$N(C(n, \varepsilon_1)) \geqslant e^{n(H-\varepsilon)}.$$

Теорема доказана.

5. Закон больших чисел для схемы Бернулли позволяет дать простое и изящное доказательство известной теоремы Вейерштрасса о приближении непрерывной функции полиномами.

Пусть $f = f(p)$ — непрерывная функция на отрезке $[0, 1]$. Введем полиномы

$$B_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k p^k q^{n-k},$$

называемые *полиномами Бернштейна* по имени автора приводимого доказательства теоремы Вейерштрасса.

Если $\xi_1, \ldots, \xi_n$ — последовательность независимых бернуллиевских случайных величин с $\mathsf{P}\{\xi_i = 1\} = p$, $\mathsf{P}\{\xi_i = 0\} = q$ и $S_n = \xi_1 + \ldots + \xi_n$, то

$$\mathsf{M} f(S_n/n) = B_n(p).$$

Поскольку непрерывная на отрезке $[0, 1]$ функция $f = f(p)$ равномерно непрерывна, то для всякого $\varepsilon > 0$ найдется $\delta > 0$ такое, что $|f(x) - f(y)| \leqslant \varepsilon$, коль скоро $|x - y| \leqslant \delta$. Ясно также, что такая функция ограничена, $|f(x)| \leqslant M < \infty$.

Учитывая это и неравенство (5), находим

$$\begin{aligned} |f(p) - B_n(p)| &= \left| \sum_{k=0}^{n} \left[f(p) - f\left(\frac{k}{n}\right)\right] C_n^k p^k q^{n-k} \right| \leqslant \\ &\leqslant \sum_{\left\{k:\ \left|\frac{k}{n} - p\right| \leqslant \delta\right\}} \left| f(p) - f\left(\frac{k}{n}\right) \right| C_n^k p^k q^{n-k} + \\ &+ \sum_{\left\{k:\ \left|\frac{k}{n} - p\right| > \delta\right\}} \left| f(p) - f\left(\frac{k}{n}\right) \right| C_n^k p^k q^{n-k} \leqslant \\ &\leqslant \varepsilon + 2M \sum_{\left\{k:\ \left|\frac{k}{n} - p\right| > \delta\right\}} C_n^k p^k q^{n-k} \leqslant \varepsilon + \frac{2M}{4n\delta^2} = \varepsilon + \frac{M}{2n\delta^2}. \end{aligned}$$

Отсюда

$$\lim_{n \to \infty} \max_{0 \leqslant p \leqslant 1} |f(p) - B_n(p)| = 0,$$

что и составляет утверждение теоремы Вейерштрасса.

6. **Задачи.**

1. Пусть ξ и η — случайные величины с коэффициентом корреляции ρ. Показать справедливость следующего двумерного аналога неравенства Чебышева:

$$\mathsf{P}\{|\xi-\mathsf{M}\xi|\geqslant\varepsilon\sqrt{\mathsf{D}\xi}\ \text{или}\ |\eta-\mathsf{M}\eta|\geqslant\varepsilon\sqrt{\mathsf{D}\eta}\}\leqslant\frac{1}{\varepsilon^2}(1+\sqrt{1-\rho^2}).$$

(Указание. Воспользоваться результатом задачи 8 из § 4.)

2. Пусть $f=f(x)$ — неотрицательная четная функция, неубывающая при положительных x. Тогда для случайной величины ξ с $|\xi(\omega)|\leqslant C$

$$\frac{\mathsf{M}f(\xi)-f(\varepsilon)}{f(C)}\leqslant\mathsf{P}\{|\xi-\mathsf{M}\xi|\geqslant\varepsilon\}\leqslant\frac{\mathsf{M}f(\xi-\mathsf{M}\xi)}{f(\varepsilon)}.$$

В частности, для $f(x)=x^2$

$$\frac{\mathsf{M}\xi^2-\varepsilon^2}{C^2}\leqslant\mathsf{P}\{|\xi-\mathsf{M}\xi|\geqslant\varepsilon\}\leqslant\frac{\mathsf{D}\xi}{\varepsilon^2}.$$

3. Пусть $\xi_1,\ldots,\xi_n$ — последовательность независимых случайных величин с $\mathsf{D}\xi_i\leqslant C$. Тогда

$$\mathsf{P}\left\{\left|\frac{\xi_1+\ldots+\xi_n}{n}-\frac{\mathsf{M}(\xi_1+\ldots+\xi_n)}{n}\right|\geqslant\varepsilon\right\}\leqslant\frac{C}{n\varepsilon^2}. \tag{15}$$

(С теми же оговорками, какие были сделаны к соотношению (8), из неравенства (15) следует справедливость закона больших чисел в более общей ситуации, нежели в схеме Бернулли.)

4. Пусть $\xi_1,\ldots,\xi_n$ — независимые бернуллиевские случайные величины с $\mathsf{P}\{\xi_i=1\}=p>0$, $\mathsf{P}\{\xi_i=-1\}=1-p$. Имеет место следующая *оценка Бернштейна*: существует $a>0$ такое, что

$$\mathsf{P}\left\{\left|\frac{S_n}{n}-(2p-1)\right|\geqslant\varepsilon\right\}\leqslant 2e^{-a\varepsilon^2 n},$$

где $S_n=\xi_1+\ldots+\xi_n$ и $\varepsilon>0$.

§ 6. Схема Бернулли. II. Предельные теоремы (локальная, Муавра — Лапласа, Пуассона)

1. Как и в предыдущем параграфе, пусть

$$S_n=\xi_1+\ldots+\xi_n.$$

Тогда

$$\mathsf{M}\frac{S_n}{n}=p, \tag{1}$$

и в силу (4.14)

$$\mathsf{M}\left(\frac{S_n}{n}-p\right)^2=\frac{pq}{n}. \tag{2}$$

Из формулы (1) следует, что $\frac{S_n}{n} \sim p$, где знак эквивалентности $\sim$ получил точную интерпретацию в законе больших чисел в виде оценки вероятностей $\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\}$. Естественно думать, что аналогичным образом вытекающему из (2) соотношению

$$\left|\frac{S_n}{n} - p\right| \sim \sqrt{\frac{pq}{n}} \tag{3}$$

также можно придать точный вероятностный смысл, рассматривая, например, вероятности типа

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant x\sqrt{\frac{pq}{n}}\right\}, \qquad x \in R^1,$$

или, что то же, вероятности

$$\mathsf{P}\left\{\left|\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}\right| \leqslant x\right\}$$

(поскольку $\mathsf{M}S_n = np$ и $\mathsf{D}S_n = npq$).

Если обозначить, как и выше, для $n \geqslant 1$

$$P_n(k) = C_n^k p^k q^{n-k}, \qquad 0 \leqslant k \leqslant n,$$

то вероятность

$$\mathsf{P}\left\{\left|\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}\right| \leqslant x\right\} = \sum_{\left\{k:\ \left|\frac{k-np}{\sqrt{npq}}\right| \leqslant x\right\}} P_n(k). \tag{4}$$

Поставим задачу об отыскании удобных асимптотических формул при $n \to \infty$ для вероятностей $P_n(k)$ и их сумм для тех k, которые удовлетворяют условиям в правой части (4).

Следующий результат дает ответ не только для этих значений k (т. е. таких, что $|k - np| = O(\sqrt{npq})$), но и для тех, которые удовлетворяют условию $|k - np| = o(npq)^{2/3}$.

Локальная предельная теорема. *Пусть $0 < p < 1$, тогда равномерно по всем k таким, что $|k - np| = o(npq)^{2/3}$*

$$P_n(k) \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}, \tag{5}$$

т. е. при $n \to \infty$

$$\sup_{\{k:\ |k-np| \leqslant \varphi(n)\}} \left|\frac{P_n(k)}{\frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}} - 1\right| \to 0, \tag{6}$$

где $\varphi(n) = o(npq)^{2/3}$.

Доказательство существенно использует формулу Стирлинга (2.6)

$$n! = \sqrt{2\pi n}\, e^{-n} n^n (1 + R(n)),$$

где $R(n) \to 0$, $n \to \infty$.

Тогда, если $n \to \infty$, $k \to \infty$, $n - k \to \infty$, то

$$C_n^k = \frac{n!}{k!\,(n-k)!} = \frac{\sqrt{2\pi n}\, e^{-n} n^n}{\sqrt{2\pi k \cdot 2\pi (n-k)}\, e^{-k} k^k \cdot e^{-(n-k)} (n-k)^{n-k}} \times$$

$$\times \frac{1 + R(n)}{(1 + R(k))(1 + R(n-k))} = \frac{1}{\sqrt{2\pi n \frac{k}{n}\left(1 - \frac{k}{n}\right)}} \cdot \frac{1 + \varepsilon(n, k, n-k)}{\left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}},$$

где очевидным образом определяемая функция $\varepsilon = \varepsilon(n, k, n-k) \to 0$ при $n \to \infty$, $k \to \infty$, $n - k \to \infty$.

Поэтому

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{1}{\sqrt{2\pi n \frac{k}{n}\left(1 - \frac{k}{n}\right)}} \frac{p^k (1-p)^{n-k}}{\left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}} (1 + \varepsilon).$$

Обозначим $\hat{p} = \frac{k}{n}$. Тогда

$$P_n(k) = \frac{1}{\sqrt{2\pi n \hat{p}(1-\hat{p})}} \left(\frac{p}{\hat{p}}\right)^k \left(\frac{1-p}{1-\hat{p}}\right)^{n-k} (1 + \varepsilon) =$$

$$= \frac{1}{\sqrt{2\pi n \hat{p}(1-\hat{p})}} \exp\left\{k \ln \frac{p}{\hat{p}} + (n-k) \ln \frac{1-p}{1-\hat{p}}\right\} \cdot (1 + \varepsilon) =$$

$$= \frac{1}{\sqrt{2\pi n \hat{p}(1-\hat{p})}} \exp\left\{n \left[\frac{k}{n} \ln \frac{p}{\hat{p}} + \left(1 - \frac{k}{n}\right) \ln \frac{1-p}{1-\hat{p}}\right]\right\} (1 + \varepsilon) =$$

$$= \frac{1}{\sqrt{2\pi n \hat{p}(1-\hat{p})}} \exp\{-nH(\hat{p})\} (1 + \varepsilon),$$

где

$$H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}.$$

Рассматриваемые значения k таковы, что $|k - np| = o(npq)^{2/3}$, а значит, $p - \hat{p} \to 0$, $n \to \infty$.

Поскольку, для $0 < x < 1$

$$H'(x) = \ln \frac{x}{p} - \ln \frac{1-x}{1-p},$$

$$H''(x) = \frac{1}{x} + \frac{1}{1-x},$$

то, представив $H(\hat{p})$ в виде $H(p+(\hat{p}-p))$ и воспользовавшись формулой Тейлора, найдем, что для достаточно больших n

$$H(\hat{p})=H(p)+H'(p)(\hat{p}-p)+\frac{1}{2}H''(p)(\hat{p}-p)^2+O(|\hat{p}-p|^3)=$$
$$=\frac{1}{2}\left(\frac{1}{p}+\frac{1}{q}\right)(\hat{p}-p)^2+O(|\hat{p}-p|^3).$$

Следовательно,

$$P_n(k)=\frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}}\exp\left\{-\frac{n}{2pq}(\hat{p}-p)^2+nO(|\hat{p}-p|^3)\right\}(1+\varepsilon).$$

Заметим, что

$$\frac{n}{2pq}(\hat{p}-p)^2=\frac{n}{2pq}\left(\frac{k}{n}-p\right)^2=\frac{(k-np)^2}{2npq}.$$

Поэтому

$$P_n(k)=\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}}(1+\varepsilon'(n,\ k,\ n-k)),$$

где

$$1+\varepsilon'(n,\ k,\ n-k)=(1+\varepsilon(n,\ k,\ n-k))\,e^{nO(|p-\hat{p}|^3)}\sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}}$$

и, как легко видеть,

$$\sup|\varepsilon'(n,\ k,\ n-k)|\to 0,\qquad n\to\infty,$$

если sup брать по тем k, для которых

$$|k-np|\leqslant\varphi(n),\quad \varphi(n)=o(npq)^{2/3}.$$

Теорема доказана.

Следствие. Утверждению локальной предельной теоремы можно придать следующую эквивалентную форму: для всех $x\in R^1$ таких, что $x=o(npq)^{1/6}$, а $np+x\sqrt{npq}$ — целые числа из множества $\{0,\ 1,\ \ldots,\ n\}$

$$P_n(np+x\sqrt{npq})\sim\frac{1}{\sqrt{2\pi npq}}e^{-\frac{x^2}{2}},\tag{7}$$

т. е. при $n\to\infty$

$$\sup_{\{x:\,|x|\leqslant\psi(n)\}}\left|\frac{P_n(np+x\sqrt{npq})}{\frac{1}{\sqrt{2\pi npq}}e^{-\frac{x^2}{2}}}-1\right|\to 0,\tag{8}$$

где $\psi(n)=o(npq)^{1/6}$.

С учетом замечаний, сделанных по поводу формулы (5.8), полученные результаты на вероятностном языке можно переформулировать следующим образом:

$$\mathsf{P}\{S_n=k\}\sim\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}},\quad |k-np|=o(npq)^{2/3},\tag{9}$$

$$\mathsf{P}\left\{\frac{S_n-np}{\sqrt{npq}}=x\right\}\sim\frac{1}{\sqrt{2\pi npq}}e^{-x^2/2},\quad x=o(npq)^{1/6}.\tag{10}$$

(В последней формуле $np+x\sqrt{npq}$ предполагаются принимающими значения $0, 1, \ldots, n$.)

Если положить $t_k=\frac{k-np}{\sqrt{npq}}$ и $\Delta t_k=t_{k+1}-t_k=\frac{1}{\sqrt{npq}}$, то последней формуле можно придать такой вид:

$$\mathsf{P}\left\{\frac{S_n-np}{\sqrt{npq}}=t_k\right\}\sim\frac{\Delta t_k}{\sqrt{2\pi}}e^{-t_k^2/2},\quad t_k=o(npq)^{1/6}.\tag{11}$$

Ясно, что $\Delta t_k=\frac{1}{\sqrt{npq}}\to 0$, $n\to\infty$, и множество точек $\{t_k\}$ как бы «заполняет» всю числовую прямую. Естественно поэтому думать, что (11) можно использовать для получения «интегральной» формулы

$$\mathsf{P}\left\{a<\frac{S_n-np}{\sqrt{npq}}\leqslant b\right\}\sim\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx,$$

$$-\infty<a\leqslant b<\infty.$$

Перейдем к точным формулировкам.

2. Пусть для $-\infty<a\leqslant b<\infty$

$$P_n(a, b]=\sum_{a<x\leqslant b}P_n\left(np+x\sqrt{npq}\right),$$

где суммирование распространяется по тем x, для которых $np+x\sqrt{npq}$ — целые числа.

Из локальной теоремы следует (см. также (11)), что для всех t_k, определенных из равенства $k=np+t_k\sqrt{npq}$ и удовлетворяющих условию $|t_k|\leqslant T<\infty$,

$$P_n\left(np+t_k\sqrt{npq}\right)=\frac{\Delta t_k}{\sqrt{2\pi}}e^{-t_k^2/2}[1+\varepsilon(t_k, n)],\tag{12}$$

где

Следовательно, для фиксированных a и b таких, что $-T \leqslant \leqslant a \leqslant b \leqslant T$,

$$\sum_{a<t_k\leqslant b} P_n(np+t_k\sqrt{npq}) =$$

$$= \sum_{a<t_k\leqslant b} \frac{\Delta t_k}{\sqrt{2\pi}} e^{-\frac{t_k^2}{2}} + \sum_{a<t_k\leqslant b} \varepsilon(t_k, n) \frac{\Delta t_k}{\sqrt{2\pi}} e^{-\frac{t_k^2}{2}} =$$

$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx + R_n^{(1)}(a, b) + R_n^{(2)}(a, b), \quad (14)$$

где

$$R_n^{(1)}(a, b) = \sum_{a<t_k\leqslant b} \frac{\Delta t_k}{\sqrt{2\pi}} e^{-t_k^2/2} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx,$$

$$R_n^{(2)}(a, b) = \sum_{a<t_k\leqslant b} \varepsilon(t_k, n) \frac{\Delta t_k}{\sqrt{2\pi}} e^{-\frac{t_k^2}{2}}.$$

Из известных свойств интегральных сумм

$$\sup_{-T\leqslant a\leqslant b\leqslant T} |R_n^{(1)}(a, b)| \to 0, \qquad n\to\infty. \quad (15)$$

Ясно также, что

$$\sup_{-T\leqslant a\leqslant b\leqslant T} |R_n^{(2)}(a, b)| \leqslant \sup_{|t_k|\leqslant T} |\varepsilon(t_k, n)| \cdot \sum_{|t_k|\leqslant T} \frac{\Delta t_k}{\sqrt{2\pi}} e^{-t_k^2/2} \leqslant$$

$$\leqslant \sup_{|t_k|\leqslant T} |\varepsilon(t_k, n)| \times$$

$$\times \left[\frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-x^2/2}\,dx + \sup_{-T\leqslant a\leqslant b\leqslant T} |R_n^{(1)}(a, b)|\right] \to 0, \quad (16)$$

где сходимость правой части к нулю следует из (15) и того известного из математического анализа факта, что

$$\frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-\frac{x^2}{2}}\,dx \leqslant \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\,dx = 1. \quad (17)$$

Обозначим

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

Тогда из (14) — (16) следует, что

$$\sup_{-T \leqslant a \leqslant b \leqslant T} |P_n(a, b] - (\Phi(b) - \Phi(a))| \to 0, \qquad n \to \infty. \tag{18}$$

Покажем сейчас, что этот результат справедлив не только для конечных T, но и для $T = \infty$. В силу (17) для заданного $\varepsilon > 0$ можно найти такое конечное $T = T(\varepsilon)$, что

$$\frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-x^2/2}\, dx > 1 - \frac{\varepsilon}{4}. \tag{19}$$

Согласно (18) можно найти также такое N, что для всех $n > N$ и $T = T(\varepsilon)$

$$\sup_{-T \leqslant a \leqslant b \leqslant T} |P_n(a, b] - (\Phi(b) - \Phi(a))| < \frac{\varepsilon}{4}. \tag{20}$$

Отсюда и из (19) следует, что

$$P_n(-T, T] > 1 - \frac{\varepsilon}{2},$$

и, следовательно,

$$P_n(-\infty, T] + P_n(T, \infty) \leqslant \frac{\varepsilon}{2},$$

где $P_n(-\infty, T] = \lim\limits_{S \downarrow -\infty} P_n(S, T]$ и $P_n(T, \infty) = \lim\limits_{S \uparrow \infty} P_n(T, S]$.

Таким образом, для любых $-\infty \leqslant a \leqslant -T < T \leqslant b \leqslant \infty$

$$\left| P_n(a, b] - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx \right| \leqslant \left| P_n(-T, T] - \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-x^2/2}\, dx \right| +$$

$$+ \left| P_n(a, -T] - \frac{1}{\sqrt{2\pi}} \int_a^{-T} e^{-x^2/2}\, dx \right| + \left| P_n(T, b] - \frac{1}{\sqrt{2\pi}} \int_T^b e^{-x^2/2}\, dx \right| \leqslant$$

$$\leqslant \frac{\varepsilon}{4} + P_n(-\infty, -T] + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-T} e^{-x^2/2}\, dx + P_n(T, \infty) +$$

$$+ \frac{1}{\sqrt{2\pi}} \int_T^{\infty} e^{-x^2/2}\, dx \leqslant \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{8} + \frac{\varepsilon}{8} = \varepsilon.$$

Вместе с (18) отсюда легко выводится, что равномерно по всем $-\infty \leqslant a < b \leqslant \infty$ $P_n(a, b]$ стремится к $\Phi(b) - \Phi(a)$.

Итак, доказана

Интегральная теорема Муавра—Лапласа. *Пусть* $0 < p < 1$,

$$P_n(k) = C_n^k p^k q^{n-k}, \quad P_n(a, b] = \sum_{a < x \leqslant b} P_n(np + x\sqrt{npq}).$$

Тогда

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|P_n(a,\ b]-\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-x^2/2}\,dx\right|\to 0,\qquad n\to\infty. \quad (21)$$

С точностью до тех же самых замечаний, которые были сделаны по поводу соотношения (5.8), результат (21) можно на вероятностном языке сформулировать следующим образом:

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|\mathsf{P}\left\{a<\frac{S_n-\mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}\leqslant b\right\}-\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-x^2/2}\,dx\right|\to 0,\ n\to\infty.$$

Из этой формулы сразу следует, что при любых $-\infty\leqslant A<B\leqslant\infty$ при $n\to\infty$

$$\mathsf{P}\{A<S_n\leqslant B\}-\left[\Phi\left(\frac{B-np}{\sqrt{npq}}\right)-\Phi\left(\frac{A-np}{\sqrt{npq}}\right)\right]\to 0. \qquad (22)$$

П р и м е р. Правильная кость подбрасывается 12 000 раз. Спрашивается, какова вероятность P того, что число шестерок будет лежать в интервале (1800, 2100].

Искомая вероятность равна

$$P=\sum_{1800<k\leqslant 2100} C_{12\,000}^k\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{12\,000-k}.$$

Понятно, что точное вычисление этой суммы представляет весьма трудоемкую работу. Если же воспользоваться интегральной теоремой, то найдем, что интересующая нас вероятность P примерно равна ($n=12\,000$, $p=\frac{1}{6}$, $a=1800$, $b=2100$)

$$\Phi\left(\frac{2100-2000}{\sqrt{12\,000\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right)-\Phi\left(\frac{1800-2000}{\sqrt{12\,000\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right)=$$
$$=\Phi(\sqrt{6})-\Phi(-2\sqrt{6})\approx\Phi(2{,}449)-\Phi(-4{,}898)\approx 0{,}992,$$

где значения $\Phi(2{,}449)$ и $\Phi(-4{,}898)$ взяты из таблиц для функции $\Phi(x)$ (так называемой нормальной функции распределения, см. далее п. 6).

3. Нанесем биномиальные вероятности $P_n(np+x\sqrt{npq})$ (x предполагается таким, что $np+x\sqrt{npq}$ — целое число) на графике (рис. 9).

Тогда локальная теорема говорит о том, что для $x=o(npq)^{1/6}$ вероятности $P_n(np+x\sqrt{npq})$ хорошо «ложатся» на кривую

$\frac{1}{\sqrt{2\pi pq}} e^{-x^2/2}$. Интегральная же теорема говорит о том, что вероятность $P_n(a, b] = \mathsf{P}\{a\sqrt{npq} < S_n - np \leqslant b\sqrt{npq}\} = \mathsf{P}\{np +$

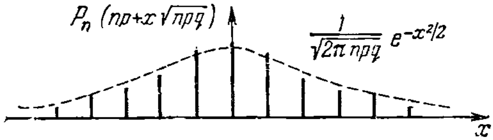


Рис. 9.

$+ a\sqrt{npq} < S_n \leqslant np + b\sqrt{npq}\}$ хорошо аппроксимируется интегралом $1/\sqrt{2\pi}\int\limits_a^b e^{-x^2/2}\,dx$.

Обозначим

$$F_n(x) = P_n(-\infty, x] \quad \left(= \mathsf{P}\left\{\frac{S_n - np}{\sqrt{npq}} \leqslant x\right\}\right).$$

Тогда из (21) следует, что

$$\sup_{-\infty \leqslant x \leqslant \infty} |F_n(x) - \Phi(x)| \to 0, \quad n \to \infty. \tag{23}$$

Естественно возникает вопрос, насколько быстро с ростом n происходит стремление к нулю в (21) и (23). Приведем (без доказательства) результат, относящийся сюда и являющийся частным случаем так называемой *теоремы Берри — Эссеена*:

$$\sup_{-\infty \leqslant x \leqslant \infty} |F_n(x) - \Phi(x)| \leqslant \frac{p^2 + q^2}{\sqrt{npq}}. \tag{24}$$

Важно подчеркнуть, что порядок сценки $(1/\sqrt{npq})$ не может быть улучшен, а это означает, что аппроксимация $F_n(x)$ с помощью функции $\Phi(x)$ может быть плохой при значениях p, близких к нулю или единице даже при больших n. Возникает поэтому вопрос о том, а нельзя ли при малых значениях p или q найти для интересующих нас вероятностей лучшую аппроксимацию, нежели так называемая нормальная, даваемая локальной и интегральной теоремами. С этой целью заметим, что, скажем, при $p = 1/2$ биномиальное распределение $\{P_n(k)\}$ имеет симметричную форму (рис. 10). Однако при малых значениях p биномиальнсе распределение приобретает асимметричную форму (см. рис. 10), и поэтому не приходится ожидать, что нормальная аппроксимация будет хорошей.

4. Оказывается, что при малых значениях p хорошую аппроксимацию для $\{P_n(k)\}$ дает так называемое пуассоновское распределение вероятностей.

Пусть теперь

$$P_n(k) = \begin{cases} C_n^k p^k q^{n-k}, & k = 0, 1, \ldots, n, \\ 0, & k = n+1, n+2, \ldots, \end{cases}$$

и предположим, что p является функцией от n, $p = p(n)$.

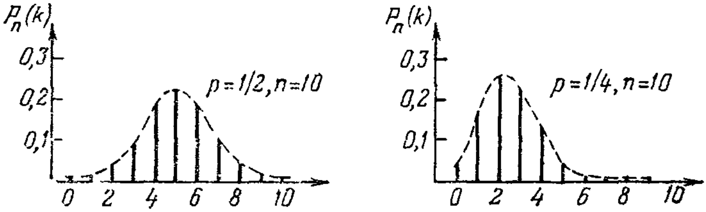


Рис. 10.

Теорема Пуассона. *Пусть $p(n) \to 0$, $n \to \infty$, причем так, что $np(n) \to \lambda$, где $\lambda > 0$. Тогда для любого $k = 0, 1, \ldots$*

$$P_n(k) \to \pi_k, \quad n \to \infty, \tag{25}$$

где

$$\pi_k = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots \tag{26}$$

Доказательство весьма просто. Поскольку по предположению $p(n) = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$, то для любого фиксированного $k = 0, 1, \ldots$ и достаточно больших n

$$P_n(k) = C_n^k p^k q^{n-k} =$$
$$= \frac{n(n-1)\ldots(n-k+1)}{k!}\left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \cdot \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k}.$$

Но

$$n(n-1)\ldots(n-k+1)\left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k =$$
$$= \frac{n(n-1)\ldots(n-k+1)}{n^k}[\lambda + o(1)]^k \to \lambda^k, \quad n \to \infty,$$

и

$$\left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k} \to e^{-\lambda}, \quad n \to \infty,$$

что и доказывает (25).

Набор чисел $\{\pi_k,\ k=0,\ 1,\ \ldots\}$ образует так называемое *пуассоновское распределение* вероятностей $\left(\pi_k \geqslant 0,\ \sum\limits_{k=0}^{\infty} \pi_k = 1\right)$. Отметим, что все рассматриваемые выше (дискретные) распределения были сосредоточены лишь в конечном числе точек. Пуассоновское распределение — это первый встретившийся нам пример (дискретного) распределения, сосредоточенного в счетном числе точек.

Приведем (без доказательства) следующий результат Ю. В. Прохорова, показывающий с какой скоростью величины $P_n(k)$ сходятся к π_k при $n \to \infty$:

$$\sum_{k=0}^{\infty} |P_n(k) - \pi_k| \leqslant \frac{2\lambda}{n} \cdot \min(2,\ \lambda). \tag{27}$$

5. Вернемся к предельной теореме Муавра — Лапласа. Покажем, как из нее следует закон больших чисел (с оговорками, сделанными к (5.8)). Поскольку

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} = \mathsf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant \varepsilon \sqrt{\frac{n}{pq}}\right\},$$

то из (21) ясно, что для $\varepsilon > 0$

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} - \frac{1}{\sqrt{2\pi}} \int\limits_{\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} e^{-x^2/2}\, dx \to 0, \quad n \to \infty, \tag{28}$$

откуда

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \to 1, \qquad n \to \infty,$$

что и составляет утверждение закона больших чисел.

Из (28)

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \sim \frac{1}{\sqrt{2\pi}} \int\limits_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} e^{-x^2/2}\, dx, \qquad n \to \infty, \tag{29}$$

в то время как неравенство Чебышева давало лишь оценку

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \frac{pq}{n\varepsilon^2}.$$

В конце § 5 было показано, что для справедливости соотношения

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha$$

неравенство Чебышева дает следующую оценку для необходимого числа наблюдений:

$$n \geqslant \frac{1}{4\varepsilon^2\alpha}.$$

Так, при $\varepsilon = 0{,}02$, $\alpha = 0{,}05$ необходимо 12 500 наблюдений. Воспользуемся теперь для решения той же задачи аппроксимацией (29).

Определим число $k(\alpha)$ из соотношения

$$\frac{1}{\sqrt{2\pi}} \int_{-k(\alpha)}^{k(\alpha)} e^{-x^2/2}\, dx = 1 - \alpha.$$

Поскольку $\varepsilon \sqrt{\frac{n}{pq}} \geqslant 2\varepsilon \sqrt{n}$, то, определяя (наименьшее целое) n из неравенства

$$2\varepsilon \sqrt{n} \geqslant k(\alpha), \tag{30}$$

получим, что

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \gtrsim 1 - \alpha. \tag{31}$$

Из (30) находим, что наименьшее целое n, удовлетворяющее неравенству

$$n \geqslant \frac{k^2(\alpha)}{4\varepsilon^2},$$

гарантирует выполнение (31), где точность аппроксимации легко может быть установлена из (24).

Беря $\varepsilon = 0{,}02$, $\alpha = 0{,}05$, находим, что на самом деле достаточно лишь 2500 наблюдений, а не 12 500, как это следовало из неравенства Чебышева. Значения $k(\alpha)$ находятся по таблицам. Приведем ряд значений $k(\alpha)$ для некоторых значений α:

α	$k(\alpha)$
0,50	0,675
0,3173	1,000
0,10	1,645
0,05	1,960
0,0454	2,000
0,01	2,576
0,0027	3,000

6. Введенная выше функция

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt, \tag{32}$$

участвующая в интегральной теореме Муавра — Лапласа, играет исключительно важную роль в теории вероятностей. Эта функция называется *нормальным* или *гауссовским распределением* вероятностей на числовой прямой с (нормальной или гауссовской) плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in R^1.$$

Мы уже встречались с (дискретными) распределениями, сосредоточенными в конечном и счетном множестве точек. Нормальное распределение принадлежит другому важному типу распределений, возникающих в теории вероятностей. Отмеченная выше его исключительная роль объясняется прежде всего тем, что при достаточно общих предположениях распределение суммы большого числа независимых случайных величин (не обязательно бернуллиевских!) хорошо аппроксимируется нормальным распределением (§ 4 гл. III).

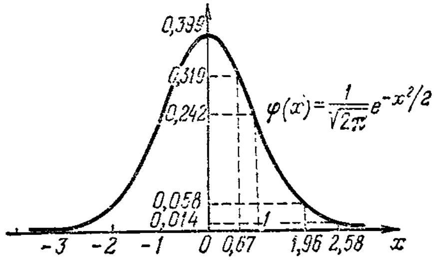


Рис. 11. График плотности $\varphi(x)$ нормального распределения.

Остановимся сейчас на некоторых простейших свойствах функций $\varphi(x)$ и $\Phi(x)$, графики которых приведены на рис. 11 и 12.

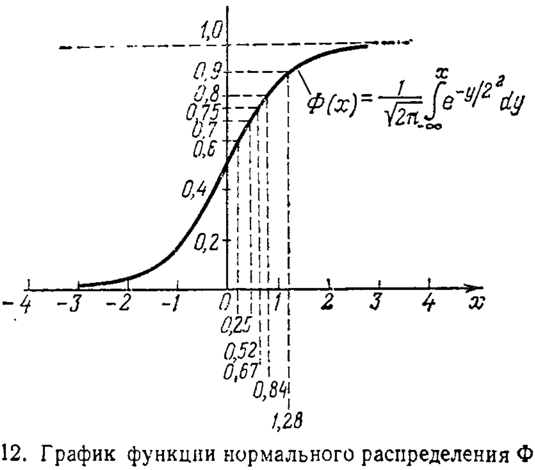


Рис. 12. График функции нормального распределения $\Phi(x)$.

Функция $\varphi(x)$ является симметричной колоколообразной кривой, убывающей с ростом $|x|$ очень быстро: так $\varphi(1)=0{,}24197$, $\varphi(2)=0{,}053991$, $\varphi(3)=0{,}004432$, $\varphi(4)=0{,}000134$, $\varphi(5)=0{,}000016$.

Максимум этой кривой достигается в точке $x=0$ и равен $(2\pi)^{-1/2}\approx$ $\approx 0{,}399$.

Кривая $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-t^2/2}\,dt$ быстро приближается с ростом x к единице: $\Phi(1)=0{,}841345$, $\Phi(2)=0{,}977250$, $\Phi(3)=0{,}998650$, $\Phi(4)=0{,}999968$, $\Phi(4{,}5)=0{,}999997$.

По поводу таблиц функций $\varphi(x)$ и $\Phi(x)$, а также других основных функций, используемых в теории вероятностей и математической статистике см. [6].

7. Задачи.

1. Пусть $n=100$, $p=1/10$, $2/10$, $3/10$, $4/10$, $5/10$. Используя таблицы, (например, из [6]) биномиального и пуассоновского распределений, сравните значения вероятностей

$$\mathsf{P}\{10<S_{100}\leqslant 12\},\qquad \mathsf{P}\{20<S_{100}\leqslant 22\},$$
$$\mathsf{P}\{33<S_{100}\leqslant 35\},\qquad \mathsf{P}\{40<S_{100}\leqslant 42\},$$
$$\mathsf{P}\{50<S_{100}\leqslant 52\}$$

с соответствующими значениями, даваемыми нормальной и пуассоновской аппроксимациями.

2. Пусть $p=1/2$ и $Z_n=2S_n-n$ (число превышений единиц над нулями в n испытаниях). Показать, что

$$\sup_j \left|\sqrt{\pi n}\, P\{Z_{2n}=j\}-e^{-j^2/4n}\right|\to 0,\qquad n\to\infty.$$

3. Доказать, что в теореме Пуассона имеет место следующая скорость сходимости:

$$\sup_k \left|P_n(k)-\frac{\lambda^k e^{-\lambda}}{k!}\right|\leqslant\frac{2\lambda^2}{n}.$$

§ 7. Оценка вероятности «успеха» в схеме Бернулли

1. В рассмотренной выше схеме Бернулли (Ω, A, P) с $\Omega=$ $=\{\omega\colon \omega=(x_1, \ldots, x_n),\ x_i=0,1\}$, $\mathscr{A}=\{A\colon A\subseteq\Omega\}$,

$$p(\omega)=p^{\Sigma x_i}q^{n-\Sigma x_i}$$

предполагалось, что число p (вероятность «успеха») известно.

Представим теперь, что p заранее неизвестно и мы хотим его определить по наблюдениям за исходами эксперимента, или, что то же, по наблюдениям за случайными величинами $\xi_1, \ldots, \xi_n$, где $\xi_i(\omega)=x_i$. Эта задача, являющаяся типичной для математической статистики, допускает различные постановки. Ниже мы рассматриваем две такие постановки: задачу *оценивания* и задачу *построения доверительных интервалов*.

Следуя обозначениям, принятым в математической статистике, неизвестный параметр p обозначим через θ, считая a priori, что значения θ принадлежат множеству $\Theta=[0, 1]$. Будем говорить также, что набор $(\Omega, \mathscr{A}, \mathsf{P}_\theta;\ \theta\in\Theta)$ с $p_\theta(\omega)=\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}$ задает *вероятностно-статистическую модель* (отвечающую «n независимым испытаниям» с вероятностью «успеха» $\theta\in\Theta$), а всякую функцию $T_n=T_n(\omega)$, принимающую значения в Θ, будем называть *оценкой*.

Если $S_n=\xi_1+\ldots+\xi_n$ и $T_n^*=\frac{S_n}{n}$, то из закона больших чисел следует, что оценка T_n^* является *состоятельной* в том смысле, что ($\varepsilon>0$)

$$\mathsf{P}_\theta\{|T_n^*-\theta|\geqslant\varepsilon\}\to 0,\qquad n\to\infty. \tag{1}$$

Кроме того, эта оценка является *несмещенной*: для всякого $\theta\in\Theta$

$$\mathsf{M}_\theta T_n^*=\theta, \tag{2}$$

где M_θ — математическое ожидание, отвечающее вероятности P_θ.

Свойство оценки быть несмещенной является вполне естественным: оно отражает тот факт, что всякая разумная оценка должна, по крайней мере «в среднем», приводить к желаемому результату. Однако легко заметить, что оценка T_n^* не является единственной несмещенной оценкой. Например, такой же будет всякая оценка

$$T_n=\frac{b_1x_1+\ldots+b_nx_n}{n},$$

где $b_1+\ldots+b_n=1$. При этом для таких оценок также будет выполняться закон больших чисел (1) (по крайней мере для неотрицательных b_i) и тем самым эти оценки T_n так же «хороши», как и T_n^*.

В этой связи возникают вопросы о том, как сравнивать различные несмещенные оценки, какую из них назвать наилучшей, оптимальной.

По самому смыслу оценок естественно было бы считать, что оценка тем лучше, чем меньше ее отклонение от оцениваемого параметра. Основываясь на этом, назовем оценку $\tilde{T}_n$ *эффективной* (в классе несмещенных оценок T_n), если

$$\mathsf{D}_\theta\tilde{T}_n=\inf_{T_n}\mathsf{D}_\theta T_n,\qquad \theta\in\Theta, \tag{3}$$

где $\mathsf{D}_\theta T_n$ — дисперсия оценки T_n, т. е. $\mathsf{M}_\theta(T_n-\theta)^2$.

Покажем, что рассмотренная выше оценка T_n^* является эффективной. Имеем

$$\mathsf{D}_\theta T_n^*=\mathsf{D}_\theta\left(\frac{S_n}{n}\right)=\frac{\mathsf{D}_\theta S_n}{n^2}=\frac{n\theta(1-\theta)}{n^2}=\frac{\theta(1-\theta)}{n}. \tag{4}$$

Поэтому, для того чтобы установить, что оценка T_n^* эффективна, достаточно показать, что

$$\inf_{T_n} \mathsf{D}_\theta T_n \geqslant \frac{\theta(1-\theta)}{n}. \tag{5}$$

При $\theta=0$ или 1 эта оценка очевидна. Пусть $\theta \in (0, 1)$ и

$$p_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}.$$

Ясно, что

$$p_\theta(\omega) = \prod_{i=1}^{n} p_\theta(x_i).$$

Обозначим

$$L_\theta(\omega) = \ln p_\theta(\omega).$$

Тогда

$$L_\theta(\omega) = \ln\theta \cdot \sum x_i + \ln(1-\theta)\sum(1-x_i)$$

и

$$\frac{\partial L_\theta(\omega)}{\partial\theta} = \frac{\sum(x_i-\theta)}{\theta(1-\theta)}.$$

Поскольку

$$1 \equiv \mathsf{M}_\theta 1 = \sum_\omega p_\theta(\omega)$$

и в силу несмещенности оценки T_n

$$\theta \equiv \mathsf{M}_\theta T_n = \sum_\omega T_n(\omega)\, p_\theta(\omega),$$

то после дифференцирования по θ получим, что

$$0 = \sum_\omega \frac{\partial p_\theta(\omega)}{\partial\theta} = \sum_\omega \frac{\left(\frac{\partial p_\theta(\omega)}{\partial\theta}\right)}{p_\theta(\omega)}\, p_\theta(\omega) = \mathsf{M}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right],$$

$$1 = \sum_\omega T_n \frac{\left(\frac{\partial p_\theta(\omega)}{\partial\theta}\right)}{p_\theta(\omega)}\, p_\theta(\omega) = \mathsf{M}_\theta\left[T_n \frac{\partial L_\theta(\omega)}{\partial\theta}\right].$$

Значит,

$$1 = \mathsf{M}_\theta\left[(T_n-\theta)\frac{\partial L_\theta(\omega)}{\partial\theta}\right]$$

и, согласно неравенству Коши — Буняковского,

$$1 \leqslant \mathsf{M}_\theta[T_n-\theta]^2 \cdot \mathsf{M}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2,$$

откуда

$$\mathrm{M}_\theta [T_n - \theta]^2 \geqslant \frac{1}{I_n(\theta)}, \tag{6}$$

где величина $I_n(\theta) = \mathrm{M}_\theta \left[\frac{\partial L_\theta(\omega)}{\partial \theta}\right]^2$ носит название *информации Фишера*.

Из (6) получаем частный случай так называемого *неравенства Рао — Крамера* для несмещенных оценок T_n

$$\inf_{T_n} \mathrm{D}_\theta T_n \geqslant \frac{1}{I_n(\theta)}. \tag{7}$$

В рассматриваемом случае

$$I_n(\theta) = \mathrm{M}_\theta \left[\frac{\partial L_\theta(\omega)}{\partial \theta}\right]^2 = \mathrm{M}_\theta \left[\frac{\sum (\xi_i - \theta)}{\theta(1-\theta)}\right]^2 = \frac{n\theta(1-\theta)}{[\theta(1-\theta)]^2} = \frac{n}{\theta(1-\theta)},$$

что и доказывает неравенство (5), из которого, как уже отмечалось, следует эффективность несмещенной оценки $T_n^* = \frac{S_n}{n}$ для неизвестного параметра θ.

2. Очевидно, что, рассматривая в качестве «точечной» оценки для θ величину T_n^*, мы совершаем некоторую ошибку. Может даже случиться, что численное значение T_n^*, подсчитанное по наблюденным значениям $x_1, \ldots, x_n$, будет довольно сильно отличаться от истинного значения θ. Поэтому целесообразно было бы указывать еще и величину погрешности.

Довольно бессмысленно надеяться, что для всех элементарных событий величины $T_n^*(\omega)$ мало отличаются от истинного значения неизвестного параметра θ. Однако из закона больших чисел мы знаем, что для всякого $\delta > 0$ при достаточно больших n вероятность события $\{|\theta - T_n^*(\omega)| > \delta\}$ будет достаточно мала.

Согласно неравенству Чебышева

$$\mathrm{P}_\theta\{|\theta - T_n^*| > \delta\} \leqslant \frac{\mathrm{D}_\theta T_n^*}{\delta^2} = \frac{\theta(1-\theta)}{n\delta^2}$$

и, значит, для всякого $\lambda > 0$

$$\mathrm{P}_\theta\left\{|\theta - T_n^*| \leqslant \lambda \sqrt{\frac{\theta(1-\theta)}{n}}\right\} \geqslant 1 - \frac{1}{\lambda^2}.$$

Если взять, к примеру, $\lambda = 3$, то с P_θ-вероятностью, большей чем 0,8888 $\left(1 - \frac{1}{3^2} = \frac{8}{9} \approx 0{,}8888\right)$, осуществится событие

$$|\theta - T_n^*| \leqslant 3 \sqrt{\frac{\theta(1-\theta)}{n}}$$

и тем более — событие

$$|\theta - T_n^*| \leqslant \frac{3}{2\sqrt{n}},$$

поскольку $\theta(1-\theta) \leqslant \frac{1}{4}$.

Таким образом,

$$\mathsf{P}_\theta\left\{|\theta - T_n^*| \leqslant \frac{3}{2\sqrt{n}}\right\} = \mathsf{P}_\theta\left\{T_n^* - \frac{3}{2\sqrt{n}} \leqslant \theta \leqslant T_n^* + \frac{3}{2\sqrt{n}}\right\} \geqslant 0{,}8888.$$

Иначе говоря, можно утверждать с вероятностью, большей чем 0,8888, что истинное значение параметра θ принадлежит интервалу $\left[T_n^* - \frac{3}{2\sqrt{n}},\ T_n^* + \frac{3}{2\sqrt{n}}\right]$. Иногда это утверждение символически записывают в такой форме:

$$\theta \simeq T_n^* \pm \frac{3}{2\sqrt{n}} \quad (\geqslant 88\%),$$

где «$\geqslant 88\%$» означает «более чем в 88% случаев».

Интервал $\left[T_n^* - \frac{3}{2\sqrt{n}},\ T_n^* + \frac{3}{2\sqrt{n}}\right]$ является примером так называемых доверительных интервалов для неизвестного параметра.

Определение. Интервал вида

$$[\psi_1(\omega),\ \psi_2(\omega)],$$

где $\psi_1(\omega)$ и $\psi_2(\omega)$ — две функции элементарных событий, назовем *доверительным интервалом надежности* $1-\delta$ (или с *уровнем значимости* δ), если для всех $\theta \in \Theta$

$$\mathsf{P}_\theta\{\psi_1(\omega) \leqslant \theta \leqslant \psi_2(\omega)\} \geqslant 1-\delta.$$

Приведенные выше рассуждения показывают, что интервал $\left[T_n^* - \frac{\lambda}{2\sqrt{n}},\ T_n^* + \frac{\lambda}{2\sqrt{n}}\right]$ имеет надежность $1 - \frac{1}{\lambda^2}$. На самом деле надежность доверительного интервала значительно выше, что связано с тем, что использованное неравенство Чебышева дает лишь грубую оценку вероятностей событий.

Для получения более точных результатов заметим, что

$$\left\{\omega:\ |\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} = \{\omega : \psi_1(T_n^*, n) \leqslant \theta \leqslant \psi_2(T_n^*, n)\},$$

где $\psi_1 = \psi_1(T_n^*, n)$ и $\psi_2 = \psi_2(T_n^*, n)$ — корни квадратного уравнения

$$(\theta - T_n^*)^2 = \frac{\lambda^2}{n}\theta(1-\theta),$$

описывающего эллипс, расположенный так, как это изображено на рис. 13. Пусть теперь

$$F_\theta^n(x) = \mathsf{P}_\theta\left\{\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leqslant x\right\}.$$

Тогда в силу (6.24)

$$\sup_x |F_\theta^n(x) - \Phi(x)| \leqslant \frac{1}{\sqrt{n\theta(1-\theta)}}.$$

Поэтому, если a priori известно, что

$$0 < \Delta \leqslant \theta \leqslant 1 - \Delta < 1,$$

где Δ — некоторая константа, то

$$\sup_x |F_\theta^n(x) - \Phi(x)| \leqslant \frac{1}{\Delta\sqrt{n}}$$

и, значит

$$\mathsf{P}_\theta\{\psi_1(T_n^*, n) \leqslant \theta \leqslant \psi_2(T_n^*, n)\} = \mathsf{P}_\theta\left\{|\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} =$$

$$= \mathsf{P}_\theta\left\{\frac{|S_n - n\theta|}{\sqrt{n\theta(1-\theta)}} \leqslant \lambda\right\} \geqslant (2\Phi(\lambda) - 1) - \frac{2}{\Delta\sqrt{n}}.$$

Пусть λ^* — то наименьшее λ, для которого

$$(2\Phi(\lambda) - 1) - \frac{2}{\Delta\sqrt{n}} \geqslant 1 - \delta^*,$$

где δ^* — заданный уровень значимости. Обозначая $\delta = \delta^* - \frac{2}{\Delta\sqrt{n}}$, находим, что λ^* есть корень уравнения

$$\Phi(\lambda) = 1 - \frac{\delta}{2}.$$

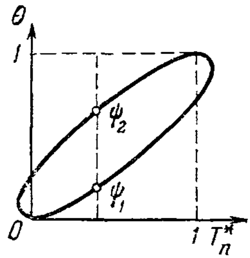


Рис. 13.

В случае больших n можно пренебречь членом $2/\Delta\sqrt{n}$ и считать, что λ^* удовлетворяет соотношению

$$\Phi(\lambda^*) = 1 - \frac{\delta^*}{2}.$$

В частности, если $\lambda^* = 3$, то $\delta^* = 0{,}9973\ \ldots$ Так что с вероятностью, примерно равной 0,9973, или после итерирования и отбрасывания членов порядка $O(n^{-3/4})$ находим, что

$$T_n^* - 3\sqrt{\frac{T_n^*(1-T_n^*)}{n}} \leqslant \theta \leqslant T_n^* + 3\sqrt{\frac{T_n^*(1-T_n^*)}{n}}. \qquad (9)$$

Отсюда следует, что доверительный интервал

$$\left[T_n^* - \frac{3}{2\sqrt{n}},\ T_n^* + \frac{3}{2\sqrt{n}}\right] \qquad (10)$$

имеет (при больших n) надежность 0,9973 (тогда как неравенство Чебышева давало надежность лишь, примерно, равную 0,8888).

Отсюда можно сделать следующий практический вывод. Пусть производится большое число N серий экспериментов, в каждой из которых по n наблюдениям оценивается параметр θ. Тогда примерно в 99,73% случаев из N в каждой серии оценка будет отличаться от истинного значения параметра не больше чем на $\frac{3}{2\sqrt{n}}$. (См. по этому поводу также конец § 5.)

3. Задачи.

1. Пусть a priori известно, что параметр θ принимает значения во множестве $\Theta_0 \subseteq [0, 1]$. Построить несмещенную оценку для параметра θ, принимающую значения лишь во множестве Θ_0.

2. В условиях предыдущей задачи найти аналог неравенства Рао — Крамера и рассмотреть вопрос об эффективных оценках.

3. В условиях первой задачи рассмотреть вопрос о построении доверительных интервалов для θ.

§ 8. Условные вероятности и математические ожидания относительно разбиений

1. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — конечное вероятностное пространство и

$$\mathscr{D} = \{D_1, \ldots, D_k\}$$

— некоторое разбиение Ω ($D_i \in \mathscr{A}$, $\mathsf{P}(D_i) > 0$, $i = 1, \ldots, k$, и $D_1 + \ldots + D_k = \Omega$). Пусть, далее, A — событие из $\mathscr{A}$ и $\mathsf{P}(A \mid D_i)$ — условная вероятность события A относительно события D_i.

С набором условных вероятностей $\{\mathsf{P}(A \mid D_i),\ i = 1, \ldots, k\}$ можно связать случайную величину

$$\pi(\omega) = \sum_{i=1}^{k} \mathsf{P}(A \mid D_i)\, I_{D_i}(\omega) \qquad (1)$$

(ср. (4.5)), принимающую на атомах разбиения D_i значения $\mathsf{P}(A \mid D_i)$. Чтобы подчеркнуть, что эта *случайная величина* связана именно с разбиением $\mathscr{D}$, ее обозначают

$$\mathsf{P}(A\,|\,\mathscr{D}) \quad \text{или} \quad \mathsf{P}(A\,|\,\mathscr{D})(\omega)$$

и называют *условной вероятностью события A относительно разбиения* $\mathscr{D}$.

Это понятие, а также вводимые далее более общие понятия условных вероятностей относительно σ-алгебр, играют важную роль в теории вероятностей, что постепенно будет раскрываться последующим изложением.

Остановимся на простейших свойствах условных вероятностей:

$$\mathsf{P}(A+B\,|\,\mathscr{D}) = \mathsf{P}(A\,|\,\mathscr{D}) + \mathsf{P}(B\,|\,\mathscr{D}); \tag{2}$$

если $\mathscr{D}$ — тривиальное разбиение, состоящее из одного множества Ω, то

$$\mathsf{P}(A\,|\,\Omega) = \mathsf{P}(A). \tag{3}$$

Определение условной вероятности $\mathsf{P}(A\,|\,\mathscr{D})$ как случайной величины дает возможность говорить о ее математическом ожидании, используя которое можно следующим компактным образом записать *формулу полной вероятности* (3.3):

$$\mathsf{MP}(A\,|\,\mathscr{D}) = \mathsf{P}(A). \tag{4}$$

Действительно, поскольку

$$\mathsf{P}(A\,|\,\mathscr{D}) = \sum_{i=1}^{k} \mathsf{P}(A\,|\,D_i)\, I_{D_i}(\omega),$$

то по определению математического ожидания (см. (4.5) и (4.6))

$$\mathsf{MP}(A\,|\,\mathscr{D}) = \sum_{i=1}^{k} \mathsf{P}(A\,|\,D_i)\,\mathsf{P}(D_i) = \sum_{i=1}^{k} \mathsf{P}(AD_i) = \mathsf{P}(A).$$

Пусть теперь $\eta = \eta(\omega)$ — случайная величина, принимающая с положительными вероятностями значения $y_1, \ldots, y_k$:

$$\eta(\omega) = \sum_{j=1}^{k} y_j I_{D_j}(\omega),$$

где $D_j = \{\omega\colon \eta(\omega) = y_j\}$. Разбиение $\mathscr{D}_\eta = \{D_1, \ldots, D_k\}$ называется разбиением, порождаемым случайной величиной η. Условную вероятность $\mathsf{P}(A\,|\,\mathscr{D}_\eta)$ будем в дальнейшем обозначать $\mathsf{P}(A\,|\,\eta)$ или $\mathsf{P}(A\,|\,\eta)(\omega)$, и называть *условной вероятностью события A относительно случайной величины* η. Условимся также под $\mathsf{P}(A\,|\,\eta = y_j)$ понимать условную вероятность $\mathsf{P}(A\,|\,D_j)$, где $D_j = \{\omega\colon \eta(\omega) = y_j\}$.

Аналогичным образом, если $\eta_1, \eta_2, \ldots, \eta_m$ — случайные величины и $\mathscr{D}_{\eta_1, \eta_2, \ldots, \eta_m}$ — разбиение, порожденное величинами $\eta_1, \eta_2, \ldots, \eta_m$ с атомами

$$D_{y_1, y_2, \ldots, y_m} = \{\omega\colon\ \eta_1(\omega) = y_1, \ \ldots, \ \eta_m(\omega) = y_m\},$$

то $\mathsf{P}(A \mid \mathscr{D}_{\eta_1, \eta_2, \ldots, \eta_m})$ обозначается $\mathsf{P}(A \mid \eta_1, \eta_2, \ldots, \eta_m)$ и называется *условной вероятностью события A относительно случайных величин* $\eta_1, \eta_2, \ldots, \eta_m$.

Пример 1. Пусть ξ и η — две независимые одинаково распределенные случайные величины, принимающие каждая значения 1 и 0 с вероятностями p и q. Найдем для $k = 0, 1, 2$ условную вероятность $\mathsf{P}(\xi + \eta = k \mid \eta)$ события $A = \{\omega\colon\ \xi + \eta = k\}$ относительно η.

С этой целью отметим сначала следующий общий полезный факт: если ξ и η — две независимые случайные величины со значениями x и y соответственно, то

$$\mathsf{P}(\xi + \eta = z \mid \eta = y) = \mathsf{P}(\xi + y = z). \tag{5}$$

В самом деле,

$$\mathsf{P}(\xi + \eta = z \mid \eta = y) = \frac{\mathsf{P}(\xi + \eta = z, \ \eta = y)}{\mathsf{P}(\eta = y)} =$$
$$= \frac{\mathsf{P}(\xi + y = z, \ \eta = y)}{\mathsf{P}(\eta = y)} = \frac{\mathsf{P}(\xi + y = z)\,\mathsf{P}(y = \eta)}{\mathsf{P}(\eta = y)} = \mathsf{P}(\xi + y = z).$$

Используя эту формулу для рассматриваемого случая, находим, что

$$\begin{aligned}\mathsf{P}(\xi + \eta = k \mid \eta) &= \mathsf{P}(\xi + \eta = k \mid \eta = 0)\, I_{\{\eta = 0\}}(\omega) + \\ &+ \mathsf{P}(\xi + \eta = k \mid \eta = 1)\, I_{\{\eta = 1\}}(\omega) = \\ &= \mathsf{P}(\xi = k)\, I_{\{\eta = 0\}}(\omega) + \mathsf{P}\{\xi = k - 1\}\, I_{\{\eta = 1\}}(\omega).\end{aligned}$$

Итак,

$$\mathsf{P}(\xi + \eta = k \mid \eta) = \begin{cases} qI_{\{\eta = 0\}}(\omega), & k = 0, \\ pI_{\{\eta = 0\}}(\omega) + qI_{\{\eta = 1\}}(\omega), & k = 1, \\ pI_{\{\eta = 1\}}(\omega), & k = 2, \end{cases} \tag{6}$$

или, что то же самое,

$$\mathsf{P}(\xi + \eta = k \mid \eta) = \begin{cases} q(1 - \eta), & k = 0, \\ p(1 - \eta) + q\eta, & k = 1, \\ p\eta, & k = 2. \end{cases} \tag{7}$$

2. Пусть $\xi = \xi(\omega)$ — случайная величина, принимающая значения в множестве $X = \{x_1, \ldots, x_l\}$:

и $\mathscr{D}=\{D_1, \ldots, D_k\}$ — некоторое разбиение. Подобно тому как для ξ по вероятностям $\mathsf{P}(A_j)$, $j=1, \ldots, l$ было определено математическое ожидание

$$\mathsf{M}\xi=\sum_{j=1}^{l} x_j \mathsf{P}(A_j), \tag{8}$$

так и с помощью условных вероятностей $\mathsf{P}(A_j \mid \mathscr{D})$, $j=1, \ldots, l$, естественно определить *условное математическое ожидание случайной величины* ξ *относительно разбиения* $\mathscr{D}$, обозначаемое $\mathsf{M}(\xi \mid \mathscr{D})$, или $\mathsf{M}(\xi \mid \mathscr{D})(\omega)$, формулой

$$\mathsf{M}(\xi \mid \mathscr{D})=\sum_{j=1}^{l} x_j \mathsf{P}(A_j \mid \mathscr{D}). \tag{9}$$

$$\begin{tikzcd}
\mathsf{P}(\cdot) \arrow[r, "(8)"] \arrow[d, "(3.1)"] & \mathsf{M}\xi \\
\mathsf{P}(\cdot \mid D) \arrow[r, "(10)"] \arrow[d, "(1)"] & \mathsf{M}(\xi \mid D) \arrow[d, "(11)"] \\
\mathsf{P}(\cdot \mid \mathscr{D}) \arrow[r, "(9)"] & \mathsf{M}(\xi \mid \mathscr{D})
\end{tikzcd}$$

Рис. 14

Согласно этому определению условное математическое ожидание $\mathsf{M}(\xi \mid \mathscr{D})(\omega)$ является случайной величиной, принимающей для всех элементарных событий ω, принадлежащих одному и тому же атому D_i, одно и то же значение $\sum_{j=1}^{l} x_j \mathsf{P}(A_j \mid D_i)$. Это замечание показывает, что к определению условного математического ожидания $\mathsf{M}(\xi \mid \mathscr{D})$ можно было бы подойти иначе. А именно, сначала определить $\mathsf{M}(\xi \mid D_i)$ — условное математическое ожидание ξ относительно события D_i формулой

$$\mathsf{M}(\xi \mid D_i)=\sum_{j=1}^{l} x_j \mathsf{P}(A_j \mid D_i)\left(=\frac{\mathsf{M}[\xi I_{D_i}]}{\mathsf{P}(D_i)}\right), \tag{10}$$

а затем положить по определению

$$\mathsf{M}(\xi \mid \mathscr{D})(\omega)=\sum_{i=1}^{k} \mathsf{M}(\xi \mid D_i) I_{D_i}(\omega) \tag{11}$$

(см. диаграмму на рис. 14).

Полезно отметить также, что значения $\mathsf{M}(\xi \mid D)$ и $\mathsf{M}(\xi \mid \mathscr{D})$ не зависят от способа представления случайной величины ξ.

Приводимые далее свойства условных математических ожиданий непосредственно вытекают из их определения:

$$\mathsf{M}(a\xi+b\eta \mid \mathscr{D})=a\mathsf{M}(\xi \mid \mathscr{D})+b\mathsf{M}(\eta \mid \mathscr{D}),\quad a,\ b \text{ — константы}; \tag{12}$$

$$\mathsf{M}(\xi \mid \Omega)=\mathsf{M}\xi; \tag{13}$$

$$\mathsf{M}(C \mid \mathscr{D})=C,\quad C \text{ — константа}; \tag{14}$$

если $\xi = I_A(\omega)$, то

$$\mathsf{M}(\xi \mid \mathscr{D}) = \mathsf{P}(A \mid \mathscr{D}). \tag{15}$$

Последнее равенство показывает, в частности, что свойства условных вероятностей можно получать непосредственно из свойств условных математических ожиданий.

Следующее важное свойство обобщает *формулу полной вероятности* (5):

$$\mathsf{M}\mathsf{M}(\xi \mid \mathscr{D}) = \mathsf{M}\xi. \tag{16}$$

Для доказательства достаточно заметить, что, согласно (5),

$$\mathsf{M}\mathsf{M}(\xi \mid \mathscr{D}) = \mathsf{M}\sum_{j=1}^{l} x_j P(A_j \mid \mathscr{D}) = \sum_{j=1}^{l} x_j \mathsf{M}\mathsf{P}(A_j \mid \mathscr{D}) = \sum_{j=1}^{l} x_j \mathsf{P}(A_j) = \mathsf{M}\xi.$$

Пусть $\mathscr{D} = \{D_1, \ldots, D_k\}$ — разбиение и $\eta = \eta(\omega)$ — некоторая случайная величина. Будем говорить, что η измерима относительно этого разбиения или $\mathscr{D}$-измерима, если $\mathscr{D}_\eta \subseteq \mathscr{D}$, т. е. $\eta = \eta(\omega)$ может быть представлена в виде

$$\eta(\omega) = \sum_{i=1}^{k} y_i I_{D_i}(\omega),$$

где y_i могут быть и равными. Иначе говоря, случайная величина $\mathscr{D}$-измерима тогда и только тогда, когда она принимает постоянные значения на атомах разбиения $\mathscr{D}$.

Пример 2. Если $\mathscr{D}$ — тривиальное разбиение, $\mathscr{D} = \{\Omega\}$, то η $\mathscr{D}$-измерима в том и только том случае, если $\eta \equiv C$, где C — постоянная. Всякая случайная величина η измерима относительно разбиения $\mathscr{D}_\eta$.

Предположим, что случайная величина η является $\mathscr{D}$-измеримой. Тогда

$$\mathsf{M}(\xi\eta \mid \mathscr{D}) = \eta\mathsf{M}(\xi \mid \mathscr{D}) \tag{17}$$

и, в частности,

$$\mathsf{M}(\eta \mid \mathscr{D}) = \eta \quad (\mathsf{M}(\eta \mid \mathscr{D}_\eta) = \eta). \tag{18}$$

Для доказательства (17) заметим, что если $\xi = \sum_{j=1}^{l} x_j I_{A_j}$, то

и, значит,

$$
\begin{aligned}
\mathsf{M}(\xi\eta\mid\mathscr{D}) &= \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \mathsf{P}(A_j D_i\mid\mathscr{D}) = \\
&= \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \sum_{m=1}^{k} \mathsf{P}(A_j D_i\mid D_m)\, I_{D_m}(\omega) = \\
&= \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \mathsf{P}(A_j D_i\mid D_i)\, I_{D_i}(\omega) = \\
&= \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \mathsf{P}(A_j\mid D_i)\, I_{D_i}(\omega).
\end{aligned} \tag{19}
$$

С другой стороны, учитывая, что $I^2_{D_i}=I_{D_i}$ и $I_{D_i}\cdot I_{D_m}=0$, $i\neq m$, получаем

$$
\begin{aligned}
\eta\mathsf{M}(\xi\mid\mathscr{D}) &= \left[\sum_{i=1}^{k} y_i I_{D_i}(\omega)\right]\cdot\left[\sum_{j=1}^{l} x_j \mathsf{P}(A_j\mid\mathscr{D})\right] = \\
&= \left[\sum_{i=1}^{k} y_i I_{D_i}(\omega)\right]\cdot\sum_{m=1}^{k}\left[\sum_{j=1}^{l} x_j \mathsf{P}(A_j\mid D_m)\right]\cdot I_{D_m}(\omega) = \\
&= \sum_{i=1}^{k}\sum_{j=1}^{l} y_i x_j \mathsf{P}(A_j\mid D_i)\cdot I_{D_i}(\omega),
\end{aligned}
$$

что вместе с (19) доказывает (17).

Установим еще одно важное свойство условных математических ожиданий. Пусть $\mathscr{D}_1$ и $\mathscr{D}_2$ — два разбиения, причем $\mathscr{D}_1 \subseteq \mathscr{D}_2$ (разбиение $\mathscr{D}_2$ «мельче» разбиения $\mathscr{D}_1$). Тогда

$$
\mathsf{M}[\mathsf{M}(\xi\mid\mathscr{D}_2)\mid\mathscr{D}_1] = \mathsf{M}(\xi\mid\mathscr{D}_1). \tag{20}
$$

Для доказательства предположим, что

$$
\mathscr{D}_1=\{D_{11},\ \ldots,\ D_{1m}\},\quad \mathscr{D}_2=\{D_{21},\ \ldots,\ D_{2n}\}.
$$

Тогда, если $\xi=\sum_{j=1}^{l} x_j I_{A_j}$, то

$$
\mathsf{M}(\xi\mid\mathscr{D}_2) = \sum_{j=1}^{l} x_j \mathsf{P}(A_j\mid\mathscr{D}_2),
$$

и достаточно лишь установить, что

$$
\mathsf{M}[\mathsf{P}(A_j\mid\mathscr{D}_2)\mid\mathscr{D}_1] = \mathsf{P}(A_j\mid\mathscr{D}_1). \tag{21}
$$

Поскольку

то

$$\begin{aligned}
\mathsf{M}[\mathsf{P}(A_j \mid \mathscr{D}_2) \mid \mathscr{D}_1] &= \sum_{q=1}^{n} \mathsf{P}(A_j \mid D_{2q})\,\mathsf{P}(D_{2q} \mid \mathscr{D}_1) = \\
&= \sum_{q=1}^{n} \mathsf{P}(A_j \mid D_{2q}) \left[\sum_{p=1}^{m} \mathsf{P}(D_{2q} \mid D_{1p})\, I_{D_{1p}}\right] = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \cdot \sum_{q=1}^{n} \mathsf{P}(A_j \mid D_{2q})\,\mathsf{P}(D_{2q} \mid D_{1p}) = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \cdot \sum_{\{q:\, D_{2q} \subseteq D_{1p}\}} \mathsf{P}(A_j \mid D_{2q})\,\mathsf{P}(D_{2q} \mid D_{1p}) = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \cdot \sum_{\{q:\, D_{2q} \subseteq D_{1p}\}} \frac{\mathsf{P}(A_j D_{2q})}{\mathsf{P}(D_{2q})} \cdot \frac{\mathsf{P}(D_{2q})}{\mathsf{P}(D_{1p})} = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \cdot \mathsf{P}(A_j \mid D_{1p}) = \mathsf{P}(A_j \mid \mathscr{D}_1),
\end{aligned}$$

что и доказывает (21).

В том случае, когда разбиение $\mathscr{D}$ порождается случайными величинами $\eta_1, \ldots, \eta_k$ ($\mathscr{D} = \mathscr{D}_{\eta_1, \ldots, \eta_k}$), условное математическое ожидание $\mathsf{M}(\xi \mid \mathscr{D}_{\eta_1, \ldots, \eta_k})$ будет обозначаться $\mathsf{M}(\xi \mid \eta_1, \ldots, \eta_k)$, или $\mathsf{M}(\xi \mid \eta_1, \ldots, \eta_k)(\omega)$, и называться *условным математическим ожиданием* ξ *относительно* $\eta_1, \ldots, \eta_k$.

Непосредственно из определения $\mathsf{M}(\xi \mid \eta)$ следует, что если ξ и η *независимы*, то

$$\mathsf{M}(\xi \mid \eta) = \mathsf{M}\xi. \tag{22}$$

Из (18) следует также, что

$$\mathsf{M}(\eta \mid \eta) = \eta. \tag{23}$$

Свойство (22) допускает следующее обобщение. Пусть случайная величина ξ не зависит от разбиения $\mathscr{D}$ (т. е. для любого $D_i \in \mathscr{D}$ случайные величины ξ и I_{D_i} независимы). Тогда

$$\mathsf{M}(\xi \mid \mathscr{D}) = \mathsf{M}\xi. \tag{24}$$

Из (20) в качестве частного случая получаем следующую полезную формулу:

$$\mathsf{M}[\mathsf{M}(\xi \mid \eta_1, \eta_2) \mid \eta_1] = \mathsf{M}(\xi \mid \eta_1). \tag{25}$$

Пример 3. Для случайных величин ξ и η, рассмотренных в примере 1, найдем $\mathsf{M}(\xi + \eta \mid \eta)$. В силу (22) и (23)

$$\mathsf{M}(\xi + \eta \mid \eta) = \mathsf{M}\xi + \eta = p + \eta.$$

Этот результат можно получить и отправляясь от (8):

$$\mathsf{M}(\xi+\eta\mid\eta)=\sum_{k=0}^{2}k\mathsf{P}(\xi+\eta=k\mid\eta)=p(1-\eta)+q\eta+2p\eta=p+\eta.$$

Пример 4. Пусть ξ и η — независимые одинаково распределенные случайные величины. Тогда

$$\mathsf{M}(\xi\mid\xi+\eta)=\mathsf{M}(\eta\mid\xi+\eta)=\frac{\xi+\eta}{2}. \tag{26}$$

Действительно, считая для простоты, что ξ и η принимают значения $1, 2, \ldots, m$, находим, что $(1\leqslant k\leqslant m,\ 2\leqslant l\leqslant 2m)$

$$\mathsf{P}(\xi=k\mid\xi+\eta=l)=\frac{\mathsf{P}(\xi=k,\ \xi+\eta=l)}{\mathsf{P}(\xi+\eta=l)}=\frac{\mathsf{P}(\xi=k,\ \eta=l-k)}{\mathsf{P}(\xi+\eta=l)}=$$

$$=\frac{\mathsf{P}(\xi=k)\,P(\eta=l-k)}{\mathsf{P}(\xi+\eta=l)}=\frac{\mathsf{P}(\eta=k)\,\mathsf{P}(\xi=l-k)}{\mathsf{P}(\xi+\eta=l)}=\mathsf{P}(\eta=k\mid\xi+\eta=l).$$

Этим доказано первое равенство в (26). Для доказательства второго достаточно заметить, что

$$2\mathsf{M}(\xi\mid\xi+\eta)=\mathsf{M}(\xi\mid\xi+\eta)+\mathsf{M}(\eta\mid\xi+\eta)=\mathsf{M}(\xi+\eta\mid\xi+\eta)=\xi+\eta.$$

3. Еще в § 1 отмечалось, что каждому разбиению $\mathscr{D}=\{D_1,\ldots,D_k\}$ конечного множества Ω соответствует алгебра $\alpha(\mathscr{D})$ подмножеств Ω. Точно так же и обратно, всякая алгебра $\mathscr{B}$ подмножеств конечного пространства Ω порождается некоторым разбиением $\mathscr{D}$ $(\mathscr{B}=\alpha(\mathscr{D}))$. Тем самым между алгебрами и разбиениями конечного пространства Ω существует взаимно однозначное соответствие. Это обстоятельство следует иметь в виду в связи с вводимым в дальнейшем понятием условного математического ожидания относительно специальных систем множеств, так называемых σ-алгебр.

В случае конечных пространств понятия алгебр и σ-алгебр совпадают. При этом оказывается, что если $\mathscr{B}$ — некоторая алгебра, то вводимое в дальнейшем (§ 7 гл. II) условное математическое ожидание $\mathsf{M}(\xi\mid\mathscr{B})$ случайной величины ξ относительно алгебры $\mathscr{B}$ просто совпадает с $\mathsf{M}(\xi\mid\mathscr{D})$ — математическим ожиданием ξ относительно разбиения $\mathscr{D}$ такого, что $\mathscr{B}=\alpha(\mathscr{D})$. В этом смысле в случае конечных пространств в дальнейшем мы не будем различать $\mathsf{M}(\xi\mid\mathscr{B})$ и $\mathsf{M}(\xi\mid\mathscr{D})$, понимая всякий раз, что $\mathsf{M}(\xi\mid\mathscr{B})$ есть по определению просто $\mathsf{M}(\xi\mid\mathscr{D})$.

4. Задачи.

1. Привести пример двух случайных величин ξ и η, которые не являются независимыми, но для которых

$$\mathsf{M}(\xi\mid\eta)=\mathsf{M}\xi.$$

(Ср. с утверждением (22).)

2. Условной дисперсией ξ относительно разбиения $\mathscr{D}$ называется случайная величина

$$\mathsf{D}(\xi \mid \mathscr{D}) = \mathsf{M}[(\xi - \mathsf{M}(\xi \mid \mathscr{D}))^2 \mid \mathscr{D}].$$

Показать, что дисперсия

$$\mathsf{D}\xi = \mathsf{MD}(\xi \mid \mathscr{D}) + \mathsf{DM}(\xi \mid \mathscr{D}).$$

3. Отправляясь от (17), доказать, что для всякой функции $f = f(\eta)$ условное математическое ожидание $\mathsf{M}(\xi \mid \eta)$ обладает следующим свойством:

$$\mathsf{M}[f(\eta)\,\mathsf{M}(\xi \mid \eta)] = \mathsf{M}[\xi f(\eta)].$$

4. Пусть ξ и η — случайные величины. Показать, что $\inf\limits_f \mathsf{M}(\eta - f(\xi))^2$ достигается на функции $f^*(\xi) = \mathsf{M}(\eta \mid \xi)$. (Таким образом, оптимальной в среднеквадратическом смысле оценкой η по ξ является условное математическое ожидание $\mathsf{M}(\eta \mid \xi)$).

5. Пусть $\xi_1, \ldots, \xi_n, \tau$ — независимые случайные величины, причем $\xi_1, \ldots, \xi_n$ одинаково распределены и τ принимает значения $1, 2, \ldots, n$. Показать, что если $S_\tau = \xi_1 + \ldots + \xi_\tau$ — сумма случайного числа случайных величин, то

$$\mathsf{M}(S_\tau \mid \tau) = \tau \mathsf{M}\xi_1, \quad \mathsf{D}(S_\tau \mid \tau) = \tau \mathsf{D}\xi_1$$

и

$$\mathsf{M}S_\tau = \mathsf{M}\tau \cdot \mathsf{M}\xi_1, \quad \mathsf{D}S_\tau = \mathsf{M}\tau \cdot \mathsf{D}\xi_1 + \mathsf{D}\tau \cdot (\mathsf{M}\xi_1)^2.$$

6. Доказать равенство (24).

§ 9. Случайное блуждание. I. Вероятности разорения и средняя продолжительность при игре с бросанием монеты

1. Значение установленных в § 6 предельных теорем для схемы Бернулли далеко не исчерпывается тем, что они дают удобные формулы для подсчета вероятностей $\mathsf{P}(S_n = k)$ и $\mathsf{P}(A < S_n \leqslant B)$. Роль этих теорем состоит также и в том, что они носят универсальный характер, т. е. остаются справедливыми не только для независимых бернуллиевских случайных величин $\xi_1, \ldots, \xi_n$, принимающих всего лишь два значения, но и для величин гораздо более общей природы. В этом смысле схема Бернулли явилась той простейшей моделью, на примере которой были подмечены многие вероятностные закономерности, присущие и гораздо более общим моделям

В настоящем и следующем параграфах будет рассмотрен ряд новых вероятностных закономерностей, подчас носящих крайне неожиданный характер. Все рассмотрения будут вестись снова для схемы Бернулли, хотя многие выводы о характере случайных колебаний остаются справедливыми для случайных блужданий более общего вида.

2. Рассмотрим схему Бернулли $(\Omega, \mathscr{A}, \mathrm{P})$, где $\Omega=\{\omega: \omega=(x_1, \ldots, x_n),\ x_i=\pm 1\}$, $\mathscr{A}$ — система всех подмножеств Ω и $p(\omega)=p^{\nu(\omega)}q^{n-\nu(\omega)}$, $\nu(\omega)=\frac{\Sigma x_i+n}{2}$. Пусть $\xi_i(\omega)=x_i$, $i=1, \ldots, n$. Тогда, как уже известно, последовательность $\xi_1, \ldots, \xi_n$ является последовательностью независимых бернуллиевских случайных величин

$$\mathrm{P}(\xi_i=1)=p, \quad \mathrm{P}(\xi_i=-1)=q, \quad p+q=1.$$

Положим $S_0=0$, $S_k=\xi_1+\ldots+\xi_k$, $1\leqslant k\leqslant n$. Последовательность $S_0, S_1, \ldots, S_n$ можно рассматривать как траекторию случайного блуждания некоторой частицы, выходящей из нуля. При этом $S_{k+1}=S_k+\xi_k$, т. е. если в момент k частица находится в точке S_k, то в момент $k+1$ она сдвигается либо на единицу вверх (с вероятностью p), либо на единицу вниз (с вероятностью q).

Пусть A и B — два целых числа, $A\leqslant 0\leqslant B$. Одна из интересных задач, связанных с рассматриваемым случайным блужданием, состоит в исследовании вопроса о том, с какой вероятностью блуждающая частица выйдет за n шагов из интервала (A, B). Интересен также вопрос о том, с какой вероятностью выход из интервала (A, B) произойдет в точке A или B.

Естественность этих вопросов становится особенно понятной, если воспользоваться следующей игровой интерпретацией. Пусть имеются два игрока (первый и второй), у которых начальные капиталы равны соответственно $(-A)$ и B. Если $\xi_i=+1$, то будем считать, что второй игрок платит единицу капитала первому; если же $\xi_i=-1$, то наоборот, первый платит второму. Таким образом, $S_k=\xi_1+\ldots+\xi_k$ можно интерпретировать как величину выигрыша первого игрока у второго (если $S_k<0$, то этот выигрыш есть на самом деле величина проигрыша первого игрока второму) за k «ходов».

В тот момент времени $k\leqslant n$, когда впервые $S_k=B$ $(S_k=A)$ капитал второго (первого) игрока становится равным нулю, иначе говоря, происходит его разорение. (Если $k<n$, то следует считать, что игра прекращается в момент времени k, хотя само блуждание остается определенным до момента n включительно.)

Прежде чем переходить к точным постановкам, введем ряд обозначений.

Пусть x — целое число из интервала $[A, B]$ и для $0\leqslant k\leqslant n$ пусть $S_k^x=x+S_k$,

$$\tau_k^x=\min\{0\leqslant l\leqslant k:\ S_l^x=A \text{ или } B\}, \tag{1}$$

где условимся считать $\tau_k^x=k$, если $A<S_l^x<B$ для всех $0\leqslant l\leqslant k$.

Для каждого $0 \leqslant k \leqslant n$ и $x \in [A, B]$ момент τ_k^x, называемый *моментом остановки* (см. § 11), является целочисленной случайной величиной, определенной на пространстве элементарных событий Ω (зависимость τ_k^x от ω явно не указывается).

Ясно, что для всех $l < k$ множество $\{\omega: \tau_k^x = l\}$ есть событие, состоящее в том, что случайное блуждание $\{S_i^x, 0 \leqslant i \leqslant k\}$, начинающееся в нулевой момент в точке x, выйдет из интервала (A, B) в момент l. Понятно также, что для $l \leqslant k$ множества $\{\omega: \tau_k^x = l, S_l^x = A\}$ и $\{\omega: \tau_k^x = l, S_l^x = B\}$ имеют смысл событий, состоящих в том, что блуждающая частица выйдет из интервала (A, B) в момент l в точках A и B соответственно.

Обозначим для всех $0 \leqslant k \leqslant n$

$$
\begin{aligned}
\mathscr{A}_k^x &= \sum_{0 \leqslant l \leqslant k} \{\omega: \tau_k^x = l, S_l^x = A\}, \\
\mathscr{B}_k^x &= \sum_{0 \leqslant l \leqslant k} \{\omega: \tau_k^x = l, S_l^x = B\},
\end{aligned} \tag{2}
$$

и пусть

$$
\alpha_k(x) = \mathsf{P}(\mathscr{A}_k^x), \quad \beta_k(x) = \mathsf{P}(\mathscr{B}_k^x)
$$

— вероятности выхода частицы за время $[0, k]$ из интервала (A, B) соответственно в точках A и B. Для этих вероятностей можно получить рекуррентные соотношения, из которых последовательно находятся $\alpha_1(x), \ldots, \alpha_n(x)$ и $\beta_1(x), \ldots, \beta_n(x)$.

Итак, пусть $A < x < B$. Ясно, что $\alpha_0(x) = \beta_0(x) = 0$. Пусть теперь $1 \leqslant k \leqslant n$. Тогда по формуле (8.5)

$$
\begin{aligned}
\beta_k(x) = \mathsf{P}(\mathscr{B}_k^x) &= \mathsf{P}(\mathscr{B}_k^x \mid S_1^x = x+1)\,\mathsf{P}(\xi_1 = 1) + \\
&+ \mathsf{P}(\mathscr{B}_k^x \mid S_1^x = x-1)\,\mathsf{P}(\xi_1 = -1) = \\
&= p\mathsf{P}(\mathscr{B}_k^x \mid S_1^x = x+1) + q\mathsf{P}(\mathscr{B}_k^x \mid S_1^x = x-1).
\end{aligned} \tag{3}
$$

Покажем, что

$$
\mathsf{P}(\mathscr{B}_k^x \mid S_1^x = x+1) = \mathsf{P}(\mathscr{B}_{k-1}^{x+1}), \quad \mathsf{P}(\mathscr{B}_k^x \mid S_1^x = x-1) = \mathsf{P}(\mathscr{B}_{k-1}^{x-1}).
$$

С этой целью заметим, что множество $\mathscr{B}_k^x$ можно представить в виде

$$
\mathscr{B}_k^x = \{\omega: (x, x+\xi_1, \ldots, x+\xi_1+\ldots+\xi_k) \in B_k^x\},
$$

где B_k^x — множество траекторий вида

$$
(x, x+x_1, \ldots, x+x_1+\ldots+x_k)
$$

с $x_1 = \pm 1$, которые за время $[0, k]$ впервые выходят из интервала (A, B) в точке B (рис. 15).

Представим множество B_k^x в виде $B_k^{x,\,x+1}+B_k^{x,\,x-1}$, где $B_k^{x,\,x+1}$ и $B_k^{x,\,x-1}$ — те траектории из B_k^x, для которых $x_1=+1$ и $x_1=-1$ соответственно.

Заметим, что каждая траектория $(x,\ x+1,\ x+1+x_2,\ \ldots$ $\ldots,\ x+1+x_2+\ldots+x_k)$ из $B_k^{x,\,x+1}$ находится во взаимно однозначном соответствии с траекторией $(x+1,\ x+1+x_2,\ \ldots,\ x+1+x_2+\ldots+x_k)$ из B_{k-1}^{x+1}. То же справедливо и для траекторий из $B_k^{x,\,x-1}$. Учитывая эти обстоятельства, а также независимость, одинаковую распределенность величин $\xi_1,\ \ldots,\ \xi_k$ и формулу (8.6), находим, что

$$\begin{aligned}\mathsf{P}\left(\mathscr{B}_k^x \mid S_1^x=x+1\right)&=\mathsf{P}\left(\mathscr{B}_k^x \mid \xi_1=1\right)=\\&=\mathsf{P}\left\{(x,\ x+\xi_1,\ \ldots,\ x+\xi_1+\right.\\&\left.+\ldots+\xi_k)\in B_k^x \mid \xi_1=1\right\}=\\&=\mathsf{P}\left\{(x+1,\ x+1+\xi_2,\ \ldots,\ x+\right.\\&\left.+1+\xi_2+\ldots+\xi_k)\in B_{k-1}^{x+1}\right\}=\\&=\mathsf{P}\left\{(x+1,\ x+1+\xi_1,\ \ldots,\ x+\right.\\&\left.+1+\xi_1+\ldots+\xi_{k-1})\in B_{k-1}^{x+1}\right\}=\\&=\mathsf{P}\left(\mathscr{B}_{k-1}^{x+1}\right).\end{aligned}$$

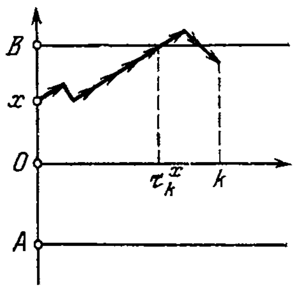


Рис. 15. Пример траектории из множества B_k^x.

Точно так же

$$\mathsf{P}\left(\mathscr{B}_k^x \mid S_1^x=x-1\right)=\mathsf{P}\left(\mathscr{B}_{k-1}^{x-1}\right).$$

Таким образом, в силу (3) для $x\in(A,\ B)$ и $k\leqslant n$

$$\beta_k(x)=p\beta_{k-1}(x+1)+q\beta_{k-1}(x-1), \tag{4}$$

где

$$\beta_l(B)=1,\quad \beta_l(A)=0,\qquad 0\leqslant l\leqslant n. \tag{5}$$

Аналогично

$$\alpha_k(x)=p\alpha_{k-1}(x+1)+q\alpha_{k-1}(x-1) \tag{6}$$

с

$$\alpha_l(A)=1,\qquad \alpha_l(B)=0,\qquad 0\leqslant l\leqslant n.$$

Поскольку $\alpha_0(x)=\beta_0(x)=0,\ x\in(A,\ B)$, то полученные рекуррентные соотношения можно (по крайней мере в принципе) использовать для отыскания вероятностей $\alpha_1(x),\ \ldots,\ \alpha_n(x)$ и $\beta_1(x),\ \ldots,\ \beta_n(x)$. Оставляя в стороне конкретное вычисление этих вероятностей, зададимся вопросом об их значениях при больших n.

С этой целью заметим, что поскольку $\mathscr{B}_{k-1}^x\subset\mathscr{B}_k^x$, $k\leqslant n$, то $\beta_{k-1}(x)\leqslant\beta_k(x)\leqslant 1$. Естественно поэтому рассчитывать (а так оно и есть, см. п. 3), что при достаточно больших n вероятность $\beta_n(x)$ близка к решению $\beta(x)$ уравнения

$$\beta(x)=p\beta(x+1)+q\beta(x-1) \tag{7}$$

с граничными условиями

$$\beta(B)=1, \quad \beta(A)=0, \tag{8}$$

получаемых формальным предельным переходом из (4) и (5).

Для решения задачи (7), (8) предположим сначала, что $p \neq q$, Нетрудно заметить, что рассматриваемое уравнение имеет два частных решения a и $b(q/p)^x$, где a и b — константы. Будем поэтому искать решение $\beta(x)$ в виде

$$\beta(x)=a+b(q/p)^x. \tag{9}$$

С учетом (8) находим, что для всех $A \leqslant x \leqslant B$

$$\beta(x)=\frac{(q/p)^x-(q/p)^A}{(q/p)^B-(q/p)^A}. \tag{10}$$

Покажем, что это есть *единственное* решение рассматриваемой задачи. С этой целью достаточно показать, что все решения задачи (7), (8) могут быть представлены в виде (9).

Пусть $\tilde{\beta}(x)$ — некоторое решение с $\tilde{\beta}(A)=0$, $\tilde{\beta}(B)=1$. Всегда можно найти такие константы $\tilde{a}$ и $\tilde{b}$, что

$$\tilde{a}+\tilde{b}(q/p)^A=\tilde{\beta}(A), \quad \tilde{a}+\tilde{b}(q/p)^{A+1}=\tilde{\beta}(A+1).$$

Тогда из (7) следует, что

$$\tilde{\beta}(A+2)=\tilde{a}+\tilde{b}(q/p)^{A+2}$$

и вообще

$$\tilde{\beta}(x)=\tilde{a}+\tilde{b}(q/p)^x.$$

Тем самым найденное решение (10) есть единственное решение рассматриваемой задачи.

Аналогичные рассуждения показывают, что единственное решение уравнения

$$\alpha(x)=p\alpha(x+1)+q\alpha(x-1), \qquad x \in (A, B) \tag{11}$$

с граничными условиями

$$\alpha(A)=1, \qquad \alpha(B)=0 \tag{12}$$

задается формулой

$$\alpha(x)=\frac{(q/p)^B-(q/p)^x}{(q/p)^B-(q/p)^A}, \quad A \leqslant x \leqslant B. \tag{13}$$

Если же $p=q=1/2$, то единственными решениями $\beta(x)$ и $\alpha(x)$ задач (7), (8) и (11), (12) являются соответственно

$$\beta(x)=\frac{x-A}{B-A} \tag{14}$$

и

$$\alpha(x)=\frac{B-x}{B-A}. \tag{15}$$

Заметим, что при любых $0\leqslant p\leqslant 1$

$$\alpha(x)+\beta(x)=1. \tag{16}$$

Величины $\alpha(x)$ и $\beta(x)$ хотелось бы назвать *вероятностями разорения первого и второго игрока* соответственно (когда начальный капитал первого есть $x-A$, а второго $B-x$) при неограниченном числе ходов, что, конечно, предполагает существование бесконечной последовательности независимых бернуллиевских случайных величин $\xi_1, \xi_2, \ldots$, где $\xi_i=+1$ трактуется как выигрыш первого игрока, а $\xi_i=-1$ — как его проигрыш. Рассмотренное в начале этого параграфа вероятностное пространство $(\Omega, \mathscr{A}, \mathrm{P})$ оказывается слишком «бедным», для того чтобы на нем существовала такая бесконечная последовательность независимых случайных величин. В дальнейшем мы увидим, что такую последовательность действительно можно построить и что величины $\beta(x)$ и $\alpha(x)$ в самом деле являются вероятностями разорения при неограниченном числе шагов.

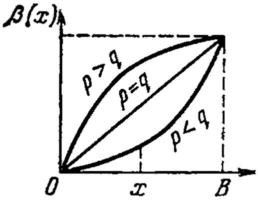


Рис. 16. График $\beta(x)$ — вероятности достижения точки B раньше точки 0, когда частица выходит из точки x.

Обратимся к некоторым следствиям, вытекающим из полученных формул.

Если положить $A=0$, $0\leqslant x\leqslant B$, то по своему смыслу функция $\beta(x)$ будет вероятностью того, что частица, вышедшая из состояния x, достигнет точки B раньше, чем точки 0. Из формул (10) и (14) следует (рис. 16), что

$$\beta(x)=\begin{cases} x/B, & p=q=1/2,\\ \dfrac{(q/p)^x-1}{(q/p)^B-1}, & p\neq q. \end{cases} \tag{17}$$

Далее, пусть $q>p$, означающее, что для первого игрока игра является неблагоприятной. Его предельная вероятность разорения $\alpha=\alpha(0)$ задается формулой

$$\alpha=\frac{(q/p)^B-1}{(q/p)^B-(q/p)^A}.$$

Предположим сейчас, что условия игры изменены: капиталы игроков по-прежнему равны $(-A)$ и B, но плата каждого игрока теперь равна 1/2, а не 1, как раньше. Иначе говоря, пусть теперь $\mathsf{P}(\xi_i = 1/2) = p$, $\mathsf{P}(\xi_i = -1/2) = q$. Обозначим в этом случае предельную вероятность разорения первого игрока через $\alpha_{1/2}$. Тогда

$$\alpha_{1/2} = \frac{(q/p)^{2B} - 1}{(q/p)^{2B} - (q/p)^{2A}},$$

и, значит,

$$\alpha_{1/2} = \alpha \cdot \frac{(q/p)^B + 1}{(q/p)^B + (q/p)^A} > \alpha,$$

если $q > p$.

Отсюда вытекает такой вывод: *если для первого игрока игра неблагоприятна (т. е. $q > p$), то увеличение ставки в два раза уменьшает вероятность его разорения.*

3. Обратимся теперь к вопросу о том, как быстро $\alpha_n(x)$ и $\beta_n(x)$ сходятся к предельным значениям $\alpha(x)$ и $\beta(x)$.

Будем считать для простоты $x = 0$ и обозначим

$$\alpha_n = \alpha_n(0), \quad \beta_n = \beta_n(0), \quad \gamma_n = 1 - (\alpha_n + \beta_n).$$

Ясно, что

$$\gamma_n = \mathsf{P}\{A < S_k < B, \quad 0 \leqslant k \leqslant n\},$$

где $\{A < S_k < B,\ 0 \leqslant k \leqslant n\}$ обозначает событие

$$\bigcap_{0 \leqslant k \leqslant n} \{A < S_k < B\}.$$

Пусть $n = rm$, где r и m — целые числа, и

$$\begin{aligned} \zeta_1 &= \xi_1 + \ldots + \xi_m, \\ \zeta_2 &= \xi_{m+1} + \ldots + \xi_{2m}, \\ &\ldots\ldots\ldots\ldots\ldots \\ \zeta_r &= \xi_{m(r-1)+1} + \ldots + \xi_{rm}. \end{aligned}$$

Тогда, если $C = |A| + B$, то нетрудно убедиться в том, что

$$\{A < S_k < B, \quad 1 \leqslant k \leqslant rm\} \subseteq \{|\zeta_1| < C, \ \ldots, \ |\zeta_r| < C\},$$

и, значит, в силу независимости величин $\zeta_1, \ldots, \zeta_r$ и их одинаковой распределенности

$$\gamma_n \leqslant \mathsf{P}\{|\zeta_1| < C, \ \ldots, \ |\zeta_r| < C\} = \prod_{i=1}^{r} \mathsf{P}\{|\zeta_i| < C\} = (\mathsf{P}\{|\zeta_1| < C\})^r. \tag{18}$$

Заметим, что $\mathsf{D}\zeta_1 = m[1 - (p-q)^2]$. Поэтому при $0 < p < 1$ для достаточно больших m

$$\mathsf{P}\{|\zeta_1| < c\} \leqslant \varepsilon_1, \tag{19}$$

где $\varepsilon_1 < 1$, поскольку если $\mathsf{P}\{|\zeta_1| \leqslant C\} = 1$, то $D\zeta_1 \leqslant C^2$.

Если же $p=0$ или $p=1$, то для достаточно больших m $\mathsf{P}\{|\zeta_1|<C\}=0$ и, следовательно, (19) выполнено при всех $0\leqslant \leqslant p\leqslant 1$.

Из (18) и (19) следует, что для достаточно больших n

$$\gamma_n \leqslant \varepsilon^n, \tag{20}$$

где $\varepsilon=\varepsilon_1^{1/m}<1$.

Согласно (16) $\alpha+\beta=1$. Поэтому

$$(\alpha-\alpha_n)+(\beta-\beta_n)=\gamma_n,$$

и так как $\alpha\geqslant\alpha_n$, $\beta\geqslant\beta_n$, то

$$0\leqslant\alpha-\alpha_n\leqslant\gamma_n\leqslant\varepsilon^n,$$
$$0\leqslant\beta-\beta_n\leqslant\gamma_n\leqslant\varepsilon^n,\quad \varepsilon<1.$$

Аналогичные оценки справедливы и для разностей $\alpha(x)-\alpha_n(x)$ и $\beta(x)-\beta_n(x)$.

4. Обратимся теперь к вопросу *средней длительности* случайного блуждания.

Пусть $m_k(x)=\mathsf{M}\tau_k^x$ — математического ожидание момента остановки τ_k^x, $k\leqslant n$. Поступая, как и при выводе рекуррентных соотношений для $\beta_k(x)$, получаем, что для $x\in(A, B)$

$$\begin{aligned}
m_k(x)=\mathsf{M}\tau_k^x&=\sum_{1\leqslant l\leqslant k} l\mathsf{P}(\tau_k^x=l)=\\
&=\sum_{1\leqslant l\leqslant k} l\cdot\left[p\mathsf{P}(\tau_k^x=l\,|\,\xi_1=1)+q\mathsf{P}(\tau_k^x=l\,|\,\xi_1=-1)\right]=\\
&=\sum_{1\leqslant l\leqslant k} l\cdot\left[p\mathsf{P}(\tau_{k-1}^{x+1}=l-1)+q\mathsf{P}(\tau_{k-1}^{x-1}=l-1)\right]=\\
&=\sum_{0\leqslant l\leqslant k-1}(l+1)\left[p\mathsf{P}(\tau_{k-1}^{x+1}=l)+q\mathsf{P}(\tau_{k-1}^{x-1}=l)\right]=\\
&=pm_{k-1}(x+1)+qm_{k-1}(x-1)+\\
&\qquad+\sum_{0\leqslant l\leqslant k-1}\left[p\mathsf{P}(\tau_{k-1}^{x+1}=l)+q\mathsf{P}(\tau_{k-1}^{x-1}=l)\right]=\\
&=pm_{k-1}(x+1)+qm_{k-1}(x-1)+1.
\end{aligned}$$

Итак, для $x\in(A, B)$ и $0\leqslant k\leqslant n$ функции $m_k(x)$ удовлетворяют рекуррентным уравнениям

$$m_k(x)=1+pm_{k-1}(x+1)+qm_{k-1}(x-1), \tag{21}$$

где $m_0(x)=0$. Из этих уравнений вместе с граничными условиями

$$m_k(A)=m_k(B)=0 \tag{22}$$

можно последовательно найти $m_1(x), \ldots, m_n(x)$.

Поскольку $m_k(x) \leqslant m_{k+1}(x)$, то существует предел

$$m(x) = \lim_{n \to \infty} m_n(x),$$

который в силу (21) удовлетворяет уравнению

$$m(x) = 1 + pm(x+1) + qm(x-1) \tag{23}$$

с граничными условиями

$$m(A) = m(B) = 0. \tag{24}$$

Чтобы найти решение этого уравнения, предположим сначала, что

$$m(x) < \infty, \quad x \in (A, B). \tag{25}$$

Тогда, если $p \neq q$, то частное решение имеет вид $\frac{x}{q-p}$ и общее решение (см. (9)) записывается в виде

$$m(x) = \frac{x}{q-p} + a + b\left(\frac{q}{p}\right)^x.$$

Отсюда с учетом граничных условий $m(A) = m(B) = 0$ находим, что

$$m(x) = \frac{1}{p-q}(B\beta(x) + A\alpha(x) - x], \tag{26}$$

где $\beta(x)$ и $\alpha(x)$ определяются из формул (10) и (13). Если же $p = q = 1/2$, то общее решение уравнения (23) имеет вид

$$m(x) = a + bx - x^2,$$

и поскольку $m(A) = m(B) = 0$, то

$$m(x) = (x - B)(x - A). \tag{27}$$

Отсюда, в частности, вытекает, что если начальные капиталы игроков равны $(B = -A)$, то

$$m(0) = B^2.$$

Возьмем $B = 10$, и пусть каждый ход в игре осуществляется через 1 с., тогда (предельное) среднее время до разорения одного из игроков довольно велико — оно равно 100 с.

Формулы (26) и (27) были получены в предположении, что $m(x) < \infty$, $x \in (A, B)$. Покажем теперь, что и на самом деле $m(x)$ конечны при всех $x \in (A, B)$. Ограничимся рассмотрением случая $x = 0$. Общий случай разбирается аналогичным образом.

Пусть $p = q = 1/2$. С последовательностью $S_0, S_1, \ldots, S_n$ и моментом остановки $\tau_n = \tau_n^0$ свяжем случайную величину S_{τ_n}, определенную следующим равенством:

$$S_{\tau_n} = \sum_{k=0}^{n} S_k I_{\{\tau_n=k\}}(\omega). \tag{28}$$

Наглядный смысл величины S_{τ_n} ясен — это есть значение случайного блуждания в момент остановки τ_n. При этом, если $\tau_n < n$, то $S_{\tau_n} = A$ или B; если же $\tau_n = n$, то $A \leqslant S_{\tau_n} \leqslant B$.

Докажем, что при $p = q = 1/2$

$$\mathsf{M}S_{\tau_n} = 0, \tag{29}$$

$$\mathsf{M}S_{\tau_n}^2 = \mathsf{M}\tau_n. \tag{30}$$

Для доказательства первого равенства заметим, что

$$\begin{aligned} \mathsf{M}S_{\tau_n} &= \sum_{k=0}^{n} \mathsf{M}[S_k I_{\{\tau_n=k\}}(\omega)] = \\ &= \sum_{k=0}^{n} \mathsf{M}[S_n I_{\{\tau_n=k\}}(\omega)] + \sum_{k=0}^{n} \mathsf{M}[(S_k - S_n) I_{\{\tau_n=k\}}(\omega)] = \\ &= \mathsf{M}S_n + \sum_{k=0}^{n} \mathsf{M}[(S_k - S_n) I_{\{\tau_n=k\}}(\omega)], \end{aligned} \tag{31}$$

где, очевидно, $\mathsf{M}S_n = 0$. Покажем, что

$$\sum_{k=0}^{n} \mathsf{M}[(S_k - S_n) I_{\{\tau_n=k\}}(\omega)] = 0.$$

С этой целью заметим, что для $0 \leqslant k < n$ $\{\tau_n > k\} = \{A < S_1 < B, \ \ldots, \ A < S_k < B\}$. Событие $\{A < S_1 < B, \ \ldots, \ A < S_k < B\}$ может быть, очевидно, представлено в виде

$$\{\omega\colon (\xi_1, \ \ldots, \ \xi_k) \in A_k\}, \tag{32}$$

где A_k — некоторое подмножество множества $\{-1, +1\}^k$. Иначе говоря, это множество определяется лишь значениями случайных величин $\xi_1, \ldots, \xi_k$ и не зависит от значений величин $\xi_{k+1}, \ldots, \xi_n$. Поскольку множество

$$\{\tau_n = k\} = \{\tau_n > k-1\} \setminus \{\tau_n > k\},$$

то оно также является множеством вида (32). В силу независимости случайных величин $\xi_1, \ldots, \xi_n$ и в силу задачи 9 к § 4 отсюда вытекает, что для любого $0 \leqslant k < n$ случайные величины $S_n - S_k$ и $I_{\{\tau_n=k\}}$ независимы, а значит,

$$\mathsf{M}[(S_n - S_k) I_{\{\tau_n=k\}}] = \mathsf{M}[S_n - S_k] \cdot \mathsf{M}I_{\{\tau_n=k\}} = 0.$$

Итак, формула (29) установлена.

Тем же методом доказывается и формула (30):

$$\mathsf{M}S_{\tau_n}^2=\sum_{k=0}^{n}\mathsf{M}S_k^2 I_{\{\tau_n=k\}}=\sum_{k=0}^{n}\mathsf{M}\left([S_n+(S_k-S_n)]^2 I_{\{\tau_n=k\}}\right)=$$

$$=\sum_{k=0}^{n}[\mathsf{M}S_n^2 I_{\{\tau_n=k\}}+2\mathsf{M}S_n(S_k-S_n)I_{\{\tau_n=k\}}+$$

$$+\mathsf{M}(S_n-S_k)^2 I_{\{\tau_n=k\}}]=\mathsf{M}S_n^2-\sum_{k=0}^{n}\mathsf{M}(S_n-S_k)^2 I_{\{\tau_n=k\}}=$$

$$=n-\sum_{k=0}^{n}(n-k)\,\mathsf{P}(\tau_n=k)=\sum_{k=0}^{n}k\mathsf{P}(\tau_n=k)=\mathsf{M}\tau_n.$$

Итак, для $p=q=1/2$ имеют место формулы (29), (30). В случае же произвольных p и q $(p+q=1)$ аналогично устанавливается, что

$$\mathsf{M}S_{\tau_n}=(p-q)\cdot\mathsf{M}\tau_n, \tag{33}$$

$$\mathsf{M}[S_{\tau_n}-\tau_n\cdot\mathsf{M}\xi_1]^2=\mathsf{D}\xi_1\cdot\mathsf{M}\tau_n, \tag{34}$$

где $\mathsf{M}\xi_1=p-q$, $\mathsf{D}\xi_1=1-(p-q)^2$.

С помощью полученных соотношений покажем, что $\lim\limits_{n\to\infty} m_n(0)=$ $=m(0)<\infty$.

Если $p=q=1/2$, то в силу (30)

$$\mathsf{M}\tau_n\leqslant\max(A^2,\ B^2). \tag{35}$$

Если же $p\neq q$, то из (33)

$$\mathsf{M}\tau_n\leqslant\frac{\max(|A|,\ B)}{|p-q|}, \tag{36}$$

откуда ясно, что $m(0)<\infty$.

Заметим также, что в случае $p=q=1/2$

$$\mathsf{M}\tau_n=\mathsf{M}S_{\tau_n}^2=A^2\cdot\alpha_n+B^2\cdot\beta_n+\mathsf{M}\left[S_n^2 I_{\{A<S_n<B\}}\right]$$

и, значит,

$$A^2\cdot\alpha_n+B^2\cdot\beta_n\leqslant\mathsf{M}\tau_n\leqslant A^2\cdot\alpha_n+B^2\cdot\beta_n+\max(A^2,\ B^2)\cdot\gamma_n.$$

Вместе с неравенствами (20) отсюда следует, что $\mathsf{M}\tau_n$ сходятся при $n\to\infty$ к предельному значению

$$m(0)=A^2\alpha+B^2\beta=A^2\cdot\frac{B}{B-A}-B^2\cdot\frac{A}{B-A}=|AB|$$

экспоненциально быстро.

Аналогичный результат справедлив и в случае $p\neq q$: экспоненциально быстро $\mathsf{M}\tau_n\to m(0)=\dfrac{\alpha A+\beta B}{p-q}$.

5. Задачи.

1. Показать, что в обобщение (33) и (34) справедливы следующие формулы:

$$\mathsf{M}S_{\tau_n^x} = x + (p-q)\,\mathsf{M}\tau_n^x,$$

$$\mathsf{M}\left[S_{\tau_n^x} - \tau_n^x \cdot \mathsf{M}\xi_1\right]^2 = \mathsf{D}\xi_1 \cdot \mathsf{M}\tau_n^x.$$

2. Исследовать вопрос о том, к чему стремятся величины $\alpha(x)$, $\beta(x)$ и $m(x)$, когда уровень $A \downarrow -\infty$.

3. Пусть в схеме Бернулли $p=q=1/2$. Каков порядок $\mathsf{M}|S_n|$ при больших n?

4. Два игрока независимым образом подбрасывают (каждый свою) симметричные моменты. Показать, что вероятность того, что у них после n подбрасываний будет одно и то же число гербов, равна $2^{-2n}\sum\limits_{k=0}^{n}(C_n^k)^2$. Вывести отсюда равенство $\sum\limits_{k=0}^{n}(C_n^k)^2 = C_{2n}^n$.

Пусть σ_n — тот первый момент, когда число гербов у одного игрока совпадает с числом гербов у другого (совершается n подбрасываний, $\sigma_n = n+1$, если указанного момента не существует). Найти $\mathsf{M}\min(\sigma_n, n)$.

§ 10. Случайное блуждание. II. Принцип отражения. Закон арксинуса

1. Как и в предыдущем параграфе, будем предполагать, что $\xi_1, \xi_2, \ldots, \xi_{2n}$ — последовательность независимых одинаково распределенных бернуллиевских случайных величин с

$$\mathsf{P}(\xi_i = 1) = p, \quad \mathsf{P}(\xi_i = -1) = q,$$

$$S_k = \xi_1 + \ldots + \xi_k, \quad 1 \leqslant k \leqslant 2n; \quad S_0 = 0.$$

Обозначим

$$\sigma_{2n} = \min\{1 \leqslant k \leqslant 2n:\ S_k = 0\},$$

полагая $\sigma_{2n} = \infty$, если $S_k \neq 0$ при всех $1 \leqslant k \leqslant 2n$.

Наглядный смысл σ_{2n} вполне понятен — это момент первого возвращения в нуль. Свойства этого момента и будут изучаться в настоящем параграфе, при этом будет предполагаться, что рассматриваемое случайное блуждание симметрично, т. е. $p=q=1/2$.

Обозначим для $0 \leqslant k \leqslant n$

$$u_{2k} = \mathsf{P}(S_{2k} = 0), \quad f_{2k} = \mathsf{P}(\sigma_{2n} = 2k). \tag{1}$$

Ясно, что $u_0 = 1$ и

Наша ближайшая цель — показать, что для $1 \leqslant k \leqslant n$ вероятность f_{2k} определяется формулой

$$f_{2k} = \frac{1}{2k} u_{2(k-1)}. \tag{2}$$

Понятно, что для $1 \leqslant k \leqslant n$

$$\{\sigma_{2n} = 2k\} = \{S_1 \neq 0,\ S_2 \neq 0,\ \ldots,\ S_{2k-1} \neq 0,\ S_{2k} = 0\},$$

и в силу симметрии

$$\begin{aligned} f_{2k} &= \mathsf{P}\{S_1 \neq 0,\ \ldots,\ S_{2k-1} \neq 0,\ S_{2k} = 0\} = \\ &= 2\mathsf{P}\{S_1 > 0,\ \ldots,\ S_{2k-1} > 0,\ S_{2k} = 0\}. \end{aligned} \tag{3}$$

Назовем *путем* длины k последовательность чисел $(S_0, \ldots, S_k)$ и обозначим через $L_k(A)$ — число путей длины k, для которых выполнено свойство A. Тогда

$$\begin{aligned} f_{2k} = 2 \sum_{(a_{2k+1}, \ldots, a_n)} & L_{2n}(S_1 > 0,\ \ldots,\ S_{2k-1} > 0,\ S_{2k} = 0, \\ & S_{2k+1} = a_{2k+1},\ \ldots,\ S_{2n} = a_{2k+1} + \ldots + a_{2n}) \cdot 2^{-2n} = \\ & = 2L_{2k}(S_1 > 0,\ \ldots,\ S_{2k-1} > 0,\ S_{2k} = 0) \cdot 2^{-2k}, \end{aligned} \tag{4}$$

где суммирование распространяется по всем наборам $(a_{2k+1}, \ldots, a_{2n})$ с $a_i = \pm 1$.

Следовательно, отыскание вероятности f_{2k} сводится к подсчету числа путей $L_{2k}(S_1 > 0, \ldots, S_{2k-1} > 0, S_{2k} = 0)$.

Лемма 1. *Пусть a, b — целые неотрицательные числа, $a - b > 0$ и $k = a + b$. Тогда*

$$L_k(S_1 > 0,\ \ldots,\ S_{k-1} > 0,\ S_k = a - b) = \frac{a-b}{k} C_k^a. \tag{5}$$

Доказательство. Действительно

$$\begin{aligned} L_k(S_1 > 0,\ \ldots,\ & S_{k-1} > 0,\ S_k = a - b) = \\ & = L_k(S_1 = 1,\ S_2 > 0,\ \ldots,\ S_{k-1} > 0,\ S_k = a - b) = \\ & = L_k(S_1 = 1,\ S_k = a - b) - L_k(S_1 = 1,\ S_k = a - b; \\ & \qquad \exists i,\ 2 \leqslant i \leqslant k - 1, \text{ такое, что } S_i \leqslant 0). \end{aligned} \tag{6}$$

Иначе говоря, число положительных путей $(S_1, S_2, \ldots, S_k)$, выходящих из точки (1, 1) и заканчивающихся в точке $(k, a - b)$ совпадает с числом всех путей, идущих из точки (1,1) в точку $(k, a - b)$ за вычетом тех путей, которые касаются или пересекают временную ось *).

*) Путь $(S_1, \ldots, S_k)$ называется *положительным* (неотрицательным), если все $S_i > 0$ $(S_i \geqslant 0)$; путь называется *касающимся* временнóй оси, если для всех $1 \leqslant j \leqslant k$, $S_j \geqslant 0$ или $S_j \leqslant 0$ и найдется такое $1 \leqslant i \leqslant k$, что $S_i = 0$, и называется *пересекающим* временнýю ось, если найдутся такие два момента времени i и j, что $S_i > 0$, а $S_j < 0$.

Заметим теперь, что

$$L_k(S_1=1,\ S_k=a-b;\ \exists i,\ 2\leqslant i\leqslant k-1,\ \text{такое, что } S_i\leqslant 0)=$$
$$=L_k(S_1=-1,\ S_k=a-b), \qquad (7)$$

т. е. число путей, идущих из точки $\alpha=(1,\ 1)$ в точку $\beta=(k,\ a-b)$ и касающихся или пересекающих временную ось, совпадает с числом всех путей, идущих из точки $\alpha^*=(1,\ -1)$ в точку $\beta=(k,\ a-b)$. Доказательство этого утверждения, носящего название *принципа отражения*, следует из легко устанавливаемого взаимно однозначного соответствия между путями $A=(S_1,\ \ldots,\ S_a,\ S_{a+1},\ \ldots,\ S_k)$, соединяющими точки α и β, и путями $B=(-S_1,\ \ldots,\ -S_a,\ S_{a+1},\ \ldots,\ S_k)$, соединяющими точки α^* и β (рис. 17); a — первая точка, где пути A и B обращаются в нуль.

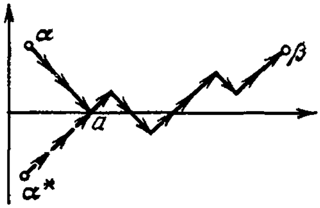


Рис. 17. К принципу отражения.

Из (6) и (7) находим

$$L_k(S_1>0,\ \ldots,\ S_{k-1}>0,\ S_k=a-b)=$$
$$=L_k(S_1=1,\ S_k=a-b)-L_k(S_1=-1,\ S_k=a-b)=$$
$$=C_{k-1}^{a-1}-C_{k-1}^{a}=\frac{a-b}{k}C_k^a,$$

что и доказывает утверждение (5).

Возвращаясь к подсчету вероятности f_{2k}, находим, что, согласно (4) и (5) (с $a=k,\ b=k-1$),

$$f_{2k}=2L_{2k}(S_1>0,\ \ldots,\ S_{2k-1}>0,\ S_{2k}=0)\cdot 2^{-2k}=$$
$$=2L_{2k-1}(S_1>0,\ \ldots,\ S_{2k-1}=1)\cdot 2^{-2k}=$$
$$=2\cdot 2^{-2k}\cdot\frac{1}{2k-1}C_{2k-1}^{k}=\frac{1}{2k}u_{2(k-1)}.$$

Итак, формула (2) доказана.

Приведем еще одно доказательство этой формулы, основанное на следующем замечании. Непосредственная проверка показывает, что

$$\frac{1}{2k}u_{2(k-1)}=u_{2(k-1)}-u_{2k}. \qquad (8)$$

В то же самое время ясно, что

$$\{\sigma_{2n}=2k\}=\{\sigma_{2n}>2(k-1)\}\setminus\{\sigma_{2n}>2k\},$$
$$\{\sigma_{2n}>2l\}=\{S_1\neq 0,\ \ldots,\ S_{2l}\neq 0\}.$$

и, значит,

$$\{\sigma_{2n}=2k\}=\{S_1\neq 0, \ldots, S_{2(k-1)}\neq 0\}\setminus\{S_1\neq 0, \ldots, S_{2k}\neq 0\}.$$

Поэтому

$$f_{2k}=\mathrm{P}\{S_1\neq 0, \ldots, S_{2(k-1)}\neq 0\}\,\mathrm{P}\{S_1\neq 0, \ldots, S_{2k}\neq 0\},$$

и, следовательно, в силу (8) для доказательства равенства $f_{2k}=\frac{1}{2k}u_{2(k-1)}$ достаточно лишь показать, что

$$L_{2k}(S_1\neq 0, \ldots, S_{2k}\neq 0)=L_{2k}(S_{2k}=0). \tag{9}$$

С этой целью заметим, что очевидным образом

$$L_{2k}(S_1\neq 0, \ldots, S_{2k}\neq 0)=2L_{2k}(S_1>0, \ldots, S_{2k}>0).$$

Поэтому для проверки (9) нужно лишь установить, что

$$2L_{2k}(S_1>0, \ldots, S_{2k}>0)=L_{2k}(S_1\geqslant 0, \ldots, S_{2k}\geqslant 0) \tag{10}$$

и

$$L_{2k}(S_1\geqslant 0, \ldots, S_{2k}\geqslant 0)=L_{2k}(S_{2k}=0). \tag{11}$$

Равенство (10) будет доказано, если показать, что между путями $A=(S_1, \ldots, S_{2k})$, у которых по крайней мере одно $S_i=0$, и положительными путями $B=(S_1, \ldots, S_{2k})$ можно установить взаимно однозначное соответствие.

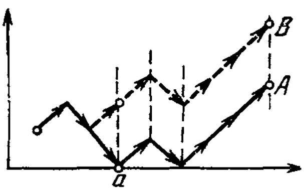


Рис. 18.

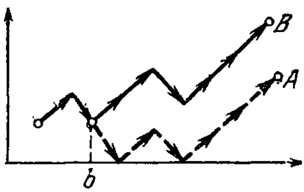


Рис. 19.

Пусть $A=(S_1, \ldots, S_{2k})$ — неотрицательный путь, у которого первое обращение в нуль происходит в точке a (т. е. $S_a=0$). Выпустим из точки $(a, 2)$ траекторию (на рис. 18 она обозначена штриховыми линиями) $(S_a+2, S_{a+1}+2, \ldots, S_{2k}+2)$. Тогда путь $B=(S_1, \ldots, S_{a-1}, S_a+2, \ldots, S_{2k}+2)$ является положительным.

Обратно, $B=(S_1, \ldots, S_{2k})$ — некоторый положительный путь и b — тот последний момент времени, для которого $S_b=1$ (рис. 19). Тогда путь $A=(S_1, \ldots, S_b, S_{b+1}-2, \ldots, S_k-2)$ является неотрицательным. Из приведенных конструкций следует, что между положительными путями и неотринательными путями, у которых по крайней мере одно $S_i=0$, существует взаимно однозначное соответствие. Тем самым формула (10) доказана.

Установим теперь справедливость равенства (11). В силу симметрии и (10) достаточно показать, что

$$L_{2k}(S_1>0, \ldots, S_{2k}>0)+L_{2k}(S_1\geqslant 0, \ldots, S_{2k}\geqslant 0$$
$$\text{и } \exists i,\ 1\leqslant i\leqslant 2k, \text{ такое, что } S_i=0)=L_{2k}(S_{2k}=0).$$

Множество путей $(S_{2k}=0)$ можно представить в виде суммы двух множеств $\mathscr{C}_1$ и $\mathscr{C}_2$, где $\mathscr{C}_1$ — те пути $(S_0, \ldots, S_{2k})$, у которых только один минимум, а $\mathscr{C}_2$ — пути, у которых минимум достигается по меньшей мере в двух точках.

Пусть $C_1 \in \mathscr{C}_1$ (рис. 20) и γ — точка минимума. Поставим пути $C_1=(S_0, S_1, \ldots, S_{2k})$ в соответствие положительный путь C_1^*, полученный следующим образом (рис. 21). Отразим траекторию $(S_0, S_1 \ldots, S_l)$ около вертикальной линии, проходящей через точку l, и полученную траекторию сместим вправо и вверх, выпустив ее из точки $(2k, 0)$. Затем сместим начало координат в точку $(l, -m)$. Полученная траектория C_1^* будет положительным путем.

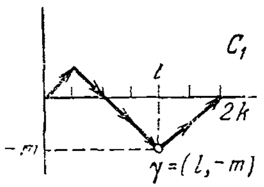


Рис. 20.

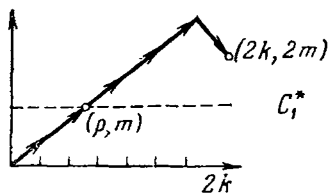


Рис. 21.

Точно так же, если путь $C_2 \in \mathscr{C}_2$, то тем же приемом ему можно поставить в соответствие некоторый неотрицательный путь C_2^*.

Обратно, пусть $C_1^*=(S_1>0, \ldots, S_{2k}>0)$ — некоторый положительный путь с $S_{2k}=2m$ (см. рис. 21). Поставим ему в соответствие путь C_1, полученный следующим образом. Пусть p — та последняя точка, где $S_p=m$. Отразим $(S_p, \ldots, S_{2m})$ около вертикальной прямой $x=p$ и сместим отраженную траекторию вниз и влево, так чтобы ее правый конец совпал с точкой $(0, 0)$. Поместим затем начало координат в левый конец полученной траектории (это будет в точности траектория, изображенная на рис. 20). Полученный путь $C_1=(S_0, \ldots, S_{2k})$ имеет минимум и $S_{2k}=0$. Аналогичная конструкция, примененная к пути $(S_1\geqslant 0, \ldots$ $\ldots, S_{2k}\geqslant 0$ и $\exists i,\ 1\leqslant i\leqslant 2k$, с $S_i=0)$, приводит к пути, у которого по меньшей мере два минимума и $S_{2k}=0$. Тем самым установлено взаимно однозначное соответствие, которое и доказывает требуемый результат (11).

Итак, равенство (9), а следовательно, и формула $f_{2k}=u_{2(k-1)}-u_{2k}=\frac{1}{2k}u_{2(k-1)}$ установлены.

Из формулы Стирлинга

$$u_{2k}=C_{2k}^{k}\cdot 2^{-2k}\sim\frac{1}{\sqrt{\pi k}},\qquad k\to\infty.$$

Поэтому

$$f_{2k}\sim\frac{1}{2\sqrt{\pi}\,k^{3/2}},\qquad k\to\infty.$$

Отсюда следует, что математическое ожидание времени первого возвращения в нуль

$$\mathsf{M}\min(\sigma_{2n},\ 2n)=\sum_{k=1}^{n}2k\mathsf{P}(\sigma_{2n}=2k)+2nu_{2n}=$$
$$=\sum_{k=1}^{n}u_{2(k-1)}+2nu_{2n}$$

является довольно-таки большим.

Более того, $\sum_{k=1}^{\infty}u_{2(k-1)}=\infty$, и, следовательно, предельное значение среднего времени возвращения блуждания в нуль (при неограниченном числе шагов) равно ∞.

Это обстоятельство поясняет многие неожиданные свойства рассматриваемого симметричного случайного блуждания. Например, естественно было бы ожидать, что за время $2n$ число ничьих при игре двух равносильных противников ($p=q=1/2$), т. е. число тех моментов времени i, для которых $S_i=0$, должно быть пропорционально $2n$. Однако на самом деле число ничьих имеет порядок $\sqrt{2n}$ (см. [69]). Отсюда вытекает, в частности, что, вопреки ожидаемому, «типичные» реализации блуждания ($S_0, S_1, \ldots, S_n$) должны иметь не синусоидальный характер (для которых примерно половину времени частица проводит на положительной стороне и другую половину — на отрицательной), а характер длинных затяжных волн. Точная формулировка утверждения дается так называемым законом арксинуса, к изложению которого мы сейчас и приступим.

2. Обозначим $P_{2k,2n}$ вероятность того, что на отрезке $[0, 2n]$ частица проводит $2k$ единиц времени на положительной стороне *).

Лемма 2. *Пусть* $u_0=1$ *и* $0\leqslant k\leqslant n$. *Тогда*

$$P_{2k,2n}=u_{2k}\cdot u_{2n-2k}.\tag{12}$$

*) Мы говорим, что в интервале $[m-1, m]$ частица находится на положительной стороне, если по крайней мере одно из значений S_{m-1} или S_m положительно.

Доказательство. Выше было установлено, что $f_{2k} = u_{2(k-1)} - u_{2k}$. Покажем, что

$$u_{2k} = \sum_{r=1}^{k} f_{2r} \cdot u_{2(k-r)}. \tag{13}$$

Поскольку $\{S_{2k}=0\} \subseteq \{\sigma_{2n} \leqslant 2k\}$, то

$$\{S_{2k}=0\} = \{S_{2k}=0\} \cap \{\sigma_{2n} \leqslant 2k\} = \sum_{1 \leqslant l \leqslant k} \{S_{2k}=0\} \cap \{\sigma_{2n}=2l\}.$$

Следовательно,

$$u_{2k} = \mathsf{P}(S_{2k}=0) = \sum_{1 \leqslant l \leqslant k} \mathsf{P}(S_{2k}=0,\ \sigma_{2n}=2l) = \sum_{1 \leqslant l \leqslant k} \mathsf{P}(S_{2k}=0 \mid \sigma_{2k}=2l)\, \mathsf{P}(\sigma_{2n}=2l).$$

Но

$$\mathsf{P}(S_{2k}=0 \mid \sigma_{2n}=2l) = \mathsf{P}(S_{2k}=0 \mid S_1 \neq 0, \ldots, S_{2l-1} \neq 0,\ S_{2l}=0) =$$
$$= \mathsf{P}(S_{2l} + (\xi_{2l+1} + \ldots + \xi_{2k}) = 0 \mid S_1 \neq 0, \ldots, S_{2l-1} \neq 0,\ S_{2l}=0) =$$
$$= \mathsf{P}(S_{2l} + (\xi_{2l+1} + \ldots + \xi_{2k}) = 0 \mid S_{2l}=0) =$$
$$= \mathsf{P}(\xi_{2l+1} + \ldots + \xi_{2k} = 0) = \mathsf{P}(S_{2(k-l)}=0).$$

Поэтому

$$u_{2k} = \sum_{1 \leqslant l \leqslant k} \mathsf{P}(S_{2(k-l)}=0)\, \mathsf{P}(\sigma_{2n}=2l),$$

что и доказывает (13).

Перейдем к доказательству формулы (12). При $k=0$ и $k=n$ ее справедливость очевидна. Пусть теперь $1 \leqslant k \leqslant n-1$. Если частица проводит $2k$ моментов времени на положительной стороне, то она проходит через нуль. Пусть $2r$ — момент первого возвращения в нуль. Возможны два случая: когда $S_k \geqslant 0$, $k \leqslant 2r$ и $S_k \leqslant 0$, $k \leqslant 2r$.

Число путей, относящихся к первому случаю, равно, как нетрудно видеть,

$$\left(\frac{1}{2} 2^{2r} f_{2r}\right) \cdot 2^{2(n-r)} P_{2(k-r),\,2(n-r)} = \frac{1}{2} \cdot 2^{2n} \cdot f_{2r} \cdot P_{2(k-r),\,2(n-r)}.$$

Во втором случае соответствующее число путей равно

$$\frac{1}{2} \cdot 2^{2n} \cdot f_{2r} \cdot P_{2k,\,2(n-r)}.$$

Следовательно, для $1 \leqslant k \leqslant n-1$

$$P_{2k,\,2n} = \frac{1}{2} \sum_{r=1}^{k} f_{2r} \cdot P_{2(k-r),\,2(n-r)} + \frac{1}{2} \sum_{r=1}^{k} f_{2r} \cdot P_{2k,\,2(n-r)}, \tag{14}$$

Предположим, что формула $P_{2k,2m}=u_{2k}\cdot u_{2m-2k}$ верна для $m=1, \ldots, n-1$. Тогда из (13) и (14) находим, что

$$P_{2k,2n}=\frac{1}{2}u_{2n-2k}\cdot\sum_{r=1}^{k}f_{2r}\cdot u_{2k-2r}+\frac{1}{2}u_{2k}\cdot\sum_{r=1}^{k}f_{2r}\cdot u_{2n-2r-2k}=$$
$$=\frac{1}{2}u_{2n-2k}\cdot u_{2k}+\frac{1}{2}u_{2k}\cdot u_{2n-2k}=u_{2k}\cdot u_{2n-2k}.$$

Лемма доказана.

Пусть теперь $\gamma(2n)$ — число единиц времени, которое частица проводит на положительной оси в интервале $[0, 2n]$. Тогда для $x<1$

$$\mathsf{P}\left\{\frac{1}{2}<\frac{\gamma(2n)}{2n}\leqslant x\right\}=\sum_{\left\{k:\ \frac{1}{2}<\frac{2k}{2n}\leqslant x\right\}}P_{2k,2n}.$$

Поскольку при $k\to\infty$

$$u_{2k}\sim\frac{1}{\sqrt{\pi k}},$$

то

$$P_{2k,2n}=u_{2k}\cdot u_{2(n-k)}\sim\frac{1}{\pi\sqrt{k(n-k)}},$$

если $k\to\infty$, $n-k\to\infty$.

Поэтому

$$\sum_{\left\{k:\ \frac{1}{2}<\frac{2k}{2n}\leqslant x\right\}}P_{2k,2n}-\sum_{\left\{k:\ \frac{1}{2}<\frac{2k}{2n}\leqslant x\right\}}\frac{1}{\pi n}\cdot\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-1/2}\to 0,\quad n\to\infty,$$

откуда

$$\sum_{\left\{k:\ \frac{1}{2}<\frac{2k}{2n}\leqslant x\right\}}P_{2k,2n}-\frac{1}{\pi}\int_{\frac{1}{2}}^{x}\frac{dt}{\sqrt{t(1-t)}}\to 0,\qquad n\to\infty.$$

Но из соображений симметрии

$$\sum_{\left\{k:\ \frac{k}{n}\leqslant\frac{1}{2}\right\}}P_{2k,2n}\to\frac{1}{2}$$

и

$$\frac{1}{\pi}\int_{1/2}^{x}\frac{dt}{\sqrt{t(1-t)}}=\frac{2}{\pi}\arcsin\sqrt{x}-\frac{1}{2}.$$

Тем самым доказана следующая

Теорема (закон арксинуса). *Вероятность того, что доля времени, проводимого частицей на положительной стороне, меньше* *или равна* x, *стремится к* $2\pi^{-1}\arcsin\sqrt{x}$:

$$\sum_{\left\{k:\ \frac{k}{n}\leqslant x\right\}} P_{2k,\,2n}\to 2\pi^{-1}\arcsin\sqrt{x}. \tag{15}$$

Заметим, что подынтегральная функция $p(t)$ в интеграле

$$\frac{1}{\pi}\int\limits_0^x \frac{dt}{\sqrt{t(1-t)}}$$

представляет U-образную кривую, уходящую в бесконечность в точках $t=0$ и $t=1$.

Отсюда следует, что при больших n

$$\mathsf{P}\left\{0<\frac{\gamma(2n)}{2n}\leqslant\Delta\right\}>\mathsf{P}\left\{\frac{1}{2}<\frac{\gamma(2n)}{2n}\leqslant\frac{1}{2}+\Delta\right\},$$

т. е. более вероятно, что доля времени, проводимой частицей на положительной стороне, будет близка к нулю или единице, нежели к естественно ожидаемому значению 1/2.

Пользуясь таблицами арксинуса и тем обстоятельством, что на самом деле скорость сходимости в (15) очень быстрая, находим, что

$$\mathsf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0{,}024\right\}\approx 0{,}1,$$

$$\mathsf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0{,}1\right\}\approx 0{,}2,$$

$$\mathsf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0{,}2\right\}\approx 0{,}3,$$

$$\mathsf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0{,}65\right\}\approx 0{,}6.$$

Таким образом, если, скажем, $n=1000$, то примерно в одном случае из десяти частица проводит всего лишь 24 единицы времени на положительной оси и, значит, бо́льшую часть времени — 976 единиц — на отрицательной оси.

3. Задачи.

1. С какой скоростью $\mathsf{M}\min(\sigma_{2n},\ 2n)\to\infty$ при $n\to\infty$?

2. Пусть $\tau_n=\min\{1\leqslant k\leqslant n:\ S_k=1\}$, считая, $\tau_n=\infty$, если $S_k<1$ при всех $1\leqslant k\leqslant n$. К чему стремится $\mathsf{M}\min(\tau_n,\ n)$ при $n\to\infty$ для симметричного $(p=q=1/2)$ и несимметричного $(p\neq q)$ блужданий?

§ 11. Мартингалы. Некоторые применения к случайному блужданию

1. Рассмотренные выше бернуллиевские случайные величины $\xi_1, \ldots, \xi_n$ образовывали последовательность *независимых* случайных величин. В этом и следующем параграфе будут введены два важных класса *зависимых* случайных величин, образующих мартингал и марковскую цепь.

Теория мартингалов будет детально излагаться в гл. VII. Сейчас же будут даны лишь необходимые определения, доказана одна теорема о сохранении мартингального свойства для моментов остановки и дано ее применение к выводу так называемой теоремы о баллотировке. В свою очередь эта последняя теорема будет использована для иного доказательства утверждения (10.5), полученного выше с применением принципа отражения.

2. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — конечное вероятностное пространство, $\mathscr{D}_1 \subseteq \mathscr{D}_2 \subseteq \ldots \subseteq \mathscr{D}_n$ — некоторая последовательность разбиений.

Определение 1. Последовательность случайных величин $\xi_1, \ldots, \xi_n$ называется *мартингалом* (относительно разбиений $\mathscr{D}_1 \subseteq \mathscr{D}_2 \subseteq \ldots \subseteq \mathscr{D}_n$), если:

1) ξ_k являются $\mathscr{D}_k$-измеримыми,

2) $\mathsf{M}(\xi_{k+1} \mid \mathscr{D}_k) = \xi_k$, $1 \leqslant k \leqslant n-1$.

Чтобы подчеркнуть, относительно какой системы разбиений случайные величины $\xi_1, \ldots, \xi_n$ образуют мартингал, будем для его обозначения использовать запись:

$$\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}, \tag{1}$$

часто опуская для простоты указание на то, что $1 \leqslant k \leqslant n$.

В том случае, когда разбиения $\mathscr{D}_k$ порождаются величинами $\xi_1, \ldots, \xi_k$, т. е.

$$\mathscr{D}_k = \mathscr{D}_{\xi_1, \ldots, \xi_k},$$

вместо того, чтобы говорить, что $\xi = (\xi_k, \mathscr{D}_k)$ — мартингал, будем просто говорить, что последовательность $\xi = (\xi_k)$ образует мартингал.

Остановимся на некоторых примерах мартингалов.

Пример 1. Пусть $\eta_1, \ldots, \eta_n$ — независимые бернуллиевские случайные величины с

$$\mathsf{P}(\eta_k = 1) = \mathsf{P}(\eta_k = -1) = 1/2,$$

$$S_k = \eta_1 + \ldots + \eta_k \quad \text{и} \quad \mathscr{D}_k = \mathscr{D}_{\eta_1, \ldots, \eta_k}.$$

Заметим, что структура разбиений $\mathscr{D}_k$ проста:

$$\mathscr{D}_1 = \{D^+, D^-\},$$

где $D^+=\{\omega\colon \eta_1=+1\}$, $D^-=\{\omega\colon \eta_1=-1\}$,

$$\mathscr{D}_2=\{D^{++},\ D^{+-},\ D^{-+},\ D^{--}\},$$

где

$$D^{++}=\{\omega\colon \eta_1=+1,\ \eta_2=+1\},\ \ldots,\ D^{--}=\{\omega\colon \eta_1=-1,\ \eta_2=-1\}$$

и т. д.

Нетрудно понять также, что $\mathscr{D}_{\eta_1,\ldots,\eta_k}=\mathscr{D}_{S_1,\ldots,S_k}$.

Покажем, что последовательность $(S_k,\ \mathscr{D}_k)$ образует мартингал. Действительно, S_k $\mathscr{D}_k$-измеримы и в силу (8.12), (8.18) и (8.24)

$$\mathsf{M}(S_{k+1}\,|\,\mathscr{D}_k)=\mathsf{M}(S_k+\eta_{k+1}\,|\,\mathscr{D}_k)=$$
$$=\mathsf{M}(S_k\,|\,\mathscr{D}_k)+\mathsf{M}(\eta_{k+1}\,|\,\mathscr{D}_k)=S_k+\mathsf{M}\eta_{k+1}=S_k.$$

Если положить $S_0=0$ и взять $D_0=\{\Omega\}$ — тривиальное разбиение, то последовательность $(S_k,\ \mathscr{D}_k)_{0\leqslant k\leqslant n}$ также будет мартингалом.

Пример 2. Пусть $\eta_1,\ \ldots,\ \eta_n$ — независимые бернуллиевские случайные величины с $\mathsf{P}(\eta_i=1)=p$, $\mathsf{P}(\eta_i=-1)=q$. Если $p\neq q$, то каждая из последовательностей $\xi=(\xi_k)$ с

$$\xi_k=\left(\frac{q}{p}\right)^{S_k},\quad \xi_k=S_k-k(p-q),\ \text{где}\ S_k=\eta_1+\ldots+\eta_k,$$

образует мартингал.

Пример 3. Пусть η — некоторая случайная величина, $\mathscr{D}_1\subseteq\ldots$ $\ldots\subseteq\mathscr{D}_n$, и

$$\xi_k=\mathsf{M}(\eta\,|\,\mathscr{D}_k). \qquad (2)$$

Тогда последовательность $\xi=(\xi_k,\ \mathscr{D}_k)$ образует мартингал. В самом деле, $\mathscr{D}_k$-измеримость $\mathsf{M}(\eta\,|\,\mathscr{D}_k)$ очевидна и, согласно (8.20),

$$\mathsf{M}(\xi_{k+1}\,|\,\mathscr{D}_k)=\mathsf{M}[\mathsf{M}(\eta\,|\,\mathscr{D}_{k+1}\,|\,\mathscr{D}_k]=\mathsf{M}(\eta\,|\,\mathscr{D}_k)=\xi_k.$$

В связи с этим заметим, что если $\xi=(\xi_k,\ \mathscr{D}_k)$ — произвольный мартингал, то в силу формулы (8.20)

$$\xi_k=\mathsf{M}(\xi_{k+1}\,|\,\mathscr{D}_k)=\mathsf{M}[\mathsf{M}(\xi_{k+2}\,|\,\mathscr{D}_{k+1})\,|\,\mathscr{D}_k]=$$
$$=\mathsf{M}(\xi_{k+2}\,|\,\mathscr{D}_k)=\ldots=\mathsf{M}(\xi_n\,|\,\mathscr{D}_k). \qquad (3)$$

Таким образом, множество всех мартингалов $\xi=(\xi_k,\ \mathscr{D}_k)$ исчерпывается мартингалами вида (2). (Заметим, что в случае бесконечных последовательностей $\xi=(\xi_k,\ \mathscr{D}_k)_{k\geqslant 1}$ это, вообще говоря, уже не так; см. задачу 7 в § 1 гл. VII).

Пример 4. Пусть $\eta_1,\ \ldots,\ \eta_n$ — последовательность независимых одинаково распределенных случайных величин, $S_k=\eta_1+\ldots$ $\ldots+\eta_k$ и $\mathscr{D}_1=\mathscr{D}_{S_1}$, $\mathscr{D}_2=\mathscr{D}_{S_1,S_2},\ \ldots,\ \mathscr{D}_n=\mathscr{D}_{S_1,\ldots,S_n}$. Покажем, что последовательность $\xi=(\xi_k,\ \mathscr{D}_k)$ с $\xi_1=\dfrac{S_n}{n}$, $\xi_2=\dfrac{S_{n-1}}{n-1},\ \ldots,\ \xi_k=$

$=\frac{S_{n+1-k}}{n+1-k}, \ldots, \xi_n=S_1$, образует мартингал. Во-первых, ясно, что $\mathscr{D}_k \subseteq \mathscr{D}_{k+1}$ и ξ_k $\mathscr{D}_k$-измеримы. Далее, в силу симметрии для $j \leqslant n-k+1$

$$\mathsf{M}(\eta_j \mid \mathscr{D}_k)=\mathsf{M}(\eta_1 \mid \mathscr{D}_k) \tag{4}$$

(ср. с (8.26)). Поэтому

$$(n-k+1)\,\mathsf{M}(\eta_1 \mid \mathscr{D}_k)=\sum_{j=1}^{n-k+1} \mathsf{M}(\eta_j \mid \mathscr{D}_k)=\mathsf{M}(S_{n-k+1} \mid \mathscr{D}_k)=S_{n-k+1},$$

а значит,

$$\xi_k=\frac{S_{n-k+1}}{n-k+1}=\mathsf{M}(\eta_1 \mid \mathscr{D}_k),$$

и мартингальность последовательности $\xi=(\xi_k, \mathscr{D}_k)$ следует из примера 3.

Пример 5. Пусть $\eta_1, \ldots, \eta_n$ — независимые бернуллиевские случайные величины с

$$\mathsf{P}(\eta_i=+1)=\mathsf{P}(\eta_i=-1)=1/2,$$

$S_k=\eta_1+\ldots+\eta_k$. Пусть A и B — два целых числа, $A<0<B$. Тогда для всякого $0<\lambda<\pi/2$ последовательность $\xi=(\xi_k, \mathscr{D}_k)$ с $\mathscr{D}_k=\mathscr{D}_{S_1, \ldots, S_k}$ и

$$\xi_k=(\cos\lambda)^{-k}\exp\left\{i\lambda\left(S_k-\frac{B+A}{2}\right)\right\} \tag{5}$$

образует комплексный мартингал (т. е. действительная и комплексная части ξ_k — мартингалы).

3. Из определения мартингала следует, что математическое ожидание $\mathsf{M}\xi_k$ одно и то же для всех k:

$$\mathsf{M}\xi_k=\mathsf{M}\xi_1.$$

Оказывается, что это свойство останется справедливым, если вместо момента k взять случайный момент.

Для формулировки этого свойства введем такое

Определение 2. Случайная величина $\tau=\tau(\omega)$, принимающая значения $1, 2, \ldots, n$, будет называться *моментом остановки* (относительно разбиений $(\mathscr{D}_k)_{1\leqslant k\leqslant n}$, $\mathscr{D}_1 \subseteq \mathscr{D}_2 \subseteq \ldots \subseteq \mathscr{D}_n$), если для любого $k=1, \ldots, n$ случайные величины $I_{\{\tau=k\}}(\omega)$ являются $\mathscr{D}_k$-измеримыми.

Если трактовать разбиение $\mathscr{D}_k$ как разбиение, порожденное наблюдениями за k шагов (например, $\mathscr{D}_k=\mathscr{D}_{\eta_1, \ldots, \eta_k}$ — разбиение, порожденное величинами $\eta_1, \ldots, \eta_k$), то $\mathscr{D}_k$-измеримость величины $I_{\{\tau=k\}}(\omega)$ означает, что осуществление или неосуществление события $\{\tau=k\}$ определяется лишь наблюдениями за k шагов (и не зависит от «будущего»).

Если $\mathscr{B}_k = \alpha(\mathscr{D}_k)$, то $\mathscr{D}_k$-измеримость величин $I_{\{\tau=k\}}(\omega)$ эквивалентна предположению, что

$$\{\tau = k\} \in \mathscr{B}_k. \tag{6}$$

С конкретными примерами моментов остановки мы уже встречались: таковыми являются моменты τ_k^x, σ_{2n}, введенные в §§ 9 и 10. Эти моменты являются частным случаем моментов остановки вида

$$\begin{aligned} \tau^A &= \min\{0 < k \leqslant n\colon\ \xi_k \in A\}, \\ \sigma^A &= \min\{0 \leqslant k \leqslant n\colon\ \xi_k \in A\}, \end{aligned} \tag{7}$$

являющихся моментами (соответственно первого после нуля и первого) достижения множества A некоторой последовательностью $\xi_0, \xi_1, \ldots, \xi_n$.

4. Теорема 1. *Пусть $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ — мартингал и τ — некоторый момент остановки относительно разбиений $(\mathscr{D}_k)_{1 \leqslant k \leqslant n}$. Тогда*

$$\mathsf{M}(\xi_\tau \mid \mathscr{D}_1) = \xi_1, \tag{8}$$

где

$$\xi_\tau = \sum_{k=1}^{n} \xi_k I_{\{\tau=k\}}(\omega) \tag{9}$$

и

$$\mathsf{M}\xi_\tau = \mathsf{M}\xi_1. \tag{10}$$

Доказательство (ср. с доказательством формулы (9.29)). Пусть $D \in \mathscr{D}_1$. Тогда, пользуясь свойствами условных математических ожиданий и (3), находим, что

$$\begin{aligned} \mathsf{M}(\xi_\tau \mid D) &= \frac{\mathsf{M}(\xi_\tau I_D)}{\mathsf{P}(D)} = \\ &= \frac{1}{\mathsf{P}(D)} \cdot \sum_{l=1}^{n} \mathsf{M}(\xi_l \cdot I_{\{\tau=l\}} \cdot I_D) = \\ &= \frac{1}{\mathsf{P}(D)} \sum_{l=1}^{n} \mathsf{M}[\mathsf{M}(\xi_n \mid \mathscr{D}_l) \cdot I_{\{\tau=l\}} \cdot I_D] = \\ &= \frac{1}{\mathsf{P}(D)} \sum_{l=1}^{n} \mathsf{M}[\mathsf{M}(\xi_n I_{\{\tau=l\}} \cdot I_D \mid \mathscr{D}_l)] = \\ &= \frac{1}{\mathsf{P}(D)} \sum_{l=1}^{n} \mathsf{M}[\xi_n I_{\{\tau=l\}} \cdot I_D] = \\ &= \frac{1}{\mathsf{P}(D)} \mathsf{M}(\xi_n I_D) = \mathsf{M}(\xi_n \mid D); \end{aligned}$$

а следовательно,

$$\mathsf{M}(\xi_\tau \mid \mathscr{D}_1) = \mathsf{M}(\xi_n \mid \mathscr{D}_1) = \xi_1.$$

Равенство $\mathsf{M}\xi_\tau = \mathsf{M}\xi_1$ следует отсюда очевидным образом.

Теорема доказана.

Следствие. Для мартингала $(S_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$ из примера 1 и любого момента остановки τ (относительно $(\mathscr{D}_k)$) справедливы формулы

$$\mathsf{M}S_\tau = 0, \quad \mathsf{M}S_\tau^2 = \mathsf{M}\tau, \tag{11}$$

называемые тождествами Вальда (ср. с (9.29) и (9.30); см. также задачу 1 и теорему 3 в § 2 гл. VII).

4. Используем теорему 1 для доказательства следующего утверждения.

Теорема 2 (теорема о баллотировке). *Пусть $\eta_1, \ldots, \eta_n$ — последовательность независимых одинаково распределенных случайных величин, принимающих неотрицательные целочисленные значения, $S_k = \eta_1 + \ldots + \eta_k$, $1 \leqslant k \leqslant n$. Тогда*

$$\mathsf{P}\{S_k < k \textit{ для всех } 1 \leqslant k \leqslant n \mid S_n\} = \left(1 - \frac{S_n}{n}\right)^+, \tag{12}$$

где $a^+ = \max(a, 0)$.

Доказательство. На множестве $\{\omega: S_n \geqslant n\}$ формула очевидна. Будем поэтому доказывать (12) для тех элементарных исходов, для которых $S_n < n$.

Рассмотрим мартингал $\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$ с $\xi_k = \dfrac{S_{n+1-k}}{n+1-k}$ и $\mathscr{D}_k = D_{S_{n+1-k}, \ldots, S_n}$, введенный в примере 4.

Определим

$$\tau = \min\{1 \leqslant k \leqslant n: \xi_k \geqslant 1\},$$

полагая $\tau = n$ на множестве $\{\xi_k < 1 \textit{ для всех } 1 \leqslant k \leqslant n\} = \left\{\max\limits_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1\right\}$. Понятно, что на этом множестве $\xi_\tau = \xi_n = S_1 = 0$ и, значит,

$$\left\{\max_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1\right\} = \left\{\max_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1,\ S_n < n\right\} \subseteq \{\xi_\tau = 0\}. \tag{13}$$

Рассмотрим теперь те исходы, для которых одновременно $\max\limits_{1\leqslant l\leqslant n} \frac{S_l}{l} \geqslant 1$ и $S_n < n$. Обозначим $\sigma = n + 1 - \tau$. Нетрудно видеть, что

$$\sigma = \max\{1 \leqslant k \leqslant n: S_k \geqslant k\}$$

и, значит, (поскольку $S_n < n$) $\sigma < n$, $S_\sigma \geqslant \sigma$ и $S_{\sigma+1} < \sigma + 1$. Следовательно, $\eta_{\sigma+1} = S_{\sigma+1} - S_\sigma < (\sigma+1) - \sigma = 1$, т. е. $\eta_{\sigma+1} = 0$. Поэтому

$\sigma \leqslant S_\sigma = S_{\sigma+1} < \sigma + 1$, а следовательно, $S_\sigma = \sigma$ и

$$\xi_\tau = \frac{S_{n+1-\tau}}{n+1-\tau} = \frac{S_\sigma}{\sigma} = 1.$$

Тем самым

$$\left\{\max_{1 \leqslant l \leqslant n} \frac{S_l}{l} \geqslant 1,\ S_n < n\right\} \subseteq \{\xi_\tau = 1\}. \tag{14}$$

Из (13) и (14) находим, что

$$\left\{\max_{1 \leqslant l \leqslant n} \frac{S_l}{l} \geqslant 1,\ S_n < n\right\} = \{\xi_\tau = 1\} \cap \{S_n < n\}.$$

Поэтому на множестве $\{S_n < n\}$

$$\mathsf{P}\left\{\max_{1 \leqslant l \leqslant n} \frac{S_l}{l} \geqslant 1 \,|\, S_n\right\} = \mathsf{P}\{\xi_\tau = 1 \,|\, S_n\} = \mathsf{M}(\xi_\tau \,|\, S_n),$$

где последнее равенство следует из того, что ξ_τ принимает лишь два значения: 0 или 1.

Заметим теперь, что $\mathsf{M}(\xi_\tau \,|\, S_n) = \mathsf{M}(\xi_\tau \,|\, \mathscr{D}_1)$ и в силу теоремы 1 $\mathsf{M}(\xi_\tau \,|\, \mathscr{D}_1) = \xi_1 = S_n/n$. Следовательно, на множестве $\{S_n < n\}$ $\mathsf{P}\{S_k < k \text{ *для всех* } 1 \leqslant k \leqslant n \,|\, S_n\} = 1 - \frac{S_n}{n}$.

Теорема доказана.

Применим эту теорему для получения другого доказательства леммы 1 из § 10 и объясним ее название как теоремы о баллотировке.

Пусть $\xi_1, \ldots, \xi_n$ — независимые бернуллиевские случайные величины с

$$\mathsf{P}(\xi_1 = 1) = \mathsf{P}(\xi_i = -1) = 1/2,$$

$S_k = \xi_1 + \ldots + \xi_k$ и a, b — целые неотрицательные числа такие, что $a - b > 0$, $a + b = n$. Покажем, что тогда

$$\mathsf{P}\{S_1 > 0, \ldots, S_n > 0 \,|\, S_n = a - b\} = \frac{a-b}{a+b}. \tag{15}$$

В самом деле, в силу симметрии

$$\begin{aligned}
&\mathsf{P}\{S_1 > 0, \ldots, S_n > 0 \,|\, S_n = a - b\} = \\
&\quad = \mathsf{P}\{S_1 < 0, \ldots, S_n < 0 \,|\, S_n = -(a-b)\} = \\
&\quad = \mathsf{P}\{S_1 + 1 < 1, \ldots, S_n + n < n \,|\, S_n + n = n - (a-b)\} = \\
&\quad = \mathsf{P}\{\eta_1 < 1, \ldots, \eta_1 + \ldots + \eta_n < n \,|\, \eta_1 + \ldots + \eta_n = n - (a-b)\} = \\
&\quad = \left[1 - \frac{n-(a-b)}{n}\right]^+ = \frac{a-b}{n} = \frac{a-b}{a+b},
\end{aligned}$$

где мы положили $\eta_k = \xi_k + 1$ и воспользовались равенством (12).

Из (15) очевидным образом выводится формула (10.5), установленная в лемме 1 § 10 с применением принципа отражения.

Будем интерпретировать $\xi_i=+1$ как голос, поданный на выборах за кандидата A, а $\xi_i=-1$ — за кандидата B. Тогда S_k есть разность числа голосов, поданных за кандидатов A и B, если в голосовании приняло участие k избирателей, а $\mathsf{P}\{S_1>0, \ldots, S_n>0 \mid S_n=a-b\}$ есть вероятность того, что кандидат A все время был впереди кандидата B, при условии, что в общей сложности A собрал a голосов, B собрал b голосов и $a-b>0$, $a+b=n$. Согласно (15) эта вероятность равна $(a-b)/n$.

5. Задачи.

1. Пусть $\mathscr{D}_0 \subseteq \mathscr{D}_1 \subseteq \ldots \subseteq \mathscr{D}_n$ — последовательность разбиений, $\mathscr{D}_0=\{\Omega\}$; η_k — $\mathscr{D}_k$-измеримая величина, $1 \leqslant k \leqslant n$. Доказать, что последовательность $\xi=(\xi_k, \mathscr{D}_k)$ с

$$\xi_k=\sum_{l=1}^{k}[\eta_l - M(\eta_l \mid \mathscr{D}_{l-1})]$$

является мартингалом.

2. Пусть случайные величины $\eta_1, \ldots, \eta_k$ таковы, что $\mathsf{M}(\eta_k \mid \eta_1, \ldots, \eta_{k-1})=0$. Доказать, что последовательность $\xi=(\xi_k)_{1\leqslant k\leqslant n}$ с $\xi_1=\eta_1$ и

$$\xi_{k+1}=\sum_{i=1}^{k}\eta_{i+1}f_i(\eta_1, \ldots, \eta_i),$$

где f_i — некоторые функции, образует мартингал.

3. Показать, что всякий мартингал $\xi=(\xi_k, \mathscr{D}_k)$ имеет некоррелированные приращения: если $a<b<c<d$, то

$$\operatorname{cov}(\xi_d-\xi_c, \xi_b-\xi_a)=0.$$

4. Пусть $\xi=(\xi_1, \ldots, \xi_n)$ — некоторая случайная последовательность такая, что ξ_k $\mathscr{D}_k$-измеримы ($\mathscr{D}_1 \subseteq \mathscr{D}_2 \subseteq \ldots \subseteq \mathscr{D}_n$). Доказать, что для того, чтобы эта последовательность была мартингалом (относительно системы разбиений $(\mathscr{D}_k)$), необходимо и достаточно, чтобы для любого момента остановки τ (относительно $(\mathscr{D}_k)$) $\mathsf{M}\xi_\tau=\mathsf{M}\xi_1$. (Выражение «для любого момента остановки» можно заменить на выражение «для любого момента остановки, принимающего два значения»).

5. Показать, что если $\xi=(\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$ — мартингал и τ — момент остановки, то для любого k

$$\mathsf{M}[\xi_n I_{\{\tau=k\}}]=\mathsf{M}[\xi_k I_{\{\tau=k\}}].$$

6. Пусть $\xi=(\xi_k, \mathscr{D}_k)$ и $\eta=(\eta_k, \mathscr{D}_k)$ — два мартингала, $\xi_1=\eta_1=0$. Доказать, что

$$\mathsf{M}\xi_n\eta_n=\sum_{k=2}^{n}\mathsf{M}(\xi_k-\xi_{k-1})(\eta_k-\eta_{k-1})$$

и, в частности,

$$\mathsf{M}\xi_n^2 = \sum_{k=2}^{n} \mathsf{M}(\xi_k - \xi_{k-1})^2.$$

7. Пусть $\eta_1, \ldots, \eta_n$ — последовательность независимых одинаково распределенных случайных величин, $\mathsf{M}\eta_i = 0$. Показать, что последовательности $\xi = (\xi_k)$ с

$$\xi_k = \left(\sum_{i=1}^{k} \eta_i\right)^2 - k\mathsf{M}\eta_i^2,$$

$$\xi_k = \frac{\exp \lambda(\eta_1 + \ldots + \eta_k)}{(\mathsf{M}\exp\lambda\eta_1)^k}$$

являются мартингалами.

8. Пусть $\eta_1, \ldots, \eta_n$ — последовательность независимых одинаково распределенных случайных величин, принимающих значения в конечном множестве Y. Пусть $f_0(y) = \mathsf{P}(\eta_1 = y)$, $y \in Y$ и $f_1(y)$ — неотрицательная функция с $\sum_{y \in Y} f_1(y) = 1$. Показать, что последовательность $\xi = (\xi_k, \mathscr{D}_k^\eta)$ с $\mathscr{D}_k^\eta = \mathscr{D}_{\eta_1, \ldots, \eta_k}$,

$$\xi_k = \frac{f_1(\eta_1) \ldots f_1(\eta_k)}{f_0(\eta_1) \ldots f_0(\eta_k)}$$

образует мартингал. (Величины ξ_k, называемые *отношениями правдоподобия*, играют исключительно важную роль в математической статистике.)

§ 12. Марковские цепи. Эргодическая теорема. Строго марковское свойство

1. В рассмотренной выше схеме Бернулли с $\Omega = \{\omega : \omega = (x_1, \ldots, x_n),\ x_i = 0,1\}$ вероятность $p(\omega)$ каждого исхода ω задавалась формулой

$$p(\omega) = p(x_1) \ldots p(x_n), \tag{1}$$

где $p(x) = p^x q^{1-x}$. При этом условии случайные величины $\xi_1, \ldots, \xi_n$ с $\xi_i(\omega) = x_i$ оказывались *независимыми* и *одинаково* распределенными с

$$\mathsf{P}(\xi_1 = x) = \ldots = \mathsf{P}(\xi_n = x) = p(x), \qquad x = 0,1.$$

Если вместо (1) положить

$$p(\omega) = p_1(x_1) \ldots p_n(x_n),$$

где $p_i(x) = p_i^x(1-p_i)^{1-x}$, $0 \leqslant p_i \leqslant 1$, то тогда случайные величины $\xi_1, \ldots, \xi_n$ также будут *независимыми*, но уже, вообще говоря, *разнораспределенными:*

$$\mathsf{P}(\xi_1 = x) = p_1(x), \ \ldots, \ \mathsf{P}(\xi_n = x) = p_n(x).$$

Рассмотрим теперь одно обобщение этих схем, приводящее к *зависимым* случайным величинам, образующим так называемую цепь Маркова.

Будем предполагать, что

$$\Omega = \{\omega\colon \ \omega = (x_0, x_1, \ldots, x_n), \ x_i \in X\},$$

где X — некоторое конечное множество. Пусть заданы также неотрицательные функции $p_0(x)$, $p_1(x, y)$, $\ldots$, $p_n(x, y)$ такие, что

$$\begin{gathered}\sum_{x \in X} p_0(x) = 1, \\ \sum_{y \in X} p_k(x, y) = 1, \quad k = 1, \ldots, n; \ y \in X.\end{gathered} \tag{2}$$

Для каждого исхода $\omega = (x_0, x_1, \ldots, x_n)$ положим

$$p(\omega) = p_0(x_0)\, p_1(x_0, x_1) \ldots p_n(x_{n-1}, x_n). \tag{3}$$

Нетрудно проверить, что $\sum\limits_{\omega \in \Omega} p(\omega) = 1$ и, следовательно, набор этих чисел $p(\omega)$ вместе с пространством Ω и системой всех его подмножеств определяет некоторую вероятностную модель, которую принято называть *моделью испытаний, связанных в цепь Маркова.*

Введем в рассмотрение случайные величины $\xi_0, \xi_1, \ldots, \xi_n$ с $\xi_i(\omega) = x_i$. Простой подсчет показывает, что

$$\begin{gathered}\mathsf{P}(\xi_0 = a) = p_0(a), \\ \mathsf{P}(\xi_0 = a_0, \ \ldots, \ \xi_k = a_k) = p_0(a_0)\, p_1(a_0, a_1) \ldots p_k(a_{k-1}, a_k).\end{gathered} \tag{4}$$

Установим теперь справедливость следующего важного свойства условных вероятностей:

$$\mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_k = a_k, \ \ldots, \ \xi_0 = a_0\} = \mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_k = a_k\} \tag{5}$$

(в предположении $\mathsf{P}(\xi_k = a_k, \ \ldots, \ \xi_0 = a_0) > 0$).

В силу (4)

$$\begin{aligned}\mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_k = a_k, \ \ldots, \ \xi_0 = a_0\} &= \\ = \frac{\mathsf{P}\{\xi_{k+1} = a_{k+1}, \ \ldots, \ \xi_0 = a_0\}}{\mathsf{P}\{\xi_k = a_k, \ \ldots, \ \xi_0 = a_0\}} &= \frac{p_0(a_0)\, p_1(a_0, a_1) \ldots p_{k+1}(a_k, a_{k+1})}{p_0(a_0) \ldots p_k(a_{k-1}, a_k)} = \\ &= p_{k+1}(a_k, a_{k+1}).\end{aligned}$$

Аналогичным образом проверяется равенство

$$\mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_k = a_k\} = p_{k+1}(a_k, a_{k+1}), \tag{6}$$

что и доказывает свойство (5).

Пусть $\mathscr{D}_k^\xi = \mathscr{D}_{\xi_0, \ldots, \xi_k}$ — разбиение, порожденное величинами $\xi_0, \ldots, \xi_k$, и $\mathscr{B}_k^\xi = \alpha(\mathscr{D}_k^\xi)$.

Тогда в соответствии с обозначениями, введенными в § 8, из (5) следует, что

$$\mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \mathscr{B}_k^\xi\} = \mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_k\} \tag{7}$$

или

$$\mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_0, \ldots, \xi_k\} = \mathsf{P}\{\xi_{k+1} = a_{k+1} \mid \xi_k\}.$$

Если воспользоваться очевидным равенством

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

то из (7) получаем, что

$$\mathsf{P}\{\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid \mathscr{B}_k^\xi\} = \\ = \mathsf{P}\{\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid \xi_k\} \tag{8}$$

или

$$\mathsf{P}\{\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid \xi_0, \ldots, \xi_k\} = \\ = \mathsf{P}\{\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid \xi_k\}. \tag{9}$$

Это равенство допускает следующую наглядную интерпретацию. Будем трактовать ξ_k как положение частицы в «настоящем», $(\xi_0, \ldots, \xi_{k-1})$ — в «прошлом» и $(\xi_{k+1}, \ldots, \xi_n)$ — в «будущем». Тогда (9) означает, что при фиксированных «прошлом» $(\xi_0, \ldots, \xi_{k-1})$ и «настоящем» ξ_k «будущее» $(\xi_{k+1}, \ldots, \xi_n)$ зависит лишь от «настоящего» ξ_k и не зависит от того, каким способом частица попала в точку ξ_k, т. е. не зависит от «прошлого» $(\xi_0, \ldots, \xi_{k-1})$.

Пусть $\text{Б} = \{\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1}\}$, $\text{Н} = \{\xi_k = a_k\}$, $\text{П} = \{\xi_{k-1} = a_{k-1}, \ldots, \xi_0 = a_0\}$. Тогда из (9) следует, что

$$\mathsf{P}(\text{Б} \mid \text{НП}) = \mathsf{P}(\text{Б} \mid \text{Н}),$$

откуда легко находим, что

$$\mathsf{P}(\text{БП} \mid \text{Н}) = \mathsf{P}(\text{Б} \mid \text{Н})\,\mathsf{P}(\text{П} \mid \text{Н}). \tag{10}$$

Иначе говоря, из (7) следует, что при фиксированном «настоящем» Н «будущее» Б и «прошлое» П оказываются независимыми. Нетрудно показать, что справедливо и обратное: из выполнения (10) для любого $k = 0, 1, \ldots, n-1$ следует выполнение свойства (7) для всякого $k = 0, 1, \ldots, n-1$.

Свойство независимости «будущего» и «прошлого», или, что то же, независимость «будущего» от «прошлого» при фиксированном «настоящем» принято называть *марковским свойством*, а соответствующую последовательность случайных величин $\xi_0, \ldots, \xi_n$ — *марковской цепью*.

Таким образом, если вероятности $p(\omega)$ элементарных событий задаются формулой (3), то последовательность $\xi = (\xi_0, \ldots, \xi_n)$ с $\xi_i(\omega) = x_i$, будет образовывать марковскую цепь.

В этой связи понятно следующее

Определение. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — некоторое (конечное) вероятностное пространство и $\xi = (\xi_0, \ldots, \xi_n)$ — последовательность случайных величин со значениями в (конечном) множестве X. Если выполнено условие (7), то последовательность $\xi = (\xi_0, \ldots, \xi_n)$ называется (конечной) *марковской цепью*.

Множество X называется *фазовым пространством* или *пространством состояний* цепи. Набор вероятностей $(p_0(x))$, $x \in X$, с $p_0(x) = \mathsf{P}(\xi_0 = x)$ называют *начальным распределением*, а матрицу $\|p_k(x, y)\|$, $x, y \in X$, с $p_k(x, y) = \mathsf{P}(\xi_k = y \mid \xi_{k-1} = x)$ — *матрицей переходных вероятностей* (из состояний x в состояния y) в момент $k = 1, \ldots, n$.

В том случае, когда переходные вероятности $p_k(x, y)$ не зависят от k, $p_k(x, y) = p(x, y)$, последовательность $\xi = (\xi_0, \ldots, \xi_n)$ называется *однородной* марковской цепью с матрицей переходных вероятностей $\|p(x, y)\|$.

Заметим, что матрица $\|p(x, y)\|$ является *стохастической*: ее элементы неотрицательны и сумма элементов любой ее строки равна единице, $\sum\limits_y p(x, y) = 1$, $x \in X$.

Будем считать, что фазовое пространство X состоит из конечного множества целочисленных точек ($X = \{0, 1, \ldots, N\}$, $X = \{0, \pm 1, \ldots, \pm N\}$ и т. д.), и обозначать, согласно традиции, $p_i = p_0(i)$ и $p_{ij} = p(i, j)$.

Понятно, что свойства однородных марковских цепей полностью определяются начальными распределениями p_i и переходными вероятностями p_{ij}. В конкретных случаях для описания эволюции цепи вместо явного выписывания матрицы $\|p_{ij}\|$ используют (ориентированный) граф, вершинами которого являются состояния из X, а стрелка

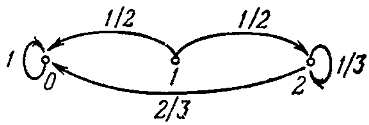
$$i \xrightarrow{p_{ij}} j,$$

идущая из состояния i в состояние j и с числом p_{ij} над ней, показывает, что из точки i возможен переход в точку j с вероятностью p_{ij}. В том случае, когда $p_{ij} = 0$, соответствующая стрелка не проводится.

Пример 1. Пусть $X = \{0, 1, 2\}$ и

$$\|p_{ij}\| = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{2}{3} & 0 & \frac{1}{3} \end{pmatrix}.$$

Этой матрице соответствует следующий граф:

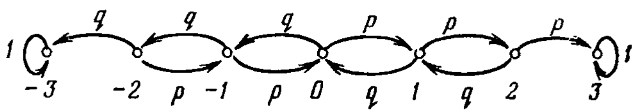


Отметим, что здесь состояние 0 называется «поглощающим»: если частица в него попала, то она в нем и остается, поскольку $p_{00} = 1$. Из состояния 1 частица с равными вероятностями переходит в соседние состояния 0 и 2, состояние 2 таково, что частица остается в нем с вероятностью 1/3 и переходит в состояние 0 с вероятностью 2/3.

Пример 2. Пусть $X = \{0, \pm 1, \ldots, \pm N\}$, $p_0 = 1$, $p_{NN} = p_{(-N)(-N)} = 1$ и для $|i| < N$

$$p_{ij} = \begin{cases} p, & j = i + 1, \\ q, & j = i - 1, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{11}$$

Переходы, соответствующие такой цепи, можно графически изобразить следующим образом ($N = 3$):



Эта цепь отвечает исследованной выше игре двух игроков, когда капитал каждого равен N и на каждом шаге первый игрок с вероятностью p выигрывает у второго $+1$ и проигрывает -1 с вероятностью q. Если трактовать состояние i как величину выигрыша первого игрока у второго, то достижение состояний N и $-N$ означает разорение второго и первого игроков соответственно.

В самом деле, если $\eta_1, \eta_2, \ldots, \eta_n$ — независимые бернуллиевские случайные величины с $\mathsf{P}(\eta_i = +1) = p$, $\mathsf{P}(\eta_i = -1) = q$, $S_0 = 0$ и $S_k = \eta_1 + \ldots + \eta_k$ — величина выигрыша первого игрока у второго, то последовательность $S_0, S_1, \ldots, S_n$ будет образовывать марковскую цепь с $p_0=1$ и матрицей переходных вероятностей (11), поскольку

$$\begin{aligned}\mathsf{P}\{S_{k+1}=j\mid S_k=i_k,\ S_{k-1}=i_{k-1},\ldots\}=\\ =\mathsf{P}\{S_k+\eta_{k+1}=j\mid S_k=i_k,\ S_{k-1}=i_{k-1},\ldots\}=\\ =\mathsf{P}\{S_k+\eta_{k+1}=j\mid S_k=i_k\}=\mathsf{P}\{\eta_{k+1}=j-i_k\}.\end{aligned}$$

Марковская цепь $S_0, S_1, \ldots, S_n$ имеет весьма простую структуру:

$$S_{k+1}=S_k+\eta_{k+1},\qquad 0\leqslant k\leqslant n-1,$$

где $\eta_1, \eta_2, \ldots, \eta_n$ — последовательность независимых случайных величин.

Те же рассуждения показывают, что если $\xi_0, \eta_1, \ldots, \eta_n$ — независимые случайные величины, то последовательность $\xi_0, \xi_1, \ldots$ $\ldots, \xi_n$ с

$$\xi_{k+1}=f_k(\xi_k,\ \eta_{k+1}),\qquad 0\leqslant k\leqslant n-1, \tag{12}$$

также образует марковскую цепь.

В этой связи полезно отметить, что так построенную марковскую цепь естественно рассматривать как вероятностный аналог (детерминированной) последовательности $x=(x_0, \ldots, x_n)$, управляемой рекуррентными соотношениями

$$x_{k+1}=f_k(x_k).$$

Приведем еще один пример марковской цепи типа (12), возникающей в задачах теории «очередей».

Пример 3. Пусть на стоянку такси в единичные моменты времени прибывают (по одной в каждый момент) машины. Если на стоянке нет ожидающих, то машина немедленно уезжает. Обозначим через η_k число ожидающих, приходящих в момент k на стоянку, и будем предполагать, что $\eta_1, \ldots, \eta_n$ — независимые случайные величины. Пусть ξ_k — длина очереди в момент k, $\xi_0=0$. Тогда, если $\xi_k=i$, то в следующий момент $k+1$ длина очереди ξ_{k+1} станет равной

$$j=\begin{cases}\eta_{k+1}, & \text{если } i=0,\\ i-1+\eta_{k+1}, & \text{если } i\geqslant 1.\end{cases}$$

Иначе говоря,

$$\xi_{k+1}=(\xi_k-1)^{+}+\eta_{k+1},\qquad 0\leqslant k\leqslant n-1,$$

где $a^{+}=\max(a,\ 0)$, и, значит, последовательность $\xi=(\xi_0, \ldots, \xi_n)$ образует цепь Маркова.

Пример 4. Этот пример относится к теории *ветвящихся процессов*. Под ветвящимся процессом с дискретным временем будем понимать последовательность случайных величин $\xi_0, \xi_1, \ldots, \xi_n$, где ξ_k интерпретируется как число частиц, существующих в момент времени k, а процесс гибели-размножения частиц происходит следующим образом: каждая частица независимо от других частиц и от «предыстории» процесса превращается в j частиц с вероятностями p_j, $j=0, 1, \ldots, M$.

Будем считать, что в начальный момент времени имеется всего лишь одна частица, $\xi_0=1$. Если в момент k было ξ_k частиц (с номерами $1, 2, \ldots, \xi_k$), то, согласно описанию, ξ_{k+1} представляется в виде случайного числа случайных величин:

$$\xi_{k+1}=\eta_1^{(k)}+\ldots+\eta_{\xi_k}^{(k)},$$

где $\eta_i^{(k)}$ — число частиц, произведенных частицей с номером i. Разумеется, если $\xi_k=0$, то и $\xi_{k+1}=0$. Считая, что все случайные величины $\eta_j^{(k)}$, $k\geqslant 0$, $j\geqslant 1$, независимы между собой, находим

$$\mathsf{P}\{\xi_{k+1}=i_{k+1} \mid \xi_k=i_k,\ \xi_{k-1}=i_{k-1}, \ldots\}=$$
$$=\mathsf{P}\{\xi_{k+1}=i_{k+1} \mid \xi_k=i_k\}=\mathsf{P}\left\{\eta_1^{(k)}+\ldots+\eta_{i_k}^{(k)}=i_{k+1}\right\}.$$

Отсюда видно, что последовательность $\xi_0, \xi_1, \ldots, \xi_n$ образует марковскую цепь.

Особый интерес представляет случай, когда каждая частица или погибает с вероятностью q, или превращается в две с вероятностью p, $p+q=1$. Для этого случая легко подсчитать, что

$$p_{ij}=\mathsf{P}\{\xi_{k+1}=j \mid \xi_k=i\}$$

задается формулой

$$p_{ij}=\begin{cases} C_i^{j/2}p^{j/2}q^{i-j/2}, & j=0, \ldots, 2i,\\ 0 & \text{в остальных случаях.}\end{cases}$$

2. Будем обозначать через $\xi=(\xi_k, \Pi, \mathbb{P})$ однородную марковскую цепь с векторами (строкой) начальных вероятностей $\Pi=(p_i)$ и матрицей переходных вероятностей $\mathbb{P}=\|p_{ij}\|$. Ясно, что

$$p_{ij}=\mathsf{P}\{\xi_1=j \mid \xi_0=i\}=\ldots=\mathsf{P}\{\xi_n=j \mid \xi_{n-1}=i\}.$$

Обозначим

$$p_{ij}^{(k)}=\mathsf{P}\{\xi_k=j \mid \xi_0=i\} \quad (=\mathsf{P}\{\xi_{k+l}=j \mid \xi_l=i\})$$

вероятность перехода за k шагов из состояния i в состояние j и

$$p_j^{(k)}=\mathsf{P}\{\xi_k=j\}$$

— вероятность нахождения частицы в момент времени k в точке j. Пусть также

$$\Pi^{(k)}=\|p_i^{(k)}\|, \quad \mathbb{P}^{(k)}=\|p_{ij}^{(k)}\|.$$

Покажем, что переходные вероятности $p_{ij}^{(k)}$ удовлетворяют *уравнению Колмогорова — Чэпмена*

$$p_{ij}^{(k+l)} = \sum_{\alpha} p_{i\alpha}^{(k)} p_{\alpha j}^{(l)}, \tag{13}$$

или в матричной форме

$$\mathbb{P}^{(k+l)} = \mathbb{P}^{(k)} \cdot \mathbb{P}^{(l)}. \tag{14}$$

Доказательство весьма просто: используя формулу полной вероятности и марковское свойство, получаем

$$p_{ij}^{(k+l)} = \mathsf{P}(\xi_{k+l} = j \mid \xi_0 = i) = \sum_{\alpha} \mathsf{P}(\xi_{k+l} = j,\ \xi_k = \alpha \mid \xi_0 = i) =$$
$$= \sum_{\alpha} \mathsf{P}(\xi_{k+l} = j \mid \xi_k = \alpha)\, \mathsf{P}(\xi_k = \alpha \mid \xi_0 = i) = \sum_{\alpha} p_{\alpha j}^{(l)} p_{i\alpha}^{(k)}.$$

Особо важны следующие два частных случая уравнений (13): *обратное уравнение*

$$p_{ij}^{(l+1)} = \sum_{\alpha} p_{i\alpha} p_{\alpha j}^{(l)} \tag{15}$$

и *прямое уравнение*

$$p_{ij}^{(k+1)} = \sum_{\alpha} p_{i\alpha}^{(k)} p_{\alpha j} \tag{16}$$

(см. рис. 22 и 23). В матричной форме прямые и обратные

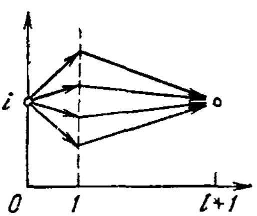


Рис. 22. К обратному уравнению.

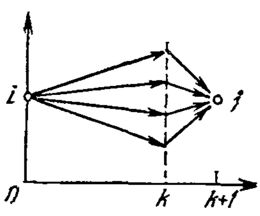


Рис. 23. К прямому уравнению.

уравнения записываются соответственно следующим образом:

$$\mathbb{P}^{(k+1)} = \mathbb{P}^{(k)} \cdot \mathbb{P}, \tag{17}$$
$$\mathbb{P}^{(k+1)} = \mathbb{P} \cdot \mathbb{P}^{(k)}. \tag{18}$$

Аналогично для (безусловных) вероятностей $p_j^{(k)}$ получаем, что

$$p_j^{(k+l)} = \sum_{\alpha} p_{\alpha}^{(k)} p_{\alpha j}^{(l)}, \tag{19}$$

или в матричной форме

$$\Pi^{(k+l)} = \Pi^{(k)} \cdot \mathbb{P}^{(l)}.$$

В частности,

$$\mathbb{\Pi}^{(k+1)} = \mathbb{\Pi}^{(k)} \cdot \mathbb{P}$$

(*прямое уравнение*) и

$$\mathbb{\Pi}^{(k+1)} = \mathbb{\Pi}^{(1)} \cdot \mathbb{P}^{(k)}$$

(*обратное уравнение*). Поскольку $\mathbb{P}^{(1)} = \mathbb{P}$, $\mathbb{\Pi}^{(1)} = \mathbb{\Pi}$, то из этих уравнений следует, что

$$\mathbb{P}^{(k)} = \mathbb{P}^k, \quad \mathbb{\Pi}^{(k)} = \mathbb{\Pi}^k.$$

Тем самым для однородных марковских цепей вероятности перехода за k шагов $p_{ij}^{(k)}$ и вероятности $p_j^{(k)}$ являются элементами k-х степеней матриц $\mathbb{P}$ и $\mathbb{\Pi}$, в связи с чем многие свойства этих цепей можно изучать методами матричного анализа.

Пример. Рассмотрим однородную марковскую цепь с двумя состояниями 0 и 1 и матрицей

$$\mathbb{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

Нетрудно подсчитать, что

$$\mathbb{P}^2 = \begin{pmatrix} p_{00}^2 + p_{01}p_{10} & p_{01}(p_{00} + p_{11}) \\ p_{10}(p_{00} + p_{11}) & p_{11}^2 + p_{01}p_{10} \end{pmatrix}$$

и (по индукции)

$$\mathbb{P}^n = \frac{1}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{11} & 1 - p_{00} \\ 1 - p_{11} & 1 - p_{00} \end{pmatrix} + \\ + \frac{(p_{00} + p_{11} - 1)^n}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{00} & -(1 - p_{00}) \\ -(1 - p_{11}) & 1 - p_{11} \end{pmatrix}$$

(в предположении, что $|p_{00} + p_{11} - 1| < 1$).

Отсюда видно, что если элементы матрицы $\mathbb{P}$ таковы, что $|p_{00} + p_{11} - 1| < 1$ (в частности, если все вероятности перехода p_{ij} положительны), то при $n \to \infty$

$$\mathbb{P}^n \to \frac{1}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{11} & 1 - p_{00} \\ 1 - p_{11} & 1 - p_{00} \end{pmatrix}, \tag{20}$$

и, значит,

$$\lim_n p_{i0}^{(n)} = \frac{1 - p_{11}}{2 - p_{00} - p_{11}}, \quad \lim_n p_{i1}^{(n)} = \frac{1 - p_{00}}{2 - p_{00} - p_{11}}.$$

Таким образом, если $|p_{00} + p_{11} - 1| < 1$, то поведение рассматриваемой марковской цепи подчиняется следующей закономерности: влияние начального состояния на вероятность нахождения частицы в том или ином состоянии исчезает с ростом времени ($p_{ij}^{(n)}$ сходятся к предельным значениям π_j, *не зависящим* от i и образующим распределение вероятностей $\pi_0 \geqslant 0$, $\pi_1 \geqslant 0$, $\pi_0 + \pi_1 = 1$); если к тому же все элементы $p_{ij}>0$, то тогда предельные значения $\pi_0>0$, $\pi_1>0$.

Следующая теорема описывает широкий класс марковских цепей, обладающих так называемым свойством *эргодичности*: пределы $\pi_j=\lim\limits_n p_{ij}^{(n)}$ не только существуют, не зависят от i, образуют распределение вероятностей $\left(\pi_j\geqslant 0,\ \sum\limits_j \pi_j=1\right)$, но и таковы, что $\pi_j>0$ при всех j (такие распределения π_j называются *эргодическими*).

Теорема 1 (эргодическая теорема). Пусть $\mathbb{P}=\|p_{ij}\|$ — *матрица переходных вероятностей марковской цепи с конечным множеством состояний* $X=\{1, 2, \ldots, N\}$.

a) *Если найдется* n_0 *такое, что*

$$\min_{i,j} p_{ij}^{(n_0)}>0, \tag{21}$$

то существуют числа $\pi_1, \ldots, \pi_N$ *такие, что*

$$\pi_j>0, \quad \sum_j \pi_j=1 \tag{22}$$

и для любого $i\in X$

$$p_{ij}^{(n)}\to\pi_j, \qquad n\to\infty. \tag{23}$$

b) *Обратно, если существуют числа* $\pi_1, \ldots, \pi_N$, *удовлетворяющие условиям* (22) *и* (23), *то найдется* n_0 *такое, что выполнено условие* (21).

c) *Числа* $(\pi_1, \ldots, \pi_N)$ *удовлетворяют системе уравнений*

$$\pi_j=\sum_\alpha \pi_\alpha p_{\alpha j}, \qquad j=1, \ldots, N. \tag{24}$$

Доказательство. a) Обозначим

$$m_j^{(n)}=\min_i p_{ij}^{(n)}, \quad M_j^{(n)}=\max_i p_{ij}^{(n)}.$$

Поскольку

$$p_{ij}^{(n+1)}=\sum_\alpha p_{i\alpha}p_{\alpha j}^{(n)}, \tag{25}$$

то

$$m_j^{(n+1)}=\min_i p_{ij}^{(n+1)}=\min_i \sum_\alpha p_{i\alpha}p_{\alpha j}^{(n)}\geqslant\min_i\sum_\alpha p_{i\alpha}\min_\alpha p_{\alpha j}^{(n)}=m_j^{(n)},$$

откуда $m_j^{(n)}\leqslant m_j^{(n+1)}$ и аналогично $M_j^{(n)}\geqslant M_j^{(n+1)}$. Поэтому для доказательства утверждения (23) достаточно показать, что

$$M_j^{(n)}-m_j^{(n)}\to 0, \quad n\to\infty, \quad j=1, \ldots, N.$$

Пусть $\varepsilon = \min\limits_{i,j} p_{ij}^{(n_0)} > 0$. Тогда

$$p_{ij}^{(n_0+n)} = \sum_\alpha p_{i\alpha}^{(n_0)} p_{\alpha j}^{(n)} = \sum_\alpha \left[p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}\right] p_{\alpha j}^{(n)} + \varepsilon \sum_\alpha p_{j\alpha}^{(n)} p_{\alpha j}^{(n)} =$$

$$= \sum_\alpha \left[p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}\right] p_{\alpha j}^{(n)} + \varepsilon p_{jj}^{(2n)}.$$

Но $p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)} \geqslant 0$, поэтому

$$p_{ij}^{(n_0+n)} \geqslant m_j^{(n)} \cdot \sum_\alpha \left[p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}\right] + \varepsilon p_{jj}^{(2n)} = m_j^{(n)} (1-\varepsilon) + \varepsilon p_{jj}^{(2n)},$$

и, значит,

$$m_j^{(n_0+n)} \geqslant m_j^{(n)} (1-\varepsilon) + \varepsilon p_{jj}^{(2n)}.$$

Аналогичным образом

$$M_j^{(n_0+n)} \leqslant M_j^{(n)} (1-\varepsilon) + \varepsilon p_{jj}^{(2n)}.$$

Объединяя эти неравенства, получаем

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \leqslant \left(M_j^{(n)} - m_j^{(n)}\right) \cdot (1-\varepsilon)$$

и, следовательно,

$$M_j^{(kn_0+n)} - m_j^{(kn_0+n)} \leqslant \left(M_j^{(n)} - m_j^{(n)}\right) (1-\varepsilon)^k \downarrow 0, \qquad k \to \infty.$$

Итак, по некоторой подпоследовательности $\{n_\beta\}$ $M_j^{(n_\beta)} - m_j^{(n_\beta)} \to 0$, $n_\beta \to \infty$. Но разность $M_j^{(n)} - m_j^{(n)}$ монотонна по n, а значит, $M_j^{(n)} - m_j^{(n)} \to 0$, $n \to \infty$.

Если обозначить $\pi_j = \lim\limits_n m_j^{(n)}$, то из полученных оценок следует, что для $n \geqslant n_0$

$$\left|p_{ij}^{(n)} - \pi_j\right| \leqslant M_j^{(n)} - m_j^{(n)} \leqslant (1-\varepsilon)^{[n/n_0]-1},$$

т. е. сходимость $p_{ij}^{(n)}$ к предельным значениям π_j происходит с геометрической скоростью.

Ясно также, что $m_j^{(n)} \geqslant m_j^{(n_0)} \geqslant \varepsilon > 0$, $n \geqslant n_0$, и, значит, $\pi_j > 0$.

b) Условие (21) непосредственно следует из (23), поскольку число состояний конечно и $\pi_j > 0$.

c) Уравнения (24) вытекают из (23) и (25).

Теорема доказана.

4. Система уравнений (24) играет большую роль в теории марковских цепей. Всякое ее неотрицательное решение $(\pi_1, \ldots, \pi_N)$, удовлетворяющее условию $\sum\limits_\alpha \pi_\alpha = 1$, принято называть *стационарным* или *инвариантным*, распределением вероятностей для марковской цепи с матрицей переходных вероятностей $\|p_{ij}\|$. Объяснение этого названия состоит в следующем.

Возьмем распределение $(\pi_1 \ldots, \pi_N)$ в качестве начального, $p_j = \pi_j$. Тогда

$$p_j^{(1)} = \sum_\alpha \pi_\alpha p_{\alpha j} = \pi_j$$

и вообще $p_j^{(n)} = \pi_j$. Иначе говоря, если в качестве начального распределения взять $(\pi_1, \ldots, \pi_N)$, то это распределение не будет изменяться со временем, т. е. для любого k

$$\mathsf{P}(\xi_k = j) = \mathsf{P}(\xi_0 = j), \quad j = 1, \ldots, N.$$

Более того, с таким начальным распределением марковская цепь $\xi = (\xi, \mathbb{П}, \mathbb{P})$ будет *стационарной*: совместное распределение вектора $(\xi_k, \xi_{k+1}, \ldots, \xi_{k+l})$ не зависит от k для любого l (предполагается, что $k + l \leqslant n$).

Условие (21) гарантирует как существование пределов $\pi_j = \lim\limits_n p_{ij}^{(n)}$, не зависящих от i, так и существование эргодического распределения, т. е. распределения с $\pi_j > 0$. Распределение $(\pi_1, \ldots, \pi_N)$ оказывается также и *стационарным* распределением. Покажем сейчас, что набор $(\pi_1, \ldots, \pi_N)$ является *единственным* стационарным распределением.

В самом деле, пусть $(\tilde{\pi}_1, \ldots, \tilde{\pi}_N)$ — еще одно стационарное распределение. Тогда

$$\tilde{\pi}_j = \sum_\alpha \tilde{\pi}_\alpha p_{\alpha j} = \ldots = \sum_\alpha \tilde{\pi}_\alpha p_{\alpha j}^{(n)},$$

и поскольку $p_{\alpha j}^{(n)} \to \pi_j$, то

$$\tilde{\pi}_j = \sum_\alpha (\tilde{\pi}_\alpha \cdot \pi_j) = \pi_j.$$

В связи с этими результатами возникают интересные и важные вопросы о достаточных, необходимых, а также необходимых и достаточных условиях, при которых: (A) *существуют* пределы $\pi_j = \lim\limits_n p_{ij}^{(n)}$, не зависящие от i; (B) пределы $(\pi_1, \ldots, \pi_N)$ образуют *распределение вероятностей*; (C) пределы $(\pi_1, \ldots, \pi_N)$ образуют *эргодическое* распределение вероятностей; (D) *существует* и при том *единственное* стационарное распределение вероятностей.

Все эти вопросы будут детально исследованы в гл. VIII для марковских цепей не только с конечным, но и счетным множеством состояний.

Отметим, что стационарное распределение вероятностей (и к тому же единственное) может существовать и для неэргодических цепей.

Действительно, если

$$\mathbb{P}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

то

$$\mathbb{P}^{2n}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \mathbb{P}^{2n+1}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},$$

и, следовательно, пределы $\lim\limits_n p_{ij}^{(n)}$ не существуют. В то же самое время система

$$\pi_j=\sum_\alpha \pi_\alpha p_{\alpha j},\qquad j=1,\ 2,$$

превращается в систему

$$\pi_1=\pi_2,$$
$$\pi_2=\pi_1,$$

единственное решение $(\pi_1,\ \pi_2)$ которой, удовлетворяющее условию $\pi_1+\pi_2=1$, есть $(^1/_2,\ ^1/_2)$.

Отметим также, что для рассмотренного выше примера система (24) имеет вид

$$\pi_0=\pi_0 p_{00}+\pi_1 p_{10},$$
$$\pi_1=\pi_0 p_{01}+\pi_1 p_{11},$$

откуда, учитывая условие $\pi_0+\pi_1=1$, находим, что единственное стационарное распределение $(\pi_0,\ \pi_1)$ совпадает с уже найденным:

$$\pi_0=\frac{1-p_{11}}{2-p_{00}-p_{11}},\qquad \pi_1=\frac{1-p_{00}}{2-p_{00}-p_{11}}.$$

Рассмотрим теперь некоторые следствия, вытекающие из эргодической теоремы.

Пусть A — некоторая группа состояний, $A\subseteq X$ и

$$I_A(x)=\begin{cases}1, & x\in A,\\ 0, & x\notin A.\end{cases}$$

Рассмотрим величину

$$\nu_A(n)=\frac{I_A(\xi_0)+\ldots+I_A(\xi_n)}{n+1}$$

— долю времени, проводимого частицей во множестве A. Поскольку

$$\mathsf{M}[I_A(\xi_k)\,|\,\xi_0=i]=\mathsf{P}(\xi_k\in A\,|\,\xi_0=i)=\sum_{j\in A}p_{ij}^{(k)}\left(=p_i^{(k)}(A)\right),$$

то

$$\mathsf{M}[\nu_A(n)\,|\,\xi_0=i]=\frac{1}{n+1}\sum_{k=0}^{n}p_i^{(k)}(A)$$

и, в частности,

$$\mathsf{M}[\nu_{\{j\}}(n) \mid \xi_0 = i] = \frac{1}{n+1} \sum_{k=0}^{n} p_{ij}^{(k)}.$$

Из анализа известно (см. также лемму 1 в § 3 гл. IV), что если последовательность $a_n \to a$, то $\frac{a_0 + \ldots + a_n}{n+1} \to a$, $n \to \infty$. Поэтому если $p_{ij}^{(k)} \to \pi_j$, $k \to \infty$, то

$$\mathsf{M}\nu_{\{j\}}(n) \to \pi_j, \quad \mathsf{M}\nu_A(n) \to \pi_A, \text{ где } \pi_A = \sum_{j \in A} \pi_j.$$

Для эргодических цепей на самом деле можно доказать большее, а именно, что для величин $I_A(\xi_0), \ldots, I_A(\xi_n), \ldots$ справедлив.

Закон больших чисел. *Если $\xi_0, \xi_1, \ldots$ — конечная эргодическая марковская цепь, то для всякого $\varepsilon > 0$ и произвольного начального распределения*

$$\mathsf{P}\{|\nu_A(n) - \pi_A| > \varepsilon\} \to 0, \quad n \to \infty. \tag{26}$$

Прежде чем переходить к доказательству, заметим, что непосредственное применение результатов § 5 к бернуллиевским величинам $I_A(\xi_0), \ldots, I_A(\xi_n), \ldots$ невозможно, поскольку они, вообще говоря, являются зависимыми. Однако доказательство можно провести по тому же пути, что и в случае независимых величин, если снова воспользоваться неравенством Чебышева и тем обстоятельством, что для эргодических цепей с конечным числом состояний найдется такое $0 < \rho < 1$, что

$$|p_{ij}^{(n)} - \pi_j| \leqslant C \cdot \rho^n. \tag{27}$$

Рассмотрим состояния i и j (они могут и совпадать) и покажем, что ($\varepsilon > 0$)

$$\mathsf{P}\{|\nu_{\{j\}}(n) - \pi_j| > \varepsilon \mid \xi_0 = i\} \to 0, \quad n \to \infty. \tag{28}$$

В силу неравенства Чебышева

$$\mathsf{P}\{|\nu_{\{j\}}(n) - \pi_j| > \varepsilon \mid \xi_0 = i\} \leqslant \frac{\mathsf{M}\{|\nu_{\{j\}}(n) - \pi_j|^2 \mid \xi_0 = i\}}{\varepsilon^2}.$$

Поэтому надо лишь показать, что

$$\mathsf{M}\{|\nu_{\{j\}}(n) - \pi_j|^2 \mid \xi_0 = i\} \to 0, \quad n \to \infty.$$

Простой подсчет показывает, что

$$\mathsf{M}\{|\nu_{\{j\}}(n) - \pi_j|^2 \mid \xi_0 = i\} = \frac{1}{(n+1)^2} \cdot \mathsf{M}\left\{\sum_{k=0}^{n} [I_{\{j\}}(\xi_k) - \pi_j]^2 \,\middle|\, \xi_0 = i\right\} =$$

где

$$m_{ij}^{(k,\,l)}=\mathsf{M}\{[I_{\{j\}}(\xi_k)\,I_{\{j\}}(\xi_l)]\,|\,\xi_0=i\}-$$
$$-\pi_j\cdot\mathsf{M}[I_{\{j\}}(\xi_k)\,|\,\xi_0=i]-\pi_j\cdot\mathsf{M}[I_{\{j\}}(\xi_l)\,|\,\xi_0=i]+\pi_j^2=$$
$$=p_{ij}^{(s)}\cdot p_{jj}^{(t)}-\pi_j\cdot p_{ij}^{(k)}-\pi_j\cdot p_{ij}^{(l)}+\pi_j^2,$$
$$s=\min(k,\ l)\ \text{и}\ t=|k-l|.$$

В силу (27)

$$p_{ij}^{(n)}=\pi_j+\varepsilon_{ij}^{(n)},\qquad |\varepsilon_{ij}^{(n)}|\leqslant C\rho^n.$$

Поэтому

$$|m_{ij}^{(k,\,l)}|\leqslant C_1[\rho^s+\rho^t+\rho^k+\rho^l],$$

где C_1 — некоторая постоянная. Следовательно,

$$\frac{1}{(n+1)^2}\sum_{k=0}^{n}\sum_{l=0}^{n}m_{ij}^{(k,\,l)}\leqslant\frac{C_1}{(n+1)^2}\sum_{k=0}^{n}\sum_{l=0}^{n}[\rho^s+\rho^t+\rho^k+\rho^l]\leqslant$$
$$\leqslant\frac{4C_1}{(n+1)^2}\cdot\frac{2(n+1)}{1-\rho}=\frac{8C_1}{(n+1)(1-\rho)}\to 0,\qquad n\to\infty,$$

откуда и следует справедливость соотношения (28), из которого очевидным образом вытекает требуемое соотношение (26).

5. В § 9 для случайного блуждания $S_0,\ S_1,\ \ldots$, порожденного схемой Бернулли, были выведены рекуррентные уравнения для вероятностей и математических ожиданий времени выхода на ту или иную границу. Аналогичные уравнения сейчас будут выведены и для марковских цепей.

Пусть $\xi=(\xi_0,\ \ldots,\ \xi_n)$ — марковская цепь с матрицей переходных вероятностей $\|p_{ij}\|$ и фазовым пространством $X=\{0,\ \pm1,\ \ldots$ $\ldots,\ \pm N\}$. Пусть A и B — два целых числа, $-N\leqslant A\leqslant 0\leqslant B\leqslant N$ и $x\in X$. Обозначим через $\mathscr{B}_{k+1}$ множество тех траекторий $(x_0,\ x_1,\ \ldots,\ x_k)$, $x_i\in X$, которые впервые выходят из интервала $(A,\ B)$ через верхнюю границу, т. е. покидают множество $(A,\ B)$, попадая в множество $(B,\ B+1,\ \ldots,\ N)$.

Положим для $A\leqslant x\leqslant B$

$$\beta_k(x)=\mathsf{P}\{(\xi_0,\ \ldots,\ \xi_k)\in\mathscr{B}_{k+1}\,|\,\xi_0=x\}.$$

С целью отыскания этих вероятностей (первого выхода марковской цепи из множества $(A,\ B)$ через верхнюю границу) воспользуемся методом, примененным при выводе обратных уравнений.

Имеем

$$\beta_k(x)=\mathsf{P}\{(\xi_0,\ \ldots,\ \xi_k)\in\mathscr{B}_{k+1}\,|\,\xi_0=x\}=$$
$$=\sum_{y}p_{xy}\cdot\mathsf{P}\{(\xi_0,\ \ldots,\ \xi_k)\in\mathscr{B}_{k+1}\,|\,\xi_0=x,\ \xi_1=y\},$$

где, как нетрудно убедиться, опираясь на марковское свойство и однородность цепи, что

$$\begin{aligned}\mathsf{P}\{(\xi_0, \ldots, \xi_k) \in \mathscr{B}_{k+1} \mid \xi_0 = x,\ \xi_1 = y\} &= \\ = \mathsf{P}\{(x,\ y,\ \xi_2, \ldots, \xi_k) \in \mathscr{B}_{k+1} \mid \xi_0 = x,\ \xi_1 = y\} &= \\ = \mathsf{P}\{(y,\ \xi_2, \ldots, \xi_k) \in \mathscr{B}_k \mid \xi_1 = y\} &= \\ = \mathsf{P}\{(y,\ \xi_1, \ldots, \xi_{k-1}) \in \mathscr{B}_k \mid \xi_0 = y\} &= \beta_{k-1}(y).\end{aligned}$$

Поэтому для $A < x < B$ и $1 \leqslant k \leqslant n$

$$\beta_k(x) = \sum_y p_{xy}\beta_{k-1}(y).$$

При этом ясно, что

$$\beta_k(x) = 1, \quad x = B,\ B+1, \ldots, N$$

и

$$\beta_k(x) = 0, \quad x = -N, \ldots, A.$$

Аналогичным образом выводятся и уравнения для $\alpha_k(x)$ — вероятностей первого выхода из интервала (A, B) через нижнюю границу.

Пусть $\tau_k = \min\{0 \leqslant l \leqslant k\colon\ \xi_l \notin (A,\ B)\}$, причем $\tau_k = k$, если множество $\{\cdot\} = \varnothing$. Тогда тот же самый метод, примененный к $m_k(x) = \mathsf{M}(\tau_k \mid \xi_0 = x)$, приводит к следующим рекуррентным уравнениям:

$$m_k(x) = 1 + \sum_y m_{k-1}(y)\,p_{xy}$$

(здесь $1 \leqslant k \leqslant n$, $A < x < B$). При этом

$$m_k(x) = 0, \quad x \notin (A,\ B).$$

Понятно, что если матрица переходных вероятностей задается формулой (11), то уравнения для $\alpha_k(x)$, $\beta_k(x)$ и $m_k(x)$ превращаются в соответствующие уравнения из § 9, где они получены, по существу, тем же самым методом, что и здесь.

Наиболее интересны применения выведенных уравнений в предельном случае, когда блуждание осуществляется неограниченно во времени. Так же, как и в § 9, соответствующие уравнения можно получить формальным предельным переходом из выведенных выше уравнений, полагая $k \to \infty$.

Для примера рассмотрим марковскую цепь с состояниями $\{0, 1, \ldots, B\}$ и переходными вероятностями

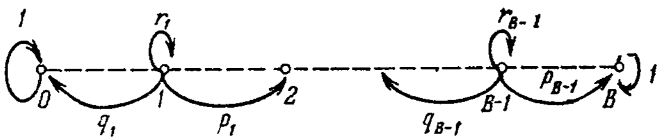
$$p_{00} = 1, \quad p_{BB} = 1$$

и для $1 \leqslant i \leqslant B-1$

$$p_{ij} = \begin{cases} p_i > 0, & j = i+1, \\ r_i, & j = i, \\ q_i > 0, & j = i-1, \end{cases}$$

где $p_i + r_i + q_i = 1$.

Этой цепи соответствует граф



Отсюда видно, что состояния 0 и B являются «поглощающими», в любом же другом состоянии i частица остается с вероятностью r_i, переходит на единицу вправо с вероятностью p_i и влево с вероятностью q_i.

Найдем $\alpha(x) = \lim\limits_{k \to \infty} \alpha_k(x)$ — предельную вероятность того, что частица, выходящая из точки x, достигнет нулевого состояния раньше, чем состояния B. Предельным переходом при $k \to \infty$ в уравнениях для $\alpha_k(x)$ получим, что для $0 < j < B$

$$\alpha(j) = q_j \alpha(j-1) + r_j \alpha(j) + p_j \alpha(j+1)$$

с граничными условиями

$$\alpha(0) = 1, \quad \alpha(B) = 0.$$

Поскольку $r_j = 1 - q_j - p_j$, то

$$p_j(\alpha(j+1) - \alpha(j)) = q_j(\alpha(j) - \alpha(j-1))$$

и, следовательно,

$$\alpha(j+1) - \alpha(j) = \rho_j(\alpha(1) - 1),$$

где

$$\rho_j = \frac{q_1 \ldots q_j}{p_1 \ldots p_j}, \quad \rho_0 = 1.$$

Но

$$\alpha(j+1) - 1 = \sum_{i=0}^{j} (\alpha(i+1) - \alpha(i)).$$

Поэтому

$$\alpha(j+1)-1=(\alpha(1)-1)\cdot\sum_{i=1}^{j}\rho_i.$$

Если $j=B-1$, то $\alpha(j+1)=\alpha(B)=0$, и, значит,

$$\alpha(1)-1=-\frac{1}{\sum_{i=1}^{B-1}\rho_i},$$

откуда

$$\alpha(1)=\frac{\sum_{i=1}^{B-1}\rho_i}{\sum_{i=0}^{B-1}\rho_i} \quad \text{и} \quad \alpha(j)=\frac{\sum_{i=j}^{B-1}\rho_i}{\sum_{i=1}^{B-1}\rho_i}, \qquad j=1,\ldots,B.$$

(Ср. с соответствующими результатами § 9.)

Пусть теперь $m(x)=\lim_k m_k(x)$ — предельное значение среднего времени блуждания до попадания в одно из состояний 0 или B. Тогда $m(0)=m(B)=0$,

$$m(x)=1+\sum_y m(y)\,p_{xy}$$

и, следовательно, для рассматриваемого примера

$$m(j)=1+q_j m(j-1)+r_j m(j)+p_j m(j+1)$$

для всех $j=1,\ldots,B-1$. Чтобы найти $m(j)$, обозначим

$$M(j)=m(j)-m(j-1), \qquad j=0,\ 1,\ \ldots,\ B.$$

Тогда

$$p_j M(j+1)=q_j M(j)-1, \qquad j=1,\ldots,B-1,$$

и последовательно находим, что

$$M(j+1)=\rho_j M(1)-R_j,$$

где

$$\rho_j=\frac{q_1\ldots q_j}{p_1\ldots p_j}, \qquad R_j=\frac{1}{p_j}\left[1+\frac{q_j}{p_{j-1}}+\ldots+\frac{q_j\ldots q_2}{p_j\ldots p_1}\right].$$

Поэтому

$$m(i)=m(j)-m(0)=\sum_{i=0}^{j-1}M(i+1)=$$

$$=\sum_{i=0}^{j-1}(\rho_i m(1)-R_i)=m(1)\sum_{i=0}^{j-1}\rho_i-\sum_{i=0}^{j-1}R_i.$$

Осталось лишь найти $m(1)$. Но $m(B)=0$, значит,

$$m(1)=\frac{\sum_{i=0}^{B-1} R_i}{\sum_{i=0}^{B-1} \rho_i},$$

и для $1<j\leqslant B$

$$m(j)=\sum_{i=0}^{j-1}\rho_i\cdot\frac{\sum_{i=0}^{B-1} R_i}{\sum_{i=0}^{B-1} \rho_i}-\sum_{i=0}^{j-1} R_i.$$

(Ср. с соответствующими результатами из § 9, полученными там для случая $r_i=0$, $p_i=p$, $q_i=q$.)

6. В этом пункте будет рассмотрено одно усиление марковского свойства (8), заключающееся в том, что оно остается справедливым при замене момента времени k на случайный момент (см. далее теорему 2). Важность этого так называемого *строго марковского свойства* будет проиллюстрирована, в частности, на примере вывода рекуррентных соотношений (38), играющих существенную роль для классификации состояний марковских цепей (гл. VIII).

Пусть $\xi=(\xi_0, \ldots, \xi_n)$ — однородная марковская цепь с матрицей переходных вероятностей $\|p_{ij}\|$, $\mathscr{D}^\xi=(\mathscr{D}_k^\xi)_{0\leqslant k\leqslant n}$ — система разбиений, $\mathscr{D}_k^\xi=\mathscr{D}_{\xi_0,\ldots,\xi_k}$. Через $\mathscr{B}_k^\xi$ будем обозначать алгебру $\alpha(\mathscr{D}_k^\xi)$, порожденную разбиением $\mathscr{D}_k^\xi$.

Придадим прежде всего марковскому свойству (8) несколько иную форму. Пусть $B\in\mathscr{B}_k^\xi$. Покажем, что тогда

$$\begin{aligned}\mathsf{P}\{\xi_n=a_n, \ldots, \xi_{k+1}=a_{k+1}\,|\,B\cap(\xi_k=a_k)\}&=\\ =\mathsf{P}\{\xi_n=a_n, \ldots, \xi_{k+1}=a_{k+1}\,|\,\xi_k=a_k\}&\end{aligned}\tag{29}$$

(предполагается, что $\mathsf{P}\{B\cap(\xi_k=a_k)\}>0$). Действительно, множество B можно представить в виде

$$B={\sum}^*\{\xi_0=a_0^*, \ldots, \xi_k=a_k^*\},$$

где суммирование $\sum^*$ распространяется по некоторым наборам $(a_0^*, \ldots, a_k^*)$. Поэтому

$$\begin{aligned}\mathsf{P}\{\xi_n=a_n, \ldots, \xi_{k+1}=a_{k+1}\,|\,B\cap(\xi_k=a_k)\}&=\\ =\frac{\mathsf{P}\{(\xi_n=a_n, \ldots, \xi_k=a_k)\cap B\}}{\mathsf{P}\{(\xi_k=a_k)\cap B\}}&=\\ =\frac{\sum^*\mathsf{P}\{(\xi_n=a_n, \ldots, \xi_k=a_k)\cap(\xi_0=a_0^*, \ldots, \xi_k=a_k^*)\}}{\mathsf{P}\{(\xi_k=a_k)\cap B\}}&.\end{aligned}\tag{30}$$

Но в силу марковского свойства

$$\mathsf{P}\{(\xi_n=a_n,\ \ldots,\ \xi_k=a_k)\cap(\xi_0=a_0^*,\ \ldots,\ \xi_k=a_k^*)\}=$$

$$=\begin{cases}\mathsf{P}\{\xi_n=a_n,\ \ldots,\ \xi_{k+1}=a_{k+1}\mid\xi_0=a_0^*,\ \ldots,\ \xi_k=a_k^*\}\times\\ \qquad\times\mathsf{P}\{\xi_0=a_0^*,\ \ldots,\ \xi_k=a_k^*\},\ \textit{если}\ a_k=a_k^*,\\ 0,\ \textit{если}\ a_k\neq a_k^*,\end{cases}$$

$$=\begin{cases}\mathsf{P}\{\xi_n=a_n,\ \ldots,\ \xi_{k+1}=a_{k+1}\mid\xi_k=a_k\}\,\mathsf{P}\{\xi_0=a_0^*,\ \ldots\\ \qquad\ldots,\ \xi_k=a_k^*\},\ \textit{если}\ a_k=a_k^*,\\ 0,\ \textit{если}\ a_k\neq a_k^*,\end{cases}$$

$$=\begin{cases}\mathsf{P}\{\xi_n=a_n,\ldots,\xi_{k+1}=a_{k+1}\mid\xi_k=a_k\}\,\mathsf{P}\{(\xi_k=a_k)\cap B\},\\ \qquad\qquad \textit{если}\ a_k=a_k^*,\\ 0,\ \textit{если}\ a_k\neq a_k^*.\end{cases}$$

Тем самым сумма $\sum^*$ в (30) равна

$$\mathsf{P}\{\xi_n=a_n,\ \ldots,\ \xi_{k+1}=a_{k+1}\mid\xi_k=a_k\}\,\mathsf{P}\{(\xi_k=a_k)\cap B\},$$

что и доказывает формулу (29).

Пусть τ — момент остановки (относительно системы разбиений $\mathscr{D}^\xi=(\mathscr{D}_k^\xi)_{0\leqslant k\leqslant n}$; см. определение 2 в § 11).

Определение. Будем говорить, что множество B из алгебры $\mathscr{B}_n^\xi$ принадлежит системе множеств $\mathscr{B}_\tau^\xi$, если для каждого $0\leqslant k\leqslant n$

$$B\cap\{\tau=k\}\in\mathscr{B}_k^\xi. \tag{31}$$

Нетрудно проверить, что совокупность таких множеств B образует алгебру (называемую алгеброй событий, наблюдаемых до момента τ).

Теорема 2. *Пусть* $\xi=(\xi_0,\ \ldots,\ \xi_n)$ — *однородная марковская цепь с матрицей переходных вероятностей* $\|p_{ij}\|$, τ — *момент остановки (относительно* $\mathscr{D}^\xi$), $B\in\mathscr{B}_\tau^\xi$ *и* $A=\{\omega\colon\ \tau+l\leqslant n\}$. *Тогда, если* $\mathsf{P}\{A\cap B\cap(\xi_\tau=a_0)\}>0$, *то*

$$\mathsf{P}\{\xi_{\tau+l}=a_l,\ \ldots,\ \xi_{\tau+1}=a_1\mid A\cap B\cap(\xi_\tau=a_0)\}=$$
$$=\mathsf{P}\{\xi_{\tau+l}=a_l,\ \ldots,\ \xi_{\tau+1}=a_1\mid A\cap(\xi_\tau=a_0)\}, \tag{32}$$

и если $\mathsf{P}\{A\cap(\xi_\tau=a_0)\}>0$, *то*

$$\mathsf{P}\{\xi_{\tau+l}=a_l,\ \ldots,\ \xi_{\tau+1}=a_1\mid A\cap(\xi_\tau=a_0)\}=p_{a_0a_1}\ \ldots\ p_{a_{l-1}a_l}. \tag{33}$$

Доказательство проведем для простоты лишь в случае $l=1$. Поскольку $B\cap(\tau=k)\in\mathscr{B}_k^{\xi}$, то, согласно (29),

$$\begin{aligned}
&\mathsf{P}\{\xi_{\tau+1}=a_1,\ A\cap B\cap(\xi_\tau=a_0)\}=\\
&=\sum_{k\leqslant n-1}\mathsf{P}\{\xi_{k+1}=a_1,\ \xi_k=a_0,\ \tau=k,\ B\}=\\
&=\sum_{k\leqslant n-1}\mathsf{P}\{\xi_{k+1}=a_1\mid\xi_k=a_0,\ \tau=k,\ B\}\,\mathsf{P}\{\xi_k=a_0,\ \tau=k,\ B\}=\\
&=\sum_{k\leqslant n-1}\mathsf{P}\{\xi_{k+1}=a_1\mid\xi_k=a_0\}\,\mathsf{P}\{\xi_k=a_0,\ \tau=k,\ B\}=\\
&=p_{a_0a_1}\cdot\sum_{k\leqslant n-1}\mathsf{P}\{\xi_k=a_0,\ \tau=k,\ B\}=p_{a_0a_1}\cdot\mathsf{P}\{A\cap B\cap(\xi_\tau=a_0)\},
\end{aligned}$$

что и доказывает одновременно (32) и (33) (в случае (33) надо взять $B=\Omega$).

Замечание. В случае $l=1$ строго марковское свойство (32), (33) эквивалентно, очевидно, тому, что для любого $C\subseteq X$

$$\mathsf{P}\{\xi_{\tau+1}\in C\mid A\cap B\cap(\xi_\tau=a_0)\}=P_{a_0}(C), \tag{34}$$

где

$$P_{a_0}(C)=\sum_{a_1\in C}p_{a_0a_1}.$$

В свою очередь (34) может быть переформулировано следующим образом: на множестве $A=\{\tau\leqslant n-1\}$

$$\mathsf{P}\{\xi_{\tau+1}\in C\mid\mathscr{B}_\tau^{\xi}\}=P_{\xi_\tau}(C), \tag{35}$$

что является одной из обычно используемых форм строго марковского свойства в общей теории однородных марковских процессов.

7. Пусть $\xi=(\xi_0,\ \ldots,\ \xi_n)$ — однородная марковская цепь с матрицей переходных вероятностей $\|p_{ij}\|$,

$$f_{ii}^{(k)}=\mathsf{P}\{\xi_k=i,\ \xi_l\neq i,\ 1\leqslant l\leqslant k-1\mid\xi_0=i\} \tag{36}$$

и для $i\neq j$

$$f_{ij}^{(k)}=\mathsf{P}\{\xi_k=j,\ \xi_l\neq j,\ 1\leqslant l\leqslant k-1\mid\xi_0=i\} \tag{37}$$

вероятности первого возвращения в состояние i и момент первого попадания в состояние j в момент времени k соответственно.

Покажем, что

n

Наглядный смысл этой формулы ясен: чтобы за n шагов попасть из состояния i в состояние j, надо сначала за k шагов $(1 \leqslant k \leqslant n)$ впервые попасть в состояние j, а затем за оставшиеся $n-k$ шагов из j попасть в j. Дадим теперь строгий вывод.

Пусть j фиксировано и

$$\tau = \min \{1 \leqslant k \leqslant n : \xi_k = j\},$$

считая $\tau = n+1$, если $\{\cdot\} = \varnothing$. Тогда $f_{ij}^{(k)} = \mathsf{P}\{\tau = k \mid \xi_0 = i\}$ и

$$\begin{aligned} p_{ij}^{(n)} = \mathsf{P}\{\xi_n = j \mid \xi_0 = i\} &= \\ &= \sum_{1 \leqslant k \leqslant n} \mathsf{P}\{\xi_n = j,\ \tau = k \mid \xi_0 = i\} = \\ &= \sum_{1 \leqslant k \leqslant n} \mathsf{P}\{\xi_{\tau+n-k} = j,\ \tau = k \mid \xi_0 = i\}, \end{aligned} \tag{39}$$

где последнее равенство следует из того, что на множестве $\{\tau = k\}$ $\xi_{\tau+n-k} = \xi_n$. Далее, для всякого $1 \leqslant k \leqslant n$ множество $\{\tau = k\} = \{\tau = k,\ \xi_\tau = j\}$. Поэтому, если $\mathsf{P}\{\xi_0 = i,\ \tau = k\} > 0$, то в силу теоремы 2

$$\begin{aligned} \mathsf{P}\{\xi_{\tau+n-k} = j \mid \xi_0 = i,\ \tau = k\} = \mathsf{P}\{\xi_{\tau+n-k} = j \mid \xi_0 = i,\ \tau = k,\ \xi_\tau = j\} &= \\ = \mathsf{P}\{\xi_{\tau+n-k} = j \mid \xi_\tau = j\} &= p_{jj}^{(n-k)} \end{aligned}$$

и, согласно (37),

$$\begin{aligned} p_{ij}^{(n)} &= \sum_{k=1}^{n} \mathsf{P}\{\xi_{\tau+n-k} = j \mid \xi_0 = i,\ \tau = k\}\, \mathsf{P}\{\tau = k \mid \xi_0 = i\} = \\ &= \sum_{k=1}^{n} p_{jj}^{(n-k)} f_{ij}^{(k)}, \end{aligned}$$

что и доказывает соотношение (38).

8. Задачи.

1. Пусть $\xi = (\xi_0, \ldots, \xi_n)$ — марковская цепь со значениями в X и $f = f(x)$ $(x \in X)$ — некоторая функция. Будет ли последовательность $(f(\xi_0), \ldots, f(\xi_n))$ образовывать марковскую цепь? Будет ли марковской цепью «обратная» последовательность $(\xi_n, \xi_{n-1}, \ldots, \xi_0)$?

2. Пусть $\mathbb{P} = \|p_{ij}\|$, $1 \leqslant i,\ j \leqslant r$, — стохастическая матрица и λ — собственное число этой матрицы, т. е. корень характеристического уравнения $\det \|\mathbb{P} - \lambda E\| = 0$. Показать, что $\lambda_0 = 1$ является собственным числом, а все остальные корни $\lambda_1, \ldots, \lambda_r$ по модулю не больше 1. Если все собственные числа $\lambda_1, \ldots, \lambda_r$ различны, то $p_{ij}^{(k)}$ допускают представление

$$p_{ij}^{(k)} = \pi_j + a_{ij}(1)\lambda_1^k + \ldots + a_{ij}(r)\lambda_r^k,$$

ГЛАВА II

МАТЕМАТИЧЕСКИЕ ОСНОВАНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. Вероятностная модель эксперимента с бесконечным числом исходов. Аксиоматика Колмогорова

1. Введенные в предшествующей главе модели позволили нам дать вероятностно-статистическое описание тех экспериментов, число исходов которых конечно. Так, тройка $(\Omega, \mathscr{A}, \mathsf{P})$ с $\Omega = \{\omega: \omega = (a_1, \ldots, a_n),\ a_i = 0,\ 1\}$, $\mathscr{A} = \{A: A \subseteq \Omega\}$ и $p(\omega) = p^{\Sigma a_i} q^{n - \Sigma a_i}$ — это модель эксперимента, состоящего в n-кратном «независимом» подбрасывании монеты с вероятностью выпадания «герба», равной p. В этой модели число $N(\Omega)$ всех исходов, т. е. число точек множества Ω, конечно и равно 2^n.

Зададимся теперь вопросом о построении вероятностной модели для эксперимента, состоящего в бесконечном «независимом» подбрасывании монеты с вероятностью выпадания «герба» на каждом шаге, равной p.

В качестве множества исходов естественно взять множество

$$\Omega = \{\omega: \ \omega = (a_1, \ a_2, \ \ldots), \ a_i = 0, \ 1\},$$

т. е. пространство всех последовательностей $\omega = (a_1, a_2, \ldots)$, элементы которых принимают два значения 0 или 1.

Чему равна мощность $N(\Omega)$ множества Ω? Хорошо известно, что всякое число $a \in [0, 1)$ может быть однозначно разложено в (содержащую бесконечное число нулей) двоичную дробь

$$a = \frac{a_1}{2} + \frac{a_2}{2^2} + \ldots \quad (a_i = 0, \ 1).$$

Отсюда ясно, что между точками ω множества Ω и точками a множества $[0, 1)$ существует взаимно однозначное соответствие, а значит, мощность множества Ω равна мощности континуума.

Таким образом, если желать строить вероятностные модели, описывающие эксперименты типа бесконечного подбрасывания монеты, то приходится привлекать к рассмотрению пространства Ω довольно сложной природы.

Попытаемся теперь понять, как разумно следовало бы задавать (приписывать) вероятности в модели бесконечного числа «независимых» подбрасываний «правильной» ($p+q=1/2$) монеты.

Поскольку в качестве Ω можно взять множество $[0, 1)$, то интересующая нас задача может рассматриваться как задача о значениях вероятностей в модели «случайного выбора точки из множества $[0, 1)$». Из соображений симметрии ясно, что все исходы должны быть «равновозможными». Но множество $[0, 1)$ несчетно, и если считать, что его вероятность равна единице, то получается, что вероятность $p(\omega)$ каждого исхода $\omega \in [0, 1)$ непременно должна быть равна нулю. Однако из такого способа задания вероятностей ($p(\omega)=0$, $\omega \in [0, 1)$) мало что следует. Дело в том, что обычно мы интересуемся не тем, с какой вероятностью произойдет тот или иной исход, а тем, какова вероятность того, что исход эксперимента будет принадлежать тому или иному заданному множеству исходов (событию) A. В элементарной теории вероятностей по вероятностям $p(\omega)$ можно было найти вероятность $\mathsf{P}(A)$ события A: $\mathsf{P}(A)=\sum_{\omega \in A} p(\omega)$. В рассматриваемом сейчас случае при $p(\omega)=0$, $\omega \in [0, 1)$, мы не можем определить, например, вероятность того, что «случайно выбранная точка из $[0, 1)$» будет принадлежать множеству $[0, 1/2)$. В то же самое время интуитивно ясно, что эта вероятность равна $1/2$.

Эти замечания подсказывают, что при построении вероятностных моделей в случае несчетных пространств Ω вероятности надо задавать не для отдельных исходов, а для некоторых множеств из Ω. Та же аргументация, что и в первой главе, показывает, что запас множеств, на которых задается вероятность, должен быть замкнутым относительно взятия объединения, пересечения и дополнения. В связи с этим полезно следующее

Определение 1. Пусть Ω — некоторое множество точек ω. Система $\mathscr{A}$ подмножеств Ω называется *алгеброй*, если

a) $\Omega \in \mathscr{A}$,

b) $A, B \in \mathscr{A} \Rightarrow A \cup B \in \mathscr{A}$, $A \cap B \in \mathscr{A}$,

c) $A \in \mathscr{A} \Rightarrow \bar{A} \in \mathscr{A}$

(заметим, что в условии b) достаточно требовать лишь, чтобы либо $A \cup B \in \mathscr{A}$, либо $A \cap B \in \mathscr{A}$, поскольку $A \cup B = \overline{\bar{A} \cap \bar{B}}$, $A \cap B = \overline{\bar{A} \cup \bar{B}}$).

Для формулировки понятия вероятностной модели нам необходимо

Определение 2. Пусть $\mathscr{A}$ — алгебра подмножеств Ω. Функция множеств $\mu=\mu(A)$, $A \in \mathscr{A}$, принимающая значения в $[0, \infty]$, называется *конечно-аддитивной мерой*, заданной на $\mathscr{A}$, если для любых двух непересекающихся множеств A и B из $\mathscr{A}$

$$\mu(A+B)=\mu(A)+\mu(B). \tag{1}$$

Конечно-аддитивная мера μ с $\mu(\Omega)<\infty$ называется конечной, а в случае $\mu(\Omega)=1$ — конечно-аддитивной вероятностной мерой, или конечно-аддитивной вероятностью.

2. Дадим теперь определение вероятностной модели (в расширенном смысле).

Определение 3. Совокупность объектов

$$(\Omega, \mathscr{A}, \mathsf{P}),$$

где

a) Ω — множество точек ω;

b) $\mathscr{A}$ — алгебра подмножеств Ω;

c) P — конечно-аддитивная вероятность на $\mathscr{A}$,

называется *вероятностной моделью в расширенном смысле.*

Оказывается, однако, что для построения плодотворной математической теории, эта вероятностная модель является слишком широкой. Поэтому приходится вводить ограничения как на классы рассматриваемых подмножеств множества Ω, так и на классы допустимых вероятностных мер.

Определение 4. Система $\mathscr{F}$ подмножеств Ω называется *σ-алгеброй*, если она является алгеброй и, кроме того, выполнено следующее свойство (усиление свойства b) из определения 1):

b*) если $A_n \in \mathscr{F}$, $n=1, 2, \ldots$, то

$$\bigcup A_n \in \mathscr{F}, \quad \bigcap A_n \in \mathscr{F}$$

(в условии b*) достаточно требовать, чтобы либо $\bigcup A_n \in \mathscr{F}$, либо $\bigcap A_n \in \mathscr{F}$).

Определение 5. Пространство Ω вместе с σ-алгеброй его подмножеств $\mathscr{F}$ называется *измеримым пространством* и обозначается $(\Omega, \mathscr{F})$.

Определение 6. Конечно-аддитивная мера μ, заданная на алгебре $\mathscr{A}$ подмножеств множества Ω, называется *счетно-аддитивной* (*σ-аддитивной*) или просто мерой, если для любых попарно непересекающихся множеств $A_1, A_2, \ldots$ из A таких, что $\sum_{n=1}^{\infty} A_n \in \mathscr{A}$

$$\mu\left(\sum_{n=1}^{\infty} A_n\right)=\sum_{n=1}^{\infty} \mu(A_n).$$

Конечно-аддитивная мера μ называется *σ-конечной*, если пространство Ω можно представить в виде

$$\Omega = \sum_{n=1}^{\infty} \Omega_n, \quad \Omega_n \in \mathscr{A}$$

с $\mu(\Omega_n) < \infty$, $n = 1, 2, \ldots$

Счетно-аддитивная мера P на алгебре $\mathscr{A}$, удовлетворяющая условию $\mathsf{P}(\Omega) = 1$, будет называться *вероятностной мерой* или *вероятностью* (определенной на множествах алгебры $\mathscr{A}$).

Остановимся на некоторых свойствах вероятностных мер.

Если $\varnothing$ — пустое множество, то

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

Если $A, B \in \mathscr{A}$, то

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Если $A, B \in \mathscr{A}$ и $B \subseteq A$, то

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

Если $A_n \in \mathscr{A}$, $n = 1, 2, \ldots$ и $\bigcup A_n \in \mathscr{A}$, то

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

Первые три свойства очевидны. Для доказательства последнего достаточно заметить, что $\sum_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} B_n$, где $B_1 = A_1$, $B_n = \bar{A}_1 \cap \ldots \cap \bar{A}_{n-1} \cap A_n$, $n \geqslant 2$, $B_i \cap B_j = \varnothing$, $i \neq j$, и, значит,

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathsf{P}\left(\sum_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathsf{P}(B_n) \leqslant \sum_{n=1}^{\infty} \mathsf{P}(A_n).$$

Приводимая ниже теорема, имеющая многочисленные применения, дает условия, при которых конечно-аддитивная функция множеств является в то же самое время и счетно-аддитивной.

Теорема. *Пусть P — конечно-аддитивная функция множеств, заданная на алгебре $\mathscr{A}$, с $\mathsf{P}(\Omega) = 1$. Тогда следующие четыре условия эквивалентны:*

1) P *σ-аддитивна* (P — *вероятность*);

2) P *непрерывна сверху, т. е. для любых множеств $A_1, A_2, \ldots \ldots \in \mathscr{A}$ таких, что $A_n \subseteq A_{n+1}$, $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$*

$$\lim_n \mathsf{P}(A_n) = \mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right);$$

3) P *непрерывна снизу, т. е. для любых множеств* $A_1, A_2, \ldots$ $\ldots \in \mathscr{A}$ *таких, что* $A_n \supseteq A_{n+1}$, $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{A}$

$$\lim_n \mathsf{P}(A_n) = \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right);$$

4) P *непрерывна в «нуле», т. е. для любых множеств* $A_1, A_2, \ldots$ $\ldots \in \mathscr{A}$ *таких, что* $A_{n+1} \subseteq A_n$, $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$

$$\lim_n \mathsf{P}(A_n) = 0.$$

Доказательство. 1) $\Longrightarrow$ 2). Поскольку

$$\bigcup_{n=1}^{\infty} A_n = A_1 + (A_2 \setminus A_1) + (A_3 \setminus A_2) + \ldots,$$

то

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathsf{P}(A_1) + \mathsf{P}(A_2 \setminus A_1) + \mathsf{P}(A_3 \setminus A_2) + \ldots =$$
$$= \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1) + \mathsf{P}(A_3) - \mathsf{P}(A_2) + \ldots =$$
$$= \lim_n \mathsf{P}(A_n).$$

2) $\Longrightarrow$ 3). Пусть $n \geqslant 1$, тогда

$$\mathsf{P}(A_n) = \mathsf{P}(A_1 \setminus (A_1 \setminus A_n)) = \mathsf{P}(A_1) - \mathsf{P}(A_1 \setminus A_n).$$

Последовательность множеств $\{A_1 \setminus A_n\}_{n \geqslant 1}$ является неубывающей (см. в следующем п. 3 таблицу) и

$$\bigcup_{n=1}^{\infty} (A_1 \setminus A_n) = A_1 \setminus \bigcap_{n=1}^{\infty} A_n.$$

Тогда в силу 2)

$$\lim_n \mathsf{P}(A_1 \setminus A_n) = \mathsf{P}\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right)$$

и, значит,

$$\lim_n \mathsf{P}(A_n) = \mathsf{P}(A_1) - \lim_n \mathsf{P}(A_1 \setminus A_n) =$$
$$= \mathsf{P}(A_1) - \mathsf{P}\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right) = \mathsf{P}(A_1) - \mathsf{P}\left(A_1 \setminus \bigcap_{n=1}^{\infty} A_n\right) =$$
$$= \mathsf{P}(A_1) - \mathsf{P}(A_1) + \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

3) $\Rightarrow$ 4) Очевидно.

4) $\Rightarrow$ 1) Пусть $A_1, A_2, \ldots \in \mathscr{A}$ попарно не пересекаются и $\sum_{n=1}^{\infty} A_n \in \mathscr{A}$. Тогда

$$\mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \mathsf{P}\left(\sum_{i=1}^{n} A_i\right) + \mathsf{P}\left(\sum_{i=n+1}^{\infty} A_i\right),$$

и поскольку $\sum_{i=n+1}^{\infty} A_i \downarrow \varnothing$, $n \to \infty$, то

$$\sum_{i=1}^{\infty} \mathsf{P}(A_i) = \lim_n \sum_{i=1}^{n} \mathsf{P}(A_i) = \lim_n \mathsf{P}\left(\sum_{i=1}^{n} A_i\right) =$$
$$= \lim_n \left[\mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) - \mathsf{P}\left(\sum_{i=n+1}^{\infty} A_i\right)\right] =$$
$$= \mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) - \lim_n \mathsf{P}\left(\sum_{i=n+1}^{\infty} A_i\right) = \mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right).$$

3. Теперь можно сформулировать ставшую общепринятой систему аксиом Колмогорова, лежащих в основе понятия вероятностного пространства.

Основное определение. *Набор объектов*

$$(\Omega, \mathscr{F}, \mathsf{P}),$$

где

a) Ω — *множество точек* ω,

b) $\mathscr{F}$ — *σ-алгебра подмножеств* Ω,

c) P — *вероятность на* $\mathscr{F}$,

называется вероятностной моделью или вероятностным пространством. При этом Ω называется пространством исходов или пространством элементарных событий, множества A из $\mathscr{F}$ — событиями, а $\mathsf{P}(A)$ — вероятностью события A.

Из данного определения видно, что аксиоматика теории вероятностей существенно опирается на аппарат теории множеств и теории меры. В связи с этим полезно дать таблицу, показывающую, как различные понятия интерпретируются в теории множеств и в теории вероятностей. Примеры наиболее важных для теории вероятностей измеримых пространств и способы задания вероятностей на них будут даны в последующих двух параграфах.

Таблица

Обозначения	Интерпретация теории множеств	Интерпретация теории вероятностей
ω	элемент, точка	исход, элементарное событие
Ω	множество точек	пространство исходов, элементарных событий; достоверное событие
$\mathscr{F}$	σ-алгебра подмножеств	σ-алгебра событий
$A \in \mathscr{F}$	множество точек	событие (если $\omega \in A$, то говорят, что наступило событие A)
$\bar{A} = \Omega \setminus A$	дополнение множества A, т. е. множество точек ω, не входящих в A	событие, состоящее в ненаступлении события A
$A \cup B$	объединение множеств A и B, т. е. множество точек ω, входящих или в A или в B	событие, состоящее в том, что произошло либо A, либо B
$A \cap B$ (или AB)	пересечение множеств A и B, т. е. множество точек ω, входящих и в A и в B	событие, состоящее в том, что одновременно произошло и A и B
$\varnothing$	пустое множество	невозможное событие
$A \cap B = \varnothing$	множества A и B не пересекаются	события A и B несовместны (не могут наступать одновременно)
$A + B$	сумма множеств, т. е. объединение непересекающихся множеств	событие, состоящее в том, что произошло одно из двух несовместных событий
$A \setminus B$	разность множеств A и B, т. е. множество точек, входящих в A, но не входящих в B	событие, состоящее в том, что произошло A, но не произошло B
$A \triangle B$	симметрическая разность множеств, т. е. множество $(A \setminus B) \cup (B \setminus A)$	событие, состоящее в том, что произошло одно из событий A или B, но не оба одновременно
$\bigcup\limits_{n=1}^{\infty} A_n$	объединение множеств $A_1, A_2, \ldots$	событие, состоящее в наступлении по крайней мере одного из событий $A_1, A_2, \ldots$
$\sum\limits_{n=1}^{\infty} A_n$	сумма, т. е. объединение попарно непересекающихся множеств $A_1, A_2, \ldots$	событие, состоящее в наступлении одного из несовместных событий $A_1, A_2, \ldots$

Продолжение

Обозначения	Интерпретация теории множеств	Интерпретация теории вероятностей
$\bigcap\limits_{n=1}^{\infty} A_n$	пересечение множеств $A_1, A_2, \ldots$	событие, состоящее в том, что одновременно произошли $A_1, A_2, \ldots$
$A_n \uparrow A$ (или $A = \lim\limits_n \uparrow A_n$)	возрастающая последовательность множеств A_n, сходящаяся к A, т. е. $A_1 \subseteq A_2 \subseteq \ldots$ и $A = \bigcup\limits_{n=1}^{\infty} A_n$	возрастающая последовательность событий, сходящихся к событию A
$A_n \downarrow A$ (или $A = \lim\limits_n \downarrow A_n$)	убывающая последовательность множеств A_n, сходящаяся к A, т. е. $A_1 \supseteq A_2 \supseteq \ldots$ и $A = \bigcap\limits_{n=1}^{\infty} A_n$	убывающая последовательность событий, сходящихся к событию A
$\overline{\lim\limits_n} A_n$ (или $\lim\sup A_n$, или $\{A_n$ б. ч.$\}$)	множество $\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$	событие, состоящее в том, что произойдет бесконечно много событий из $A_1, A_2, \ldots$
$\underline{\lim\limits_n} A_n$ (или $\lim\inf A_n$)	множество $\bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k$	событие, состоящее в том, что произойдут все события $A_1, A_2, \ldots$ за исключением, быть может, только конечного числа

4. Задачи.

1. Пусть $\Omega = \{r\colon r \in [0, 1]\}$ — множество рациональных точек а $[0, 1]$, $\mathscr{A}$ — алгебра множеств, каждое из которых является онечной суммой непересекающихся множеств A вида $\{r\colon a < r < b\}$, $\{r\colon a \leqslant r < b\}$, $\{r\colon a < r \leqslant b\}$, $\{r\colon a \leqslant r \leqslant b\}$ и $\mathsf{P}(A) = b - a$. Іоказать, что $\mathsf{P}(A)$, $A \in \mathscr{A}$, является конечно-аддитивной, но е счетно-аддитивной функцией множеств.

2. Пусть Ω — некоторое счетное множество и $\mathscr{F}$ — совокупность сех его подмножеств. Положим $\mu(A) = 0$, если A конечно и $(A) = \infty$, если A бесконечно. Показать, что функция множеств μ онечно-аддитивна, но не счетно-аддитивна.

3. Пусть μ — конечная мера на σ-алгебре $\mathscr{F}$, $A_n \in \mathscr{F}$, $n = 1$, $, \ldots$ и $A = \lim\limits_n A_n$ (т. е. $A = \overline{\lim\limits_n} A_n = \underline{\lim\limits_n} A_n$). Показать, что $(A) = \lim\limits_n \mu(A_n)$.

4. Доказать, что $\mathsf{P}(A \triangle B) = \mathsf{P}(A) + \mathsf{P}(B) - 2\mathsf{P}(A \cap B)$.

5. Показать, что «расстояния» $\rho_1(A, B)$ и $\rho_2(A, B)$, определенные по формулам

$$\rho_1(A, B) = \mathsf{P}(A \triangle B),$$

$$\rho_2(A, B) = \begin{cases} \dfrac{\mathsf{P}(A \triangle B)}{\mathsf{P}(A \cup B)}, & \text{если } \mathsf{P}(A \cup B) \neq 0, \\ 0, & \text{если } \mathsf{P}(A \cup B) = 0, \end{cases}$$

удовлетворяет неравенству треугольника.

6. Пусть μ — конечно-аддитивная мера на алгебре $\mathscr{A}$, множества $A_1, A_2, \ldots \in \mathscr{A}$, попарно не пересекаются и $A = \sum_{i=1}^{\infty} A_i \in \mathscr{A}$. Тогда $\mu(A) \geqslant \sum_{i=1}^{\infty} \mu(A_i)$.

7. Доказать, что

$$\overline{\limsup A_n} = \liminf \bar{A}_n, \quad \overline{\liminf A_n} = \limsup \bar{A}_n,$$

$$\liminf A_n \subseteq \limsup A_n, \quad \limsup (A_n \cup B_n) = \limsup A_n \cup \limsup B_n,$$

$$\limsup A_n \cap \liminf B_n \subseteq \limsup (A_n \cap B_n) \subseteq \limsup A_n \cap \limsup B_n.$$

Если $A_n \uparrow A$ или $A_n \downarrow A$, то

$$\liminf A_n = \limsup A_n.$$

8. Пусть $\{x_n\}$ — числовая последовательность и $A_n = (-\infty, x_n)$. Показать, что $x = \limsup x_n$ и $A = \limsup A_n$ связаны следующим образом: $(-\infty, x) \subseteq A \subseteq (-\infty, x]$. Иначе говоря, A равно или $(-\infty, x)$ или $(-\infty, x]$.

9. Привести пример, показывающий, что для мер, принимающих значение $+\infty$, из счетной аддитивности не вытекает, вообще говоря, непрерывность в «нуле» $\varnothing$.

§ 2. Алгебры и σ-алгебры. Измеримые пространства

1. Алгебры и σ-алгебры являются составными элементами при построении вероятностных моделей. Приведем примеры и ряд результатов, относящихся к этим объектам.

Пусть Ω — некоторое пространство элементарных событий. Очевидным образом системы множеств

$$\mathscr{F}_* = \{\varnothing, \Omega\}, \quad \mathscr{F}^* = \{A : A \subseteq \Omega\}$$

являются и алгебрами, и σ-алгебрами. При этом $\mathscr{F}_*$ — тривиальная, самая «бедная» σ-алгебра, а $\mathscr{F}^*$ — самая «богатая» σ-алгебра, состоящая из всех подмножеств Ω.

В случае конечных пространств Ω σ-алгебра $\mathscr{F}^*$ вполне обозрима, и, как правило, именно ее рассматривают в элементарной теории в качестве системы «событий». В случае же несчетных пространств класс $\mathscr{F}^*$ оказывается слишком широким, поскольку на системе таких множеств не всегда удается «согласованным образом» задать вероятность.

Если $A \subseteq \Omega$, то система

$$\mathscr{F}_A = \{A, \bar{A}, \varnothing, \Omega\}$$

является также примером алгебры (и σ-алгебры), называемой алгеброй (σ-алгеброй), порожденной множеством A.

Эта система множеств является частным случаем систем, порождаемых разбиениями. А именно, пусть

$$\mathscr{D} = \{D_1, D_2, \ldots\}$$

— некоторое *счетное* разбиение Ω на непустые множества:

$$\Omega = D_1 + D_2 + \ldots; \quad D_i \cap D_j = \varnothing, \quad i \neq j.$$

Тогда система $\mathscr{A} = \alpha(\mathscr{D})$, образованная из множеств, являющихся объединением конечного числа элементов разбиения, является алгеброй.

Следующая лемма имеет важное значение, поскольку в ней устанавливается принципиальная возможность построения наименьших алгебры и σ-алгебры, содержащих заданную систему множеств.

Лемма 1. *Пусть $\mathscr{E}$ — некоторая система множеств из Ω. Тогда существуют наименьшая алгебра, обозначаемая $\alpha(\mathscr{E})$, и наименьшая σ-алгебра, обозначаемая $\sigma(\mathscr{E})$, содержащие все множества из $\mathscr{E}$.*

Доказательство. Класс всех подмножеств $\mathscr{F}^*$ пространства Ω есть σ-алгебра. Таким образом, по крайней мере одна алгебра и σ-алгебра, содержащие $\mathscr{E}$, существуют. Образуем теперь систему $\alpha(\mathscr{E})$, $(\sigma(\mathscr{E}))$, состоящую из тех множеств, которые принадлежат любой алгебре (σ-алгебре), содержащей $\mathscr{E}$. Нетрудно проверить, что такая система есть алгебра (σ-алгебра) и к тому же наименьшая.

Замечание. Систему $\alpha(\mathscr{E})$ (соответственно $\sigma(\mathscr{E})$) часто называют наименьшей алгеброй (соответственно σ-алгеброй), порожденной системой множеств $\mathscr{E}$.

Часто возникает вопрос о том, при каких дополнительных условиях алгебра или какая-нибудь другая система множеств является в то же самое время и σ-алгеброй. Приведем несколько результатов в этом направлении.

Определение 1. Система $\mathscr{M}$ подмножеств Ω называется *монотонным классом*, если из того, что $A_n \in \mathscr{M}$, $n = 1, 2, \ldots$, и $A_n \uparrow A$ или $A_n \downarrow A$ следует, что $A \in \mathscr{M}$.

Пусть $\mathscr{E}$ — некоторая система множеств. Будем обозначать через $\mu(\mathscr{E})$ наименьший монотонный класс, содержащий $\mathscr{E}$. (Доказательство существования такого класса проводится так же, как и в лемме 1.)

Лемма 2. *Для того чтобы алгебра $\mathscr{A}$ была в то же время и σ-алгеброй, необходимо и достаточно, чтобы она была монотонным классом.*

Доказательство. Каждая σ-алгебра является, очевидным образом, монотонным классом. Пусть теперь $\mathscr{A}$ является монотонным классом и $A_n \in \mathscr{A}$, $n = 1, 2, \ldots$ Ясно, что $B_n = \bigcup_{i=1}^{n} A_i \in \mathscr{A}$ и $B_n \subseteq B_{n+1}$. Следовательно, по определению монотонного класса $B_n \uparrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$. Аналогично устанавливается, что $\bigcap_{i=1}^{\infty} A_i \in \mathscr{A}$.

Используя эту лемму, докажем справедливость следующего результата, показывающего, как, отправляясь от алгебры $\mathscr{A}$, можно с помощью монотонных предельных переходов получить σ-алгебру $\sigma(\mathscr{A})$.

Теорема 1. *Пусть $\mathscr{A}$ — алгебра. Тогда*

$$\mu(\mathscr{A}) = \sigma(\mathscr{A}). \tag{1}$$

Доказательство. Из леммы 2 $\mu(\mathscr{A}) \subseteq \sigma(\mathscr{A})$. Поэтому достаточно показать, что $\mu(\mathscr{A})$ является σ-алгеброй. Но система $\mathscr{M} = \mu(\mathscr{A})$ — монотонный класс, поэтому опять-таки по лемме 2 достаточно только установить, что $\mu(\mathscr{A})$ является алгеброй.

Возьмем $A \in \mathscr{M}$ и покажем, что тогда $\bar{A} \in \mathscr{M}$. С этой целью применим часто используемый в дальнейшем *принцип подходящих множеств*, состоящий в следующем.

Обозначим

$$\tilde{\mathscr{M}} = \{B\colon\ B \in \mathscr{M},\ \bar{B} \in \mathscr{M}\}$$

все те множества, которые обладают интересующим нас свойством. Ясно, что $\mathscr{A} \subseteq \tilde{\mathscr{M}} \subseteq \mathscr{M}$. Установим, что $\tilde{\mathscr{M}}$ — монотонный класс.

Пусть $B_n \in \tilde{\mathscr{M}}$, тогда $B_n \in \mathscr{M}$, $\bar{B}_n \in \mathscr{M}$, и поэтому

$$\lim \uparrow B_n \in \mathscr{M}, \quad \lim \uparrow \bar{B}_n \in \mathscr{M}, \quad \lim \downarrow B_n \in \mathscr{M}, \quad \lim \downarrow \bar{B}_n \in \mathscr{M}.$$

Следовательно,

$$\overline{\lim \uparrow B_n} = \lim \downarrow \bar{B}_n \in \mathscr{M}, \quad \overline{\lim \downarrow B_n} = \lim \uparrow \bar{B}_n \in \mathscr{M},$$

$$\overline{\lim \uparrow \bar{B}_n} = \lim \downarrow B_n \in \mathscr{M}, \quad \overline{\lim \downarrow \bar{B}_n} = \lim \uparrow B_n \in \mathscr{M},$$

а значит, $\tilde{\mathscr{M}}$ — монотонный класс. Но $\tilde{\mathscr{M}} \subseteq \mathscr{M}$ и $\mathscr{M}$ — наименьший монотонный класс. Поэтому $\tilde{\mathscr{M}} = \mathscr{M}$, и если $A \in \mathscr{M} = \mu(\mathscr{A})$, то и $\bar{A} \in \mathscr{M}$, т. е. класс $\mathscr{M}$ замкнут относительно операции взятия дополнения.

Покажем теперь, что класс $\mathscr{M}$ замкнут относительно взятия пересечения.

Пусть $A \in \mathscr{M}$ и

$$\mathscr{M}_A = \{B\colon\ B \in \mathscr{M},\ A \cap B \in \mathscr{M}\}.$$

Из равенств

$$\lim \downarrow (A \cap B_n) = A \cap \lim \downarrow B_n,$$
$$\lim \uparrow (A \cap B_n) = A \cap \lim \uparrow B_n$$

следует, что $\mathscr{M}_A$ — монотонный класс.

Далее, легко проверяется, что

$$(A \in \mathscr{M}_B) \Leftrightarrow (B \in \mathscr{M}_A). \tag{2}$$

Пусть теперь $A \in \mathscr{A}$, тогда поскольку $\mathscr{A}$ — алгебра, то для всякого $B \in \mathscr{A}$ множество $A \cap B \in \mathscr{A}$ и, значит,

$$\mathscr{A} \subseteq \mathscr{M}_A \subseteq \mathscr{M}.$$

Но $\mathscr{M}_A$ — монотонный класс (поскольку $\lim \uparrow AB_n = A \lim \uparrow B_n$ и $\lim \downarrow AB_n = A \lim \downarrow B_n$), а $\mathscr{M}$ — наименьший монотонный класс. Значит, $\mathscr{M}_A = \mathscr{M}$ для любого $A \in \mathscr{A}$. Но тогда из (2) вытекает, что для $A \in \mathscr{A}$ и $B \in \mathscr{M}$

$$(A \in \mathscr{M}_B) \Leftrightarrow (B \in \mathscr{M}_A = \mathscr{M}).$$

Следовательно, если $A \in \mathscr{A}$, то для любого $B \in \mathscr{M}$

$$A \in \mathscr{M}_B.$$

В силу произвольности $A \in \mathscr{A}$ отсюда следует, что

$$\mathscr{A} \subseteq \mathscr{M}_B \subseteq \mathscr{M}.$$

Значит, для всякого $B \in \mathscr{M}$

$$\mathscr{M}_B = \mathscr{M},$$

т. е. если $B \in \mathscr{M}$ и $C \in \mathscr{M}$, то $C \cap B \in \mathscr{M}$.

Итак, класс $\mathscr{M}$ замкнут относительно операций взятия дополнения и пересечения (а значит, и объединения). Следовательно, $\mathscr{M}$ — алгебра, что и доказывает теорему.

Определение 2. Пусть Ω — некоторое пространство. Класс $\mathscr{D}$ подмножеств Ω называется *d-системой*, если

a) $\Omega \in \mathscr{D}$;

b) $A, B \in \mathscr{D},\ A \subseteq B \Rightarrow B \setminus A \in \mathscr{D}$;

c) $A_n \in \mathscr{D},\ A_n \subseteq A_{n+1} \Rightarrow \bigcup A_n \in \mathscr{D}$.

Если $\mathscr{E}$ — некоторая система множеств, то через $d(\mathscr{E})$ будет обозначаться наименьшая d-система, содержащая $\mathscr{E}$.

Теорема 2. *Если $\mathscr{E}$ есть система множеств, замкнутая относительно образования пересечений, то*

$$d(\mathscr{E}) = \sigma(\mathscr{E}). \tag{3}$$

Доказательство. Каждая σ-алгебра является d-системой, и, следовательно, $d(\mathscr{E}) \subseteq \sigma(\mathscr{E})$. Поэтому, если доказать, что система $d(\mathscr{E})$ замкнута относительно взятия пересечений, то $d(\mathscr{E})$ будет σ-алгеброй и тогда, конечно, будет справедливо противоположное включение $\sigma(\mathscr{E}) \subseteq d(\mathscr{E})$.

Для доказательства снова воспользуемся принципом подходящих множеств.

Пусть

$$\mathscr{E}_1 = \{B \in d(\mathscr{E}) \colon B \cap A \in d(\mathscr{E}) \text{ для всех } A \in \mathscr{E}\}.$$

Если $B \in \mathscr{E}$, то $B \cap A \in \mathscr{E}$ для всех $A \in \mathscr{E}$ и, значит, $\mathscr{E} \subseteq \mathscr{E}_1$. Но $\mathscr{E}_1$ есть d-система. Поэтому $d(\mathscr{E}) \subseteq \mathscr{E}_1$. С другой стороны, по определению $\mathscr{E}_1 \subseteq d(\mathscr{E})$. Следовательно,

$$\mathscr{E}_1 = d(\mathscr{E}).$$

Пусть теперь

$$\mathscr{E}_2 = \{B \in d(\mathscr{E}) \colon B \cap A \in d(\mathscr{E}) \text{ для всех } A \in d(\mathscr{E})\}.$$

Снова легко проверяется, что $\mathscr{E}_2$ есть d-система. Если $B \in \mathscr{E}$, то по определению $\mathscr{E}_1$ для всех $A \in \mathscr{E}_1 = d(\mathscr{E})$ получим, что $B \cap A \in d(\mathscr{E})$. Следовательно, $\mathscr{E} \subseteq \mathscr{E}_2$ и $d(\mathscr{E}) \subseteq \mathscr{E}_2$. Но $d(\mathscr{E}) \supseteq \mathscr{E}_2$, поэтому $d(\mathscr{E}) = \mathscr{E}_2$, и, значит, для каждых A и B из $d(\mathscr{E})$ множество $A \cap B$ также принадлежит $d(\mathscr{E})$, т. е. система $d(\mathscr{E})$ замкнута относительно взятия пересечений.

Теорема доказана.

Перейдем к рассмотрению наиболее важных для теории вероятностей измеримых пространств $(\Omega, \mathscr{F})$.

2. Измеримое пространство $(R, \mathscr{B}(R))$. Пусть $R = (-\infty, \infty)$ — действительная прямая и

$$(a, b] = \{x \in R \colon a < x \leqslant b\}$$

для всех $-\infty \leqslant a < b < \infty$. Условимся под интервалом $(a, \infty]$ понимать интервал (a, ∞). (Это соглашение необходимо для того, чтобы дополнение до интервала $(-\infty, b]$ было интервалом того же вида, т. е. открытым слева и замкнутым справа.)

Обозначим через $\mathscr{A}$ систему множеств в R, состоящих из конечных сумм непересекающихся интервалов вида $(a, b]$:

$$A \in \mathscr{A}, \text{ если } A = \sum_{i=1}^{n} (a_i, b_i], \quad n < \infty.$$

Нетрудно проверить, что эта система множеств, в которую мы включаем также и пустое множество $\varnothing$, образует алгебру, которая, однако, не является σ-алгеброй, поскольку, если $A_n = (0,\ 1-1/n] \in \mathscr{A}$, то $\bigcup_n A_n = (0,\ 1) \notin \mathscr{A}$.

Пусть $\mathscr{B}(R)$ — наименьшая σ-алгебра $\sigma(\mathscr{A})$, содержащая систему $\mathscr{A}$. Эта σ-алгебра, играющая важную роль в математическом анализе, называется *борелевской* алгеброй множеств числовой прямой, а ее множества — *борелевскими*.

Если обозначить через $\mathscr{I}$ систему интервалов I вида $(a,\ b]$, а через $\sigma(\mathscr{I})$ — наименьшую σ-алгебру, содержащую $\mathscr{I}$, то нетрудно проверить, что $\sigma(\mathscr{I})$ будет совпадать с борелевской алгеброй. Иначе говоря, к борелевской алгебре можно прийти от системы $\mathscr{I}$, минуя обращение к алгебре $\mathscr{A}$, поскольку $\sigma(\mathscr{I}) = \sigma(\alpha(\mathscr{I}))$.

Заметим, что

$$(a,\ b) = \bigcup_{n=1}^{\infty} \left(a,\ b - \frac{1}{n}\right], \qquad a < b,$$

$$[a,\ b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n},\ b\right], \qquad a < b,$$

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n},\ a\right].$$

Тем самым в борелевскую алгебру наряду с интервалами вида $(a,\ b]$ входят одноточечные множества $\{a\}$, а также любое из шести множеств

$$(a,\ b),\quad [a,\ b],\quad [a,\ b),\quad (-\infty,\ b),\quad (-\infty,\ b],\quad (a,\ \infty). \tag{4}$$

Отметим также, что при конструировании борелевской алгебры $\mathscr{B}(R)$ можно было бы отправляться не от интервалов вида $(a,\ b]$, а от любого из шести указанных интервалов, поскольку все наименьшие σ-алгебры, порожденные системами каждого из интервалов вида (4), совпадают с σ-алгеброй $\mathscr{B}(R)$.

Иногда приходится иметь дело с σ-алгеброй $\mathscr{B}(\bar{R})$ множеств на расширенной числовой прямой $\bar{R} = [-\infty,\ \infty]$. Так называют наименьшую σ-алгебру, порожденную интервалами вида

$$(a,\ b] = \{x \in \bar{R}:\ a < x \leqslant b\}, \qquad -\infty \leqslant a < b \leqslant \infty,$$

где под $(-\infty,\ b]$ понимается множество $\{x \in \bar{R}:\ -\infty \leqslant x \leqslant b\}$.

З а м е ч а н и е 1. Для измеримого пространства $(R,\ \mathscr{B}(R))$ часто используются также обозначения $(R,\ \mathscr{B})$, $(R^1,\ \mathscr{B}_1)$.

З а м е ч а н и е 2. Введем на числовой прямой R метрику

$$\rho_1(x,\ y) = \frac{|x-y|}{1+|x-y|}$$

(эквивалентную обычной евклидовой метрике $|x-y|$) и обозначим через $\mathscr{B}_0(R)$ наименьшую σ-алгебру, порожденную открытыми множествами $S_\rho(x^0)=\{x\in R\colon \rho_1(x, x^0)<\rho\}$, $\rho>0$, $x^0\in R$. Тогда $\mathscr{B}_0(R)=\mathscr{B}(R)$ (см. задачу 7).

3. Измеримое пространство $(R^n, \mathscr{B}(R^n))$. Пусть $R^n=R\times\ldots\times R$ — прямое, или декартово, произведение n экземпляров (копий) числовой прямой, т. е. множество упорядоченных наборов $x=(x_1, \ldots, x_n)$, где $-\infty<x_k<\infty$, $k=1, \ldots, n$. Множество $I=I_1\times\ldots\times I_n$, где $I_k=(a_k, b_k]$, т. е. множество $\{x\in R^n\colon x_k\in I_k, k=1, \ldots, n\}$, назовем прямоугольником, I_k — сторонами этого прямоугольника. Через $\mathscr{I}$ обозначим совокупность всех прямоугольников в I. Наименьшая σ-алгебра $\sigma(\mathscr{I})$, порожденная системой прямоугольников $\mathscr{I}$, называется *борелевской* алгеброй множеств в R^n и обозначается $\mathscr{B}(R^n)$. Покажем, что к этой борелевской алгебре можно было бы прийти и иначе.

Наряду с прямоугольниками $I=I_1\times\ldots\times I_n$ рассмотрим прямоугольники $B=B_1\times\ldots\times B_n$ с борелевскими сторонами (B_k — борелевское множество числовой прямой, стоящей на k-м месте в прямом произведении $R\times\ldots\times R$). Наименьшая σ-алгебра, содержащая все прямоугольники с борелевскими сторонами, обозначается

$$\mathscr{B}(R)\otimes\ldots\otimes\mathscr{B}(R)$$

и называется *прямым произведением* σ-алгебр $\mathscr{B}(R)$. Покажем, что на самом деле

$$\mathscr{B}(R^n)=\mathscr{B}(R)\otimes\ldots\otimes\mathscr{B}(R).$$

Иначе говоря, наименьшие σ-алгебры, порожденные прямоугольниками $I=I_1\times\ldots\times I_n$ и (более широким) классом прямоугольников $B=B_1\times\ldots\times B_n$ с борелевскими сторонами, совпадают.

Доказательство существенно опирается на следующее предложение.

Лемма 3. *Пусть $\mathscr{E}$ — некоторый класс множеств из Ω, множество $B\subseteq\Omega$, и пусть по определению*

$$\mathscr{E}\cap B=\{A\cap B\colon A\in\mathscr{E}\}. \tag{5}$$

Тогда

$$\sigma(\mathscr{E}\cap B)=\sigma(\mathscr{E})\cap B. \tag{6}$$

Доказательство. Поскольку $\mathscr{E}\subseteq\sigma(\mathscr{E})$, то

$$\mathscr{E}\cap B\subseteq\sigma(\mathscr{E})\cap B. \tag{7}$$

Но $\sigma(\mathscr{E})\cap B$ является σ-алгеброй, поэтому из (7) следует, что

$$\sigma(\mathscr{E}\cap B)\subseteq\sigma(\mathscr{E})\cap B.$$

Для доказательства обратного включения снова воспользуемся принципом подходящих множеств.

Обозначим

$$\mathscr{C}_B=\{A\in\sigma(\mathscr{E}):\ A\cap B\in\sigma(\mathscr{E}\cap B)\}.$$

Поскольку $\sigma(\mathscr{E})$ и $\sigma(\mathscr{E}\cap B)$ являются σ-алгебрами, то $\mathscr{C}_B$ также σ-алгебра, причем, очевидно,

$$\mathscr{E}\subseteq\mathscr{C}_B\subseteq\sigma(\mathscr{E}),$$

откуда $\sigma(\mathscr{E})\subseteq\sigma(\mathscr{C}_B)=\mathscr{C}_B\subseteq\sigma(\mathscr{E})$ и, значит, $\sigma(\mathscr{E})=\mathscr{C}_B$. Поэтому для каждого множества $A\in\sigma(\mathscr{E})$

$$A\cap B\in\sigma(\mathscr{E}\cap B),$$

и, следовательно, $\sigma(\mathscr{E})\cap B\subseteq\sigma(\mathscr{E}\cap B)$.

Лемма доказана.

Доказательство совпадения σ-алгебр $\mathscr{B}(R^n)$ и $\mathscr{B}\otimes\ldots\otimes\mathscr{B}$. Для $n=1$ их совпадение очевидно. Докажем теперь, что они совпадают для $n=2$.

Поскольку $\mathscr{B}(R^2)\subseteq\mathscr{B}\times\mathscr{B}$, то достаточно показать, что борелевский прямоугольник $B_1\times B_2$ принадлежит $\mathscr{B}(R^2)$.

Пусть $R^2=R_1\times R_2$, где R_1 и R_2 — «первая» и «вторая» действительные прямые, $\tilde{\mathscr{B}}_1=\mathscr{B}_1\times R_2$, $\tilde{\mathscr{B}}_2=R_1\times\mathscr{B}_2$, где $\mathscr{B}_1\times R_2$ $(R_1\times\mathscr{B}_2)$ есть совокупность множеств вида $B_1\times R_2$ $(R_1\times B_2)$, с $B_1\in\mathscr{B}_1$ $(B_2\in\mathscr{B}_2)$. Пусть также $\mathscr{I}_1$ и $\mathscr{I}_2$ — совокупности интервалов в R_1 и R_2 и $\tilde{\mathscr{I}}_1=\mathscr{I}_1\times R_2$, $\tilde{\mathscr{I}}_2=R_1\times\mathscr{I}_2$. Тогда в силу (6)

$$\begin{aligned}B_1\times B_2=\tilde{B}_1\cap\tilde{B}_2\in\tilde{\mathscr{B}}_1\cap\tilde{B}_2=\sigma(\tilde{\mathscr{I}}_1)\cap\tilde{B}_2=\\=\sigma(\tilde{\mathscr{I}}_1\cap\tilde{B}_2)\subseteq\sigma(\tilde{\mathscr{I}}_1\cap\tilde{\mathscr{I}}_2)=\sigma(\mathscr{I}_1\times\mathscr{I}_2),\end{aligned}$$

что и требовалось доказать.

Случай произвольного $n>2$ рассматривается аналогичным образом.

Замечание. Пусть $\mathscr{B}_0(R^n)$ — наименьшая σ-алгебра, порожденная открытыми множествами

$$S_\rho(x^0)=\{x\in R^n:\ \rho_n(x,\ x^0)<\rho\},\quad x^0\in R^n,\quad \rho>0,$$

в метрике

$$\rho_n(x,\ x^0)=\sum_{k=1}^{n}2^{-k}\rho_1(x_k,\ x_k^0),$$

где $x=(x_1,\ \ldots,\ x_n)$, $x^0=(x_1^0,\ \ldots,\ x_n^0)$.

Тогда $\mathscr{B}_0(R^n)=\mathscr{B}(R^n)$ (задача 7).

4. Измеримое пространство $(R^\infty,\ \mathscr{B}(R^\infty))$ играет значительную роль в теории вероятностей, поскольку оно служит основой построения вероятностных моделей экспериментов с бесконечным числом шагов.

Пространство R^∞ — это пространство *упорядоченных* числовых последовательностей

$$x=(x_1, x_2, \ldots), \qquad -\infty < x_k < \infty, \qquad k=1, 2, \ldots$$

Обозначим через I_k и B_k соответственно интервалы $(a_k, b_k]$ и борелевские множества k-й числовой прямой (с координатой x_k). Рассмотрим *цилиндрические* множества

$$\mathscr{I}(I_1\times\ldots\times I_n)=\{x:\ x=(x_1, x_2, \ldots),\ x_1\in I_1, \ldots, x_n\in I_n\}, \tag{8}$$

$$\mathscr{I}(B_1\times\ldots\times B_n)=\{x:\ x=(x_1\ldots),\ x_1\in B_1, \ldots, x_n\in B_n\}, \tag{9}$$

$$\mathscr{I}(B^n)=\{x:\ (x_1, \ldots, x_n)\in B^n\}, \tag{10}$$

где B^n — борелевское множество из $\mathscr{B}(R^n)$. Каждый из «цилиндров» $\mathscr{I}(B_1\times\ldots\times B_n)$ или $\mathscr{I}(B^n)$ может рассматриваться также как цилиндр с основаниями в R^{n+1}, R^{n+2}, ..., поскольку

$$\mathscr{I}(B_1\times\ldots\times B_n)=\mathscr{I}(B_1\times\ldots\times B_n\times R),$$

$$\mathscr{I}(B^n)=\mathscr{I}(B^{n+1}),$$

где $B^{n+1}=B^n\times R$.

Отсюда следует, что как система цилиндров $\mathscr{I}(B_1\times\ldots\times B_n)$, так и система цилиндров $\mathscr{I}(B^n)$ образуют алгебры. Нетрудно проверить, что множества, составленные из объединений непересекающихся цилиндров $\mathscr{I}(I_1\times\ldots\times I_n)$, также образуют алгебру. Обозначим через $\mathscr{B}(R^\infty)$, $\mathscr{B}_1(R^\infty)$ и $\mathscr{B}_2(R^\infty)$ наименьшие σ-алгебры, содержащие все множества (8), (9) и (10) соответственно. (Часто σ-алгебру $\mathscr{B}_1(R^\infty)$ обозначают $\mathscr{B}(R)\otimes\mathscr{B}(R)\otimes\ldots$.) Понятно, что $\mathscr{B}(R^\infty)\subseteq\mathscr{B}_1(R^\infty)\subseteq\mathscr{B}_2(R^\infty)$. На самом же деле все эти три σ-алгебры совпадают.

Для доказательства обозначим для каждого $n=1, 2, \ldots$

$$\mathscr{C}_n=\{A\in R^n:\ \{x:\ (x_1, \ldots, x_n)\in A\}\in\mathscr{B}(R^\infty)\}.$$

Пусть $B^n\in\mathscr{B}(R^n)$. Тогда

$$B^n\in\mathscr{C}_n\subseteq\mathscr{B}(R^\infty).$$

Но $\mathscr{C}_n$ — σ-алгебра, а значит,

$$\mathscr{B}(R^n)\subseteq\sigma(\mathscr{C}_n)=\mathscr{C}_n\subseteq\mathscr{B}(R^\infty)$$

и, следовательно,

$$\mathscr{B}_2(R^\infty)\subseteq\mathscr{B}(R^\infty).$$

Итак, $\mathscr{B}(R^\infty)=\mathscr{B}_1(R^\infty)=\mathscr{B}_2(R^\infty)$.

В дальнейшем множества из $\mathscr{B}(R^\infty)$ будем называть борелевскими множествами (в R^∞).

Замечание. Пусть $\mathscr{B}_0(R^\infty)$ — наименьшая σ-алгебра, порожденная *открытыми* множествами в метрике

$$\rho_\infty(x, x^0) = \sum_{k=1}^{\infty} 2^{-k}\rho_1(x_k, x_k^0),$$

где $x = (x_1, x_2, \ldots)$, $x^0 = (x_1^0, x_2^0, \ldots)$. Тогда $\mathscr{B}(R^\infty) = \mathscr{B}_0(R^\infty)$ (задача 7).

Приведем несколько примеров борелевских множеств в R^∞:

(a) $\{x \in R^\infty\colon \sup x_n > a\}$,
$\{x \in R^\infty\colon \inf x_n < a\}$;

(b) $\{x \in R^\infty\colon \overline{\lim}\, x_n \leqslant a\}$,
$\{x \in R^\infty\colon \underline{\lim}\, x_n > a\}$,

где, как обычно,

$$\overline{\lim}\, x_n = \inf_n \sup_{m \geqslant n} x_m, \quad \underline{\lim}\, x_n = \sup_n \inf_{m \geqslant n} x_m;$$

(c) $\{x \in R^\infty\colon x_n \to\}$ —
множество тех $x \in R^\infty$, для которых $\lim x_n$ существует и конечен;

(d) $\{x \in R^\infty\colon \lim x_n > a\}$;

(e) $\left\{x \in R^\infty\colon \sum_{n=1}^{\infty} |x_n| > a\right\}$;

(f) $\left\{x \in R^\infty\colon \sum_{k=1}^{n} x_k = 0 \text{ по крайней мере для одного } n \geqslant 1\right\}$.

Чтобы убедиться, например, в том, что множества из (a) входят в систему $\mathscr{B}(R^\infty)$, достаточно заметить, что

$$\{x\colon \sup x_n > a\} = \bigcup_n \{x\colon x_n > a\} \in \mathscr{B}(R^\infty),$$

$$\{x\colon \inf x_n < a\} = \bigcup_n \{x\colon x_n < a\} \in \mathscr{B}(R^\infty).$$

5. Измеримое пространство $(R^T, \mathscr{B}(R^T))$, где T — произвольное множество. Пространство R^T — это совокупность действительных функций $x = (x_t)$, определенных для $t \in T$ *). В основном нас будет интересовать тот случай, когда T — некоторое несчетное подмножество числовой прямой. Для простоты и определенности можно сейчас предположить, что $T = [0, \infty)$.

---

*) В дальнейшем для функции из R^T используются также обозначения: $x = (x_t)_{t \in R^T}$, $x = (x_t)$, $t \in R^T$.

Введем в рассмотрение три типа цилиндрических множеств

$$\mathscr{I}_{t_1, \ldots, t_n}(I_1 \times \ldots \times I_n) = \{x: x_{t_1} \in I_1, \ldots, x_{t_n} \in I_n\}, \tag{11}$$

$$\mathscr{I}_{t_1, \ldots, t_n}(B_1 \times \ldots \times B_n) = \{x: x_{t_1} \in B_1, \ldots, x_{t_n} \in B_n\}, \tag{12}$$

$$\mathscr{I}_{t_1, \ldots, t_n}(B^n) = \{x: (x_{t_1}, \ldots, x_{t_n}) \in B^n\}, \tag{13}$$

где I_k — множества вида $(a_k, b_k]$, B_k — борелевские множества на числовой прямой, а B^n — борелевское множество в R^n.

Множество $\mathscr{I}_{t_1, \ldots, t_n}(I_1 \times \ldots \times I_n)$ есть не что иное, как множество тех функций, которые в моменты $t_1, \ldots, t_n$ «проходят через окна» $I_1, \ldots, I_n$, а в остальные моменты принимают произвольные значения (рис. 24).

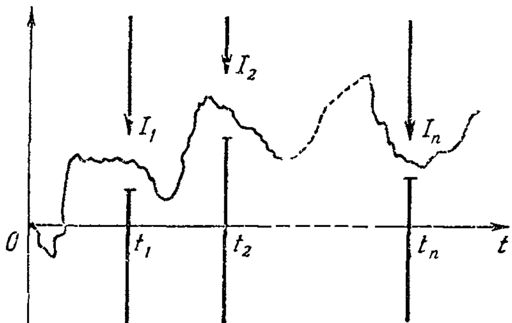


Рис. 24.

Обозначим через $\mathscr{B}(R^T)$, $\mathscr{B}_1(R^T)$ и $\mathscr{B}_2(R^T)$ наименьшие σ-алгебры, содержашие все цилиндрические множества (11), (12) и (13) соответственно. Ясно, что

$$\mathscr{B}(R^T) \subseteq \mathscr{B}_1(R^T) \subseteq \mathscr{B}_2(R^T). \tag{14}$$

На самом же деле все эти три σ-алгебры совпадают между собой. Более того, исчерпывающим образом можно описать и структуру их множеств.

Теорема 3. *Пусть T — любое несчетное множество. Тогда $\mathscr{B}(R^T) = \mathscr{B}_1(R^T) = \mathscr{B}_2(R^T)$, и любое множество $A \in \mathscr{B}(R^T)$ имеет следующую структуру: найдется не более чем счетное множество точек $t_1, t_2, \ldots$ из T и борелевское множество B из $\mathscr{B}(R^\infty)$ такое, что*

$$A = \{x: (x_{t_1}, x_{t_2}, \ldots) \in B\}. \tag{15}$$

Доказательство. Обозначим через $\mathscr{E}$ совокупность множеств вида (15) (при различных наборах $(t_1, t_2, \ldots)$ и множествах B из $\mathscr{B}(R^\infty)$). Если $A_1, A_2, \ldots \in \mathscr{E}$ и отвечающие им наборы есть $T^{(1)}=(t_1^{(1)}, t_2^{(1)}, \ldots)$, $T^{(2)}=(t_1^{(2)}, t_2^{(2)}, \ldots)$, то множество $T^{(\infty)}=\bigcup_k T^{(k)}$ можно взять в качестве единой системы такой, что все A_t будут представлены в виде

$$A_t=\{x\colon (x_{\tau_1}, x_{\tau_2}, \ldots)\in B_t\},$$

где B_t — некоторые множества из (одной и той же) σ-алгебры $\mathscr{B}(R^\infty)$, а $\tau_i\in T^{(\infty)}$.

Отсюда следует, что система множеств $\mathscr{E}$ образует σ-алгебру. Понятно, что эта σ-алгебра содержит все цилиндрические множества вида (1) и, поскольку $\mathscr{B}_2(R^T)$ есть наименьшая σ-алгебра, содержащая эти множества, то вместе с (14) это дает

$$\mathscr{B}(R^T)\subseteq\mathscr{B}_1(R^T)\subseteq\mathscr{B}_2(R^T)\subseteq\mathscr{E}. \tag{16}$$

Рассмотрим множество A из $\mathscr{E}$, представимое в виде (15). Если зафиксировать набор $(t_1, t_2, \ldots)$, то тогда те же рассуждения, что и в случае пространства $(R^\infty, \mathscr{B}(R^\infty))$, показывают, что множество A будет элементом σ-алгебры, порожденной цилиндрическими множествами (11). Но эта σ-алгебра, очевидно, принадлежит σ-алгебре $\mathscr{B}(R^T)$, что вместе с (16) и доказывает оба утверждения теоремы.

Итак, любое борелевское множество A из σ-алгебры $\mathscr{B}(R^T)$ определяется ограничениями, наложенными на функции $x=(x_t)$, $t\in T$, не более чем в счетном числе точек $t_1, t_2, \ldots$ Отсюда следует, в частности, что множества

$$A_1=\{x\colon \sup x_t<C \text{ для всех } t\in[0, 1]\},$$
$$A_2=\{x\colon x_t=0 \text{ по крайней мере для одного } t\in[0, 1]\},$$
$$A_3=\{x\colon x_t \text{ непрерывна в фиксированной точке } t_0\in[0, 1]\},$$

зависящие от «поведения» функций в несчетном числе точек, не обязаны быть борелевскими. И действительно, все три указанных множества *не принадлежат* $\mathscr{B}(R^{[0,1]})$.

Покажем это для множества A_1. Если $A_1\in\mathscr{B}(R^{[0,1]})$, то, согласно доказанной теореме, можно найти такие точки $(t_1^0, t_2^0, \ldots)$ и множество $B^0\in\mathscr{B}(R^\infty)$, что

$$\left\{x\colon \sup_t x_t<C,\ t\in[0, 1]\right\}=\left\{x\colon (x_{t_1^0}, x_{t_2^0}, \ldots)\in B^0\right\}.$$

Ясно, что функция $y_t\equiv C-1$ принадлежит A_1, и, следовательно, $(y_{t_1^0}, \ldots)\in B^0$. Образуем тогда функцию

Понятно, что

$$(y_{t_1^0}, y_{t_2^0}, \ldots) = (z_{t_1^0}, z_{t_2^0}, \ldots),$$

и, следовательно, функция $z = (z_t)$ принадлежит множеству $\{x: (x_{t_1^0}, \ldots) \in B^0\}$. Но в то же время ясно, что она не принадлежит множеству $\{x: \sup x_t < C\}$. Полученное противоречие показывает, что $A_1 \notin \mathscr{B}(R^{[0, 1]})$.

В связи с неизмеримостью множеств A_1, A_2 и A_3 по отношению к σ-алгебре $\mathscr{B}(R^{[0, 1]})$ в пространстве всех функций $x = (x_t)$, $t \in [0, 1]$, естественно рассмотреть более узкие классы функций, где эти множества могут оказаться измеримыми. Интуитивно понятно, что так будет, если в качестве исходного пространства рассмотреть, например, пространство непрерывных функций.

6. Измеримое пространство $(C, \mathscr{B}(C))$. Пусть $T = [0, 1]$ и C — пространство непрерывных функций $x = (x_t)$, $0 \leqslant t \leqslant 1$. Относительно равномерной метрики $\rho(x, y) = \sup\limits_{t \in T} |x_t - y_t|$ это пространство является метрическим. В C можно ввести две σ-алгебры: $\mathscr{B}(C)$ — σ-алгебру, порожденную цилиндрическими множествами, и σ-алгебру $\mathscr{B}_0(C)$, порожденную открытыми (в метрике $\rho(x, y)$) множествами. Покажем, что на самом деле обе эти σ-алгебры совпадают: $\mathscr{B}(C) = \mathscr{B}_0(C)$.

Пусть $B = \{x: x_{t_0} < b\}$ — некоторое цилиндрическое множество. Нетрудно убедиться, что это множество является открытым. Отсюда вытекает, что $\{x: x_{t_1} < b_1, \ldots, x_{t_n} < b_n\} \in \mathscr{B}_0(C)$ и, значит, $\mathscr{B}(C) \subseteq \mathscr{B}_0(C)$.

Обратно, рассмотрим множество $B_\rho = \{y: y \in S_\rho(x^0)\}$, где x^0 — некоторая функция из C и $S_\rho(x^0) = \{x \in C: \sup\limits_{t \in T} |x_t - x_t^0| < \rho\}$ — открытая сфера с центром в x^0. В силу непрерывности функций из C

$$B_\rho = \{y \in C: y \in S_\rho(x^0)\} = \{y \in C: \max_t |y_t - x_t^0| < \rho\} =$$

$$= \bigcap_{t_k} \left\{y \in C: \left|y_{t_k} - x_{t_k}^0\right| < \rho\right\} \in \mathscr{B}(C), \tag{17}$$

где t_k — рациональные точки отрезка $[0, 1]$. Поэтому $\mathscr{B}_0(C) \subseteq \mathscr{B}(C)$.

Следующим важным примером является

7. Измеримое пространство $(D, \mathscr{B}(D))$, где D — пространство функций $x = (x_t)$, $t \in [0, 1]$, являющихся непрерывными справа ($x_t = x_{t+}$ для всех $t < 1$) и имеющих пределы слева (в любой точке $t > 0$).

Так же, как и в случае пространства C, в D можно ввести метрику $d(x, y)$ так, что σ-алгебра $\mathscr{B}_0(D)$, порожденная открытыми множествами, будет совпадать с σ-алгеброй $\mathscr{B}(D)$, порожденной цилиндрическими множествами. Эта метрика $d(x, y)$, введенная

А. В. Скороходом, определяется следующим образом:

$$d(x, y) = \inf\{\varepsilon > 0:\ \exists\lambda \in \Lambda:\ \sup_t |x_t - y_{\lambda(t)}| + \sup_t |t - \lambda(t)| \leqslant \varepsilon\}, \tag{18}$$

где Λ — множество строго возрастающих непрерывных на $[0, 1]$ функций $\lambda = \lambda(t)$ с $\lambda(0) = 0$, $\lambda(1) = 1$.

8. Измеримое пространство $\left(\prod_{t \in T} \Omega_t, \boxed{\otimes}_{t \in T} \mathscr{F}_t\right)$. Наряду с пространством $(R^T, \mathscr{B}(R^T))$, являющимся прямым произведением T копий числовой прямой с системой борелевских множеств, в теории вероятностей рассматривают также измеримые пространства $\left(\prod_{t \in T} \Omega_t, \boxed{\otimes}_{t \in T} \mathscr{F}_t\right)$, образованные следующим образом.

Пусть T — произвольный набор индексов и $(\Omega_t, \mathscr{F}_t)$ — измеримые пространства, $t \in T$. Обозначим $\Omega = \prod_{t \in T} \Omega_t$ — множество всех функций $\omega = (\omega_t)$, $t \in T$, таких что $\omega_t \in \Omega_t$ для каждого $t \in T$.

Совокупность цилиндрических множеств

$$\mathscr{I}_{t_1, \ldots, t_n}(B_1 \times \ldots \times B_n) = \{\omega:\ \omega_{t_1} \in B_1, \ \ldots, \ \omega_{t_n} \in B_n\},$$

где $B_{t_i} \in \mathscr{F}_{t_i}$, образует, как нетрудно показать, алгебру. Наименьшую σ-алгебру, содержащую все цилиндрические множества, обозначают $\boxed{\otimes}_{t \in T} \mathscr{F}_t$, а измеримое пространство $\left(\prod \Omega_t, \boxed{\otimes} \mathscr{F}_t\right)$ называют *прямым произведением* измеримых пространств $(\Omega_t, \mathscr{F}_t)$, $t \in T$.

9. Задачи.

1. Пусть $\mathscr{B}_1$ и $\mathscr{B}_2$ — σ-алгебры подмножеств пространства Ω. Будут ли σ-алгебрами системы множеств

$$\mathscr{B}_1 \cap \mathscr{B}_2 \equiv \{A:\ A \in \mathscr{B}_1 \quad \text{и} \quad A \in \mathscr{B}_2\},$$
$$\mathscr{B}_1 \cup \mathscr{B}_2 \equiv \{A:\ A \in \mathscr{B}_1 \ \text{или} \ A \in \mathscr{B}_2\}?$$

2. Пусть $\mathscr{D} = \{D_1, D_2, \ldots\}$ — некоторое счетное разбиение Ω и $\mathscr{B} = \sigma(\mathscr{D})$. Будет ли число множеств, составляющих $\mathscr{B}$, также счетно?

3. Показать, что

$$\mathscr{B}(R^n) \otimes \mathscr{B}(R) = \mathscr{B}(R^{r+1}).$$

4. Доказать, что множества (b) — (f) (см. п. 4) принадлежат $\mathscr{B}(R^\infty)$.

5. Доказать, что множества A_2 и A_3 (см. п. 5) не принадлежат $\mathscr{B}(R^{[0, 1]})$.

6. Доказать, что функция (15) действительно задает метрику.

7. Доказать, что $\mathscr{B}_0(R^n)=\mathscr{B}(R^n)$, $n\geqslant 1$, и $\mathscr{B}_0(R^\infty)=\mathscr{B}(R^\infty)$.

8. Пусть $C=C[0, \infty)$ — пространство непрерывных функций $x=(x_t)$, определенных для $t\geqslant 0$. Показать, что относительно метрики

$$\rho(x, y)=\sum_{n=1}^{\infty} 2^{-n}\min\left[\sup_{0\leqslant t\leqslant n}|x_t-y_t|, 1\right], \quad x, y\in C$$

это пространство является полным сепарабельным метрическим пространством и σ-алгебра $\mathscr{B}_0(C)$, порожденная открытыми множествами, совпадает с σ-алгеброй $\mathscr{B}(C)$, порожденной цилиндрическими множествами.

§ 3. Способы задания вероятностных мер на измеримых пространствах

1. Измеримое пространство $(R, \mathscr{B}(R))$. Пусть $\mathsf{P}=\mathsf{P}(A)$ — вероятностная мера, определенная на борелевских множествах A числовой прямой. Возьмем $A=(-\infty, x]$, и положим

$$F(x)=\mathsf{P}(-\infty, x], \qquad x\in R. \tag{1}$$

Так определенная функция обладает следующими свойствами:

1) $F(x)$ — *неубывающая функция*;

2) $F(-\infty)=0$, $F(+\infty)=1$, *где*

$$F(-\infty)=\lim_{x\downarrow-\infty}F(x), \quad F(+\infty)=\lim_{x\uparrow\infty}F(x);$$

3) $F(x)$ *непрерывна справа и имеет пределы слева в каждой точке* $x\in R$.

Первое свойство очевидно, последние два вытекают из свойства непрерывности вероятностной меры.

Определение 1. Всякая функция $F=F(x)$, удовлетворяющая перечисленным условиям 1) — 3), называется *функцией распределения* (на числовой прямой R).

Итак, каждой вероятностной мере P на $(R, \mathscr{B}(R))$ соответствует (в силу (1)) некоторая функция распределения. Оказывается, что имеет место и обратное утверждение.

Теорема 1. *Пусть $F=F(x)$ — некоторая функция распределения на числовой прямой R. Тогда на $(R, \mathscr{B}(R))$ существует и притом единственная вероятностная мера P такая, что для любых $-\infty\leqslant a<b<\infty$*

$$\mathsf{P}(a, b]=F(b)-F(a). \tag{2}$$

Доказательство. Пусть $\mathscr{A}$ — алгебра множеств A из R, являющихся конечными суммами непересекающихся интервалов вида $(a, b]$:

$$A=\sum_{k=1}^{n}(a_k, b_k].$$

Определим на этих множествах функцию множеств P_0, полагая

$$\mathsf{P}_0(A)=\sum_{k=1}^{n}[F(b_k)-F(a_k)], \qquad A\in\mathscr{A}. \tag{3}$$

На алгебре $\mathscr{A}$ эта формула определяет и, очевидно, однозначно некоторую конечно-аддитивную функцию множеств. Поэтому, если показать, что на этой алгебре эта функция к тому же счетно-аддитивна, то существование и единственность требуемой меры P на $\mathscr{B}(R)$ будет непосредственно вытекать из следующего общего результата теории меры (приводимого без доказательства).

Теорема Каратеодори. *Пусть* Ω — *некоторое пространство,* $\mathscr{A}$ — *алгебра его подмножеств и* $\mathscr{B}=\sigma(\mathscr{A})$ — *наименьшая* σ*-алгебра, содержащая* $\mathscr{A}$. *Пусть* μ_0 — σ*-конечная мера на* $(\Omega, \mathscr{A})$. *Тогда существует и притом единственная мера* μ на $(\Omega, \mathscr{B}(\mathscr{A}))$, *являющаяся продолжением* μ_0, *т. е. такая, что*

$$\mu(A)=\mu_0(A), \qquad A\in\mathscr{A}.$$

Итак, покажем, что функция P_0 счетно-аддитивна на алгебре $\mathscr{A}$. Согласно теореме из § 1 для этого достаточно проверить непрерывность P_0 в $\varnothing$, т. е. проверить, что

$$\mathsf{P}_0(A_n)\downarrow 0, \qquad A_n\downarrow\varnothing, \qquad A_n\in\mathscr{A}.$$

Пусть $A_1, A_2, \ldots$ — некоторая выбранная последовательность множеств из $\mathscr{A}$ со свойством $A_n\downarrow\varnothing$. Предположим сначала, что все множества A_n принадлежат некоторому замкнутому интервалу $[-N, N]$, $N<\infty$. Поскольку A_n состоят из конечного числа сумм интервалов вида $(a, b]$ и поскольку в силу непрерывности справа функций $F(x)$

$$\mathsf{P}_0(a', b]=F(b)-F(a')\to F(b)-F(a)=\mathsf{P}_0(a, b]$$

при $a'\downarrow a$, то для каждого A_n найдется множество $B_n\in\mathscr{A}$ такое, что его замыкание $[B_n]\subseteq A_n$ и

$$\mathsf{P}_0(A_n)-\mathsf{P}_0(B_n)\leqslant\varepsilon\cdot 2^{-n},$$

где ε — некоторое заранее заданное число, большее нуля.

По предположению $\bigcap A_n=\varnothing$, а значит, и $\bigcap[B_n]=\varnothing$. Но множества $[B_n]$ замкнуты, поэтому найдется такое конечное $n_0=n_0(\varepsilon)$, что

$$\bigcap_{n=1}^{n_0}[B_n]=\varnothing. \tag{4}$$

(В самом деле, $[-N, N]$ — компакт, а система множеств $\{[-N, N]\diagdown[B_n]\}_{n\geqslant 1}$ образует *открытое покрытие* этого компакта. Тогда по лемме Гейне — Бореля существует конечное подпокрытие:

$$\bigcup_{n=1}^{n_0}([-N, N]\diagdown[B_n]) = [-N, N]$$

а значит, $\bigcap_{n=1}^{n_0}[B_n] = \varnothing$).

Учитывая (4) и то, что $A_{n_0} \subseteq A_{n_0-1} \subseteq \ldots \subseteq A_1$, находим

$$\mathsf{P}_0(A_{n_0}) = \mathsf{P}_0\left(A_{n_0}\diagdown\bigcap_{k=1}^{n_0} B_k\right) + \mathsf{P}_0\left(\bigcap_{k=1}^{n_0} B_k\right) =$$

$$= \mathsf{P}_0\left(A_{n_0}\diagdown\bigcap_{k=1}^{n_0} B_k\right) \leqslant \mathsf{P}_0\left(\bigcup_{k=1}^{n_0}(A_k\diagdown B_k)\right) \leqslant$$

$$\leqslant \sum_{k=1}^{n_0} \mathsf{P}_0(A_k\diagdown B_k) \leqslant \sum_{k=1}^{n_0} \varepsilon\cdot 2^{-k} \leqslant \varepsilon.$$

Поэтому $\mathsf{P}_0(A_n)\downarrow 0$, $n\to\infty$.

Откажемся теперь от предположения, что все $A_n \subseteq [-N, N]$ для некоторого N. Зададим $\varepsilon > 0$ и выберем такое N, что $\mathsf{P}_0[-N, N] > 1-\varepsilon/2$. Тогда, поскольку

$$A_n = A_n\cap[-N, N] + A_n\cap\overline{[-N, N]},$$

то

$$\mathsf{P}_0(A_n) = \mathsf{P}_0(A_n[-N, N] + \mathsf{P}_0(A_n\cap\overline{[-N, N]}) \leqslant$$
$$\leqslant \mathsf{P}_0(A_n\cap[-N, N]) + \varepsilon/2$$

и, применяя предшествующие рассуждения (с заменой A_n на $A_n\cap[-N, N]$), получаем, что для достаточно больших n $\mathsf{P}_0(A_n\cap$ $\cap[-N, N])\leqslant\varepsilon/2$. Тем самым снова $\mathsf{P}_0(A_n)\downarrow 0$, $n\to\infty$. Теорема доказана.

Итак, между вероятностными мерами P на $(R, \mathscr{B}(R))$ и функциями распределения F на числовой прямой R существует взаимно однозначное соответствие. Меру P, построенную по функции F, принято называть вероятностной мерой Лебега — Стилтьеса, отвечающей функции распределения F.

Особо важен случай, когда

$$F(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leqslant x \leqslant 1, \\ 1, & x > 1. \end{cases}$$

В этом случае соответствующую вероятностную меру (обозначим ее λ) называют *мерой Лебега* на отрезке $[0, 1]$. Ясно, что $\lambda(a, b] =$

$=b-a$. Иначе говоря, мера Лебега интервала $(a, b]$ (а также любого из интервалов (a, b), $[a, b]$, $[a, b)$) равна просто его длине $b-a$.

Обозначим

$$\mathscr{B}([0, 1]) = \{A \cap [0, 1]: A \in \mathscr{B}(R)\}$$

совокупность борелевских множеств отрезка $[0, 1]$. Наряду с этими множествами часто приходится рассматривать так называемые лебеговские множества отрезка $[0, 1]$. Будем говорить, что множество $\Lambda \subseteq [0, 1]$ относится к системе $\overline{\mathscr{B}}([0, 1])$, если можно найти такие борелевские множества A и B, что $A \subseteq \Lambda \subseteq B$ и $\lambda(B \setminus A) = 0$. Нетрудно проверить, что система $\overline{\mathscr{B}}([0, 1])$ является σ-алгеброй. Именно ее и называют *системой лебеговских множеств отрезка* $[0, 1]$. Ясно, что $\mathscr{B}([0, 1]) \subseteq \overline{\mathscr{B}}([0, 1])$.

Меру λ, определенную пока лишь на множествах из $\mathscr{B}([0, 1])$, естественным образом можно продолжить и на систему лебеговских множеств $\overline{\mathscr{B}}([0, 1])$. А именно, если $\Lambda \in \overline{\mathscr{B}}([0, 1])$ и $A \subseteq \Lambda \subseteq B$, где $A, B \in \mathscr{B}([0, 1])$, $\lambda(B \setminus A) = 0$, то положим $\lambda(\Lambda) = \lambda(A)$. Так определенная функция множеств $\lambda = \lambda(\Lambda)$, $\Lambda \in \overline{\mathscr{B}}([0, 1])$ является, как нетрудно проверить, вероятностной мерой на $([0, 1], \overline{\mathscr{B}}([0, 1]))$. Ее также называют *лебеговской мерой* (на системе лебеговских множеств).

Замечание. Проведенная процедура пополнения (продолжения) меры применяется и оказывается полезной не только в рассмотренном случае. Например, пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — некоторое вероятностное пространство. Обозначим через $\overline{\mathscr{F}}^{\mathsf{P}}$ совокупность всех подмножеств A пространства Ω, для которых можно найти также множества B_1 и B_2 из $\mathscr{F}$, что $B_1 \subseteq A \subseteq B_2$ и $\mathsf{P}(B_2 \setminus B_1) = 0$. Естественным образом (с помощью равенства $\mathsf{P}(A) = P(B_1)$) вероятностная мера определяется и для множеств $A \in \overline{\mathscr{F}}^{\mathsf{P}}$. Полученное таким образом новое вероятностное пространство $(\Omega, \overline{\mathscr{F}}^{\mathsf{P}}, \mathsf{P})$ называется пополнением пространства $(\Omega, \mathscr{F}, \mathsf{P})$ относительно меры P.

Если вероятностная мера P такова, что $\overline{\mathscr{F}}^{\mathsf{P}} = \mathscr{F}$, то она называется *полной*, а соответствующее пространство $(\Omega, \mathscr{F}, \mathsf{P})$ — *полным вероятностным пространством*.

Устанавливаемое равенством $\mathsf{P}(a, b] = F(b) - F(a)$ соответствие между вероятностными мерами P и функциями распределения F дает возможность конструирования разных вероятностных мер с помощью задания соответствующих функций распределения.

Дискретные меры. Пусть функция распределения $F = F(x)$ является (рис. 25) кусочно-постоянной, меняющей свои значения в точках $x_1, x_2, \ldots$ ($\Delta F(x_i) > 0$, где $\Delta F(x) = F(x) - F(x-)$). Соответствующая этой функции вероятностная мера P сосредоточена в точках $x_1, x_2, \ldots$:

$$\mathsf{P}(\{x_k\}) = \Delta F(x_k) > 0, \qquad \sum_k \mathsf{P}(\{x_k\}) = 1.$$

Набор чисел $(p_1, p_2, \ldots)$, где $p_k = \mathsf{P}(\{x_k\})$, называют *дискретным распределением вероятностей*.

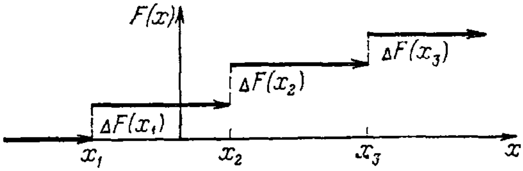


Рис. 25.

Приведем таблицу наиболее употребительных типов дискретных вероятностных распределений с соответствующими наименованиями (табл. 1).

Таблица 1

Распределение	Вероятности p_k	Параметры
Дискретное равномерное	$\frac{1}{N}$, $k = 1, 2, \ldots, N$	$N = 1, 2, \ldots$
Бернуллиевское	$p_1 = p$, $p_0 = q$	$0 \leqslant p \leqslant 1$, $q = 1 - p$
Биномиальное	$C_n^k p^k q^{n-k}$, $k = 0, 1, \ldots, n$	$0 \leqslant p \leqslant 1$, $q = 1 - p$, $n = 1, 2, \ldots$
Пуассоновское	$\frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, \ldots$	$\lambda > 0$
Геометрическое	$q^{k-1}p$, $k = 1, 2, \ldots$	$0 < p \leqslant 1$, $q = 1 - p$
Отрицательно-биномиальное	$C_{k-1}^{r-1} p^r q^{k-r}$, $k = r, r+1, \ldots$	$0 < p \leqslant 1$, $q = 1 - p$, $r = 1, 2, \ldots$

Абсолютно-непрерывные меры. Пусть существует неотрицательная функция $f = f(t)$, $t \in R$, такая, что

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \tag{5}$$

где под интегралом сейчас понимается интеграл в смысле Римана, а в общем случае — в смысле Лебега (см. § 6).

Функция $f = f(x)$, $x \in R$, называется *плотностью* функции распределения $F = F(x)$ (плотностью распределения вероятностей или просто плотностью).

Понятно, что всякая неотрицательная функция $f=f(x)$, интегрируемая по Риману и такая, что $\int_{-\infty}^{\infty} f(x)\,dx=1$, определяет формулой (5) некоторую функцию распределения. В табл. 2 приведены особо важные для теории вероятностей и математической статистики примеры разных типов плотностей $f=f(x)$ с указанием их наименований и параметров (плотность $f(x)$ считается равной нулю для неуказанных в таблице значений x).

Сингулярные меры. Так называют меры, функции распределения которых непрерывны, но точки их роста образуют множество *нулевой меры Лебега*. Не останавливаясь подробно на этом случае, ограничимся лишь примером такой функции.

Возьмем отрезок [0, 1] и построим функцию $F(x)$ с помощью следующего приема, принадлежащего Г. Кантору.

Разделим отрезок [0, 1] на три равные части и положим (рис. 26)

$$F_1(x)=\begin{cases} 1/2 & x\in(1/3,\ 2/3), \\ 0 & x=0, \\ 1 & x=1, \end{cases}$$

доопределяя ее в остальных точках с помощью линейной интерполяции.

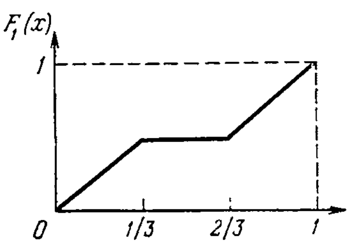


Рис. 26

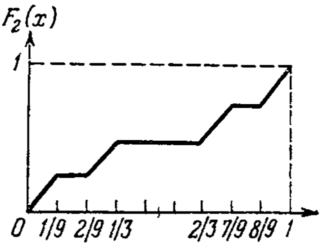


Рис. 27.

Далее, каждый из интервалов [0, 1/3] и [2/3, 1] снова делим на три части и определяем функцию (рис. 27)

$$F_2(x)=\begin{cases} 1/2 & x\in(1/3,\ 2/3), \\ 1/4 & x\in(1/9,\ 2/9), \\ 3/4 & x\in(7/9,\ 8/9), \\ 0 & x=0, \\ 1 & x=1 \end{cases}$$

Таблица

Тип распределения	Плотность	Параметры
Равномерное на $[a, b]$	$\frac{1}{b-a}, \quad a \leqslant x \leqslant b$	$a, b \in R; \quad a < b$
Нормальное, или гауссовское	$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in R$	$m \in R, \quad \sigma > 0$
Гамма	$\frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, \quad x \geqslant 0$	$\alpha > 0, \quad \beta > 0$
Бета	$\frac{x^{r-1}(1-x)^{s-1}}{\beta(r, s)}, \quad 0 \leqslant x \leqslant 1$	$r > 0, \quad s > 0$
Экспоненциальное (гамма-распределение с $\alpha = 1$, $\beta = 1/\lambda$)	$\lambda e^{-\lambda x}, \quad x \geqslant 0$	$\lambda > 0$
Двустороннее экспоненциальное распределение	$\frac{\lambda}{2} e^{-\lambda \lvert x \rvert}, \quad x \in R$	$\lambda > 0$
Хи-квадрат, χ^2 (гамма распределение с $\alpha = n/2$, $\beta = 2$)	$\frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}},$ $x \geqslant 0$	$n = 1, 2, \ldots$
Стьюдента, t	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}},$ $x \in R$	$n = 1, 2, \ldots$
F	$\frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{\left(1+\frac{mx}{n}\right)^{\frac{m+n}{2}}}$ $x \geqslant 0$	$m, n = 1, 2, \ldots$
Коши	$\frac{\theta}{\pi(x^2+\theta^2)}, \quad x \in R$	$\theta > 0.$

со значениями в остальных точках, полученными линейной интерполяцией.

Продолжая этот процесс, построим последовательность функций $F_n(x)$, $n=1, 2, \ldots$, которые сходятся к некоторой неубывающей непрерывной функции $F(x)$ (называемой канторовской), точки роста которой (x — точка роста $F(x)$, если $F(x+\varepsilon)-F(x-\varepsilon)>0$ для любого $\varepsilon>0$) сбразуют множество лебеговской меры нуль. Действительно, из конструкции $F(x)$ видно, что общая длина интервалов $(1/3, 2/3)$, $(1/9, 2/9)$, $(7/9, 8/9)$, ..., на которых функция принимает постоянные значения, равна

$$\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\ldots=\frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n=1. \tag{6}$$

Обозначим через $\mathscr{N}$ множество точек роста канторовской функции $F(x)$. Из (6) следует, что $\lambda(\mathscr{N})=0$. В то же самое время, если μ — мера, соответствующая канторовской функции $F(x)$, то $\mu(\mathscr{N})=1$. (В этом случае говорят, что мера *сингулярна* по отношению к лебеговской мере λ.)

Не останавливаясь более на вопросе о возможных типах функций распределения, ограничимся лишь замечанием о том, что на самом деле указанными *тремя типами* исчерпываются все функции. Точнее, произвольная функция распределения может быть представлена в виде $p_1F_1+p_2F_2+p_3F_3$, где F_1 — дискретная, F_2 — абсолютно непрерывная, F_3 — сингулярная функции распределения, p_i — неотрицательные числа, $p_1+p_2+p_3=1$.

2. Теорема 1 устанавливает взаимно однозначное соответствие между вероятностными мерами на $(R, \mathscr{B}(R))$ и функциями распределения на R. Анализ доказательства этой теоремы показывает, что на самом деле справедлив более общий результат, позволяющий, в частности, ввести так называемую меру Лебега на всей числовой прямой.

Пусть μ — некоторая σ-конечная мера на $(\Omega, \mathscr{A})$, где $\mathscr{A}$ — алгебра подмножеств Ω. Оказывается, что утверждение теоремы Каратеодори о продолжении меры μ с алгебры $\mathscr{A}$ на наименьшую σ-алгебру $\sigma(\mathscr{A})$ остается справедливым и для σ-конечных мер, что и дает возможность обобщения теоремы 1.

Назовем *мерой Лебега — Стилтьеса* на $(R, \mathscr{B}(R))$ всякую (счетно-аддитивную) меру μ такую, что для любого ограниченного интервала I его мера $\mu(I)<\infty$. *Обобщенной функцией распределения* на числовой прямой R назовем всякую неубывающую непрерывную справа функцию $G=G(x)$ со значениями в $(-\infty, \infty)$.

Теорема 1 допускает обобщение в том смысле, что формула

$$\mu(a, b]=G(b)-G(a), \qquad a<b,$$

снова устанавливает взаимно однозначное соответствие между мерами Лебега — Стилтьеса μ и обобщенными функциями распределения G.

В самом деле, если $G(+\infty)-G(-\infty)<\infty$, то доказательство, примененное в теореме 1, проходит без всяких изменений, поскольку этот случай сводится к случаю, когда $G(+\infty)-G(-\infty)=1$ и $G(-\infty)=0$.

Пусть теперь $G(+\infty)-G(-\infty)=\infty$. Положим

$$G_n(x)=\begin{cases} G(x), & |x|\leqslant n, \\ G(n), & x=n, \\ G(-n), & x=-n. \end{cases}$$

Определим на алгебре $\mathscr{A}$ конечно-аддитивную меру μ_0 так, что $\mu_0(a, b]=G(b)-G(a)$, и пусть μ_n — уже построенные (по теореме 1) счетно-аддитивные меры, соответствующие функциям $G_n(x)$.

Очевидно, что на $\mathscr{A}$ $\mu_n\uparrow\mu_0$. Пусть теперь $A_1, A_2, \ldots$ — непересекающиеся множества из $\mathscr{A}$ и $A\equiv\sum A_n\in\mathscr{A}$. Тогда (задача 6 из § 1)

$$\mu_0(A)\geqslant\sum_{n=1}^{\infty}\mu_0(A_n).$$

И если $\sum\limits_{n=1}^{\infty}\mu_0(A_n)=\infty$, то $\mu_0(A)=\sum\limits_{n=1}^{\infty}\mu_0(A_n)$. Предположим теперь, что $\sum\mu_0(A_n)<\infty$. Тогда

$$\mu_0(A)=\lim_n\mu_n(A)=\lim_n\sum_{k=1}^{\infty}\mu_n(A_k).$$

Согласно сделанному предположению $\sum\mu_0(A_n)<\infty$. Поэтому

$$0\leqslant\mu_0(A)-\sum_{k=1}^{\infty}\mu_0(A_k)=\lim_n\left[\sum_{k=1}^{\infty}(\mu_n(A_k)-\mu_0(A_k))\right]\leqslant 0,$$

поскольку $\mu_n\leqslant\mu_0$.

Итак, σ-конечная конечно-аддитивная мера μ_0 является счетно-аддитивной на $\mathscr{A}$, и, значит, (по теореме Каратеодори) она может быть продолжена до счетно-аддитивной меры μ на $\sigma(\mathscr{A})$.

Особо важен тот случай, когда $G(x)=x$. Отвечающая этой обобщенной функции распределения мера λ называется мерой Лебега на $(R, \mathscr{B}(R))$. Как и в случае отрезка $[0, 1]$ на числовой прямой R, вводится система лебеговских множеств $\overline{\mathscr{B}}(R)$ ($\Lambda\in\overline{\mathscr{B}}(R)$, если существуют борелевские множества A и B такие, что $A\subseteq\Lambda\subseteq B$, $\lambda(B\setminus A)=0$), для которых определяется также лебеговская мера $\bar\lambda$ ($\bar\lambda(\Lambda)=\lambda(A)$, если $A\subseteq\Lambda\subseteq B$, $\Lambda\in\overline{\mathscr{B}}(R)$ и $\lambda(B\setminus A)=0$).

3. Измеримое пространство $(R^n, \mathscr{B}(R^n))$. Как и в случае действительной прямой, предположим, что P — некоторая вероятностная мера на $(R^n, \mathscr{B}(R^n))$.

Обозначим

$$F_n(x_1, \ldots, x_n) = \mathsf{P}((-\infty, x_1] \times \ldots \times (-\infty, x_n]),$$

или, в более компактной форме,

$$F_n(x) = \mathsf{P}(-\infty, x],$$

где $x = (x_1, \ldots, x_n)$, $(-\infty, x] = (-\infty, x_1] \times \ldots \times (-\infty, x_n]$.

Введем разностный оператор $\Delta_{a_i, b_i}: R^n \to R$, действующий по формуле $(a_i \leqslant b_i)$

$$\Delta_{a_i, b_i} F_n(x_1, \ldots, x_n) = F_n(x_1, \ldots, x_{i-1}, b_i, x_{i+1} \ldots) - \\ - F_n(x_1, \ldots, x_{i-1}, a_i, x_{i+1} \ldots)$$

Простой подсчет показывает, что

$$\Delta_{a_1 b_1} \ldots \Delta_{a_n b_n} F_n(x_1 \ldots x_n) = \mathsf{P}(a, b], \tag{7}$$

где $(a, b] = (a_1, b_1] \times \ldots \times (a_n, b_n]$. Отсюда, в частности, видно, что, в отличие от одномерного случая, вероятность $\mathsf{P}(a, b]$, вообще говоря, не равна разности $F_n(b) - F_n(a)$.

Поскольку $\mathsf{P}(a, b] \geqslant 0$, то из (7) следует, что для любых $a = (a_1 \ldots a_n)$, $b = (b_1 \ldots b_n)$,

$$\Delta_{a_1 b_1} \ldots \Delta_{a_n b_n} F_n(x_1, \ldots, x_n) \geqslant 0. \tag{8}$$

Из непрерывности вероятности P вытекает также, что функция $F_n(x_1, \ldots, x_n)$ непрерывна справа по совокупности переменных, т. е. если $x^{(k)} \downarrow x$, $x^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})$, то

$$F_n(x^{(k)}) \downarrow F_n(x), \quad k \to \infty. \tag{9}$$

Ясно также, что

$$F_n(+\infty, \ldots, +\infty) = 1 \tag{10}$$

и

$$\lim_{x \downarrow y} F_n(x_1, \ldots, x_n) = 0, \tag{11}$$

если по крайней мере одна из координат у y принимает значение $-\infty$.

Определение 2. Всякую функцию $F = F_n(x_1 \ldots x_n)$, удовлетворяющую условиям (8) — (11), будем называть *n-мерной функцией распределения* (в пространстве R^n).

Используя те же самые рассуждения, что и в теореме 1, можно доказать справедливость следующего результата.

Теорема 2. *Пусть $F = F_n(x_1, \ldots, x_n)$ — некоторая функция распределения в R^n. Тогда на $(R^n, \mathscr{B}(R^n))$ существует и притом единственная вероятностная мера* P *такая, что*

$$\mathsf{P}(a, b] = \Delta_{a_1 b_1} \ldots \Delta_{a_n b_n} F_n(x_1, \ldots, x_n). \tag{12}$$

Приведем некоторые примеры n-мерных функций распределения.

Пусть $F^1, \ldots, F^n$ — одномерные функции распределения (на R) и

$$F_n(x_1, \ldots, x_n) = F^1(x_1) \ldots F^n(x_n).$$

Ясно, что эта функция непрерывна справа и удовлетворяет условиям (10), (11). Нетрудно проверить также, что

$$\Delta_{a_1 b_1} \ldots \Delta_{a_n b_n} F_n(x_1, \ldots, x_n) = \prod [F^k(b_k) - F^k(a_k)] \geqslant 0.$$

Следовательно, $F_n(x_1, \ldots, x_n)$ — некоторая функция распределения.

Особо важен случай, когда

$$F^k(x_k) = \begin{cases} 0, & x_k < 0, \\ x_k, & 0 \leqslant x_k \leqslant 1, \\ 1, & x_k > 1. \end{cases}$$

В этом случае для всех $0 \leqslant x_k \leqslant 1$, $k = 1, \ldots, n$,

$$F_n(x_1, \ldots, x_n) = x_1 \ldots x_n.$$

Соответствующую этой n-мерной функции распределения вероятностную меру называют *n-мерной мерой Лебега на* $[0, 1]^n$.

Большой запас n-мерных функций распределения получается в виде

$$F_n(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_n(t_1, \ldots, t_n)\, dt_1 \ldots dt_n,$$

где $f_n(t_1, \ldots, t_n)$ — неотрицательные функции с

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_n(t_1, \ldots, t_n)\, dt_1 \ldots dt_n = 1,$$

а интегралы понимаются в смысле Римана (и в более общем случае — в смысле Лебега). Функции $f = f_n(t_1, \ldots, t_n)$ называют *плотностями* n-мерной функции распределения, n-мерной плотностью распределения вероятностей, или просто n-мерными плотностями.

В случае $n = 1$ функция

с $\sigma > 0$ есть плотность (невырожденного) *гаусссвского*, или *нормального распределения.* Существуют естественные аналоги этой плотности и в случае $n > 1$.

Пусть $\mathbb{R} = \|r_{ij}\|$ — некоторая неотрицательно определенная симметрическая матрица порядка $n \times n$:

$$\sum_{i,\, j=1}^{n} r_{ij}\lambda_i\lambda_j \geqslant 0, \quad \lambda_i \in \mathbb{R}, \quad i = 1, \ldots, n,$$
$$r_{ij} = r_{ji}$$

В том случае, когда $\mathbb{R}$ — положительно определенная матрица, ее $|\mathbb{R}| \equiv \det \mathbb{R} > 0$, и, следовательно, определена обратная матрица $A = \|a_{ij}\|$.

Тогда функция

$$f_n(x_1, \ldots, x_n) = \frac{|A|^{1/2}}{(2\pi)^{n/2}} e^{-1/2 \sum a_{ij}(x_i - m_i)(x_j - m_j)}, \tag{13}$$

где $m_i \in R$, $i = 1, \ldots, n$, обладает тем свойством, что интеграл (Римана) от нее по всему пространству равен 1 (это будет доказано в § 13) и, следовательно в силу ее положительности она является плотностью.

Эта функция называется *плотностью n-мерного* (невырожденного) *гауссовского*, или *нормального распределения* (с вектором средних значений $m = (m_1, \ldots, m_n)$ и матрицей ковариаций $\mathbb{R} = A^{-1}$).

В случае $n = 2$ плотность $f_2(x_1, x_2)$ может быть приведена к виду

$$f_2(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \times \left[\frac{(x_1-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\}, \tag{14}$$

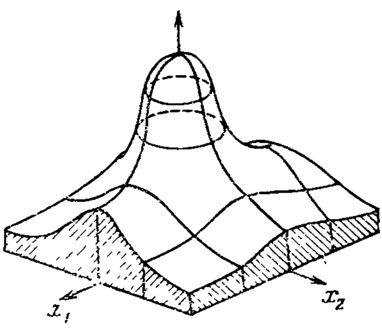


Рис. 28. Плотность двумерного нормального распределения.

где $\sigma_i > 0$, $|\rho| < 1$. (Смысл параметров m_i, σ_i и ρ будет объяснен в § 8.) Приводимый рис. 28 дает представление о виде двумерной гауссовской плотности.

Замечание. Как и в случае $n = 1$, теорема 2 допускает обобщение на (аналогичным образом определяемые) меры Лебега — Стилтьеса в $(R^n, \mathscr{B}(R^n))$ и обобщенные функции распределения в R^n. В том случае, когда обобщенная функция распределения $G_n(x_1 \ldots x_n)$ равна $x_1 \ldots x_n$, соответствующая мера называется *мерой Лебега* на борелевских множествах пространства R^n. Ясно, что для нее

$$\lambda(a,\ b] = \prod_{i=1}^{n} (b_i - a_i),$$

т. е. мера Лебега «прямоугольника»

$$(a,\ b] = (a_1,\ b_1] \times \ldots \times (a_n,\ b_n]$$

равна его «объему».

4. **Измеримое пространство** $(R^\infty,\ \mathscr{B}(R^\infty))$. В случае пространств R^n, $n \geqslant 1$, вероятностные меры строились по следующей схеме: сначала для элементарных множеств — прямоугольников вида $(a,\ b]$, затем естественным образом на множествах вида $A = \sum (a_i,\ b_i]$ и, наконец, с помощью теоремы Каратеодори — на множествах из $\mathscr{B}(R^n)$.

Аналогичная схема построения вероятностных мер «работает» и в случае пространства $(R^\infty,\ \mathscr{B}(R^\infty))$.

Обозначим через

$$\mathscr{I}_n(B) = \{x \in R^\infty\colon\ (x_1,\ \ldots,\ x_n) \in B\}, \qquad B \in \mathscr{B}(R^n),$$

цилиндрическое множество в пространстве R^∞ с «основанием» $B \in \mathscr{B}(R^n)$. Как мы сейчас увидим, именно цилиндрические множества естественно считать теми *элементарными множествами* в R^∞, по значениям вероятностей которых определяется вероятностная мера на множествах из $\mathscr{B}(R^\infty)$.

Пусть P — некоторая вероятностная мера на $(R^\infty,\ \mathscr{B}(R^\infty))$. Обозначим для $n = 1,\ 2,\ \ldots$

$$P_n(B) = \mathsf{P}(\mathscr{I}_n(B)), \qquad B \in \mathscr{B}(R^n). \tag{15}$$

Последовательность вероятностных мер $P_1,\ P_2,\ \ldots,$ определенных соответственно на $(R,\ \mathscr{B}(R))$, $(R^2,\ \mathscr{B}(R^2))$, $\ldots$, обладает следующим очевидным *свойством согласованности*: для любого $n = 1,\ 2,\ \ldots$ и $B \in \mathscr{B}(R^n)$

$$P_{n+1}(B \times R) = P_n(B). \tag{16}$$

Весьма примечательно, что имеет место и обратный результат.

Теорема 3 (теорема Колмогорова о продолжении меры в $(R^\infty,\ \mathscr{B}(R^\infty))$). *Пусть $P_1,\ P_2,\ \ldots$ — последовательность вероятностных мер на $(R,\ \mathscr{B}(R))$, $(R^2,\ \mathscr{B}(R^2))$, $\ldots$, обладающих свойством согласованности* (16). *Тогда существует и притом единственная вероятностная мера P на $(R^\infty,\ \mathscr{B}(R^\infty))$ такая, что для каждого $n = 1,\ 2,\ \ldots$*

Доказательство. Пусть $B^n \in \mathscr{B}(R^n)$ и $\mathscr{J}_n(B^n)$ — цилиндр с «основанием» B^n. Припишем этому цилиндру меру $\mathsf{P}(\mathscr{J}_n(B^n))$, полагая $\mathsf{P}(\mathscr{J}_n(B^n)) = P_n(B^n)$.

Покажем, что в силу условия согласованности такое определение является корректным, т. е. значение $P(\mathscr{J}_n(B^n))$ не зависит от способа представления цилиндрического множества $\mathscr{J}_n(B^n)$. В самом деле, пусть один и тот же цилиндр представлен двумя способами:

$$\mathscr{J}_n(B^n) = \mathscr{J}_{n+k}(B^{n+k}).$$

Отсюда следует, что если $(x_1, \ldots, x_{n+k}) \in R^{n+k}$, то

$$(x_1, \ldots, x_n) \in B^n \Leftrightarrow (x_1, \ldots, x_{n+k}) \in B^{n+k}, \tag{18}$$

и, значит, в силу (16) и (18)

$$\begin{aligned} P_n(B^n) &= P_{n+1}((x_1, \ldots, x_{n+1}) : (x_1, \ldots, x_n) \in B^n) = \ldots = \\ &= P_{n+k}((x_1, \ldots, x_{n+k}) : (x_1, \ldots, x_n) \in B^n) = \\ &= P_{n+k}(B^{n+k}). \end{aligned}$$

Обозначим $\mathscr{A}(R^\infty)$ совокупность всех цилиндрических множеств $\hat{B}^n = \mathscr{J}_n(B^n)$, $B^n \in \mathscr{B}(R^n)$, $n = 1, 2, \ldots$

Пусть теперь $\hat{B}_1, \ldots, \hat{B}_k$ — непересекающиеся множества из $\mathscr{A}(R^\infty)$. Без ограничения общности можно считать, что все они таковы, что для некоторого n $\hat{B}_i = \mathscr{J}_n(B_i^n)$, $i = 1, \ldots, k$, где $B_1^n, \ldots, B_k^n$ — непересекающиеся множества из $\mathscr{B}(R^n)$. Тогда

$$\mathsf{P}\left(\sum_{i=1}^k \hat{B}_i\right) = \mathsf{P}\left(\sum_{i=1}^k \mathscr{J}_n(B_i^n)\right) = P_n\left(\sum_{i=1}^k B_i^n\right) = \sum_{i=1}^n P_n(B_i^n) = \sum_{i=1}^n \mathsf{P}(\hat{B}_i),$$

т. е. функция множеств P конечно-аддитивна на алгебре $\mathscr{A}(R^\infty)$.

Покажем, что P непрерывна в «нуле», т. е. если последовательность множеств $\hat{B}_n \downarrow \varnothing$, $n \to \infty$, то $\mathsf{P}(\hat{B}_n) \to 0$, $n \to \infty$. Предположим противное, т. е. пусть $\lim\limits_n P(\hat{B}_n) = \delta > 0$. Без ограничения общности можно считать, что последовательность $\{\hat{B}_n\}$ такова, что

$$\hat{B}_n = \{x\colon (x_1, \ldots, x_n) \in B_n\}, \quad B_n \in \mathscr{B}(R^n).$$

Воспользуемся следующим свойством (см. задачу 9) вероятностных мер P_n на $(R^n, \mathscr{B}(R^n))$: если $B_n \in \mathscr{B}(R^n)$, то для заданного $\delta > 0$ можно найти такой компакт $A_n \in \mathscr{B}(R^n)$, что $A_n \subseteq B_n$ и

$$P_n(B_n \setminus A_n) \leqslant \delta / 2^{n+1}.$$

Поэтому, если

$$\hat{A}_n = \{x\colon (x_1, \ldots, x_n) \in A_n\},$$

то

$$\mathsf{P}(\hat{B}_n \setminus A_n) = P_n(B_n \setminus A_n) \leqslant \delta/2^{n+1}.$$

Образуем множество $\hat{C}_n = \bigcap_{k=1}^{n} A_k$, и пусть C_n таковы, что

$$\hat{C}_n = \{x\colon (x_1, \ldots, x_n) \in C_n\}.$$

Тогда, учитывая, что множества $\hat{B}_n$ убывают, находим

$$\mathsf{P}(\hat{B}_n \setminus \hat{C}_n) \leqslant \sum_{k=1}^{n} P(\hat{B}_n \setminus \hat{A}_k) \leqslant \sum_{k=1}^{n} \mathsf{P}(\hat{B}_k \setminus \hat{A}_k) \leqslant \delta/2.$$

Но по предположению $\lim_n \mathsf{P}(\hat{B}_n) = \delta > 0$, значит, $\lim_n \mathsf{P}(\hat{C}_n) \geqslant \delta/2 > 0$. Покажем, что это противоречит тому, что $\hat{C}_n \downarrow \varnothing$.

Действительно, выберем в множествах $\hat{C}_n$ по точке $\hat{x}^{(n)} = \left(x_1^{(n)}, x_2^{(n)}, \ldots\right)$. Тогда для каждого $n \geqslant 1$ $\left(x_1^{(n)}, \ldots, x_n^{(n)}\right) \in C_n$.

Пусть (n_1) — некоторая подпоследовательность последовательности (n) такая, что $x_1^{(n_1)} \to x_1^0$, где x_1^0 — некоторая точка в C_1. (Такая подпоследовательность существует, поскольку все $x_1^{(n)} \in C_1$, а C_1 — компакт). Из последовательности (n_1) выберем подпоследовательность (n_2) такую, что $\left(x_1^{(n_2)}, x_2^{(n_2)}\right) \to \left(x_1^0, x_2^0\right) \in C_2$. Аналогичным образом пусть $\left(x_1^{(n_k)}, \ldots, x_k^{(n_k)}\right) \to \left(x_1^0, \ldots, x_k^0\right) \in C_k$. Образуем, наконец, диагональную последовательность (m_k), где m_k есть k-й член в последовательности (n_k). Тогда для любого $i = 1, 2, \ldots$ $x_i^{(m_k)} \to x_i^0$ при $m_k \to \infty$, причем точка $\left(x_1^0, x_2^0, \ldots\right) \in \hat{C}_n$ для любого $n = 1, 2, \ldots$, что, очевидно, противоречит предположению о том, что $\hat{C}_n \downarrow \varnothing$, $n \to \infty$. Теорема доказана.

Замечание. В рассмотренном сейчас случае пространство R^∞ есть счетное произведение прямых, $R^\infty = R \times R \times \ldots$ Естественно поставить вопрос о том, а верна ли теорема 3 для случая, когда вместо $(R^\infty, \mathscr{B}(R^\infty))$ берется прямое произведение измеримых пространств $(\Omega_i, \mathscr{F}_i)$, $i = 1, 2, \ldots$

В приведенном выше доказательстве можно усмотреть, что единственное свойство числовой прямой топологического характера, которое было существенно использовано, состояло в том, что в любом множестве из $\mathscr{B}(R^n)$ можно найти компакт, вероятностная мера которого сколь угодно близка к вероятностной мере этого множества. Известно, однако, что это свойство присуще не только пространствам $(R^n, \mathscr{B}(R^n))$, но и любым полным сепарабельным метрическим пространствам с σ-алгебрами, порожденными открытыми множествами.

Таким образом, теорема 3 остается справедливой, если считать, что $P_1, P_2, \ldots$ — последовательность согласованных вероятностных мер на $(\Omega_1, \mathscr{F}_1)$, $(\Omega_1\times\Omega_2, \mathscr{F}_1\otimes\mathscr{F}_2), \ldots$, где $(\Omega_i, \mathscr{F}_i)$ — полные сепарабельные метрические пространства с σ-алгебрами $\mathscr{F}_i$, порожденными открытыми множествами, а вместо $(R^\infty, \mathscr{B}(R^\infty))$ рассмотреть пространство $(\Omega_1\times\Omega_2\times\ldots, \mathscr{F}_1\otimes\mathscr{F}_2\otimes\ldots)$.

В § 9 (теорема 2) будет показано, что результат теоремы 3 также остается справедливым и в случае произвольных измеримых пространств $(\Omega_i, \mathscr{F}_i)$, если меры P_n сконструированы некоторым специальным образом. В общем же случае (без каких-либо предположений топологического характера о структуре рассматриваемых измеримых пространств или о структуре семейства мер $\{P_n\}$) теорема 3 может быть и неверна, что показывает следующий пример.

Рассмотрим пространство $\Omega=(0, 1]$, которое, очевидно, не является полным, и построим в нем последовательность σ-алгебр $\mathscr{F}_1\subseteq\mathscr{F}_2\subseteq\ldots$ по следующей схеме. Пусть для всех $n=1, 2, \ldots$

$$\varphi_n(\omega)=\begin{cases}1, & 0<\omega<1/n,\\ 0, & 1/n\leqslant\omega\leqslant 1,\end{cases}$$

$$\mathscr{C}_n=\{A\in\Omega:\ A=\{\omega:\ \varphi_n(\omega)\in B\},\ B\in\mathscr{B}(R)\}$$

и $\mathscr{F}_n=\sigma\{\mathscr{C}_1, \ldots, \mathscr{C}_n\}$ — наименьшая σ-алгебра, содержащая системы множеств $\mathscr{C}_1, \ldots, \mathscr{C}_n$. Ясно, что $\mathscr{F}_1\subseteq\mathscr{F}_2\subseteq\ldots$ Пусть $\mathscr{F}=\sigma(\bigcup\mathscr{F}_n)$ — наименьшая σ-алгебра, содержащая все $\mathscr{F}_n$. Рассмотрим измеримое пространство $(\Omega, \mathscr{F}_n)$ и определим на нем вероятностную меру P_n следующим образом:

$$P_n\{\omega:\ (\varphi_1(\omega), \ldots, \varphi_n(\omega))\in B^n\}=\begin{cases}1, & \textit{если } (1, \ldots, 1)\in B^n,\\ 0 & \textit{в противном случае},\end{cases}$$

где $B^n\in\mathscr{B}(R^n)$. Нетрудно убедиться в том, что семейство мер $\{P_n\}$ является согласованным: если $A\in\mathscr{F}_n$, то $P_{n+1}(A)=P_n(A)$. Можно, однако, утверждать, что на $(\Omega, \mathscr{F})$ не существует вероятностной меры P такой, чтобы ее *сужение* $\mathsf{P}\,|\,\mathscr{F}_n$ (т. е. мера P, рассматриваемая лишь на множествах из $\mathscr{F}_n$) совпадало с P_n, $n=1, 2, \ldots$ В самом деле, допустим, что такая вероятностная мера P существует. Тогда

$$\mathsf{P}\{\omega:\ \varphi_1(\omega)=\ldots=\varphi_n(\omega)=1\}=P_n\{\omega:\ \varphi_1(\omega)=\ldots=\varphi_n(\omega)=1\}=1 \tag{19}$$

для любого $n=1, 2, \ldots$ Но

$$\{\omega:\ \varphi_1(\omega)=\ldots=\varphi_n(\omega)=1\}=(0, 1/n)\downarrow\varnothing,$$

что противоречит (19) и предположению о счетной аддитивности (а значит, и непрерывности в «нуле» $\varnothing$) функции множеств P.

Приведем теперь пример вероятностной меры в $(R^\infty, \mathscr{B}(R^\infty))$. Пусть $F_1(x)$, $F_2(x)$, ... — последовательность одномерных функций распределения. Определим функции $G(x)=F_1(x)$, $G_2(x_1, x_2) = F_1(x_1)F_2(x_2)$, ... и соответствующие им вероятностные меры на $(R, \mathscr{B}(R))$, $(R^2, \mathscr{B}(R^2))$, ... обозначим P_1, P_2, ... Тогда из теоремы 3 следует, что в $(R^\infty, \mathscr{B}(R^\infty))$ существует такая мера P, что

$$\mathsf{P}\{x \in R^\infty:\ (x_1, \ldots, x_n) \in B\} = P_n(B), \qquad B \in \mathscr{B}(R^n)$$

и, в частности,

$$\mathsf{P}\{x \in R^\infty:\ x_1 \leqslant a_1, \ldots, x_n \leqslant a_n\} = F_1(a_1) \ldots F_n(a_n).$$

Возьмем в качестве $F_i(x)$ — бернуллиевское распределение:

$$F_i(x) = \begin{cases} 0, & x < 0, \\ q, & 0 \leqslant x < 1, \\ 1, & x \geqslant 1. \end{cases}$$

Тогда можно утверждать, что в пространстве Ω всех числовых последовательностей $x=(x_1, x_2, \ldots)$, $x_i = 0, 1$, с σ-алгеброй его борелевских подмножеств существует вероятностная мера P такая, что для любого

$$\mathsf{P}\{x:\ x_1 = a_1, \ldots, x_n = a_n\} = p^{\Sigma a_i} q^{n - \Sigma a_i}.$$

Заметим, что именно этого результата нам не хватало в первой главе, чтобы сформулировать закон больших чисел в форме (1.5.8).

5. **Измеримые пространства** $(R^T, \mathscr{B}(R^T))$. Пусть T — произвольное множество индексов $t \in T$ и R_t — числовая прямая, соответствующая индексу t. Рассмотрим произвольный конечный неупорядоченный набор $\tau = [t_1, \ldots, t_n]$ различных индексов t_i, $t_i \in T$, $n \geqslant 1$, и пусть P_τ — вероятностная мера на $(R^\tau, \mathscr{B}(R^\tau))$ с $R^\tau = R_{t_1} \times \ldots \times R_{t_n}$.

Будем говорить, что семейство вероятностных мер $\{P_\tau\}$, где τ пробегает множество всех конечных неупорядоченных наборов, является *согласованным*, если для любых наборов $\tau = [t_1, \ldots, t_n]$ и $\sigma = [s_1, \ldots, s_k]$ таких, что $\sigma \subseteq \tau$

$$P_\sigma\{(x_{s_1}, \ldots, x_{s_k}) : (x_{s_1}, \ldots, x_{s_k}) \in B\} = \\ = P_\tau\{(x_{t_1}, \ldots, x_{t_n}) : (x_{s_1}, \ldots, x_{s_k}) \in B\} \qquad (20)$$

для любого $B \in \mathscr{B}(R^\sigma)$.

Теорема 4 (теорема Колмогорова о продолжении меры в $(R^T, \mathscr{B}(R^T))$). *Пусть $\{P_\tau\}$ — семейство согласованных вероятностных мер на $(R^\tau, \mathscr{B}(R^\tau))$. Тогда существует и притом единственная вероятностная мера P на $(R^T, \mathscr{B}(R^T))$ такая, что*

$$\mathsf{P}\{x \in R^T:\ (x_{t_1}, \ldots, x_{t_n}) \in B\} = P_{[t_1, \ldots, t_n]}(B) \qquad (21)$$

для всех неупорядоченных наборов $\tau=[t_1, \ldots, t_n]$ *различных индексов* $t_i \in T$ и $B \in \mathscr{B}(R^\tau)$.

Доказательство. Пусть множество $B \in \mathscr{B}(R^T)$. Согласно теореме из § 2 найдется такое счетное множество $S=\{s_1, s_2, \ldots\} \subseteq T$, что $B \in \mathscr{B}(R^S)$, где $R^S=R_{s_1}\times R_{s_2}\times\ldots$

На множествах $B \in \mathscr{B}(R^T)$ определим функцию (множеств) P, полагая

$$\mathsf{P}(B)=P_S(B), \tag{22}$$

где P_S — та вероятностная мера, существование которой гарантируется теоремой 3. Мы утверждаем, что P — именно та мера, существование которой утверждается в теореме. С этой целью надо, во-первых, проверить, что определение (22) корректно, т. е. приводит к одному и тому же значению $\mathsf{P}(B)$ при разных способах представления множества B, и, во-вторых, что эта функция множеств счетно-аддитивна.

Итак, пусть $B \in \mathscr{B}(R^{S_1})$ и $B \in \mathscr{B}(R^{S_2})$. Ясно, что тогда $B \in \mathscr{B}(R^{S_1 \cup S_2})$, и поэтому достаточно показать, что если $S \subseteq S'$, $B \in \mathscr{B}(R^S)$, то $P_S(B)=P_{S'}(B)$. Но $\mathscr{B}(R^S) \subseteq \mathscr{B}(R^{S'})$ и на множествах B вида

$$B=\{x \in R^S\colon (x_{s_1}, \ldots, x_{s_k}) \in A\},\quad A \in \mathscr{B}(R^k),$$

меры $P_S(B)$ и $P_{S'}(B)$ совпадают в силу условия согласованности. Поэтому в силу теоремы 3 они совпадают и для всех множеств $B \in \mathscr{B}(R^S)$, что и доказывает корректность определения $\mathsf{P}(B)$.

Пусть теперь $\{B_n\}$ — некоторая последовательность попарно непересекающихся множеств из $\mathscr{B}(R^T)$. Тогда найдется такое счетное множество $S \subseteq T$, что для любого n $B_n \in \mathscr{B}(R^S)$. Тогда поскольку P_S — вероятностные меры, то

$$\mathsf{P}\left(\sum B_n\right)=P_S\left(\sum B_n\right)=\sum P_S(B_n)=\sum \mathsf{P}(B_n).$$

Теорема доказана.

Замечание 1. Подчеркнем, что T — *любое* множество индексов. При этом в силу замечания к теореме 3 настоящая теорема остается в силе, если вместо числовых прямых R_t рассматривать любые полные сепарабельные метрические пространства Ω_t (с σ-алгебрами, порожденными открытыми множествами).

Замечание 2. Исходное семейство вероятностных мер $\{P_\tau\}$ предполагалось заданным для всех *неупорядоченных* наборов $\tau=[t_1, \ldots, t_n]$ различных индексов. Иногда в качестве исходного берут семейство вероятностных мер $\{P_\tau\}$, где τ пробегает множество всех *упорядоченных* наборов $\tau=(t_1, \ldots, t_n)$ различных индексов. В этом случае для справедливости теоремы 4 к условию

(20) надо тогда добавить еще одно *условие согласованности*:

$$P_{(t_1, \ldots, t_n)}(A_{t_1} \times \ldots \times A_{t_n}) = P_{(t_{i_1}, \ldots, t_{i_n})}(A_{t_{i_1}} \times \ldots \times A_{t_{i_n}}), \qquad (23)$$

где $(i_1, \ldots, i_n)$ — произвольная перестановка чисел $(1, \ldots, n)$, $A_{t_i} \in \mathscr{B}(R_{t_i})$, очевидность которого как необходимого условия существования вероятностной меры P следует из (21) (с заменой $P_{[t_1, \ldots, t_n]}(B)$ на $P_{(t_1, \ldots, t_n)}(B)$).

В дальнейшем мы всегда будем предполагать, что рассматриваемые наборы τ являются *неупорядоченными*. Если T — множество на числовой прямой (или некоторое вполне упорядоченное множество), то без ограничения общности можно считать, что рассматриваемые наборы $\tau = [t_1, \ldots, t_n]$ таковы, что $t_1 < t_2 < \ldots < t_n$. Таким образом, все «конечномерные», вероятности достаточно задавать лишь для таких наборов $\tau = [t_1, \ldots, t_n]$, у которых $t_1 < t_2 < \ldots < t_n$.

Рассмотрим сейчас тот случай, когда $T = [0, \infty)$. В этом случае R^T есть пространство всех действительных функций $x = (x_t)_{t \geq 0}$. Важным примером вероятностной меры на $(R^{[0, \infty)}, \mathscr{B}(R^{[0, \infty)}))$ является так называемая винеровская мера, строящаяся следующим образом.

Рассмотрим семейство $\{\varphi_t(y \mid x)\}_{t \geq 0}$ гауссовских плотностей (по y при фиксированном x)

$$\varphi_t(y \mid x) = \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/2t}, \qquad y \in R,$$

и определим для каждого набора $\tau = [t_1, \ldots, t_n]$, $t_1 < t_2 < \ldots < t_n$ и множества $B = I_1 \times \ldots \times I_n$, $I_k = (a_k, b_k)$, меру $P_\tau(B)$ по формуле

$$P_\tau(I_1 \times \ldots \times I_n) = \\ = \int_{I_1} \ldots \int_{I_n} \varphi_{t_1}(a_1 \mid 0)\, \varphi_{t_2 - t_1}(a_2 \mid a_1) \ldots \varphi_{t_n - t_{n-1}}(a_n \mid a_{n-1})\, da_1 \ldots da_n \qquad (24)$$

(интегрирование понимается в смысле Римана). Затем для каждого цилиндрического множества $\mathscr{I}_{t_1 \ldots t_n}(I_1 \times \ldots \times I_n) = \{x \in R^T: x_{t_1} \in I_1, \ldots, x_{t_n} \in I_n\}$ определим функцию множеств P, полагая

$$\mathsf{P}(\mathscr{I}_{t_1 \ldots t_n}(I_1 \times \ldots \times I_n)) = P_{[t_1 \ldots t_n]}(I_1 \times \ldots \times I_n).$$

Наглядный смысл такого способа приписывания меры цилиндрическому множеству $\mathscr{I}_{t_1 \ldots t_n}(I_1 \times \ldots \times I_n)$ состоит в следующем.

Множество $\mathscr{I}_{t_1 \ldots t_n}(I_1 \times \ldots \times I_n)$ — это множество всех функций, проходящих в моменты $t_1, \ldots, t_n$ через «окна» $I_1, \ldots, I_n$ (см. рис. 24 § 2). Будем интерпретировать $\varphi_{t_k - t_{k-1}}(a_k \mid a_{k-1})$ как вероятность того, что частица, выходящая из точки a_{k-1} за время

$t_k - t_{k-1}$ попадет в окрестность точки a_k. Тогда то, что в (24) рассматривается произведение плотностей, означает определенную независимость приращений смещений движущейся «частицы» на интервалах времени $[0, t_1]$, $[t_1, t_2]$, ..., $[t_{n-1}, t_n]$.

Так построенное семейство мер $\{P_\tau\}$ является, как нетрудно проверить, согласованным и, следовательно, может быть продолжено до меры на $(R^{[0, \infty)}, \mathscr{B}(R^{[0, \infty)}))$. Полученная таким образом мера играет важную роль в теории вероятностей. Эта мера была введена Н. Винером и называется *винеровской мерой*.

6. Задачи.

1. Пусть $F(x) = \mathsf{P}(-\infty, x]$. Показать справедливость следующих формул:

$$\mathsf{P}(a, b] = F(b) - F(a), \qquad \mathsf{P}(a, b) = F(b-) - F(a),$$
$$\mathsf{P}[a, b] = F(b) - F(a-), \qquad \mathsf{P}[a, b) = F(b-) - F(a-),$$
$$\mathsf{P}\{x\} = F(x) - F(x-),$$

где $F(x-) = \lim_{y \uparrow x} F(y)$.

2. Убедиться в справедливости формулы (7).

3. Провести доказательство теоремы 2.

4. Показать, что функция распределения $F = F(x)$ на R имеет не более чем счетное число точек разрыва. Справедлив ли соответствующий результат для функций распределения в R^n?

5. Показать, что каждая из функций

$$G(x, y) = \begin{cases} 1, & x + y \geqslant 0, \\ 0, & x + y < 0, \end{cases}$$

$$G(x, y) = [x + y] - \text{целая часть } x + y,$$

является непрерывной справа, возрастающей по каждой переменной, но не является (обобщенной) функцией распределения в R^2.

6. Пусть μ — мера Лебега — Стилтьеса, отвечающая непрерывной обобщенной функции распределения. Показать, что если множество A не более чем счетно, то $\mu(A) = 0$.

7. Пусть c — мощность континуума. Показать, что мощность борелевских множеств в R^n равна c, а лебеговских — 2^c.

8. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — некоторое вероятностное пространство и $\mathscr{A}$ — алгебра подмножеств Ω такая, что $\sigma(\mathscr{A}) = \mathscr{F}$. Используя принцип подходящих множеств, доказать, что для всякого $\varepsilon > 0$ и $B \in \mathscr{F}$ можно найти такое множество $A \in \mathscr{A}$, что

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

9. Пусть P — вероятностная мера в $(R^n, \mathscr{B}(R^n))$. Используя задачу 8, доказать, что для всякого $\varepsilon > 0$ и $B \in \mathscr{B}(R^n)$ можно

найти такой компакт $A \in \mathscr{B}(R^n)$, что $A \subseteq B$ и

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

(Этот результат используется в доказательстве теоремы 1.)

10. Проверить согласованность мер, задаваемых формулами (21).

§ 4. Случайные величины. I

1. Пусть $(\Omega, \mathscr{F})$ — некоторое измеримое пространство и $(R, \mathscr{B}(R))$ — числовая прямая с системой борелевских множеств $\mathscr{B}(R)$.

Определение 1. Действительная функция $\xi = \xi(\omega)$, определенная на $(\Omega, \mathscr{F})$, называется *$\mathscr{F}$-измеримой функцией* или *случайной величиной*, если для любого $B \in \mathscr{B}(R)$

$$\{\omega\colon\ \xi(\omega) \in B\} \in \mathscr{F}, \tag{1}$$

или, что то же самое, если прообраз $\xi^{-1}(B) \equiv \{\omega\colon \xi(\omega) \in B\}$ является измеримым множеством в Ω.

В том случае, когда $(\Omega, \mathscr{F}) = (R^n, \mathscr{B}(R^n))$, $\mathscr{B}(R^n)$-измеримые функции называют *борелевскими*.

Простейшим примером случайной величины является индикатор $I_A(\omega)$ любого (измеримого) множества $A \in \mathscr{F}$.

Случайная величина ξ, представимая в виде

$$\xi(\omega) = \sum_{i=1}^{\infty} x_i I_{A_i}(\omega), \tag{2}$$

где $\sum A_i = \Omega$, $A_i \in \mathscr{F}$, будет называться *дискретной*. Если же в (2) сумма конечна, то такая случайная величина будет называться *простой*.

Следуя той же интерпретации, что и в § 4 главы I, можно сказать, что случайная величина есть некоторая числовая характеристика эксперимента, значения которой зависят от «случая» ω. При этом требование измеримости (1) важно, и вот по какой причине. Если на $(\Omega, \mathscr{F})$ задана вероятностная мера P, то тогда имеет смысл говорить о вероятности события $\{\omega\colon\ \xi(\omega) \in B\}$, состоящего в том, что значения случайной величины принадлежат некоторому борелевскому множеству B.

В этой связи дадим такое

Определение 2. Вероятностная мера P_ξ на $(R, \mathscr{B}(R))$ с

$$P_\xi(B) = \mathsf{P}\{\omega\colon\ \xi(\omega) \in B\}, \quad B \in \mathscr{B}(R),$$

называется *распределением вероятностей случайной величины* ξ на $(R, \mathscr{B}(R))$.

Определение 3. Функция

$$F_\xi(x)=\mathsf{P}\{\omega\colon\ \xi(\omega)\leqslant x\},\quad x\in R,$$

называется *функцией распределения случайной величины* ξ.

Для дискретной случайной величины мера P_ξ сосредоточена не более чем в счетном числе точек и может быть представлена в виде

$$P_\xi(B)=\sum_{\{k\colon\ x_k\in B\}} p(x_k), \tag{3}$$

где $p(x_k)=\mathsf{P}\{\xi=x_k\}=\Delta F_\xi(x_k)$.

Очевидно, что верно и обратное: если P_ξ представимо в виде (3), то ξ является *дискретной* случайной величиной.

Случайная величина ξ называется *непрерывной*, если ее функция распределения $F_\xi(x)$ непрерывна по $x\in R$.

Случайная величина ξ называется *абсолютно непрерывной*, если существует неотрицательная функция $f=f_\xi(x)$, называемая плотностью, такая, что

$$F_\xi(x)=\int_{-\infty}^{x} f_\xi(y)\,dy,\qquad x\in R, \tag{4}$$

(интеграл понимается в смысле Римана, а в более общем случае — в смысле Лебега; см. далее § 6).

2. Установление того, что некоторая функция $\xi=\xi(\omega)$ является случайной величиной, требует проверки выполнимости свойства (1) для всех множеств $B\in\mathscr{F}$. Следующая лемма показывает, что класс таких «пробных» множеств может быть сужен.

Лемма 1. *Пусть $\mathscr{E}$ — некоторая система множеств такая, что $\sigma(\mathscr{E})=\mathscr{B}(R)$. Для того чтобы некоторая функция $\xi=\xi(\omega)$ была $\mathscr{F}$-измеримой, необходимо и достаточно, чтобы*

$$\{\omega\colon\ \xi(\omega)\in E\}\in\mathscr{F} \tag{5}$$

для всех $E\in\mathscr{E}$.

Доказательство. Необходимость очевидна. Для доказательства достаточности опять воспользуемся принципом подходящих множеств.

Пусть $\mathscr{D}$ — система тех борелевских множеств D из $\mathscr{B}(R)$, для которых $\xi^{-1}(D)\in\mathscr{F}$. Операция «взятия прообраза» сохраняет, как нетрудно проверить теоретико-множественные операции объединения, пересечения и дополнения:

$$\begin{gathered}\xi^{-1}\Big(\bigcup_\alpha B_\alpha\Big)=\bigcup_\alpha \xi^{-1}(B_\alpha),\\ \xi^{-1}\Big(\bigcap_\alpha B_\alpha\Big)=\bigcap_\alpha \xi^{-1}(B_\alpha),\\ \overline{\xi^{-1}(B_\alpha)}=\xi^{-1}(\overline{B_\alpha}).\end{gathered} \tag{6}$$

Отсюда следует, что система $\mathscr{D}$ является σ-алгеброй. Значит,

$$\mathscr{E} \subseteq \mathscr{D} \subseteq \mathscr{B}(R)$$

и

$$\sigma(\mathscr{E}) \subseteq \sigma(\mathscr{D}) = \mathscr{D} \subseteq \mathscr{B}(R).$$

Но $\sigma(\mathscr{E}) = \mathscr{B}(R)$, следовательно, $\mathscr{D} = \mathscr{B}(R)$.

Следствие. Для того чтобы $\xi = \xi(\omega)$ была случайной величиной, необходимо и достаточно, чтобы для любых $x \in R$

$$\{\omega\colon\ \xi(\omega) < x\} \in \mathscr{F},$$

или

$$\{\omega\colon\ \xi(\omega) \leqslant x\} \in \mathscr{F}.$$

Доказательство сразу следует из того, что каждая из систем множеств

$$\mathscr{E}_1 = \{x\colon\ x < c,\ c \in R\},$$
$$\mathscr{E}_2 = \{x\colon\ x \leqslant c,\ c \in R\}$$

порождает σ-алгебру $\mathscr{B}(R)$, $\sigma(\mathscr{E}_1) = \sigma(\mathscr{E}_2) = \mathscr{B}(R)$ (см. § 2).

Приводимая ниже лемма дает возможность конструирования случайных величин как функций от других случайных величин.

Лемма 2. *Пусть $\varphi = \varphi(x)$ — борелевская функция, а $\xi = \xi(\omega)$ — случайная величина. Тогда сложная функция $\eta = \varphi \circ \xi$, т. е. функция $\eta(\omega) = \varphi(\xi(\omega))$, является также случайной величиной.*

Доказательство следует из того, что для $B \in \mathscr{B}(R)$

$$\{\omega\colon\ \eta(\omega) \in B\} = \{\omega\colon\ \varphi(\xi(\omega)) \in B\} = \{\omega\colon\ \xi(\omega) \in \varphi^{-1}(B)\} \in \mathscr{F}, \quad (7)$$

поскольку $\varphi^{-1}(B) \in \mathscr{B}(R)$.

Таким образом, если ξ — случайная величина, то такие функции, как, скажем, ξ^n, $\xi^+ = \max(\xi, 0)$, $\xi^- = -\min(\xi, 0)$, $|\xi|$ также являются случайными величинами, поскольку функции x^n, x^+, x^-, $|x|$ являются борелевскими (задача 4).

3. Отправляясь от заданной системы случайных величин $\{\xi_n\}$, можно из них строить новые функции, например $\sum\limits_{k=1}^{\infty} |\xi_k|$, $\overline{\lim}\, \xi_n$, $\underline{\lim}\, \xi_n$ и т. д. Заметим, что эти функции принимают свои значения, вообще говоря, уже в расширенной числовой прямой $\bar{R} = [-\infty, \infty]$. Поэтому целесообразно несколько расширить класс $\mathscr{F}$-измеримых функций, допуская, чтобы они принимали также значения $\pm\infty$.

Определение 4. Функция $\xi = \xi(\omega)$, определенная на $(\Omega, \mathscr{F})$ и принимающая значения в $\bar{R} = [-\infty, \infty]$, будет называться *расширенной случайной величиной*, если для любого борелевского множества $B \in \mathscr{B}(R)$ выполнено условие (1).

Следующая теорема, несмотря на ее простоту, является ключевой при построении интеграла Лебега (§ 6).

Теорема 1. a) *Для любой (в том числе и расширенной) случайной величины* $\xi = \xi(\omega)$ *найдется последовательность простых случайных величин* $\xi_1, \xi_2, \ldots$ *таких, что* $|\xi_n| \leqslant \xi$ *и* $\xi_n(\omega) \to \xi(\omega)$, $n \to \infty$, *для всех* $\omega \in \Omega$.

b) *Если к тому же* $\xi(\omega) \geqslant 0$, *то найдется последовательность простых случайных величин* $\xi_1, \xi_2, \ldots$ *таких, что* $\xi_n(\omega) \uparrow \xi(\omega)$, $n \to \infty$, *для всех* $\omega \in \Omega$.

Доказательство. Начнем с доказательства второго утверждения. Положим для $n = 1, 2, \ldots$

$$\xi_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\left\{\frac{k-1}{2^n} \leqslant \xi(\omega) < \frac{k}{2^n}\right\}}(\omega) + nI_{\{\xi(\omega) \geqslant n\}}(\omega).$$

Непосредственно проверяется, что построенная последовательность $\xi_n(\omega)$ такова, что $\xi_n(\omega) \uparrow \xi(\omega)$ для всех $\omega \in \Omega$. Из этого утверждения вытекает также справедливость первого утверждения, если только заметить, что ξ может быть представлена в виде $\xi = \xi^+ - \xi^-$. Теорема доказана.

Покажем теперь, что класс расширенных случайных величин замкнут относительно поточечной сходимости. С этой целью заметим прежде всего, что если $\xi_1, \xi_2, \ldots$ — последовательность расширенных случайных величин, то функции $\sup \xi_n$, $\inf \xi_n$, $\overline{\lim}\, \xi_n$ и $\underline{\lim}\, \xi_n$ также являются случайными величинами (быть может, расширенными). Следует это непосредственно из того, что

$$\{\omega: \sup \xi_n > x\} = \bigcup_n \{\omega: \xi_n > x\} \in \mathscr{F},$$

$$\{\omega: \inf \xi_n < x\} = \bigcup_n \{\omega: \xi_n < x\} \in \mathscr{F}$$

и $\overline{\lim}\, \xi_n = \inf\limits_n \sup\limits_{m \geqslant n} \xi_m$, $\underline{\lim}\, \xi_n = \sup\limits_n \inf\limits_{m \geqslant n} \xi_m$.

Теорема 2. *Пусть* $\xi_1, \xi_2, \ldots$ — *последовательность расширенных случайных величин и* $\xi(\omega) = \lim \xi_n(\omega)$. *Тогда* $\xi(\omega)$ *также является расширенной случайной величиной.*

Доказательство сразу следует из сделанного выше замечания и того, что

$$\begin{aligned}\{\omega: \xi(\omega) < x\} &= \{\omega: \lim \xi_n(\omega) < x\} = \\ &= \left\{\omega: \overline{\lim}\, \xi_n(\omega) = \underline{\lim}\, \xi_n(\omega)\right\} \cap \left\{\overline{\lim}\, \xi_n(\omega) < x\right\} = \\ &= \Omega \cap \left\{\overline{\lim}\, \xi_n(\omega) < x\right\} = \left\{\overline{\lim}\, \xi_n(\omega) < x\right\} \in \mathscr{F}.\end{aligned}$$

4. Остановимся еще на некоторых свойствах простейших функций от случайных величин, рассматриваемых на измеримом пространстве $(\Omega, \mathscr{F})$ и принимающих, быть может, значения в расширенной числовой прямой $\bar{R}=[-\infty, \infty]$ *).

Если ξ и η — две случайные величины, то $\xi+\eta$, $\xi-\eta$, $\xi\eta$ и ξ/η также являются случайными величинами $\Big($в предположении, что они определены, т. е. не возникает неопределенностей типа $\infty-\infty$, $\frac{\infty}{\infty}$, $\frac{a}{0}\Big)$.

В самом деле, пусть $\{\xi_n\}$ и $\{\eta_n\}$ — последовательности случайных величин, сходящиеся к ξ и η (см. теорему 1). Тогда

$$\xi_n \pm \eta_n \to \xi \pm \eta,$$
$$\xi_n\eta_n \to \xi\eta,$$
$$\frac{\xi_n}{\eta_n+\frac{1}{n}I_{\{\eta_n=0\}}(\omega)} \to \frac{\xi}{\eta}.$$

Каждая из функций в левых частях этих соотношений является простой случайной величиной. Поэтому в силу теоремы 2 предельные функции $\xi\pm\eta$, $\xi\eta$ и ξ/η также являются случайными величинами.

5. Пусть ξ — случайная величина. Рассмотрим множества из $\mathscr{F}$ вида $\{\omega: \xi(\omega)\in B\}$, $B\in\mathscr{B}(R)$. Наименьшую σ-алгебру, порожденную такими множествами, называют *σ-алгеброй, порожденной случайной величиной* ξ. Будем ее обозначать $\mathscr{F}_\xi$.

Если φ — некоторая борелевская функция, то из леммы 2 следует, что функция $\eta=\varphi\circ\xi$ также является случайной величиной, причем $\mathscr{F}_\xi$-измеримой, т. е. такой, что $\{\omega: \eta(\omega)\in B\}\in\mathscr{F}_\xi$, $B\in\mathscr{B}(R)$ (см. (7)). Оказывается, что справедлив и обратный результат.

Теорема 3. *Пусть η — $\mathscr{F}_\xi$-измеримая случайная величина. Тогда найдется такая борелевская функция φ, что $\eta=\varphi\circ\xi$, т. е. для каждого $\omega\in\Omega$ $\eta(\omega)=\varphi(\xi(\omega))$.*

Доказательство. Пусть Φ — класс всех $\mathscr{F}_\xi$-измеримых функций $\eta=\eta(\omega)$, а $\tilde{\Phi}_\xi$ — класс $\mathscr{F}_\xi$-измеримых функций, представимых в виде $\varphi\circ\xi$, где φ — некоторая борелевская функция. Ясно, что $\tilde{\Phi}_\xi \subseteq \Phi_\xi$. Утверждение теоремы состоит в том, что на самом деле $\tilde{\Phi}_\xi=\Phi_\xi$.

---

*) В дальнейшем принимаются обычные соглашения относительно арифметических операций в $\bar{R}$: если $a\in R$, то $a\pm\infty=\pm\infty$, $\frac{a}{\pm\infty}=0$; $a\cdot\infty=\infty$, если $a>0$ и $a\cdot\infty=-\infty$, если $a<0$, $0\cdot(\pm\infty)=0$, $\infty+\infty=\infty$, $-\infty-\infty=-\infty$.

Пусть $A \in \mathscr{F}_\xi$ и $\eta(\omega) = I_A(\omega)$. Покажем, что $\eta = \tilde{\Phi}_\xi$. Действительно, если $A \in \mathscr{F}_\xi$, то найдется $B \in \mathscr{B}(R)$ такое, что $A = \{\omega\colon \xi(\omega) \in B\}$. Обозначим

$$\chi_B(x) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B. \end{cases}$$

Тогда $I_A(\omega) = \chi_B(\xi(\omega)) \in \tilde{\Phi}_\xi$. Отсюда следует, что и любая простая $\mathscr{F}_\xi$-измеримая функция $\sum_{i=1}^{n} c_i I_{A_i}(\omega)$, $A_i \in \mathscr{F}_\xi$, также принадлежит классу $\tilde{\Phi}_\xi$.

Пусть теперь η — произвольная $\mathscr{F}_\xi$-измеримая функция. По теореме 1 найдется последовательность простых $\mathscr{F}_\xi$-измеримых функций $\{\eta_n\}$ таких, что $\eta_n(\omega) \to \eta(\omega)$, $n \to \infty$, $\omega \in \Omega$. Как только что было установлено, существуют такие борелевские функции $\varphi_n = \varphi_n(x)$, что $\eta_n(\omega) = \varphi_n(\xi(\omega))$. При этом $\varphi_n(\xi(\omega)) \to \eta(\omega)$, $n \to \infty$, $\omega \in \Omega$.

Обозначим $B = \{x \in R\colon \lim\limits_n \varphi_n(x)$ существует$\}$. Это множество является борелевским. Поэтому функция

$$\varphi(x) = \begin{cases} \lim\limits_n \varphi_n(x), & x \in B, \\ 0, & x \notin B \end{cases}$$

также является борелевской (см. задачу 7).

Но тогда, очевидно, $\eta(\omega) = \lim\limits_n \varphi_n(\xi(\omega)) = \varphi(\xi(\omega))$ для всех $\omega \in \Omega$. Следовательно, $\tilde{\Phi}_\xi = \Phi_\xi$.

6. Рассмотрим вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$, в котором σ-алгебра $\mathscr{F}$ порождается некоторым конечным или счетным разбиением $\mathscr{D} = \{D_1, D_2, \ldots\}$, $\sum D_i = \Omega$, $\mathsf{P}(D_i) > 0$. Будем при этом предполагать, что D_i являются атомами относительно меры P, т. е. если $A \subseteq D_i$, $A \in \mathscr{F}$, то или $\mathsf{P}(A) = 0$ или $\mathsf{P}(D_i \setminus A) = 0$.

Лемма 3. *Пусть* ξ — $\mathscr{F}$*-измеримая функция, где* $\mathscr{F} = \sigma(\mathscr{D})$. *Тогда* ξ *постоянна на атомах разбиения, т. е.* ξ *представима в виде*

$$\xi(\omega) = \sum_{k=1}^{\infty} x_k I_{D_k}(\omega) \qquad (\mathsf{P}\text{-}п.\ н.). \tag{8}$$

(Запись «$\xi = \eta$ (P-*п. н.*)» означает, что $\mathsf{P}(\xi \neq \eta) = 0$.)

Доказательство. Пусть D — атом разбиения относительно меры P. Покажем, что на этом множестве случайная величина ξ (P-п. н.) постоянна, т. е. $\mathsf{P}\{D \cap (\xi \neq \text{const})\} = 0$.

Обозначим $K = \sup\{x \in R\colon \mathsf{P}\{D \cap (\xi < x)\} = 0\}$. Тогда

$$\mathsf{P}\{D \cap (\xi < K)\} = \mathsf{P}\Bigg[\bigcup_{\substack{r < K \\ r \text{ — рационально}}} \{\omega \in D\colon \xi(\omega) < r\}\Bigg] = 0,$$

поскольку, если $\mathsf{P}\{D\cap(\xi<x)\}=0$, то и $\mathsf{P}\{D\cap(\xi<y)\}=0$ для всех $y\leqslant x$.

Пусть $x>K$, тогда $\mathsf{P}\{D\cap(\xi<x)\}>0$ и, значит, $\mathsf{P}\{D\cap(\xi\geqslant x)\}=0$, поскольку D — атом. Поэтому

$$\mathsf{P}\{D\cap(\xi>K)\}=\mathsf{P}\Bigg[\bigcup_{\substack{r>K\\ r\text{ — рационально}}}\{\omega\in D:\ \xi\geqslant r\}\Bigg]=0.$$

Итак,

$$\mathsf{P}\{D\cap(\xi>K)\}=\mathsf{P}\{D\cap(\xi<K)\}=0$$

и, значит, $\mathsf{P}\{D\cap(\xi\neq K)\}=0$.

Общее утверждение (8) следует отсюда в силу того, что $\sum D_i=\Omega$. Лемма доказана.

7. Задачи.

1. Показать, что случайная величина ξ непрерывна, если и только если $\mathsf{P}(\xi=x)=0$ для всех $x\in R$.

2. Если $|\xi|$ является $\mathscr{F}$-измеримой, то верно ли, что ξ также $\mathscr{F}$-измерима?

3. Показать, что функция $\xi=\xi(\omega)$ является расширенной случайной величиной тогда и только тогда, когда $\{\omega:\ \xi(\omega)\in\bar{B}\}\in\mathscr{F}$ для всех $\bar{B}\in\mathscr{B}(\bar{R})$.

4. Доказать, что функции x^n, $x^+=\max(x, 0)$, $x^-=-\min(x, 0)$, $|x|=x^++x^-$ являются борелевскими.

5. Если ξ и η — $\mathscr{F}$-измеримы, то $\{\omega:\ \xi(\omega)=\eta(\omega)\}\in\mathscr{F}$.

6. Пусть ξ и η — две случайные величины на $(\Omega, \mathscr{F})$ и множество $A\in\mathscr{F}$. Тогда функция

$$\zeta(\omega)=\xi(\omega)\cdot I_A+\eta(\omega)\,I_{\bar{A}}$$

также является случайной величиной.

7. Пусть $\xi_1, \ldots, \xi_n$ — случайные величины и $\varphi(x_1, \ldots, x_n)$ — борелевская функция. Показать, что функция $\varphi(\xi_1(\omega), \ldots, \xi_n(\omega))$ также является случайной величиной.

8. Пусть ξ и η — две случайные величины, принимающие значения $1, 2, \ldots, N$. Предположим, что $\mathscr{F}_\xi=\mathscr{F}$. Показать, что существует такая перестановка $(i_1, i_2, \ldots, i_N)$ чисел $(1, 2, \ldots, N)$, что для каждого $j=1, 2, \ldots, N$ $\{\omega:\ \xi=j\}=\{\omega:\ \eta=i_j\}$.

§ 5. Случайные элементы

1. Наряду со случайными величинами в теории вероятностей и ее приложениях рассматривают случайные объекты более общей природы, например случайные «точки», векторы, функции, процессы, поля, множества, меры и т. д. В связи с этим желательно иметь понятие случайного объекта произвольной природы.

Определение 1. Пусть $(\Omega, \mathscr{F})$ и $(E, \mathscr{E})$—два измеримых пространства. Будем говорить, что функция $X = X(\omega)$, определенная на Ω и принимающая значения в E, есть $\mathscr{F}/\mathscr{E}$-*измеримая функция*, или *случайный элемент* (со значениями в E), если для любого $B \in \mathscr{E}$

$$\{\omega: X(\omega) \in B\} \in \mathscr{F}. \tag{1}$$

Иногда случайные элементы (со значениями в E) называют также E-значными случайными величинами.

Рассмотрим частные случаи этого определения.

Если $(E, \mathscr{E}) = (R, \mathscr{B}(R))$, то определение случайного элемента совпадает с определением случайной величины (§ 4).

Пусть $(E, \mathscr{E}) = (R^n, \mathscr{B}(R^n))$. Тогда случайный элемент $X(\omega)$ есть «случайная точка» в R^n. Если π_k — проекция R^n на k-ю координатную ось, то $X(\omega)$ можно представить в виде

$$X(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega)), \tag{2}$$

где $\xi_k = \pi_k \circ X$.

Из условия (1) вытекает, что ξ_k — обычные случайные величины. Действительно, для любого $B \in \mathscr{B}(R)$

$$\begin{aligned}\{\omega: \xi_k(\omega) \in B\} &= \\ &= \{\omega: \xi_1(\omega) \in R, \ldots, \xi_{k-1} \in R, \xi_k \in B, \xi_{k+1} \in R, \ldots\} = \\ &= \{\omega: X(\omega) \in (R \times \ldots \times R \times B \times R \times \ldots \times R)\} \in \mathscr{F},\end{aligned}$$

поскольку множество $R \times \ldots \times R \times B \times R \times \ldots \times R \in \mathscr{B}(R^n)$.

Определение 2. Всякий упорядоченный набор случайных величин $(\eta_1(\omega), \ldots, \eta_n(\omega))$ будем называть *n-мерным случайным вектором.*

В соответствии с этим определением всякий случайный элемент $X(\omega)$ со значениями в R^n является n-мерным случайным вектором. Справедливо и обратное: всякий случайный вектор $X(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega))$ есть случайный элемент в R^n. Действительно, если $B_k \in \mathscr{B}(R)$, $k = 1, \ldots, n$, то

$$\{\omega: X(\omega) \in (B_1 \times \ldots \times B_n)\} = \prod_{k=1}^{n} \{\omega: \xi_k(\omega) \in B_k\} \in \mathscr{F}.$$

Но наименьшая σ-алгебра, содержащая множества $B_1 \times \ldots \times B_n$, совпадает с $\mathscr{B}(R^n)$. Тогда из очевидного обобщения леммы 1 из § 3 сразу получаем, что для любого $B \in \mathscr{B}(R^n)$ множество $\{\omega: X(\omega) \in B\}$ принадлежит $\mathscr{F}$.

Пусть $(E, \mathscr{E}) = (\boldsymbol{Z}, \mathscr{B}(\boldsymbol{Z}))$, где $\boldsymbol{Z}$ — множество комплексных чисел $z = x + iy$, $x, y \in R$, а $\mathscr{B}(\boldsymbol{Z})$ — наименьшая σ-алгебра, содержащая множества вида $\{z: z = x + iy, a_1 < x \leqslant b_1, a_2 < y \leqslant b_2\}$. Из предыдущего рассмотрения следует, что комплекснозначная случайная величина $Z(\omega)$ представляется в виде $Z(\omega) = X(\omega) + iY(\omega)$, где $X(\omega)$ и $Y(\omega)$ — случайные величины. Поэтому $Z(\omega)$ называют также *комплексными случайными величинами*.

Пусть $(E, \mathscr{E}) = (R^T, \mathscr{B}(R^T))$, где T — некоторое подмножество числовой прямой. В этом случае всякий случайный элемент $X = X(\omega)$, представимый, очевидно, в виде $X = (\xi_t)_{t \in T}$ с $\xi_t = \pi_t \circ X$, называют случайной функцией с временны́м интервалом T.

Так же как и для случайных векторов, устанавливается, что всякая случайная функция является в то же самое время случайным процессом в смысле следующего определения.

Определение 3. Пусть T — некоторое подмножество числовой прямой. Совокупность случайных величин $X = (\xi_t)_{t \in T}$ называется *случайным процессом с временны́м интервалом* T.

Если $T = \{1, 2, \ldots\}$, то $X = (\xi_1, \xi_2, \ldots)$ называют *случайным процессом с дискретным временем* или *случайной последовательностью*.

Если $T = [0, 1]$, $(-\infty, \infty)$, $[0, \infty)$, ..., $X = (\xi_t)_{t \in T}$ называют *случайным процессом с непрерывным временем*.

Используя структуру σ-алгебр $\mathscr{B}(R^T)$ (§ 2), нетрудно показать, что всякий случайный процесс $X = (\xi_t)_{t \in T}$ (в смысле определения 3) является в то же самое время случайной функцией пространства $(R^T, \mathscr{B}(R^T))$.

Определение 4. Пусть $X = (\xi_t)_{t \in T}$ — случайный процесс. Для каждого фиксированного $\omega \in \Omega$ функция $(\xi_t(\omega))_{t \in T}$ называется *реализацией* или *траекторией* процесса, соответствующей исходу ω.

По аналогии с определением 2 § 4 естественно следующее

Определение 5. Пусть $X = (\xi_t)_{t \in T}$ — случайный процесс. Вероятностная мера P_X на $(R^T, \mathscr{B}(R^T))$ с

$$P_X(B) = \mathsf{P}\{\omega\colon X(\omega) \in B\}, \quad B \in \mathscr{B}(R^T),$$

называется *распределением вероятностей процесса* X. Вероятности

$$P_{t_1, \ldots, t_n}(B) \equiv \mathsf{P}\{\omega\colon (\xi_{t_1}, \ldots, \xi_{t_n}) \in B\}$$

с $t_1 < t_2 < \ldots < t_n$, $t_i \in T$, называются конечномерными вероятностями (или распределениями вероятностей). Функции

$$F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) \equiv \mathsf{P}\{\omega\colon \xi_{t_1} \leqslant x_1, \ldots, \xi_{t_n} \leqslant x_n\}$$

с $t_1 < t_2 < \ldots < t_n$, $t_i \in T$, называются конечномерными функциями распределения.

Пусть $(E, \mathscr{E}) = (C, \mathscr{B}_0(C))$, где C — пространство непрерывных функций $x = (x_t)_{t \in T}$ на $T = [0, 1]$ с σ-алгеброй $\mathscr{B}_0(C)$, порожденной открытыми множествами (§ 2). Покажем, что всякий случайный элемент X пространства $(C, \mathscr{B}_0(C))$ есть в то же самое время случайный процесс с непрерывными траекториями в смысле определения 3.

В самом деле, согласно § 2 множество $A=\{x\in C\colon x_t<a\}$ есть открытое множество в $\mathscr{B}_0(C)$. Поэтому

$$\{\omega\colon \xi_t(\omega)<a\}=\{\omega\colon X(\omega)\in A\}\in\mathscr{F}.$$

С другой стороны, пусть $X=(\xi_t(\omega))_{t\in T}$ есть случайный процесс (в смысле определения 3), траектории которого при каждом $\omega\in\Omega$ являются непрерывными функциями. В соответствии с (2.14)

$$\{x\in C\colon x\in S_\rho(x^0)\}=\bigcap_{t_k}\{x\in C\colon |x_{t_k}-x^0_{t_k}|<\rho\},$$

где t_k — рациональные точки отрезка $[0,\ 1]$. Поэтому

$$\{\omega\colon X(\omega)\in S_\rho(X^0\omega))\}=\bigcap_{t_k}\{\omega\colon |\xi_{t_k}(\omega)-\xi^0_{t_k}(\omega)|<\rho\}\in\mathscr{F},$$

а значит, и $\{\omega\colon X(\omega)\in B\}\in\mathscr{F}$ для любого $B\in\mathscr{B}_0(C)$.

Аналогичные рассуждения показывают также, что всякий случайный элемент пространства $(D,\ \mathscr{B}_0(D))$ может рассматриваться как случайный процесс с траекториями из пространства функций без разрывов второго рода, и наоборот.

2. Пусть $(\Omega,\ \mathscr{F},\ \mathsf{P})$ — вероятностное пространство и $(E_\alpha,\ \mathscr{E}_\alpha)$ — измеримые пространства, где индекс α принадлежит некоторому (произвольному) множеству $\mathfrak{A}$.

Определение 6. Будем говорить, что $\mathscr{F}/\mathscr{E}_\alpha$-измеримые функции $(X_\alpha(\omega))$, $\alpha\in\mathfrak{A}$, независимы (или независимы в совокупности), если для любого конечного набора индексов $\alpha_1,\ldots,\alpha_n$ случайные элементы $X_{\alpha_1},\ldots,X_{\alpha_n}$ независимы, т. е.

$$\mathsf{P}(X_{\alpha_1}\in B_{\alpha_1},\ldots,X_{\alpha_n}\in B_{\alpha_n})=\mathsf{P}(X_{\alpha_1}\in B_{\alpha_1})\ldots\mathsf{P}(X_{\alpha_n}\in B_{\alpha_n}),\quad(3)$$

где $B_\alpha\in\mathscr{E}_\alpha$.

Пусть $\mathfrak{A}=\{1,\ 2,\ldots,n\}$, ξ_α — случайные величины, $\alpha\in\mathfrak{A}$, и

$$F_\xi(x_1,\ldots,x_n)=\mathsf{P}(\xi_1\leqslant x_1,\ldots,\xi_n\leqslant x_n)$$

— n-мерная функция распределения вектора $\xi=(\xi_1,\ldots,\xi_n)$. Пусть $F_{\xi_i}(x_i)$ — функция распределения случайной величины ξ_i, $i=1,\ldots,n$.

Теорема. *Для того чтобы случайные величины $\xi_1,\ldots,\xi_n$ были независимы, необходимо и достаточно, чтобы для всех $(x_1,\ldots,x_n)\in R^n$*

$$F_\xi(x_1,\ldots,x_n)=F_{\xi_1}(x_1)\ldots F_{\xi_n}(x_n).\quad(4)$$

Доказательство. Необходимость очевидна. Для доказательства достаточности положим $a=(a_1,\ldots,a_n)$, $b=(b_1,\ldots,b_n)$

$$P_\xi(a,\ b]=\mathsf{P}\{\omega\colon a_1<\xi_1\leqslant b_1,\ldots,a_n<\xi_n\leqslant b_n\}$$
$$P_{\xi_i}(a_i,\ b_i]=\mathsf{P}\{a_i<\xi_i\leqslant b_i\}.$$

Тогда в силу (4) и (3.7)

$$P_\xi(a, b] = \prod_{i=1}^{n} [F_{\xi_i}(b_i) - F_{\xi_i}(a_i)] = \prod_{i=1}^{n} P_{\xi_i}(a_i, b_i]$$

и, значит,

$$\mathsf{P}\{\xi_1 \in I_1, \ldots, \xi_n \in I_n\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i \in I_i\}, \tag{5}$$

где $I_i = (a_i, b_i]$.

Зафиксируем $I_2, \ldots, I_n$ и покажем, что для любого $B_1 \in \mathscr{B}(R)$

$$\mathsf{P}\{\xi_1 \in B_1, \xi_2 \in I_2, \ldots, \xi_n \in I_n\} = \mathsf{P}\{\xi_1 \in B_1\} \prod_{i=2}^{n} \mathsf{P}\{\xi_i \in I_i\}. \tag{6}$$

Пусть $\mathscr{M}$ — совокупность множеств из $\mathscr{B}(R)$, для которых выполнено (6). В $\mathscr{M}$ входит, очевидно, алгебра $\mathscr{A}$ множеств, состоящих из сумм непересекающихся интервалов вида $I_1 = (a_1, b_1]$. Поэтому $\mathscr{A} \subseteq \mathscr{M} \subseteq \mathscr{B}(R)$. Из счетной аддитивности (а следовательно, и непрерывности) вероятностной меры следует также, что система $\mathscr{M}$ является монотонным классом. Поэтому (см. п. 1 § 2)

$$\mu(\mathscr{A}) \subseteq \mathscr{M} \subseteq \mathscr{B}(R).$$

Но согласно теореме 1 из § 2 $\mu(\mathscr{A}) = \sigma(\mathscr{A}) = \mathscr{B}(R)$. Поэтому $\mathscr{M} = \mathscr{B}(R)$.

Итак, (6) доказано. Фиксируя теперь $B_1, I_3, \ldots, I_n$, тем же методом доказываем справедливость (6) с заменой I_2 на борелевское множество B_2. Продолжая этот процесс, очевидным образом приходим к требуемому равенству

$$\mathsf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \mathsf{P}(\xi_1 \in B_1) \ldots \mathsf{P}(\xi_n \in B_n),$$

где $B_i \in \mathscr{B}(R)$. Теорема доказана.

3. Задачи.

1. Пусть $\xi_1, \ldots, \xi_n$ — дискретные случайные величины. Показать, что они независимы тогда и только тогда, когда для любых действительных $x_1, \ldots, x_n$

$$\mathsf{P}(\xi_1 = x_1, \ldots, \xi_n = x_n) = \prod_{i=1}^{n} \mathsf{P}(\xi_i = x_i).$$

2. Провести доказательство того, что всякая случайная функция (в смысле определения 1) есть случайный процесс (в смысле определения 3), и наоборот.

3. Пусть $X_1, \ldots, X_n$ — случайные элементы со значениями в $(E_1, \mathscr{E}_1), \ldots (E_n, \mathscr{E}_n)$ соответственно. Пусть, далее, $(E_1', \mathscr{E}_1'), \ldots, (E_n', \mathscr{E}_n')$ — измеримые пространства и $g_1, \ldots, g_n$ являются $\mathscr{E}_1/\mathscr{E}_1', \ldots, \mathscr{E}_n/\mathscr{E}_n'$-изме-

римыми функциями соответственно. Показать, что если $X_1, \ldots, X_n$ независимы, то независимы также и случайные элементы $g_1 \circ X_1, \ldots, g_n \circ X_n$.

§ 6. Интеграл Лебега. Математическое ожидание

1. В том случае, когда $(\Omega, \mathscr{F}, \mathsf{P})$ — конечное вероятностное пространство и $\xi = \xi(\omega)$ — простая случайная величина,

$$\xi(\omega) = \sum_{k=1}^{n} x_k I_{A_k}(\omega), \tag{1}$$

понятие математического ожидания $\mathsf{M}\xi$ было определено в § 4 гл. I. Та же самая конструкция математического ожидания $\mathsf{M}\xi$ от простых случайных величин ξ используется и в случае произвольного вероятностного пространства $(\Omega, \mathscr{F}, \mathsf{P})$. А именно, по определению полагается

$$\mathsf{M}\xi = \sum_{k=1}^{n} x_k \mathsf{P}(A_k). \tag{2}$$

Это определение корректно (в том смысле, что значение $\mathsf{M}\xi$ не зависит от способа представления ξ в виде (1)), что показывается точно так же, как и в случае конечных вероятностных пространств. Аналогичным образом устанавливаются простейшие свойства математического ожидания (см. п. 5 § 4 гл. I).

Цель этого параграфа — дать определение и изучить свойства математического ожидания $\mathsf{M}\xi$ произвольной случайной величины. С точки зрения анализа математическое ожидание $\mathsf{M}\xi$ есть не что иное, как интеграл Лебега от $\mathscr{F}$-измеримой функции $\xi = \xi(\omega)$ по мере P, для которого (наряду с $\mathsf{M}\xi$) используются также следующие обозначения: $\int_\Omega \xi(\omega)\,\mathsf{P}(d\omega)$ или $\int_\Omega \xi\, d\mathsf{P}$.

2. Пусть $\xi = \xi(\omega)$ — неотрицательная случайная величина. Построим последовательность простых неотрицательных случайных величин $\{\xi_n\}_{n \geq 1}$ таких, что $\xi_n(\omega) \uparrow \xi(\omega)$, $n \to \infty$, для каждого $\omega \in \Omega$ (см. теорему 1 в § 4).

Поскольку $\mathsf{M}\xi_n \leq \mathsf{M}\xi_{n+1}$ (ср. со свойством 3) из п. 5 § 4 гл. I), то существует $\lim_n \mathsf{M}\xi_n$, который может принимать и значение $+\infty$.

Определение 1. *Интегралом Лебега* от неотрицательной случайной величины $\xi = \xi(\omega)$, или ее *математическим ожиданием*, называется величина

$$\mathsf{M}\xi \equiv \lim_n \mathsf{M}\xi_n. \tag{3}$$

Чтобы это определение было корректным, надо показать, что значение этого предела не зависит от выбора аппроксимирующей последовательности $\{\xi_n\}$. Иначе говоря, надо показать, что если $\xi_n \uparrow \xi$ и $\eta_m \uparrow \xi$, где $\{\eta_m\}$ — последовательность простых функций, то

$$\lim_n \mathsf{M}\xi_n = \lim_m \mathsf{M}\eta_m. \tag{4}$$

Лемма 1. *Пусть η и ξ_n — простые случайные величины, $n \geqslant 1$, причем*

$$\xi_n \uparrow \xi \geqslant \eta.$$

Тогда

$$\lim_m \mathsf{M}\xi_n \geqslant \mathsf{M}\eta. \tag{5}$$

Доказательство. Пусть $\varepsilon > 0$ и

$$A_n = \{\omega:\ \xi_n \geqslant \eta - \varepsilon\}.$$

Ясно, что $A_n \uparrow \Omega$ и

$$\xi_n = \xi_n I_{A_n} + \xi_n I_{\bar{A}_n} \geqslant \xi_n I_{A_n} \geqslant (\eta - \varepsilon) I_{A_n}.$$

Поэтому, используя свойства математических ожиданий от простых случайных величин, находим, что

$$\mathsf{M}\xi_n \geqslant \mathsf{M}(\eta - \varepsilon) I_{A_n} = \mathsf{M}\eta I_{A_n} - \varepsilon \mathsf{P}(A_n) =$$
$$= \mathsf{M}\eta - \mathsf{M}\eta I_{\bar{A}_n} - \varepsilon \mathsf{P}(A_n) \geqslant \mathsf{M}\eta - C\mathsf{P}(\bar{A}_n) - \varepsilon,$$

где $C = \max_\omega \eta(\omega)$. Отсюда в силу произвольности $\varepsilon > 0$ вытекает требуемое неравенство (5). Из этой леммы следует, что $\lim_n \mathsf{M}\xi_n \geqslant \lim_n \mathsf{M}\eta_m$ и по симметрии $\lim_m \mathsf{M}\eta_m \geqslant \lim_n \mathsf{M}\xi_n$, что и доказывает (4).

Часто оказывается полезным следующее

Замечание 1. Для математического ожидания $\mathsf{M}\xi$ от неотрицательной случайной величины ξ имеет место следующее представление:

$$\mathsf{M}\xi = \sup_{\{s \in S:\ s \leqslant \xi\}} \mathsf{M}s, \tag{6}$$

где $S = \{s\}$ — множество простых неотрицательных случайных величин (задача 1).

Итак, для неотрицательных случайных величин математическое ожидание определено. Перейдем теперь к общему случаю.

Пусть ξ — случайная величина и $\xi^+ = \max(\xi, 0)$, $\xi^- = -\min(\xi, 0)$.

Определение 2. Говорят, что математическое ожидание $\mathsf{M}\xi$ случайной величины ξ *существует*, или *определено*, если по крайней мере одна из величин $\mathsf{M}\xi^+$ или $\mathsf{M}\xi^-$ конечна:

$$\min(\mathsf{M}\xi^+, \mathsf{M}\xi^-) < \infty.$$

В этом случае *по определению* полагают

$$\mathsf{M}\xi \equiv \mathsf{M}\xi^{+} - \mathsf{M}\xi^{-}.$$

Математическое ожидание $\mathsf{M}\xi$ называют иначе *интегралом Лебега* (от функции ξ по вероятностной мере P).

Определение 3. Говорят, что *математическое ожидание* случайной величины ξ *конечно*, если $\mathsf{M}\xi^{+} < \infty$ и $\mathsf{M}\xi^{-} < \infty$.

Поскольку $|\xi| = \xi^{+} + \xi^{-}$, то конечность $\mathsf{M}\xi$, или $|\mathsf{M}\xi| < \infty$, эквивалентна тому, что $\mathsf{M}|\xi| < \infty$. (В этом смысле интегрирование по Лебегу носит «абсолютный» характер.)

Замечание 2. Наряду с математическим ожиданием $\mathsf{M}\xi$ важными числовыми характеристиками случайной величины ξ являются величины $\mathsf{M}\xi^{r}$ (если они определены) и $\mathsf{M}|\xi|^{r}$, $r > 0$, называемые соответственно *моментом* r-го порядка (r-м моментом) и *абсолютным моментом* r-го порядка (r-м моментом) случайной величины ξ.

Замечание 3. В данном выше определении интеграла Лебега $\int\limits_{\Omega} \xi(\omega)\,\mathsf{P}(d\omega)$ предполагалось, что мера P является вероятностной ($\mathsf{P}(\Omega) = 1$), а $\mathscr{F}$-измеримые функции (случайные величины) ξ принимают значения в $R = (-\infty, \infty)$. Предположим теперь, что μ — произвольная мера, заданная на измеримом пространстве $(\Omega, \mathscr{F})$ и принимающая, быть может, значение $+\infty$, а $\xi = \xi(\omega)$ — $\mathscr{F}$-измеримая функция со значениями в $\overline{R} = [-\infty, \infty]$ (расширенная случайная величина). В этом случае интеграл Лебега $\int\limits_{\Omega} \xi(\omega)\,\mu(d\omega)$ определяется тем же самым способом: сначала для неотрицательных простых ξ (по формуле (2) с заменой P на μ), затем для произвольных неотрицательных ξ и в общем случае по формуле

$$\int\limits_{\Omega} \xi(\omega)\,\mu(d\omega) = \int\limits_{\Omega} \xi^{+}\mu(d\omega) - \int\limits_{\Omega} \xi^{-}\mu(d\omega),$$

если только не возникает неопределенности вида $\infty - \infty$.

Для математического анализа особо важен случай, когда $(\Omega, \mathscr{F}) = (R, \mathscr{B}(R))$, а μ—мера Лебега. В этом случае интеграл $\int\limits_{R} \xi(x)\,\mu(dx)$ обозначают $\int\limits_{R} \xi(x)\,dx$, или $\int\limits_{-\infty}^{\infty} \xi(x)\,dx$, или $(L)\int\limits_{-\infty}^{\infty} \xi(x)\,dx$, чтобы подчеркнуть отличие этого интеграла от интеграла Римана $(R)\int\limits_{-\infty}^{\infty} \xi(x)\,dx$. Если же мера μ (Лебега — Стилтьеса) соответствует некоторой обобщенной функции распределения $G = G(x)$, то интеграл $\int\limits_{R} \xi(x)\,\mu(dx)$ называют также *интегралом Лебега — Стилтьеса* и обозначают $(L\text{-}S)\int\limits_{R} \xi(x)\,G(dx)$, чтобы отличать его от соответсдвующего интеграла Римана — Стилтьеса $(R\text{-}S)\int\limits_R \xi(x)\,G(dx)$ (см. далее п. 10).

Из дальнейшего (свойство D) станет ясно, что если $\mathsf{M}\xi$ определено, то определены также математические ожидания $\mathsf{M}(\xi I_A)$ для любого $A \in \mathscr{F}$. Для $\mathsf{M}(\xi I_A)$, или, что то же, $\int\limits_\Omega \xi I_A\,d\mathsf{P}$, часто используются обозначения $\mathsf{M}(\xi;\,A)$ и $\int\limits_A \xi\,d\mathsf{P}$. Интеграл $\int\limits_A \xi\,d\mathsf{P}$ принято называть *интегралом Лебега от* ξ *по мере* P *на множестве* A.

Аналогично и в случае произвольной меры μ вместо $\int\limits_\Omega \xi\cdot I_A\,d\mu$ пишем $\int\limits_A \xi\,d\mu$. В частности, если μ — n-мерная мера Лебега — Стилтьеса, $A=(a_1,\,b_1]\times\ldots\times(a_n,\,b_n]$, то вместо $\int\limits_A \xi\,d\mu$ используем запись $\int\limits_{a_1}^{b_1}\ldots\int\limits_{a_n}^{b_n}\xi(x_1,\,\ldots,\,x_n)\,\mu(dx_1,\ldots,\,dx_n)$. Если μ — мера Лебега, то вместо $\mu(dx_1,\,\ldots,\,dx_n)$ пишем просто $dx_1\,\ldots\,dx_n$.

2. Свойства математического ожидания $\mathsf{M}\xi$ случайных величин ξ.

A. *Пусть* c — постоянная *и* $\mathsf{M}\xi$ *существует. Тогда* $\mathsf{M}(c\xi)$ *также существует и*

$$\mathsf{M}(c\xi)=c\mathsf{M}\xi.$$

B. *Пусть* $\xi\leqslant\eta$, *тогда*

$$\mathsf{M}\xi\leqslant\mathsf{M}\eta$$

в том смысле, что

$$\textit{если } -\infty<\mathsf{M}\xi, \textit{ то } -\infty<\mathsf{M}\eta \textit{ и } \mathsf{M}\xi\leqslant\mathsf{M}\eta$$

или

$$\textit{если } \mathsf{M}\eta<\infty, \textit{ то } \mathsf{M}\xi<\infty \textit{ и } \mathsf{M}\xi\leqslant\mathsf{M}\eta.$$

C. *Если* $\mathsf{M}\xi$ *существует, то*

$$|\mathsf{M}\xi|\leqslant\mathsf{M}|\xi|.$$

D. *Если* $\mathsf{M}\xi$ *существует, то для каждого* $A\in\mathscr{F}$ $\mathsf{M}(\xi I_A)$ *также существует; если* $\mathsf{M}\xi$ *конечно, то* $\mathsf{M}(\xi I_A)$ *также конечно.*

E. *Если* ξ *и* η — *неотрицательные случайные величины, или такие, что* $\mathsf{M}|\xi|<\infty$, $\mathsf{M}|\eta|<\infty$, *то*

$$\mathsf{M}(\xi+\eta)=\mathsf{M}\xi+\mathsf{M}\eta.$$

(По поводу обобщения этого свойства см. задачу 2).

Приведем доказательство свойств A — E.

A. Для простых случайных величин утверждение очевидно. Пусть $\xi \geqslant 0$, $\xi_n \uparrow \xi$, где ξ_n — простые случайные величины, и $c \geqslant 0$. Тогда $c\xi_n \uparrow c\xi$ и, значит,

$$\mathsf{M}(c\xi) = \lim_n \mathsf{M}(c\xi_n) = c \lim_n \mathsf{M}\xi_n = c\mathsf{M}\xi.$$

В общем случае надо рассмотреть представление $\xi = \xi^+ - \xi^-$ и заметить, что для $c \geqslant 0$, $(c\xi)^+ = c\xi^+$, $(c\xi)^- = c\xi^-$, а для $c < 0$ $(c\xi)^+ = -c\xi^-$, $(c\xi)^- = -c\xi^+$.

B. Если $0 \leqslant \xi \leqslant \eta$, то $\mathsf{M}\xi$ и $\mathsf{M}\eta$ определены и неравенство $\mathsf{M}\xi \leqslant \mathsf{M}\eta$ сразу следует из формулы (6). Пусть теперь $\mathsf{M}\xi > -\infty$, тогда $\mathsf{M}\xi^- < \infty$. Если $\xi \leqslant \eta$, то $\xi^+ \leqslant \eta^+$ и $\xi^- \geqslant \eta^-$. Поэтому $\mathsf{M}\eta^- \leqslant \mathsf{M}\xi^- < \infty$, следовательно, $\mathsf{M}\eta$ определено и $\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^- \leqslant \mathsf{M}\eta^+ - \mathsf{M}\eta^- = \mathsf{M}\eta$. Аналогичным образом рассматривается случай, когда $\mathsf{M}\eta < \infty$.

C. Поскольку $-|\xi| \leqslant \xi \leqslant |\xi|$, то из свойств **A** и **B**

$$-\mathsf{M}|\xi| \leqslant \mathsf{M}\xi \leqslant \mathsf{M}|\xi|,$$

т. е. $|\mathsf{M}\xi| \leqslant \mathsf{M}|\xi|$.

D. Следует из **B** и того, что

$$(\xi I_A)^+ = \xi^+ I_A \leqslant \xi^+, \quad (\xi I_A)^- = \xi^- I_A \leqslant \xi^-.$$

E. Пусть $\xi \geqslant 0$, $\eta \geqslant 0$ и пусть $\{\xi_n\}$ и $\{\eta_n\}$ — последовательности простых функций таких, что $\xi_n \uparrow \xi$, $\eta_n \uparrow \eta$. Тогда $\mathsf{M}(\xi_n + \eta_n) = \mathsf{M}\xi_n + \mathsf{M}\eta_n$ и $\mathsf{M}(\xi_n + \eta_n) \uparrow \mathsf{M}(\xi + \eta)$, $\mathsf{M}\xi_n \uparrow \mathsf{M}\xi$, $\mathsf{M}\eta_n \uparrow \mathsf{M}\eta$ и, значит, $\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta$. Случай, когда $\mathsf{M}|\xi| < \infty$, $\mathsf{M}|\eta| < \infty$, сводится к рассмотренному, если воспользоваться тем, что $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$, $\xi^+ \leqslant |\xi|$, $\xi^- \leqslant |\xi|$ и $\eta^+ \leqslant |\eta|$, $\eta^- \leqslant |\eta|$.

Следующая группа утверждений относительно математических ожиданий связана с понятием «Р-почти наверное». Будем говорить, что *некоторое свойство выполнено «Р-почти наверное», если существует множество $\mathscr{N} \in \mathscr{F}$ с $\mathsf{P}(\mathscr{N}) = 0$ такое, что это свойство выполнено для каждой точки $\omega \in \Omega \setminus \mathscr{N}$.* Вместо слов «Р-почти наверное» часто говорят «Р-почти всюду» или просто «почти наверное» (п. н.), «почти всюду» (п. в.).

F. Если $\xi = 0$ (п. н.), то $\mathsf{M}\xi = 0$.

В самом деле, если ξ — простая случайная величина, $\xi = \sum x_k I_{A_k}(\omega)$ и $x_k \neq 0$, то по условию $\mathsf{P}(A_k) = 0$, а значит, $\mathsf{M}\xi = 0$. Если же $\xi \geqslant 0$ и $0 \leqslant s \leqslant \xi$, где s — простая случайная величина, то $s = 0$ (п. н.), а следовательно, $\mathsf{M}s = 0$ и $\mathsf{M}\xi = \sup_{\{s \in S:\, s \leqslant \xi\}} \mathsf{M}s = 0$. Общий случай сводится к рассмотренному обычным переходом к представлению $\xi = \xi^+ - \xi^-$ с учетом того, что $\xi^+ \leqslant |\xi|$, $\xi^- \leqslant |\xi|$ и $|\xi| = 0$ (п. н.).

G. Если $\xi = \eta$ (п. н.) и $\mathsf{M}|\xi| < \infty$, то $\mathsf{M}|\eta| < \infty$ и $\mathsf{M}\xi = \mathsf{M}\eta$ (см. также задачу 3).

В самом деле, пусть $\mathscr{N}=\{\omega\colon \xi\neq\eta\}$. Тогда $\mathsf{P}(\mathscr{N})=0$ и $\xi=\xi I_{\mathscr{N}}+\xi I_{\bar{\mathscr{N}}}$, $\eta=\eta I_{\mathscr{N}}+\eta I_{\bar{\mathscr{N}}}=\eta I_{\mathscr{N}}+\xi I_{\bar{\mathscr{N}}}$. По свойствам **E** и **F** $\mathsf{M}\xi=\mathsf{M}\xi I_{\mathscr{N}}+\mathsf{M}\xi I_{\bar{\mathscr{N}}}=\mathsf{M}\xi I_{\bar{\mathscr{N}}}=\mathsf{M}\eta I_{\bar{\mathscr{N}}}$. Но $\mathsf{M}\eta I_{\mathscr{N}}=0$, поэтому по свойству **E** $\mathsf{M}\xi=\mathsf{M}\eta I_{\bar{\mathscr{N}}}+\mathsf{M}\eta I_{\mathscr{N}}=\mathsf{M}\eta$.

H. Пусть $\xi\geqslant 0$ и $\mathsf{M}\xi=0$. Тогда $\xi=0$ (п. н.).

Для доказательства обозначим $A=\{\omega\colon \xi(\omega)>0\}$, $A_n=\{\omega\colon \xi(\omega)\geqslant 1/n\}$. Ясно, что $A_n\uparrow A$ и $0\leqslant \xi\cdot I_{A_n}\leqslant \xi\cdot I_A$. Поэтому по свойству **B**

$$0\leqslant \mathsf{M}\xi I_{A_n}\leqslant \mathsf{M}\xi=0.$$

Следовательно,

$$0=\mathsf{M}\xi I_{A_n}\geqslant \frac{1}{n}\mathsf{P}(A_n)$$

и, значит, $\mathsf{P}(A_n)=0$ для всех $n\geqslant 1$. Но $\mathsf{P}(A)=\lim \mathsf{P}(A_n)$ и, следовательно, $\mathsf{P}(A)=0$.

I. Пусть ξ и η таковы, что $\mathsf{M}|\xi|<\infty$, $\mathsf{M}|\eta|<\infty$ и для всех $A\in\mathscr{F}$ $\mathsf{M}(\xi I_A)\leqslant\mathsf{M}(\eta I_A)$. Тогда $\xi\leqslant\eta$ (п. н.).

В самом деле, пусть $B=\{\omega\colon \xi(\omega)>\eta(\omega)\}$. Тогда $\mathsf{M}(\eta I_B)\leqslant\mathsf{M}(\xi I_B)\leqslant\mathsf{M}(\eta I_B)$ и, значит, $\mathsf{M}(\xi I_B)=\mathsf{M}(\eta I_B)$. В силу свойства **E** $\mathsf{M}((\xi-\eta)I_B)=0$ и по свойству **H** $(\xi-\eta)I_B=0$ (п. н.), откуда $\mathsf{P}(B)=0$.

J. Пусть ξ — расширенная случайная величина и $\mathsf{M}|\xi|<\infty$. Тогда $|\xi|<\infty$ (п. н.). Действительно, пусть $A=\{\omega\colon |\xi(\omega)|=\infty\}$ и $\mathsf{P}(A)>0$. Тогда $\mathsf{M}|\xi|\geqslant\mathsf{M}(|\xi| I_A)=\infty\cdot\mathsf{P}(A)=\infty$, что противоречит предположению $\mathsf{M}|\xi|<\infty$. (См. также задачу 4.)

3. В этом пункте будут рассмотрены основные теоремы *о предельном переходе* под знаком математического ожидания (интеграла Лебега).

Теорема 1 (о монотонной сходимости). *Пусть $\eta, \xi, \xi_1, \xi_2, \ldots$ — случайные величины.*

a) *Если $\xi_n\geqslant\eta$ для всех $n\geqslant 1$, $\mathsf{M}\eta>-\infty$ и $\xi_n\uparrow\xi$, то*

$$\mathsf{M}\xi_n\uparrow\mathsf{M}\xi.$$

b) *Если $\xi_n\leqslant\eta$ для всех $n\geqslant 1$, $\mathsf{M}\eta<\infty$ и $\xi_n\downarrow\xi$, то*

$$\mathsf{M}\xi_n\downarrow\mathsf{M}\xi.$$

Доказательство. a) Предположим сначала, что $\eta\geqslant 0$. Пусть для каждого $k\geqslant 1$ $\{\xi_k^{(n)}\}_{n\geqslant 1}$ — последовательность простых функций таких, что $\xi_k^{(n)}\uparrow\xi_k$, $n\to\infty$. Обозначим $\zeta^{(n)}=\max\limits_{1\leqslant k\leqslant n}\xi_k^{(n)}$. Тогда

$$\zeta^{(n-1)}\leqslant\zeta^{(n)}=\max_{1\leqslant k\leqslant n}\xi_k^{(n)}\leqslant\max_{1\leqslant k\leqslant n}\xi_k=\xi_n.$$

Пусть $\zeta=\lim\limits_n\zeta^{(n)}$. Поскольку для $1\leqslant k\leqslant n$

$$\xi_k^{(n)}\leqslant\zeta^{(n)}\leqslant\xi_n,$$

то, переходя к пределу при $n\to\infty$, получим, что для любого $k\geqslant 1$

$$\xi_k\leqslant\zeta\leqslant\xi,$$

а значит, $\xi=\zeta$.

Случайные величины $\zeta^{(n)}$ простые и $\zeta^{(n)}\uparrow\zeta$. Поэтому

$$\mathsf{M}\xi=\mathsf{M}\zeta=\lim \mathsf{M}\zeta^{(n)}\leqslant\lim \mathsf{M}\xi_n.$$

С другой стороны, очевидно, что, поскольку $\xi_n\leqslant\xi_{n+1}\leqslant\xi$, то

$$\lim \mathsf{M}\xi_n\leqslant \mathsf{M}\xi.$$

Тем самым $\lim \mathsf{M}\xi_n=\mathsf{M}\xi$.

Пусть теперь η — произвольная случайная величина с $\mathsf{M}\eta>-\infty$.

Если $\mathsf{M}\eta=\infty$, то в силу **B** $\mathsf{M}\xi_n=\mathsf{M}\xi=\infty$ и утверждение доказано. Пусть $\mathsf{M}\eta<\infty$. Тогда вместе с условием $\mathsf{M}\eta>-\infty$ получаем, что $\mathsf{M}|\eta|<\infty$. Ясно, что $0\leqslant\xi_n-\eta\uparrow\xi-\eta$ для всех $\omega\in\Omega$. Поэтому, согласно доказанному, $\mathsf{M}(\xi_n-\eta)\uparrow\mathsf{M}(\xi-\eta)$ и, значит, (по свойству **E** и задаче 2)

$$\mathsf{M}\xi_n-\mathsf{M}\eta\uparrow\mathsf{M}\xi-\mathsf{M}\eta.$$

Но $\mathsf{M}|\eta|<\infty$, поэтому $\mathsf{M}\xi_n\uparrow\mathsf{M}\xi$, $n\to\infty$.

Доказательство утверждения b) следует из a), если вместо исходных величин рассмотреть величины со знаком минус.

Следствие. Пусть $\{\eta_n\}_{n\geqslant 1}$ — последовательность неотрицательных случайных величин. Тогда

$$\mathsf{M}\sum_{n=1}^{\infty}\eta_n=\sum_{n=1}^{\infty}\mathsf{M}\eta_n.$$

Доказательство следует из свойства **E** (см. также задачу (2) теоремы о монотонной сходимости и того замечания, что $\sum\limits_{n=1}^{k}\eta_n\uparrow$ $\uparrow\sum\limits_{n=1}^{\infty}\eta_n$, $k\to\infty$.

Теорема 2 (лемма Фату). *Пусть $\eta, \xi_1, \xi_2, \ldots$ — случайные величины.*

a) *Если $\xi_n\geqslant\eta$ для всех $n\geqslant 1$ и $\mathsf{M}\eta>-\infty$, то*

$$\mathsf{M}\,\underline{\lim}\,\xi_n\leqslant\underline{\lim}\,\mathsf{M}\xi_n.$$

b) *Если $\xi_n\leqslant\eta$ для всех $n\geqslant 1$ и $\mathsf{M}\eta<\infty$, то*

$$\overline{\lim}\,\mathsf{M}\xi_n\leqslant\mathsf{M}\,\overline{\lim}\,\xi_n.$$

c) *Если $|\xi_n|\leqslant\eta$ для всех $n\geqslant 1$ и $\mathsf{M}\eta<\infty$, то*

$$\mathsf{M}\,\underline{\lim}\,\xi_n\leqslant\underline{\lim}\,\mathsf{M}\xi_n\leqslant\overline{\lim}\,\mathsf{M}\xi_n\leqslant\mathsf{M}\,\overline{\lim}\,\xi_n. \tag{7}$$

Доказательство. а) Пусть $\zeta_n = \inf\limits_{m \geqslant n} \xi_m$, тогда

$$\underline{\lim}\, \xi_n = \lim_n \inf_{m \geqslant n} \xi_m = \lim_n \zeta_n.$$

Ясно, что $\zeta_n \uparrow \underline{\lim}\, \xi_n$ и $\zeta_n \geqslant \eta$ для всех $n \geqslant 1$. Тогда из теоремы 1

$$\mathsf{M}\, \underline{\lim}\, \xi_n = \mathsf{M} \lim_n \zeta_n = \lim_n \mathsf{M}\zeta_n = \underline{\lim}_n\, \mathsf{M}\zeta_n \leqslant \underline{\lim}\, \mathsf{M}\xi_n,$$

что и доказывает утверждение а). Второе утверждение следует из первого. Третье — есть следствие первых двух.

Теорема 3 (Лебега о мажорируемой сходимости). *Пусть* $\eta, \xi, \xi_1, \xi_2, \ldots$ — *случайные величины такие, что* $|\xi_n| \leqslant \eta$, $\mathsf{M}\eta < \infty$ *и* $\xi_n \to \xi$ (*п. н.*). *Тогда* $\mathsf{M}|\xi| < \infty$,

$$\mathsf{M}\xi_n \to \mathsf{M}\xi \tag{8}$$

и

$$\mathsf{M}|\xi_n - \xi| \to 0 \tag{9}$$

при $n \to \infty$.

Доказательство. В силу леммы Фату справедлива формула (7). По предположению $\underline{\lim}\, \xi_n = \overline{\lim}\, \xi_n = \xi$ (п. н.). Поэтому по свойству **G**

$$\mathsf{M}\, \underline{\lim}\, \xi_n = \underline{\lim}\, \mathsf{M}\xi_n = \overline{\lim}\, \mathsf{M}\xi_n = \mathsf{M}\, \overline{\lim}\, \xi_n = \mathsf{M}\xi,$$

что и доказывает (8). Ясно также, что $|\xi| \leqslant \eta$. Поэтому $\mathsf{M}|\xi| < \infty$.

Утверждение (9) доказывается так же, если только заметить, что $|\xi_n - \xi| \leqslant 2\eta$.

Следствие. Пусть $\eta, \xi, \xi_1, \ldots$ — случайные величины такие, что $|\xi_n| \leqslant \eta$, $\xi_n \to \xi$ (п. н.) и $\mathsf{M}\eta^p < \infty$ для некоторого $p > 0$. Тогда $\mathsf{M}|\xi|^p < \infty$ и $\mathsf{M}|\xi - \xi_n|^p \to 0$, $n \to \infty$.

Для доказательства достаточно заметить, что $|\xi| \leqslant \eta$, $|\xi - \xi_n|^p \leqslant (|\xi| + |\xi_n|)^p \leqslant (2\eta)^p$.

Условие «$|\xi_n| \leqslant \eta$, $\mathsf{M}\eta < \infty$», входящее в лемму Фату, теорему о мажорируемой сходимости и обеспечивающее выполнение формул (7) — (9), можно несколько ослабить. Для формулировки соответствующего результата (теорема 4) введем

Определение 4. Семейство случайных величин $\{\xi_n\}_{n \geqslant 1}$ называется *равномерно интегрируемым*, если

$$\sup_n \int_{\{|\xi_n| > c\}} |\xi_n|\, \mathsf{P}(d\omega) \to 0, \quad c \to \infty, \tag{10}$$

или (в других обозначениях)

Ясно, что если случайные величины ξ_n, $n \geqslant 1$, таковы, что $|\xi_n| \leqslant \eta$, $\mathsf{M}\eta < \infty$, то семейство $\{\xi_n\}_{n \geqslant 1}$ будет равномерно интегрируемым.

Теорема 4. *Пусть $\{\xi_n\}_{n \geqslant 1}$ — семейство равномерно интегрируемых случайных величин.*

a) *Тогда*

$$\mathsf{M} \varliminf \xi_n \leqslant \varliminf \mathsf{M}\xi_n \leqslant \varlimsup \mathsf{M}\xi_n \leqslant \mathsf{M} \varlimsup \xi_n.$$

b) *Если к тому же $\xi_n \to \xi$ (п. н.), тогда случайная величина ξ интегрируема и*

$$\mathsf{M}\xi_n \to \mathsf{M}\xi, \qquad n \to \infty,$$
$$\mathsf{M}|\xi_n - \xi| \to 0, \quad n \to \infty.$$

Доказательство. a) Для всякого $c > 0$

$$\mathsf{M}\xi_n = \mathsf{M}\left[\xi_n I_{\{\xi_n < -c\}}\right] + \mathsf{M}\left[\xi_n I_{\{\xi_n \geqslant -c\}}\right]. \tag{12}$$

В силу равномерной интегрируемости для всякого $\varepsilon > 0$ можно c выбрать столь большим, что

$$\sup_n \left|\mathsf{M}\left[\xi_n I_{\{\xi_n < -c\}}\right]\right| < \varepsilon. \tag{13}$$

В силу леммы Фату

$$\varliminf \mathsf{M}\left[\xi_n I_{\{\xi_n \geqslant -c\}}\right] \geqslant \mathsf{M}\left[\varliminf \xi_n I_{\{\xi_n \geqslant -c\}}\right].$$

Но $\xi_n I_{\{\xi_n \geqslant -c\}} \geqslant \xi_n$, поэтому

$$\varliminf \mathsf{M}\left[\xi_n I_{\{\xi_n \geqslant -c\}}\right] \geqslant \mathsf{M}\left[\varliminf \xi_n\right]. \tag{14}$$

Из (12) — (14) находим, что

$$\varliminf \mathsf{M}\xi_n \geqslant \mathsf{M}\left[\varliminf \xi_n\right] - \varepsilon.$$

В силу произвольности $\varepsilon > 0$ отсюда следует, что $\varliminf \mathsf{M}\xi_n \geqslant \mathsf{M} \varliminf \xi_n$. Аналогично доказывается неравенство с верхними пределами. Утверждение b) вытекает из a) так же, как и в теореме 3.

Аналогичным образом доказывается, что $\varlimsup \mathsf{M}\xi_n \leqslant \mathsf{M} \varlimsup \xi_n$. Что же касается утверждений b), то они доказываются так же, как соответствующие утверждения в теореме 3.

Наиболее полно значение понятия равномерной интегрируемости раскрывается в следующей теореме, дающей необходимое и достаточное условие для предельного перехода под знаком математического ожидания.

Теорема 5. *Пусть $0 \leqslant \xi_n \to \xi$ и $\mathsf{M}\xi_n < \infty$. Тогда $\mathsf{M}\xi_n \to \mathsf{M}\xi < \infty$ тогда и только тогда, когда семейство случайных величин $\{\xi_n\}_{n \geqslant 1}$ равномерно интегрируемо.*

Доказательство. Достаточность следует из утверждения b) теоремы 4. Для доказательства необходимости рассмотрим (не более чем счетное) множество $A=\{a\colon \mathsf{P}(\xi=a)>0\}$. Тогда для каждого $a\notin A$ $\xi_n I_{\{\xi_n<a\}}\to\xi I_{\{\xi<a\}}$, причем семейство величин $\{\xi_n I_{\{\xi_n<a\}}\}_{n\geqslant 1}$ будет равномерно интегрируемым. Поэтому в силу «достаточности» $\mathsf{M}\xi_n I_{\{\xi_n<a\}}\to\mathsf{M}\xi I_{\{\xi<a\}}$, $a\notin A$, а значит,

$$\mathsf{M}\xi_n I_{\{\xi_n\geqslant\alpha\}}\to\mathsf{M}\xi I_{\{\xi\geqslant a\}},\quad a\notin A,\quad n\to\infty. \tag{15}$$

Зафиксируем $\varepsilon>0$ и выберем сначала $a_0\in A$ столь большим, что $\mathsf{M}\xi I_{\{\xi\geqslant a_0\}}<\varepsilon/2$, а затем N_0 столь большим, что для всех $n\geqslant N_0$

$$\mathsf{M}\xi_n I_{\{\xi_n\geqslant a_0\}}\leqslant\mathsf{M}\xi I_{\{\xi\geqslant a_0\}}+\varepsilon/2,$$

и значит, $\mathsf{M}\xi_n I_{\{\xi_n\geqslant a_0\}}\leqslant\varepsilon$. Выберем, наконец, $a_1\geqslant a_0$ и столь большим, что для всех $n\leqslant N_0$ $\mathsf{M}\xi_n I_{\{\xi_n\geqslant a_1\}}\leqslant\varepsilon$. Тогда

$$\sup_n \mathsf{M}\xi_n I_{\{\xi_n\geqslant a_1\}}\leqslant\varepsilon,$$

что и доказывает равномерную интегрируемость семейства случайных величин $\{\xi_n\}_{n\geqslant 1}$.

4. Остановимся на некоторых критериях равномерной интегрируемости.

Прежде всего заметим, что если $\{\xi_n\}$ — семейство равномерно интегрируемых случайных величин, то

$$\sup_n \mathsf{M}|\xi_n|<\infty. \tag{16}$$

В самом деле, для фиксированного $\varepsilon>0$ и достаточно больших $c>0$

$$\sup_n \mathsf{M}|\xi_n|=\sup_n\left[\mathsf{M}\left(|\xi_n|I_{\{|\xi_n|\geqslant c\}}\right)+\mathsf{M}\left(|\xi_n|I_{\{|\xi_n|<c\}}\right)\right]\leqslant$$
$$\leqslant\sup_n \mathsf{M}\left(|\xi_n|I_{\{|\xi_n|\geqslant c\}}\right)+\sup_n \mathsf{M}\left(|\xi_n|I_{(|\xi_n|<c)}\right)\leqslant\varepsilon+c,$$

что и доказывает (16).

Оказывается, что условие (16) вместе с так называемым условием «равномерной непрерывности» является необходимым и достаточным для равномерной интегрируемости.

Лемма 2. *Для того чтобы семейство случайных величин* $\{\xi_n\}_{n\geqslant 1}$ *было равномерно интегрируемо, необходимо и достаточно, чтобы* $\mathsf{M}|\xi_n|$, $n\geqslant 1$, *были равномерно ограничены (т. е. выполнено условие* (16)) *и чтобы* $\mathsf{M}\{|\xi_n|I_A\}$, $n\geqslant 1$, *были равномерно непрерывны (т. е.* $\sup\mathsf{M}\{|\xi_n|I_A\}\to 0$, *когда* $\mathsf{P}(A)\to 0$).

Доказательство. Необходимость. Условие (16) было проверено выше. Далее,

$$\mathsf{M}\{|\xi_n| I_A\} = \mathsf{M}\{|\xi_n| I_{A\cap\{|\xi_n|\geqslant c\}}\} + \mathsf{M}\{|\xi_n| I_{A\cap\{|\xi_n|<c\}}\} \leqslant \leqslant \mathsf{M}\{|\xi_n| I_{\{|\xi_n|\geqslant c\}}\} + c\mathsf{P}(A). \tag{17}$$

Выберем c столь большим, что $\sup\limits_n \mathsf{M}\{|\xi_n| I_{\{|\xi_n|\geqslant c\}}\} \leqslant \varepsilon/2$. Тогда, если $\mathsf{P}(A) \leqslant \varepsilon/2c$, то из (17)

$$\sup_n \mathsf{M}\{|\xi_n| I_A\} \leqslant \varepsilon,$$

что и доказывает равномерную непрерывность.

Достаточность. Пусть $\varepsilon > 0$ и $\delta > 0$ таково, что из условия $\mathsf{P}(A) < \delta$ следует, что равномерно по n $\mathsf{M}(|\xi_n| I_A) \leqslant \varepsilon$. Поскольку для всякого $c > 0$

$$\mathsf{M}|\xi_n| \geqslant \mathsf{M}|\xi_n| I_{\{|\xi_n|\geqslant c\}} \geqslant c\mathsf{P}\{|\xi_n| \geqslant c\}$$

(ср. с неравенством Чебышева), то

$$\sup_n \mathsf{P}\{|\xi_n| \geqslant c\} \leqslant \frac{1}{c} \sup \mathsf{M}|\xi_n| \to 0, \quad c \to \infty,$$

а значит, для достаточно больших c в качестве множества A можно взять любое из множеств $\{|\xi_n| \geqslant c\}$, $n \geqslant 1$. Поэтому $\sup \mathsf{M}(|\xi_n| I_{\{|\xi_n|\geqslant c\}}) \leqslant \varepsilon$, что и доказывает равномерную интегрируемость. Лемма доказана.

В следующем предложении дается удобное достаточное условие равномерной интегрируемости.

Лемма 3. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность интегрируемых случайных величин и $G = G(t)$ — неотрицательная возрастающая функция, определенная для $t \geqslant 0$, такая, что*

$$\lim_{t\to\infty} \frac{G(t)}{t} = \infty, \tag{18}$$

$$\sup_n \mathsf{M}[G(|\xi_n|)] < \infty. \tag{19}$$

Тогда семейство случайных величин $\{\xi_n\}_{n\geqslant 1}$ является равномерно интегрируемым.

Доказательство. Пусть $\varepsilon > 0$, $M = \sup\limits_n \mathsf{M}[G(|\xi_n|)]$, $a = \frac{M}{\varepsilon}$. Выберем c столь большим, что $\frac{G(t)}{t} \geqslant a$ для $t \geqslant c$. Тогда

$$\mathsf{M}[|\xi_n| I_{\{|\xi_n|\geqslant c\}}] \leqslant \frac{1}{a} \mathsf{M}[G(|\xi_n|) \cdot I_{\{|\xi_n|\geqslant c\}}] \leqslant \frac{M}{a} = \varepsilon$$

равномерно по всем $n \geqslant 1$.

5. Если ξ и η — независимые простые случайные величины, то, как и в п. 5 § 4 гл. I, доказывается, что $\mathsf{M}\xi\eta = \mathsf{M}\xi \cdot \mathsf{M}\eta$. Установим теперь справедливость аналогичного утверждения в общем случае (см. также задачу 5).

Теорема 6. *Пусть* ξ *и* η — *независимые случайные величины с* $\mathsf{M}|\xi| < \infty$, $\mathsf{M}|\eta| < \infty$. *Тогда* $\mathsf{M}|\xi\eta| < \infty$ и

$$\mathsf{M}\xi\eta = \mathsf{M}\xi \cdot \mathsf{M}\eta. \tag{20}$$

Доказательство. Пусть сначала $\xi \geqslant 0$, $\eta \geqslant 0$. Положим

$$\xi_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{\left\{\frac{k}{n} \leqslant \xi(\omega) < \frac{k+1}{n}\right\}},$$

$$\eta_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{\left\{\frac{k}{n} \leqslant \eta(\omega) < \frac{k+1}{n}\right\}}.$$

Тогда $\xi_n \leqslant \xi$, $|\xi_n - \xi| \leqslant \frac{1}{n}$ и $\eta_n \leqslant \eta$, $|\eta_n - \eta| \leqslant 1/n$. Поскольку $\mathsf{M}\xi < \infty$, $\mathsf{M}\eta < \infty$, то по теореме Лебега о мажорируемой сходимости

$$\lim \mathsf{M}\xi_n = \mathsf{M}\xi, \quad \lim \mathsf{M}\eta_n = \mathsf{M}\eta.$$

Далее, в силу независимости ξ и η

$$\mathsf{M}\xi_n\eta_n = \sum_{k,\, l \geqslant 0} \frac{kl}{n^2} \mathsf{M} I_{\left\{\frac{k}{n} \leqslant \xi < \frac{k+1}{n}\right\}} I_{\left\{\frac{l}{n} \leqslant \eta < \frac{l+1}{n}\right\}} =$$
$$= \sum_{k,\, l \geqslant 0} \frac{kl}{n^2} \mathsf{M} I_{\left\{\frac{k}{n} \leqslant \xi < \frac{k+1}{n}\right\}} \cdot \mathsf{M} I_{\left\{\frac{l}{n} \leqslant \eta < \frac{l+1}{n}\right\}} = \mathsf{M}\xi_n \cdot \mathsf{M}\eta_n.$$

Заметим теперь, что

$$|\mathsf{M}\xi\eta - \mathsf{M}\xi_n\eta_n| \leqslant \mathsf{M}|\xi\eta - \xi_n\eta_n| \leqslant \mathsf{M}[|\xi| \cdot |\eta - \eta_n|] +$$
$$+ \mathsf{M}[|\eta_n| \cdot |\xi - \xi_n|] \leqslant \frac{1}{n} \mathsf{M}\xi + \frac{1}{n} \mathsf{M}\left(\eta + \frac{1}{n}\right) \to 0, \quad n \to \infty.$$

Поэтому $\mathsf{M}\xi\eta = \lim_n \mathsf{M}\xi_n\eta_n = \lim \mathsf{M}\xi_n \cdot \lim \mathsf{M}\eta_n = \mathsf{M}\xi \cdot \mathsf{M}\eta$, причем $\mathsf{M}\xi\eta < \infty$.

Общий случай сводится к рассмотренному, если воспользоваться представлениями $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$, $\xi\eta = \xi^+\eta^+ - \xi^-\eta^+ - \xi^+\eta^- + \xi^-\eta^-$. Теорема доказана.

6. Приводимые в этом пункте неравенства для математических ожиданий систематически применяются и в теории вероятностей, и в математическом анализе.

Неравенство Чебышева. *Пусть ξ — неотрицательная случайная величина, тогда для всякого $\varepsilon > 0$*

$$\mathsf{P}(\xi \geqslant \varepsilon) \leqslant \frac{\mathsf{M}\xi}{\varepsilon}. \tag{21}$$

Доказательство сразу следует из того, что

$$\mathsf{M}\xi \geqslant \mathsf{M}[\xi \cdot I_{\{\xi \geqslant \varepsilon\}}] \geqslant \varepsilon \mathsf{M} I_{\{\xi \geqslant \varepsilon\}} = \varepsilon \mathsf{P}(\xi \geqslant \varepsilon).$$

Из (21) получаем следующие разновидности неравенства Чебышева, если ξ — произвольная случайная величина, то

$$\mathsf{P}(\xi \geqslant \varepsilon) \leqslant \frac{\mathsf{M}\xi^2}{\varepsilon^2} \tag{22}$$

и

$$\mathsf{P}(|\xi - \mathsf{M}\xi| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}\xi^2}{\varepsilon^2}, \tag{23}$$

где $\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2$ — дисперсия случайной величины ξ.

Неравенство Коши — Буняковского. *Пусть ξ и η таковы, что $\mathsf{M}\xi^2 < \infty$, $\mathsf{M}\eta^2 < \infty$. Тогда $\mathsf{M}|\xi\eta| < \infty$ и*

$$(\mathsf{M}|\xi\eta|)^2 \leqslant \mathsf{M}\xi^2 \cdot \mathsf{M}\eta^2. \tag{24}$$

Доказательство. Будем предполагать, что $\mathsf{M}\xi^2 > 0$, $\mathsf{M}\eta^2 > 0$. Тогда, обозначая $\tilde{\xi} = \frac{\xi}{\sqrt{\mathsf{M}\xi^2}}$, $\tilde{\eta} = \frac{\eta}{\sqrt{\mathsf{M}\eta^2}}$, находим, что поскольку $2|\tilde{\xi}\tilde{\eta}| \leqslant \tilde{\xi}^2 + \tilde{\eta}^2$, то

$$2\mathsf{M}|\tilde{\xi}\tilde{\eta}| \leqslant \mathsf{M}\tilde{\xi}^2 + \mathsf{M}\tilde{\eta}^2 = 2,$$

т. е. $\mathsf{M}|\tilde{\xi}\tilde{\eta}| \leqslant 1$, что и доказывает (24).

Если же, скажем, $\mathsf{M}\xi^2 = 0$, то тогда по свойству **I** $\xi = 0$ (п. н.) и по свойству **F** $\mathsf{M}\xi\eta = 0$, т. е. (24) также выполнено.

Неравенство Иенсена. Пусть $g = g(x)$ — выпуклая книзу борелевская функция и $\mathsf{M}|\xi| < \infty$. Тогда

$$g(\mathsf{M}\xi) \leqslant \mathsf{M}g(\xi). \tag{25}$$

Доказательство. Если функция $g = g(x)$ выпуклая книзу, то для каждого $x_0 \in R$ найдется число $\lambda(x_0)$ такое, что для всех $x \in R$

$$g(x) \geqslant g(x_0) + (x - x_0) \cdot \lambda(x_0). \tag{26}$$

Полагая $x = \xi$ и $x_0 = \mathsf{M}\xi$, из (26) находим, что

$$g(\xi) \geqslant g(\mathsf{M}\xi) + (\xi - \mathsf{M}\xi) \cdot \lambda(\mathsf{M}\xi)$$

и, следовательно, $\mathsf{M}g(\xi) \geqslant g(\mathsf{M}\xi)$.

Из неравенства Иенсена выводится целая серия полезных неравенств. Получим, к примеру,

Неравенство Ляпунова. *Если* $0<s<t$, *то*

$$(\mathsf{M}|\xi|^s)^{1/s} \leqslant (\mathsf{M}|\xi|^t)^{1/t}. \tag{27}$$

Для доказательства обозначим $r=t/s$. Тогда, полагая $\eta=|\xi|^s$ и применяя неравенство Иенсена к функции $g(x)=|x|^r$, находим, $|\mathsf{M}\eta|^r \leqslant \mathsf{M}|\eta|^r$, т. е.

$$(\mathsf{M}|\xi|^s)^{t/s} \leqslant \mathsf{M}|\xi|^t,$$

что и доказывает (27).

Из неравенства Ляпунова вытекает следующая цепочка неравенств между абсолютными моментами:

$$\mathsf{M}|\xi| \leqslant (\mathsf{M}|\xi|^2)^{1/2} \leqslant \ldots \leqslant (\mathsf{M}|\xi|^n)^{1/n}. \tag{28}$$

Неравенство Гёльдера. *Пусть* $1<p<\infty$, $1<q<\infty$ *и* $\frac{1}{p}+\frac{1}{q}=1$. *Если* $\mathsf{M}|\xi|^p<\infty$, $\mathsf{M}|\eta|^q<\infty$, *то* $\mathsf{M}|\xi\eta|<\infty$ *и*

$$\mathsf{M}|\xi\eta| \leqslant (\mathsf{M}|\xi|^p)^{1/p}(\mathsf{M}|\eta|^q)^{1/q}. \tag{29}$$

Если $\mathsf{M}|\xi|^p=0$ или $\mathsf{M}|\eta|^q=0$, то (24) следует немедленно, так же как и в случае неравенства Коши — Буняковского (являющегося частным случаем неравенства Гёльдера при $p=q=2$).

Пусть теперь $\mathsf{M}|\xi|^p>0$, $\mathsf{M}|\eta|^q>0$ и

$$\tilde{\xi}=\frac{\xi}{(\mathsf{M}|\xi|^p)^{1/p}}, \quad \tilde{\eta}=\frac{\eta}{(\mathsf{M}|\eta|^q)^{1/q}}.$$

Воспользуемся неравенством

$$x^a y^b \leqslant ax+by, \tag{30}$$

справедливым для положительных x, y, a, b, $a+b=1$ и вытекающим непосредственно из свойства выпуклости кверху логарифмической функции:

$$\ln[ax+by] \geqslant a\ln x + b\ln y = \ln x^a y^b.$$

Тогда, полагая $x=\tilde{\xi}^p$, $y=\tilde{\eta}^q$, $a=\frac{1}{p}$, $b=\frac{1}{q}$, находим, что

$$\tilde{\xi}\tilde{\eta} \leqslant \frac{1}{p}\tilde{\xi}^p+\frac{1}{q}\tilde{\eta}^q,$$

откуда

$$\mathsf{M}\tilde{\xi}\tilde{\eta} \leqslant \frac{1}{p}\mathsf{M}\tilde{\xi}^p+\frac{1}{q}\mathsf{M}\tilde{\eta}^q=\frac{1}{p}+\frac{1}{q}=1,$$

что и доказывает (29).

Неравенство Минковского. *Если* $\mathsf{M}|\xi|^p<\infty$, $\mathsf{M}|\eta|^p<\infty$, $1\leqslant p<\infty$, *то* $\mathsf{M}|\xi+\eta|^p<\infty$ *и*

$$(\mathsf{M}|\xi+\eta|^p)^{1/p}\leqslant(\mathsf{M}|\xi|^p)^{1/p}+(\mathsf{M}|\eta|^p)^{1/p}. \tag{31}$$

Установим прежде всего следующее неравенство: если $a,\ b>0$ и $p\geqslant 1$, то

$$(a+b)^p\leqslant 2^{p-1}(a^p+b^p). \tag{32}$$

В самом деле, рассмотрим функцию $F(x)=(a+x)^p-2^{p-1}\times\times(a^p+x^p)$. Тогда

$$F'(x)=p(a+x)^{p-1}-2^{p-1}px^{p-1},$$

и поскольку $p\geqslant 1$, то $F'(a)=0$, $F'(x)>0$ для $x<a$ и $F'(x)<0$ для $x>a$. Поэтому

$$F(b)\leqslant\max F(x)=F(a)=0,$$

что и дает неравенство (32).

В соответствии с этим неравенством

$$|\xi+\eta|^p\leqslant(|\xi|+|\eta|)^p\leqslant 2^{p-1}(|\xi|^p+|\eta|^p) \tag{33}$$

и, значит, если $\mathsf{M}|\xi|^p<\infty$, $\mathsf{M}|\eta|^p<\infty$, то $\mathsf{M}|\xi+\eta|^p<\infty$.

Если $p=1$, то неравенство (31) следует из (33).

Будем теперь предполагать, что $p>1$. Возьмем $q>1$ таким, что $\frac{1}{p}+\frac{1}{q}=1$. Тогда

$$|\xi+\eta|^p=|\xi+\eta|\cdot|\xi+\eta|^{p-1}\leqslant|\xi|\cdot|\xi+\eta|^{p-1}+|\eta|\,|\xi+\eta|^{p-1}. \tag{34}$$

Заметим, что $(p-1)q=p$. Поэтому

$$\mathsf{M}(|\xi+\eta|^{p-1})^q=\mathsf{M}|\xi+\eta|^p<\infty,$$

и, значит, в силу неравенства Гёльдера

$$\mathsf{M}(|\xi|\,|\xi+\eta|^{p-1})\leqslant(\mathsf{M}|\xi|^p)^{1/p}(\mathsf{M}|\xi+\eta|^{(p-1)q})^{1/q}=$$
$$=(\mathsf{M}|\xi|^p)^{1/p}(\mathsf{M}|\xi+\eta|^p)^{1/q}<\infty.$$

Точно так же и

$$\mathsf{M}(|\eta|\,|\xi+\eta|^{p-1})\leqslant(\mathsf{M}|\eta|^p)^{1/p}(\mathsf{M}|\xi+\eta|^p)^{1/q}.$$

Поэтому в силу (34)

$$\mathsf{M}|\xi+\eta|^p\leqslant(\mathsf{M}|\xi+\eta|^p)^{1/q}((\mathsf{M}|\xi|^p)^{1/p}+(\mathsf{M}|\eta|^p)^{1/p}). \tag{35}$$

Если $\mathsf{M}|\xi+\eta|^p=0$, то требуемое неравенство (31) очевидно. Пусть теперь $\mathsf{M}|\xi+\eta|^p>0$. Тогда из (35) находим

$$(\mathsf{M}|\xi+\eta|^p)^{1-\frac{1}{q}}\leqslant(\mathsf{M}|\xi|^p)^{1/p}+(\mathsf{M}|\eta|^p)^{1/p},$$

что и дает требуемое неравенство (31), поскольку $1-\frac{1}{q}=\frac{1}{p}$.

7. Пусть ξ — случайная величина, для которой определено математическое ожидание $\mathsf{M}\xi$. Тогда, согласно свойству **D**, определена функция множеств

$$Q(A) \equiv \int_A \xi \, d\mathsf{P}, \qquad A \in \mathscr{F}. \tag{36}$$

Покажем, что эта функция является счетно-аддитивной.

Предположим сначала, что ξ — неотрицательная случайная величина. Если $A_1, A_2, \ldots$ — попарно непересекающиеся множества из $\mathscr{F}$ и $A = \sum A_n$, то в силу следствия к теореме 1

$$\begin{aligned} Q(A) &= \mathsf{M}(\xi \cdot I_A) = \mathsf{M}(\xi \cdot I_{\Sigma A_n}) = \mathsf{M}(\textstyle\sum \xi \cdot I_{A_n}) = \\ &= \textstyle\sum \mathsf{M}(\xi \cdot I_{A_n}) = \sum Q(A_n). \end{aligned}$$

Если же ξ — произвольная случайная величина, для которой $\mathsf{M}\xi$ определено, то счетная аддитивность $Q(A)$ следует из представления

$$Q(A) = Q^+(A) - Q^-(A), \tag{37}$$

где

$$Q^+(A) = \int_A \xi^+ \, d\mathsf{P}, \quad Q^-(A) = \int_A \xi^- \, d\mathsf{P},$$

установленной счетной аддитивности для неотрицательных случайных величин и того факта, что $\min(Q^+(\Omega), Q^-(\Omega)) < \infty$.

Итак, если $\mathsf{M}\xi$ определено, то функция множеств $Q = Q(A)$ является мерой со знаком — счетно-аддитивной функцией множеств, представимой в виде $Q = Q_1 - Q_2$, где по крайней мере одна из мер Q_1 или Q_2 конечна.

Покажем, что функция множеств $Q = Q(A)$ обладает следующим важным свойством *абсолютной непрерывности* относительно меры P:

$$\text{если } \mathsf{P}(A) = 0, \text{ то } Q(A) = 0 \quad (A \in \mathscr{F})$$

(это свойство кратко записывают в виде: $Q \ll \mathsf{P}$).

Для доказательства достаточно рассмотреть случай неотрицательных случайных величин. Если $\xi = \sum_{k=1}^{n} x_k I_{A_k}$ — простая неотрицательная случайная величина и $\mathsf{P}(A) = 0$, то

$$Q(A) = M(\xi \cdot I_A) = \sum_{k=1}^{n} x_k \mathsf{P}(A_k \cap A) = 0.$$

Если же $\{\xi_n\}_{n \geqslant 1}$ — последовательность неотрицательных простых функций таких, что $\xi_n \uparrow \xi \geqslant 0$, то по теореме о монотонной сходимости

$$\mathsf{Q}(A) = \mathsf{M}(\xi \cdot I_A) = \lim \mathsf{M}(\xi_n \cdot I_A) = 0,$$

поскольку $\mathsf{M}(\xi_n \cdot I_A) = 0$ для любого $n \geq 1$ и A с $\mathsf{P}(A) = 0$.

Итак, интеграл Лебега $\mathsf{Q}(A) = \int\limits_A \xi \, d\mathsf{P}$, рассматриваемый как функция множеств $A \in \mathscr{F}$, является мерой со знаком, абсолютно-непрерывной относительно меры P ($\mathsf{Q} \ll \mathsf{P}$). Весьма замечательно, что имеет место и обратный результат.

Теорема Радона — Никодима. *Пусть $(\Omega, \mathscr{F})$ — измеримое пространство, μ — σ-конечная мера и λ — мера со знаком (т. е. $\lambda = \lambda_1 - \lambda_2$, где по крайней мере одна из мер λ_1 или λ_2 конечна), являющаяся абсолютно непрерывной относительно μ. Тогда существует $\mathscr{F}$-измеримая функция $f = f(\omega)$, принимающая значения в $\bar{R} = [-\infty, \infty]$ такая, что*

$$\lambda(A) = \int\limits_A f(\omega) \, \mu(d\omega), \qquad A \in \mathscr{F}. \tag{38}$$

С точностью до множеств μ-меры нуль функция $f(\omega)$ единственна: если $h = h(\omega)$ — другая $\mathscr{F}$-измеримая функция такая, что $\lambda(A) = \int\limits_A h(\omega) \, \mu(d\omega)$, $A \in \mathscr{F}$, то $\mu\{\omega: f(\omega) \neq h(\omega)\} = 0$.

Если λ — мера, то $f = f(\omega)$ принимает значения в $\bar{R}_+ = [0, \infty]$.

Замечание. Функция $f = f(\omega)$ в представлении (38) называется *производной Радона — Никодима* или *плотностью* меры λ относительно меры μ, и обозначается $\frac{d\lambda}{d\mu}$ или $\frac{d\lambda}{d\mu}(\omega)$.

Теорема Радона — Никодима, приводимая без доказательства, будет играть ключевую роль в конструкции условных математических ожиданий (§ 7).

8. Если $\xi = \sum\limits_{i=1}^{n} x_i I_{A_i}$ — простая случайная величина, то

$$\mathsf{M}g(\xi) = \sum g(x_i) \, \mathsf{P}(A_i) = \sum g(x_i) \, \Delta F_\xi(x_i). \tag{39}$$

Иначе говоря, для подсчета математического ожидания функции от (простой) случайной величины ξ нет надобности знать всю вероятностную меру P, а достаточно знать распределение вероятностей P_ξ или, что эквивалентно, функцию распределения F_ξ случайной величины ξ.

Следующая важная теорема обобщает это свойство.

Теорема 7 (о замене переменных под знаком интеграла Лебега). *Пусть $(\Omega, \mathscr{F})$ и $(E, \mathscr{E})$ — два измеримых пространства и $X = X(\omega)$ — $\mathscr{F}/\mathscr{E}$-измеримая функция со значениями в E. Пусть P — вероятностная мера на $(\Omega, \mathscr{F})$ и P_X — вероятностная мера* *на $(E, \mathscr{E})$, индуцируемая $X = X(\omega)$:*

$$P_X(A) = \mathsf{P}\{\omega\colon X(\omega) \in A\}, \quad A \in \mathscr{E}. \tag{40}$$

Тогда для всякой $\mathscr{E}$-измеримой функции $g = g(x)$, $x \in E$

$$\int_A g(x)\,P_X(dx) = \int_{X^{-1}(A)} g(X(\omega))\,\mathsf{P}(d\omega), \qquad A \in \mathscr{E} \tag{41}$$

(в том смысле, что если существует один из интегралов, то определен и второй, и они совпадают).

Доказательство. Пусть множество $A \in \mathscr{E}$ и $g(x) = I_B(x)$, где $B \in \mathscr{E}$. Тогда искомое соотношение (41) превращается в равенство

$$P_X(AB) = \mathsf{P}(X^{-1}(A) \cap X^{-1}(B)), \tag{42}$$

справедливость которого следует из (40) и замечания, что $X^{-1}(A) \cap \cap X^{-1}(B) = X^{-1}(A \cap B)$.

Из (42) вытекает, что (41) справедливо для неотрицательных простых функций $g = g(x)$, а значит, в силу теоремы о монотонной сходимости (41) справедливо и для произвольных неотрицательных $\mathscr{E}$-измеримых функций.

В общем же случае надо представить функцию g в виде $g^+ - g^-$ и заметить, что, поскольку для функций g^+ и g^- равенство (41) справедливо и если, например, $\int_A g^+(x)\,P_X(dx) < \infty$, то и $\int_{X^{-1}(A)} g^+(X(\omega))\,\mathsf{P}(d\omega) < \infty$, а значит, из существования $\int_A g(x)\,P_X(dx)$ следует существование интеграла $\int_{X^{-1}(A)} g(X(\omega))\,\mathsf{P}(d\omega)$.

Следствие. Пусть $(E, \mathscr{E}) = (R, \mathscr{B}(R))$ и $\xi = \xi(\omega)$ — случайная величина с распределением вероятностей P_ξ. Тогда, если $g = g(x)$ — борелевская функция и существует любой из интегралов $\int_A g(x)\,P_\xi(dx)$ или $\int_{\xi^{-1}(A)} g(\xi(\omega))\,\mathsf{P}(d\omega)$, то

$$\int_A g(x)\,P_\xi(dx) = \int_{\xi^{-1}(A)} g(\xi(\omega))\,\mathsf{P}(d\omega).$$

В частности, при $A = \Omega$ получаем, что

$$\mathsf{M}g(\xi(\omega)) = \int_\Omega g(\xi(\omega))\,\mathsf{P}(d\omega) = \int_R g(x)\,P_\xi(dx). \tag{43}$$

Мера P_ξ однозначно восстанавливается по функции распределения F_ξ (теорема 1 в § 3). Поэтому интегралы Лебега $\int_R g(x)\,P_\xi(dx)$ часто обозначают $\int_R g(x)\,F_\xi(dx)$ и называют *интегралами Лебега —*

Стилтьеса (по мере, соответствующей функции распределения $F_\xi(x)$).

Рассмотрим случай, когда функция распределения $F_\xi(x)$ имеет плотность $f_\xi(x)$, т. е. пусть

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(y)\,dy, \tag{44}$$

где $f_\xi = f_\xi(x)$ — неотрицательная борелевская функция, а интеграл понимается как интеграл Лебега по лебеговской мере на множество $(-\infty, x]$ (см. замечание 2 в п. 1). В предположении (44) формула (43) принимает следующий вид:

$$\mathsf{M}g(\xi(\omega)) = \int_{-\infty}^{\infty} g(x) f_\xi(x)\,dx, \tag{45}$$

где интеграл понимается как интеграл Лебега от функции $g(x) f_\xi(x)$ по лебеговской мере. В самом деле, если $g(x) = I_B(x)$, $B \in \mathscr{B}(R)$, то требуемая формула превращается в равенство

$$P_\xi(B) = \int_B f_\xi(x)\,dx, \qquad B \in \mathscr{B}(R), \tag{46}$$

справедливость которого следует из теоремы 1 § 3 и формулы

$$F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(x)\,dx.$$

В общем случае доказательство то же, что и в теореме 7.

9. Рассмотрим специальный случай измеримых пространств $(\Omega, \mathscr{F})$ с мерой μ, где $\Omega = \Omega_1 \times \Omega_2$, $\mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2$, а мера $\mu = \mu_1 \times \mu_2$ — есть прямое произведение конечных мер μ_1 и μ_2 (т. е. такая мера на $\mathscr{F}$, что

$$\mu_1 \times \mu_2 (A \times B) = \mu_1(A_1)\,\mu_2(B), \qquad A \in \mathscr{F}_1, \quad B \in \mathscr{F}_2;$$

существование такой меры будет следовать из доказательства теоремы 8).

Приводимая далее теорема играет ту же самую роль, что и известная теорема из анализа о сведении двойного интеграла Римана к повторному.

Теорема 8 (теорема Фубини). *Пусть $\xi = \xi(\omega_1, \omega_2)$ является $\mathscr{F}_1 \otimes \mathscr{F}_2$-измеримой функцией, интегрируемой по мере $\mu_1 \times \mu_2$:*

$$\int_{\Omega_1 \times \Omega_2} |\xi(\omega_1, \omega_2)|\,d(\mu_1 \times \mu_2) < \infty. \tag{47}$$

Тогда интегралы $\int_{\Omega_1} \xi(\omega_1, \omega_2)\,\mu_1(d\omega_1)$ *и* $\int_{\Omega_2} \xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)$

1) *определены для всех* ω_2 *и* ω_1; 2) *являются* $\mathscr{F}_2$- и $\mathscr{F}_1$-*измеримыми функциями, соответственно,*

$$\mu_2\left\{\omega_2\colon \int_{\Omega_1} |\xi(\omega_1, \omega_2)|\,\mu_1(d\omega_1)=\infty\right\}=0,$$
$$\mu_1\left\{\omega_1\colon \int_{\Omega_2} |\xi(\omega_1, \omega_2)|\,\mu_2(d\omega_2)=\infty\right\}=0 \tag{48}$$

и 3)

$$\int_{\Omega_1\times\Omega_2} \xi(\omega_1, \omega_2)\,d(\mu_1\times\mu_2)=\int_{\Omega_1}\left[\int_{\Omega_2}\xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=$$
$$=\int_{\Omega_2}\left[\int_{\Omega_1}\xi(\omega_1, \omega_2)\,\mu_1(d\omega_1)\right]\mu_2(d\omega_2). \tag{49}$$

Доказательство. Покажем прежде всего, что для любого фиксированного $\omega_1\in\Omega_1$ функция $\xi_{\omega_1}(\omega_2)=\xi(\omega_1, \omega_2)$ является $\mathscr{F}_2$-измеримой по ω_2.

Пусть $F\in\mathscr{F}_1\otimes\mathscr{F}_2$ и $\xi(\omega_1, \omega_2)=I_F(\omega_1, \omega_2)$. Обозначим $F_{\omega_1}=\{\omega_2\in\Omega_2\colon (\omega_1, \omega_2)\in F\}$ — *сечение* множества F в точке ω_1, и пусть $\mathscr{C}_{\omega_1}=\{F\in\mathscr{F}\colon F_{\omega_1}\in\mathscr{F}_2\}$. Надо показать, что для любого ω_1 $\mathscr{C}_{\omega_1}=\mathscr{F}$.

Если $F=A\times B$, $A\in\mathscr{F}_1$, $B\in\mathscr{F}_2$, то

$$(A\times B)_{\omega_1}=\begin{cases} B, & \text{если } \omega_1\in A,\\ \varnothing, & \text{если } \omega_1\notin A.\end{cases}$$

Поэтому прямоугольники с измеримыми сторонами принадлежат $\mathscr{C}_{\omega_1}$. Далее, если $F\in\mathscr{F}$, то $(\bar F)_{\omega_1}=\overline{F_{\omega_1}}$, а если $\{F^n\}_{n\geq 1}$ — множества из $\mathscr{F}$, то $(\bigcup F^n)_{\omega_1}=\bigcup F^n_{\omega_1}$. Отсюда следует, что $\mathscr{C}_{\omega_1}=\mathscr{F}$.

Пусть теперь $\xi(\omega_1, \omega_2)\geq 0$. Тогда, поскольку для каждого ω_1 функция $\xi(\omega_1, \omega_2)$ является $\mathscr{F}_2$-измеримой, то определен интеграл $\int_{\Omega_2}\xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)$. Покажем, что этот интеграл является $\mathscr{F}_1$-измеримой функцией и

$$\int_{\Omega_1}\left[\int_{\Omega_2}\xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\int_{\Omega_1\times\Omega_2}\xi(\omega_1, \omega_2)\,d(\mu_1\times\mu_2). \tag{50}$$

Предположим, что $\xi(\omega_1, \omega_2)=I_{A\times B}(\omega_1, \omega_2)$, $A\in\mathscr{F}_1$, $B\in\mathscr{F}_2$. Тогда, поскольку $I_{A\times B}(\omega_1, \omega_2)=I_A(\omega_1)\,I_B(\omega_2)$, то

$$\int_{\Omega_2} I_{A\times B}(\omega_1, \omega_2)\,\mu_2(d\omega_2)=I_A(\omega_1)\int_{\Omega_2} I_B(\omega_2)\,\mu_2(d\omega_2) \tag{51}$$

и, следовательно, интеграл в левой части (51) является $\mathscr{F}_1$-измеримой функцией.

Пусть теперь $\xi(\omega_1, \omega_2)=I_F(\omega_1, \omega_2)$, $F\in\mathscr{F}=\mathscr{F}_1\otimes\mathscr{F}_2$. Покажем, что интеграл $f(\omega_1)=\int_{\Omega_2} I_F(\omega_1, \omega_2)\,\mu_2(d\omega_2)$ является $\mathscr{F}$-измеримым. С этой целью обозначим $\mathscr{C}=\{F\in\mathscr{F}\colon f(\omega_1)-\mathscr{F}_1\text{-измеримa}\}$. Согласно доказанному множества $A\times B$ принадлежат $\mathscr{C}$ $(A\in\mathscr{F}_1,\ B\in\mathscr{F}_2)$, а значит, и алгебра $\mathscr{A}$, образованная из конечных сумм непересекающихся множеств такого вида, также принадлежит $\mathscr{C}$. Из теоремы о монотонной сходимости следует, что система $\mathscr{C}$ является монотонным классом, $\mathscr{C}=\mu(\mathscr{C})$. Поэтому в силу включений $\mathscr{A}\subseteq\mathscr{C}\subseteq\mathscr{F}$ и теоремы 1 из § 2 $\mathscr{F}=\sigma(\mathscr{A})=\mu(\mathscr{A})\subseteq\mu(\mathscr{C})=\mathscr{C}\subseteq\mathscr{F}$, т. е. $\mathscr{C}=\mathscr{F}$.

Наконец, если $\xi(\omega_1,\ \omega_2)$ — произвольная неотрицательная $\mathscr{F}$-измеримая функция, то $\mathscr{F}_1$-измеримость интеграла $\int\limits_{\Omega_2}\xi(\omega_1,\ \omega_2)\times\mu_2(d\omega_2)$ следует из теоремы о монотонной сходимости и теоремы 2 § 4.

Покажем сейчас, что мера $\mu=\mu_1\times\mu_2$, определенная на $\mathscr{F}=\mathscr{F}_1\otimes\mathscr{F}_2$ и обладающая свойством $\mu_1\times\mu_2(A\times B)=\mu_1(A)\cdot\mu_2(B)$, $A\in\mathscr{F}_1$, $B\in\mathscr{F}_2$, действительно существует и единственная.

Положим для $F\in\mathscr{F}$

$$\mu(F)=\int\limits_{\Omega_1}\left[\int\limits_{\Omega_2}I_{F_{\omega_1}}(\omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1).$$

Как было показано, внутренний интеграл является $\mathscr{F}_1$-измеримой функцией и, следовательно, функция множеств $\mu(F)$ действительно определена для $F\in\mathscr{F}$. Ясно, что если $F=A\times B$, то $\mu(A\times B)=\mu_1(A)\,\mu_2(B)$. Пусть теперь $\{F^n\}$ — непересекающиеся множества в $\mathscr{F}$. Тогда

$$\begin{aligned}\mu\left(\sum F^n\right)&=\int\limits_{\Omega_1}\left[\int\limits_{\Omega_2}I_{(\Sigma F^n)_{\omega_1}}(\omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\\&=\int\limits_{\Omega_1}\sum_n\left[\int\limits_{\Omega_2}I_{F^n_{\omega_1}}(\omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\\&=\sum_n\int\limits_{\Omega_1}\left[\int\limits_{\Omega_2}I_{F^n_{\omega_1}}(\omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\sum_n\mu(F^n),\end{aligned}$$

т. е. μ является мерой (σ-конечной) на $\mathscr{F}$.

Из теоремы Каратеодори следует, что эта мера μ является единственной мерой со свойством $\mu(A\times B)=\mu_1(A)\,\mu_2(B)$.

Установим теперь формулу (50). Если $\xi(\omega_1,\ \omega_2)=I_{A\times B}(\omega_1,\ \omega_2)$, $A\in\mathscr{F}_1$, $B\in\mathscr{F}_2$, то

$$\int\limits_{\Omega_1\times\Omega_2}I_{A\times B}(\omega_1,\ \omega_2)\,d(\mu_1\times\mu_2)=\mu_1\times\mu_2(A\times B),\tag{52}$$

и так как $I_{A\times B}(\omega_1,\ \omega_2)=I_A(\omega_1)\,I_B(\omega_2)$, то

$$\begin{aligned}&\int\limits_{\Omega_1}\left[\int\limits_{\Omega_2}I_{A\times B}(\omega_1,\ \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\\&\qquad=\int\limits_{\Omega_1}\left[I_A(\omega_1)\int\limits_{\Omega_2}I_B(\omega_1,\ \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\mu_1(A)\,\mu_2(B).\end{aligned}\tag{53}$$

Но по определению меры $\mu_1\times\mu_2$

$$\mu_1\times\mu_2(A\times B)=\mu_1(A)\,\mu_2(B).$$

Поэтому из (52) и (53) следует справедливость (50) для $\xi(\omega_1, \omega_2) = I_{A\times B}(\omega_1, \omega_2)$.

Пусть теперь $\xi(\omega_1, \omega_2)=I_F(\omega_1, \omega_2)$, $F\in\mathscr{F}$. Функция множеств

$$\lambda(F)=\int\limits_{\Omega_1\times\Omega_2} I_F(\omega_1, \omega_2)\,d(\mu_1\times\mu_2), \qquad F\in\mathscr{F}$$

является, очевидно, σ-конечной мерой. Нетрудно проверить также, что таковой же является функция множеств

$$\nu(F)=\int\limits_{\Omega_1}\left[\int\limits_{\Omega_2} I_F(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1).$$

Как было установлено выше, λ и ν совпадают на множестве вида $F=A\times B$, а значит, и на алгебре $\mathscr{A}$. Отсюда по теореме Каратеодори следует, что λ и ν совпадают для всех $F\in\mathscr{F}$.

Перейдем теперь к доказательству собственно утверждений теоремы Фубини. В силу (47)

$$\int\limits_{\Omega_1\times\Omega_2}\xi^+(\omega_1, \omega_2)\,d(\mu_1\times\mu_2)<\infty, \qquad \int\limits_{\Omega_1\times\Omega_2}\xi^-(\omega_1, \omega_2)\,d(\mu_1\times\mu_2)<\infty.$$

Согласно доказанному интеграл $\int\limits_{\Omega_2}\xi^+(\omega_1, \omega_2)\,\mu_2(d\omega_2)$ является $\mathscr{F}_1$-измеримой функцией от ω_1 и

$$\int\limits_{\Omega_1}\left[\int\limits_{\Omega_2}\xi^+(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1)=\int\limits_{\Omega_1\times\Omega_2}\xi^+(\omega_1, \omega_2)\,d(\mu_1\times\mu_2)<\infty.$$

Поэтому в силу задачи 4 (см. также свойство **J** в п. 2)

$$\int\limits_{\Omega_2}\xi^+(\omega_1, \omega_2)\,\mu_2(d\omega_2)<\infty\ (\mu_1\text{-п. н.}).$$

Точно так же и

$$\int\limits_{\Omega_2}\xi^-(\omega_1, \omega_2)\,\mu_1(d\omega_1)<\infty,\ (\mu_1\text{-п. н.}),$$

а значит,

$$\int\limits_{\Omega_2}|\xi(\omega_1, \omega_2)|\,\mu_2(d\omega_2)<\infty\ (\mu_1\text{-п. н.}).$$

Ясно, что за исключением некоторого множества $\mathscr{N}$, имеющего μ_1-меру нуль,

$$\int\limits_{\Omega_2}\xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)=$$

$$=\int\limits_{\Omega_2}\xi^+(\omega_1, \omega_2)\,\mu_2(d\omega_2)-\int\limits_{\Omega_2}\xi^-(\omega_1, \omega_2)\,\mu_2(d\omega_2). \qquad (54)$$

Полагая входящие сюда интегралы равными нулю для $\omega_1 \in \mathscr{N}$, можем считать, что (54) выполнено для всех $\omega_1 \in \Omega_1$. Тогда, интегрируя (54) по мере μ_1 и учитывая (50), получим, что

$$\int_{\Omega_1}\left[\int_{\Omega_2} \xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1) =$$
$$= \int_{\Omega_1}\left[\int_{\Omega_2} \xi^+(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1) -$$
$$- \int_{\Omega_1}\left[\int_{\Omega_2} \xi^-(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1) = \int_{\Omega_1\times\Omega_2} \xi^+(\omega_1, \omega_2)\,d(\mu_1\times\mu_2) -$$
$$- \int_{\Omega_1\times\Omega_2} \xi^-(\omega_1, \omega_2)\,d(\mu_1\times\mu_2) = \int_{\Omega_1\times\Omega_2} \xi(\omega_1, \omega_2)\,d(\mu_1\times\mu_2).$$

Аналогичным образом устанавливается первое соотношение в (48) и равенство

$$\int_{\Omega_1\times\Omega_2} \xi(\omega_1, \omega_2)\,d(\mu_1\times\mu_2) = \int_{\Omega_2}\left[\int_{\Omega_1} \xi(\omega_1, \omega_2)\,\mu_1(d\omega_1)\right]\mu_2(d\omega_2).$$

Теорема доказана.

С л е д с т в и е. Если $\int_{\Omega_1}\left[\int_{\Omega_2} |\xi(\omega_1, \omega_2)|\,\mu_2(d\omega_2)\right]\mu_1(d\omega_1) < \infty$, то утверждения теоремы Фубини также выполнены.

Действительно, при сформулированном условии из (50) следует (47), а значит, справедливы и все утверждения теоремы Фубини.

П р и м е р. Пусть (ξ, η) — пара случайных величин, распределение которых имеет двумерную плотность $f_{\xi\eta}(x, y)$, т. е.

$$\mathsf{P}((\xi, \eta) \in B) = \int_B f_{\xi\eta}(x, y)\,dx\,dy, \quad B \in \mathscr{B}(R^2),$$

где $f_{\xi\eta}(x, y)$ — неотрицательная $\mathscr{B}(R^2)$-измеримая функция, а интеграл понимается как интеграл Лебега по двумерной лебеговской мере.

Покажем, что тогда одномерные распределения для ξ и η также имеют плотности $f_\xi(x)$ и $f_\eta(y)$, причем

$$f_\xi(x) = \int_{-\infty}^{\infty} f_{\xi\eta}(x, y)\,dy$$

и (55)

$$f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi\eta}(x, y)\,dx.$$

В самом деле, если $A \in \mathscr{B}(R)$, то по теореме Фубини

$$\mathsf{P}(\xi \in A) = \mathsf{P}((\xi, \eta) \in A \times R) = \int\limits_{A \times R} f_{\xi\eta}(x, y)\,dx\,dy = \\ = \int\limits_A \left[\int\limits_R f_{\xi\eta}(x, y)\,dy\right] dx,$$

что и доказывает как наличие плотности распределения вероятностей у ξ, так и первую формулу (55). Аналогично доказывается вторая формула.

Согласно теореме из § 5, для того чтобы случайные величины ξ и η были независимы, необходимо и достаточно, чтобы

$$F_{\xi\eta}(x, y) = F_\xi(x) F_\eta(y), \qquad (x, y) \in R^2.$$

Покажем, что в случае наличия двумерной плотности $f_{\xi\eta}(x, y)$ величины ξ и η независимы тогда и только тогда, когда

$$f_{\xi\eta}(x, y) = f_\xi(x) f_\eta(y) \tag{56}$$

(равенство понимается почти наверное относительно двумерной лебеговской меры).

В самом деле, если выполнено (56), то по теореме Фубини

$$F_{\xi\eta}(x, y) = \\ = \int\limits_{(-\infty, x] \times (-\infty, y]} f_{\xi\eta}(x, y)\,dx\,dy = \int\limits_{(-\infty, x] \times (-\infty, y]} f_\xi(x) f_\eta(y)\,dx\,dy = \\ = \int\limits_{(-\infty, x]} f_\xi(x)\,dx \left(\int\limits_{(-\infty, y]} f_\eta(y)\,dy\right) = F_\xi(x) F_\eta(y)$$

и, следовательно, ξ и η независимы.

Обратно, если они независимы и имеют плотность $f_{\xi\eta}(x, y)$, то опять-таки по теореме Фубини

$$\int\limits_{(-\infty, x] \times (-\infty, y]} f_{\xi\eta}(x, y)\,dx\,dy = \\ = \left(\int\limits_{(-\infty, x]} f_\xi(x)\,dx\right) \left(\int\limits_{(-\infty, y]} f_\eta(y)\,dy\right) = \\ = \int\limits_{(-\infty, x] \times (-\infty, y]} f_\xi(x) f_\eta(y)\,dx\,dy.$$

Отсюда следует, что для любого $B \in \mathscr{B}(R^2)$

$$\int\limits_B f_{\xi\eta}(x, y)\,dx\,dy = \int\limits_B f_\xi(x) f_\eta(y)\,dx\,dy,$$

и из свойства I легко вывести, что выполнено (56).

10. В этом пункте будет рассмотрен вопрос о соотношениях между интегралами Лебега и Римана.

Прежде всего отметим, что конструкция интеграла Лебега не зависит от того, на каком измеримом пространстве $(\Omega, \mathscr{F})$ заданы подлежащие интегрированию функции. В то же время интеграл Римана для абстрактных пространств не определяется вовсе, а для случая пространств $\Omega = R^n$ он определяется последовательным образом: сначала для R^1, а затем с соответствующими изменениями переносится на случай $n > 1$.

Подчеркнем, что в основу построения интегралов Римана и Лебега положены разные идеи. Первый шаг в конструкции Римана состоит в том, что точки $x \in R^1$ группируются по признаку их близости на оси x. В конструкции же Лебега (для $\Omega = R^1$) точки $x \in R^1$ группируются по другому признаку — по близости значений, подлежащих интегрированию функций. Следствием этих разных подходов является то, что соответствующие интегральные суммы Римана будут иметь предел лишь для не «слишком» разрывных функций, в то время как лебеговские интегральные суммы будут сходиться к предельным значениям для более широкого класса функций.

Напомним определение интеграла Римана — Стилтьеса. Пусть $G = G(x)$ — некоторая обобщенная функция распределения на R (см. п. 2 § 3), μ — соответствующая ей мера Лебега — Стилтьеса, и пусть $g = g(x)$ — ограниченная функция, обращающаяся в нуль вне отрезка $[a, b]$.

Рассмотрим разбиение $\mathscr{P} = \{x_0, \ldots, x_n\}$,

$$a = x_0 < x_1 < \ldots < x_n = b,$$

отрезка $[a, b]$ и составим верхние и нижние суммы

$$\overline{\sum}_{\mathscr{P}} = \sum_{i=1}^{n} \bar{g}_i [G(x_{i+1}) - G(x_i)], \quad \underline{\sum}_{\mathscr{P}} = \sum_{i=1}^{n} \underline{g}_i [G(x_{i+1}) - G(x_i)],$$

где

$$\bar{g}_i = \sup_{x_{i-1} < y \leqslant x_i} g(y), \quad \underline{g}_i = \inf_{x_{i-1} < y \leqslant x_i} g(y).$$

Определим простые функции $\bar{g}_{\mathscr{P}}(x)$ и $\underline{g}_{\mathscr{P}}(x)$, полагая на $x_{i-1} < x \leqslant x_i$

$$\bar{g}_{\mathscr{P}}(x) = \bar{g}_i, \quad \underline{g}_{\mathscr{P}}(x) = \underline{g}_i,$$

и определяя $\bar{g}_{\mathscr{P}}(a) = \underline{g}_{\mathscr{P}}(a) = g(a)$. Ясно, что тогда

$$\overline{\sum}_{\mathscr{P}} = (L\text{-}S) \int_a^b \bar{g}_{\mathscr{P}}(x)\, G(dx)$$

и

$$\underline{\sum}_{\mathscr{T}} = (L\text{-}S)\int_a^b \underline{g}_{\mathscr{T}}(x)\,G(dx).$$

Пусть теперь $\{\mathscr{T}_k\}$ — последовательность разбиений таких, что $\mathscr{T}_k \subseteq \mathscr{T}_{k+1}$. Тогда

$$\bar{g}_{\mathscr{T}_1} \geqslant \bar{g}_{\mathscr{T}_2} \geqslant \ldots \geqslant g \geqslant \ldots \geqslant \underline{g}_{\mathscr{T}_2} \geqslant \underline{g}_{\mathscr{T}_1},$$

и если $|g(x)| \leqslant C$, то по теореме о мажорируемой сходимости

$$\begin{aligned} \lim_{k\to\infty} \overline{\sum}_{\mathscr{T}_k} &= (L\text{-}S)\int_a^b \bar{g}(x)\,G(dx), \\ \lim_{k\to\infty} \underline{\sum}_{\mathscr{T}_k} &= (L\text{-}S)\int_a^b \underline{g}(x)\,G(dx), \end{aligned} \tag{57}$$

где $\bar{g}(x) = \lim\limits_k \bar{g}_{\mathscr{T}_k}(x)$, $\underline{g}(x) = \lim\limits_k \underline{g}_{\mathscr{T}_k}(x)$.

Если пределы $\lim\limits_k \overline{\sum}_{\mathscr{T}_k}$ и $\lim\limits_k \underline{\sum}_{\mathscr{T}_k}$ *конечны, совпадают и их общее значение не зависит от выбора последовательности разбиений* $\{\mathscr{T}_k\}$, то говорят, что функция $g = g(x)$ *интегрируема по Риману — Стилтьесу*, а соответствующее общее значение пределов обозначается

$$(R\text{-}S)\int_a^b g(x)\,G(dx). \tag{58}$$

В том случае, когда $G(x) = x$, этот интеграл называется *интегралом Римана* и обозначается

$$(R)\int_a^b g(x)\,dx.$$

Пусть теперь $(L\text{-}S)\int_a^b g(x)\,G(dx)$ — соответствующий интеграл Лебега — Стилтьеса (см. замечание 2 в п. 2).

Теорема 9. *Если функция $g = g(x)$ непрерывна на $[a, b]$, то она интегрируема по Риману — Стилтьесу и*

$$(R\text{-}S)\int_a^b g(x)\,G(dx) = (L\text{-}S)\int_a^b g(x)\,G(dx). \tag{59}$$

Доказательство. Так как функция $g(x)$ непрерывна, то $\bar{g}(x) = g(x) = \underline{g}(x)$. Поэтому в силу (57) $\lim\limits_{k\to\infty} \overline{\sum}_{\mathscr{T}_k} = \lim\limits_{k\to\infty} \underline{\sum}_{\mathscr{T}_k}$

Таким образом, непрерывная функция $g=g(x)$ интегрируема по Риману — Стилтьесу и, более того, ее интеграл совпадает (опять-таки в силу (57)) с интегралом Лебега — Стилтьеса.

Рассмотрим несколько подробнее вопрос о соотношении между интегралами Римана и Лебега в случае лебеговской меры на прямой R.

Теорема 10. *Пусть $g=g(x)$ — ограниченная функция на $[a, b]$.*

a) *Функция $g=g(x)$ интегрируема по Риману на $[a, b]$ тогда и только тогда, когда она непрерывна почти всюду (относительно меры Лебега $\bar{\lambda}$ на $\overline{\mathscr{B}}([a, b])$).*

b) *Если $g=g(x)$ интегрируема по Риману, то она интегрируема по Лебегу и*

$$(R)\int_a^b g(x)\,dx=(L)\int_a^b g(x)\,\bar{\lambda}(dx). \tag{60}$$

Доказательство. a) Пусть функция $g=g(x)$ интегрируема по Риману. Тогда, согласно (57),

$$(L)\int_a^b \bar{g}(x)\,\bar{\lambda}(dx)=(L)\int_a^b \underline{g}(x)\,\bar{\lambda}(dx).$$

Но $\underline{g}(x)\leqslant g(x)\leqslant \bar{g}(x)$, поэтому в силу свойства **H**

$$\underline{g}(x)=g(x)=\bar{g}(x) \qquad (\bar{\lambda}\text{-п. н.}), \tag{61}$$

откуда нетрудно вывести, что функция $g(x)$ непрерывна почти всюду (относительно меры $\bar{\lambda}$).

Обратно, пусть функция $g=g(x)$ непрерывна почти всюду (относительно меры $\bar{\lambda}$). Тогда выполнено (61) и, следовательно, $g(x)$ отличается от измеримой (по Борелю) функции $\bar{g}(x)$ лишь на множестве $\mathscr{N}$ с $\bar{\lambda}(\mathscr{N})=0$. Но тогда

$$\{x\colon g(x)\leqslant c\}=\{x\colon g(x)\leqslant c\}\cap\overline{\mathscr{N}}+\{x\colon g(x)\leqslant c\}\cap\mathscr{N}=$$
$$=\{x\colon \bar{g}(x)\leqslant c\}\cap\overline{\mathscr{N}}+\{x\colon g(x)\leqslant c\}\cap\mathscr{N}.$$

Ясно, что множество $\{x\colon \bar{g}(x)\leqslant c\}\cap\overline{\mathscr{N}}\in\overline{\mathscr{B}}([a, b])$, а множество $\{x\colon g(x)\leqslant c\}\cap\mathscr{N}$ является подмножеством множества $\mathscr{N}$, имеющего лебеговскую меру $\bar{\lambda}$, равную нулю и, следовательно, также принадлежащего $\overline{\mathscr{B}}([a, b])$. Тем самым $g(x)$ $\overline{\mathscr{B}}([a, b])$-измерима и как ограниченная функция интегрируема по Лебегу. Поэтому по свойству **G**

$$(L)\int_a^b \bar{g}(x)\,\bar{\lambda}(dx)=(L)\int_a^b \underline{g}(x)\,\bar{\lambda}(dx)=(L)\int_a^b g(x)\,\bar{\lambda}(dx),$$

что и завершает доказательство утверждения a).

b) Если функция $g=g(x)$ интегрируема по Лебегу, то согласно а) она непрерывна ($\bar{\lambda}$-п. н.). Выше было показано, что тогда $g(x)$ интегрируема по Лебегу и ее интегралы Римана и Лебега совпадают.

Теорема доказана.

Замечание. Пусть μ — некоторая мера Лебега — Стилтьеса на $\mathscr{B}([a, b])$. Обозначим $\overline{\mathscr{B}}_\mu([a, b])$ систему подмножеств $\Lambda \subseteq [a, b]$, для которых найдутся множества A и B из $\mathscr{B}([a, b])$ такие, что $A \subseteq \Lambda \subseteq B$ и $\mu(B \setminus A)=0$. Пусть $\bar{\mu}$ — *продолжение* меры μ на $\overline{\mathscr{B}}_\mu([a, b])$ ($\bar{\mu}(\Lambda)=\mu(A)$ для Λ таких, что $A \subseteq \Lambda \subseteq B$ и $\mu(B \setminus A)=0$). Тогда утверждение теоремы останется в силе, если вместо лебеговской меры $\bar{\lambda}$ рассмотреть меру $\bar{\mu}$, а вместо интегралов Римана и Лебега рассмотреть соответствующие интегралы Римана — Стилтьеса и Лебега — Стилтьеса по мере $\bar{\mu}$.

11. Задачи.

1. Доказать представление (6).

2. Показать, что справедливо следующее обобщение свойства **E**. Пусть ξ и η — случайные величины, для которых определены $\mathsf{M}\xi$ и $\mathsf{M}\eta$ и выражение $\mathsf{M}\xi+\mathsf{M}\eta$ имеет смысл (не имеет вида $\infty-\infty$ или $-\infty+\infty$). Тогда

$$\mathsf{M}(\xi+\eta)=\mathsf{M}\xi+\mathsf{M}\eta.$$

3. Обобщить свойство **G**, показав, что если $\xi=\eta$ (п. н.) и $\mathsf{M}\xi$ существует, то $\mathsf{M}\eta$ также существует и $\mathsf{M}\eta=\mathsf{M}\xi$.

4. Пусть ξ — расширенная случайная величина, μ — σ-конечная мера, $\int_\Omega |\xi|\,d\mu<\infty$. Показать, что тогда $|\xi|<\infty$ (μ-п. н.) (ср. со свойством **J**).

5. Пусть μ — σ-конечная мера, ξ и η — расширенные случайные величины, для которых $\mathsf{M}\xi$ и $\mathsf{M}\eta$ определены. Тогда, если для всех $A \in \mathscr{F}$ $\int_A \xi\,d\mathsf{P} \leqslant \int_A \eta\,d\mathsf{P}$, то $\xi \leqslant \eta$ (μ-п. н.). (Ср. со свойством **I**.)

6. Пусть ξ и η — независимые неотрицательные случайные величины. Показать, что тогда $\mathsf{M}\xi\eta=\mathsf{M}\xi\cdot\mathsf{M}\eta$.

7. Используя лемму Фату, показать, что

$$\mathsf{P}(\underline{\lim}\, A_n) \leqslant \underline{\lim}\, \mathsf{P}(A_n), \quad \mathsf{P}(\overline{\lim}\, A_n) \leqslant \overline{\lim}\, \mathsf{P}(A_n).$$

8. Привести пример, показывающий, что в теореме о мажорируемой сходимости условие «$|\xi_n| \leqslant \eta$, $\mathsf{M}\eta<\infty$» не может быть, вообще говоря, ослаблено.

9. Привести пример, показывающий, что в лемме Фату условие «$\xi_n \leqslant \eta$, $\mathsf{M}\eta>-\infty$» не может быть, вообще говоря, отброшено.

10. Доказать справедливость следующих вариантов леммы Фату. Пусть семейство случайных величин $\{\xi_n^+\}_{n\geqslant 1}$ равномерно интегрируемо и $\mathsf{M}\,\overline{\lim}\,\xi_n$ существует. Тогда

$$\overline{\lim}\, \mathsf{M}\xi_n \leqslant \mathsf{M}\,\overline{\lim}\,\xi_n$$

Пусть $\xi_n \leqslant \eta_n$, $n \geqslant 1$, где семейство $\{\eta_n\}_{n \geqslant 1}$ равномерно интегрируемо и η_n сходятся п. н. (или только по вероятности — см. далее § 10) к некоторой случайной величине η. Тогда $\overline{\lim}\,\mathsf{M}\xi_n \leqslant \leqslant \mathsf{M}\,\overline{\lim}\,\xi_n$.

11. Функция Дирихле

$$d(x) = \begin{cases} 1, & x \text{ — иррациональное,} \\ 0, & x \text{ — рациональное,} \end{cases}$$

определенная на $[0, 1]$, интегрируема по Лебегу, но не интегрируема по Риману. Почему?

12. Привести пример последовательности интегрируемых по Риману функций $\{f_n\}_{n \geqslant 1}$, заданных на $[0, 1]$ и таких, что $|f_n| \leqslant 1$, $f_n \to f$ почти всюду по мере Лебега, но f не интегрируема по Риману.

13. Пусть $(a_{i,j};\ i,\ j \geqslant 1)$ — последовательность действительных чисел таких, что $\sum\limits_{i,j} |a_{i,j}| < \alpha$. Вывести из теоремы Фубини, что

$$\sum_{(i,j)} a_{ij} = \sum_i \left(\sum_j a_{ij}\right) = \sum_j \left(\sum_i a_{ij}\right). \tag{62}$$

14. Привести пример последовательности $(a_{ij};\ i,\ j \geqslant 1)$, для которой $\sum\limits_{i,j} |a_{ij}| = \infty$ и равенства в (62) несправедливы.

15. Отправляясь от простых функций и используя теоремы о предельных переходах под знаком интеграла Лебега, доказать справедливость следующего результата *об интегрировании с помощью подстановки.*

Пусть $h = h(y)$ — неубывающая непрерывно дифференцируемая функция на интервале $[a, b]$, а $f(x)$ — интегрируемая (по мере Лебега) функция на интервале $[h(a), h(b)]$. Тогда функция $f(h(y))\,h'(y)$ интегрируема на $[a, b]$ и

$$\int_{h(a)}^{h(b)} f(x)\,dx = \int_a^b f(h(y))\,h'(y)\,dy.$$

16. Пусть ξ — неотрицательная случайная величина с функцией распределения $F_\xi(x)$. Показать, что

$$\mathsf{M}\xi = \int_0^\infty [1 - F_\xi(x)]\,dx$$

и для любой константы $c \geqslant 0$

$$\mathsf{M}\min(\xi, c) = \int_0^c [1 - F_\xi(x)]\,dx.$$

17. Пусть $\xi, \xi_1, \xi_2, \ldots$ — неотрицательные интегрируемые случайные величины такие, что $\mathsf{M}\xi_n \to \mathsf{M}\xi$ и для всякого $\varepsilon > 0$ $\mathsf{P}(\xi - \xi_n > \varepsilon) \to 0$. Показать, что тогда $\mathsf{M}|\xi_n - \xi| \to 0$, $n \to \infty$.

18. Пусть ξ, η, ζ и ξ_n, η_n, ζ_n, $n \geqslant 1$, — случайные величины такие, что

$$\xi_n \xrightarrow{\mathsf{P}} \xi, \quad \eta_n \xrightarrow{\mathsf{P}} \eta, \quad \zeta_n \xrightarrow{\mathsf{P}} \zeta, \quad \eta_n \leqslant \xi_n \leqslant \zeta_n, \quad n \geqslant 1,$$
$$\mathsf{M}\zeta_n \to \mathsf{M}\zeta, \quad \mathsf{M}\eta_n \to \mathsf{M}\eta,$$

и математические ожидания $\mathsf{M}\xi$, $\mathsf{M}\eta$, $\mathsf{M}\zeta$ конечны. Показать, что тогда справедлива следующая *лемма Пратта*: $\mathsf{M}\xi_n \to \mathsf{M}\xi$.

Если к тому же $\eta_n \leqslant 0 \leqslant \zeta_n$, то $\mathsf{M}|\xi_n - \xi| \to 0$.

Вывести отсюда, что если $\xi_n \xrightarrow{\mathsf{P}} \xi$, $\mathsf{M}|\xi_n| \to \mathsf{M}|\xi|$ и $\mathsf{M}|\xi| < \infty$, то $\mathsf{M}|\xi_n - \xi| \to 0$.

§ 7. Условные вероятности и условные математические ожидания относительно σ-алгебр

1. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство, и событие $A \in \mathscr{F}$ таково, что $\mathsf{P}(A) > 0$. Как и в случае конечных вероятностных пространств, *условной вероятностью события B относительно A* (обозначение: $\mathsf{P}(B \mid A)$) будем называть величину $\frac{\mathsf{P}(BA)}{\mathsf{P}(A)}$, а условной вероятностью события B относительно конечного или счетного разбиения $\mathscr{D} = \{D_1, D_2, \ldots\}$ с $\mathsf{P}(D_i) > 0$, $i \geqslant 1$ (обозначение: $\mathsf{P}(B \mid \mathscr{D})$) назовем случайную величину, равную $\mathsf{P}(B \mid D_i)$ для $\omega \in D_i$, $i \geqslant 1$.

Аналогичным образом, если ξ — случайная величина, для которой определено $\mathsf{M}\xi$, то *условным математическим ожиданием ξ относительно события A* с $\mathsf{P}(A) > 0$ (обозначение: $\mathsf{M}(\xi \mid A)$) будем называть величину $\frac{\mathsf{M}(\xi I_A)}{\mathsf{P}(A)}$ (ср. с (I.8.10)).

Случайная величина $\mathsf{P}(B \mid \mathscr{D})$ является, очевидно, измеримой относительно σ-алгебры $\mathscr{G} = \sigma(\mathscr{D})$, в связи с чем ее обозначают также $\mathsf{P}(B \mid \mathscr{G})$ (см. § 8 гл. I).

В теории вероятностей приходится, однако, сталкиваться с необходимостью рассмотрения условных вероятностей относительно событий, имеющих *нулевую* вероятность.

Рассмотрим, например, следующий эксперимент. Пусть ξ — случайная величина, равномерно распределенная на $[0, 1]$. Если $\xi = x$, то подбрасывается монета, у которой вероятность появления «герба» равна x, а «решетки» — $(1-x)$. Пусть ν — число появлений «герба» при n независимых подбрасываниях такой монеты. Спрашивается, чему равна «условная вероятность $P(\nu = k \mid \xi = x)$»? Поскольку $\mathsf{P}(\xi = x) = 0$, то интересующая нас «условная вероятность $\mathsf{P}(\nu = k \mid \xi = x)$» пока не определена, хотя интуитивно понятно, что эта вероятность должна быть равна $C_n^k x^k (1-x)^{n-k}$, Дадим теперь общее определение условного математического ожидания (и, в частности, условной вероятности) относительно σ-алгебр $\mathscr{G}$, $\mathscr{G} \subseteq \mathscr{F}$ и сравним его с определением, данным в § 8 гл. I для случая конечных вероятностных пространств.

2. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство, $\mathscr{G}$ — некоторая σ-алгебра, $\mathscr{G} \subseteq \mathscr{F}$ ($\mathscr{G}$ — *σ-подалгебра* $\mathscr{F}$) и $\xi = \xi(\omega)$ — случайная величина. Напомним, что, согласно § 6, математическое ожидание $\mathsf{M}\xi$ определялось в два этапа: сначала для неотрицательных случайных величин ξ, а затем в общем случае с помощью равенства

$$\mathsf{M}\xi = \mathsf{M}\xi^{+} - \mathsf{M}\xi^{-}$$

и только в предположении, что

$$\min(\mathsf{M}\xi^{-}, \ \mathsf{M}\xi^{+}) < \infty.$$

Подобная двухэтапная конструкция применяется и при определении условных математических ожиданий $\mathsf{M}(\xi \mid \mathscr{G})$.

Определение 1. 1) *Условным математическим ожиданием неотрицательной* случайной величины ξ *относительно σ-алгебры* $\mathscr{G}$ называется неотрицательная расширенная случайная величина, обозначаемая $\mathsf{M}(\xi \mid \mathscr{G})$ или $\mathsf{M}(\xi \mid \mathscr{G})(\omega)$, такая, что

a) $\mathsf{M}(\xi \mid \mathscr{G})$ является $\mathscr{G}$-измеримой;

b) для любого $A \in \mathscr{G}$

$$\int_A \xi \, dP = \int_A \mathsf{M}(\xi \mid \mathscr{G}) \, d\mathsf{P}. \tag{1}$$

2) *Условное математическое ожидание* $\mathsf{M}(\xi \mid \mathscr{G})$, или $\mathsf{M}(\xi \mid \mathscr{G})(\omega)$, *произвольной* случайной величины ξ *относительно σ-алгебры* $\mathscr{G}$ считается определенным, если P-п. н.

$$\min(\mathsf{M}(\xi^{+} \mid \mathscr{G}), \ \mathsf{M}(\xi^{-} \mid \mathscr{G})) < \infty,$$

и задается формулой

$$\mathsf{M}(\xi \mid \mathscr{G}) \equiv \mathsf{M}(\xi^{+} \mid \mathscr{G}) - \mathsf{M}(\xi^{-} \mid \mathscr{G}),$$

причем на множестве (нулевой вероятности) тех элементарных событий, для которых $\mathsf{M}(\xi^{+} \mid \mathscr{G}) = \mathsf{M}(\xi^{-} \mid \mathscr{G}) = \infty$, разность $\mathsf{M}(\xi^{+} \mid \mathscr{G}) - \mathsf{M}(\xi^{-} \mid \mathscr{G})$ определяется произвольно, например полагается равной нулю.

Прежде всего покажем, что для неотрицательных случайных величин $\mathsf{M}(\xi \mid \mathscr{G})$ действительно существует. Согласно (6.36) функция множеств

$$\mathsf{Q}(A) = \int_A \xi \, d\mathsf{P}, \qquad A \in \mathscr{G}, \tag{2}$$

является мерой на $(\Omega, \mathscr{G})$, которая абсолютно непрерывна относительно меры P (рассматриваемой на $(\Omega, \mathscr{G})$, $\mathscr{G} \subseteq \mathscr{F}$). Поэтому (по теореме Радона — Никодима) существует такая неотрицательная

$\mathcal{G}$-измеримая расширенная случайная величина $\mathsf{M}(\xi\,|\,\mathcal{G})$, что

$$Q(A)=\int_A \mathsf{M}(\xi\,|\,\mathcal{G})\,d\mathsf{P}. \tag{3}$$

Из (2) и (3) следует соотношение (1).

Замечание 1. В соответствии с теоремой Радона — Никодима условное математическое ожидание $\mathsf{M}(\xi\,|\,\mathcal{G})$ определяется однозначно лишь с точностью до множеств P-меры нуль. Иначе говоря, в качестве $\mathsf{M}(\xi\,|\,\mathcal{G})$ можно взять любую $\mathcal{G}$-измеримую функцию $f(\omega)$, называемую вариантом условного математического ожидания, для которой $Q(A)=\int_A f(\omega)d\mathsf{P}$, $A\in\mathcal{G}$.

Отметим также, что в соответствии с замечанием к теореме Радона-Никодима

$$\mathsf{M}(\xi\,|\,\mathcal{G})\equiv\frac{dQ}{d\mathsf{P}}(\omega), \tag{4}$$

т. е. условное математическое ожидание есть не что иное, как производная Радона-Никодима меры Q относительно меры P (рассматриваемых на $(\Omega,\mathcal{G})$).

Замечание 2. В связи с соотношением (1) заметим, что мы не можем, вообще говоря, положить $\mathsf{M}(\xi\,|\,\mathcal{G})=\xi$, поскольку случайная величина ξ *не обязана быть* $\mathcal{G}$-измеримой.

Замечание 3. Предположим, что случайная величина ξ такова, что для нее существует $\mathsf{M}\xi$. Тогда $\mathsf{M}(\xi\,|\,\mathcal{G})$ можно было бы определить как такую $\mathcal{G}$-измеримую функцию, для которой справедливо (1). Обычно именно так и поступают. Приводимое нами определение $\mathsf{M}(\xi\,|\,\mathcal{G})\equiv\mathsf{M}(\xi^+\,|\,\mathcal{G})-\mathsf{M}(\xi^-\,|\,\mathcal{G})$ обладает тем преимуществом, что в случае тривиальной σ-алгебры $\mathcal{G}=\{\varnothing,\Omega\}$ оно превращается в определение $\mathsf{M}\xi$ и при этом оно не предполагает существования $\mathsf{M}\xi$. (Например, если ξ — случайная величина с $\mathsf{M}\xi^+=\infty$, $\mathsf{M}\xi^-=\infty$, а $\mathcal{G}=\mathcal{F}$, то $\mathsf{M}\xi$ не определено, но в смысле определения 1 $\mathsf{M}(\xi\,|\,\mathcal{G})$ существует и есть просто $\xi=\xi^+-\xi^-$).

Определение 2. Пусть $B\in\mathcal{F}$. Условное математическое ожидание $\mathsf{M}(I_B\,|\,\mathcal{G})$ обозначается $\mathsf{P}(B\,|\,\mathcal{G})$, или $\mathsf{P}(B\,|\,\mathcal{G})(\omega)$, и называется *условной вероятностью события B относительно σ-алгебры $\mathcal{G}$, $\mathcal{G}\subseteq\mathcal{F}$*.

Из определений 1 и 2 следует, что для каждого фиксированного $B\in\mathcal{F}$ условная вероятность $\mathsf{P}(B\,|\,\mathcal{G})$ есть такая случайная величина, что:

а) $\mathsf{P}(B\,|\,\mathcal{G})$ является $\mathcal{G}$-измеримой,

в) для любого $A\in\mathcal{G}$

$$\mathsf{P}(A\cap B)=\int_A \mathsf{P}(B\,|\,\mathcal{G})\,d\mathsf{P}. \tag{5}$$

Определение 3. Пусть ξ — случайная величина и $\mathscr{G}_\eta$ — σ-алгебра, порожденная некоторым случайным элементом η. Тогда $\mathsf{M}(\xi \mid \mathscr{G}_\eta)$, если оно определено, обозначается $\mathsf{M}(\xi \mid \eta)$ или $\mathsf{M}(\xi \mid \eta)(\omega)$ и называется *условным математическим ожиданием* ξ *относительно* η.

Условная вероятность $\mathsf{P}(B \mid \mathscr{G}_\eta)$ обозначается $\mathsf{P}(B \mid \eta)$ или $\mathsf{P}(B \mid \eta)(\omega)$, и называется *условной вероятностью события* B *относительно* η.

3. Покажем, что данное здесь определение $\mathsf{M}(\xi \mid \mathscr{G})$ согласуется с определением условного математического ожидания § 8 гл. 1.

Пусть $\mathscr{D} = \{D_1, D_2, \dots\}$ — некоторое конечное или счетное разбиение с атомами D_i относительно вероятности P (т. е. $\mathsf{P}(D_i) > 0$, и если $A \subseteq D_i$, то или $\mathsf{P}(A) = 0$, или $\mathsf{P}(D_i \mid A) = 0$).

Теорема 1. *Если* $\mathscr{G} = \sigma(\mathscr{D})$ *и* ξ — *случайная величина, для которой* $\mathsf{M}\xi$ *определено, то*

$$\mathsf{M}(\xi \mid \mathscr{G}) = \frac{\mathsf{M}(\xi I_{D_i})}{\mathsf{P}(D_i)} \quad (\mathsf{P}\text{-п. н. на } D_i) \tag{6}$$

или, что то же,

$$\mathsf{M}(\xi \mid \mathscr{G}) = \frac{1}{\mathsf{P}(D_i)} \int_{D_i} \xi \, d\mathsf{P} \quad (\mathsf{P}\text{-п. н. на } D_i).$$

(Запись «$\xi = \eta$ (P-п. н. на A)», или «$\xi = \eta$ (A; P-п. н.)» означает, что $\mathsf{P}(A \cap \{\xi \neq \eta\}) = 0$.

Доказательство. Согласно лемме 3 из § 4 на D_i $\mathsf{M}(\xi \mid \mathscr{G}) = K_i$, где K_i — постоянная. Но

$$\int_{D_i} \xi \, d\mathsf{P} = \int_{D_i} \mathsf{M}(\xi \mid \mathscr{G}) d\mathsf{P} = K_i \mathsf{P}(D_i),$$

откуда

$$K_i = \frac{1}{\mathsf{P}(D_i)} \int_{D_i} \xi \, d\mathsf{P}.$$

Теорема доказана.

Таким образом, введенное в гл. I понятие условного математического ожидания $\mathsf{M}(\xi \mid \mathscr{D})$ относительно конечного разбиения $\mathscr{D} = \{D_1, \dots, D_n\}$ является частным случаем понятия условного математического ожидания относительно σ-алгебры $\mathscr{G} = \sigma(\mathscr{D})$.

4. **Свойства условных математических ожиданий.** (Будем предполагать, что для всех рассматриваемых случайных величин математические ожидания определены и σ-алгебра $\mathscr{G} \subseteq \mathscr{F}$).

A*. *Если* C — *постоянная и* $\xi = C$ (п. н.), *то* $\mathsf{M}(\xi \mid \mathscr{G}) = C$ (п. н.).

B*. *Если* $\xi \leqslant \eta$ (п. н), *то* $\mathsf{M}(\xi \mid \mathscr{G}) \leqslant \mathsf{M}(\eta \mid \mathscr{G})$ (п. н.).

C*. $|\mathsf{M}(\xi \mid \mathscr{G})| \leqslant \mathsf{M}(|\xi| \mid \mathscr{G})$ (п. н.).

D*. *Если* a, b — *постоянные и* $a\mathsf{M}\xi + b\mathsf{M}\eta$ *определено, то*

$$\mathsf{M}(a\xi + b\eta \mid \mathscr{G}) = a\mathsf{M}(\xi \mid \mathscr{G}) + b\mathsf{M}(\eta \mid \mathscr{G}) \quad (\text{п. н.}).$$

E*. *Пусть* $\mathscr{F}_* = \{\varnothing, \Omega\}$ — *тривиальная σ-алгебра. Тогда*

$$\mathsf{M}(\xi \mid \mathscr{F}_*) = \mathsf{M}\xi \quad (\text{п. н.}).$$

F*. $\mathsf{M}(\xi \mid \mathscr{F}) = \xi$ (п. н.).

G*. $\mathsf{M}(\mathsf{M}(\xi \mid \mathscr{G})) = \mathsf{M}\xi$.

H*. *Если* $\mathscr{G}_1 \subseteq \mathscr{G}_2$, *то*

$$\mathsf{M}[\mathsf{M}(\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1] = \mathsf{M}(\xi \mid \mathscr{G}_1) \quad (\text{п. н.}).$$

I*. *Если* $\mathscr{G}_1 \supseteq \mathscr{G}_2$, *то*

$$\mathsf{M}[\mathsf{M}(\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1] = \mathsf{M}(\xi \mid \mathscr{G}_2) \quad (\text{п. н.}).$$

J*. *Пусть случайная величина* ξ, *для которой* $\mathsf{M}\xi$ *определено, не зависит от σ-алгебры* $\mathscr{G}$ (*т. е. не зависит от* I_B, $B \in \mathscr{G}$). *Тогда*

$$\mathsf{M}(\xi \mid \mathscr{G}) = \mathsf{M}\xi \quad (\text{п. н.}).$$

K*. *Пусть* η — $\mathscr{G}$-*измеримая случайная величина*, $\mathsf{M}|\eta| < \infty$ *и* $\mathsf{M}|\xi\eta| < \infty$. Тогда

$$\mathsf{M}(\xi\eta \mid \mathscr{G}) = \eta \mathsf{M}(\xi \mid \mathscr{G}) \quad (\text{п. н.}).$$

Приведем доказательства этих свойств.

A*. Функция, равная постоянной, измерима относительно $\mathscr{G}$. Поэтому остается лишь проверить равенство

$$\int_A \xi \, d\mathsf{P} = \int_A C \, d\mathsf{P}, \quad A \in \mathscr{G}.$$

Но в силу предположения $\xi = C$ (п. н.) и свойства **G** из § 6 это равенство выполнено очевидным образом.

B*. Если $\xi \leqslant \eta$ (п. н.), то по свойству **B** из § 6

$$\int_A \xi \, d\mathsf{P} \leqslant \int_A \eta \, d\mathsf{P}, \quad A \in \mathscr{G},$$

а значит,

$$\int_A \mathsf{M}(\xi \mid \mathscr{G}) d\mathsf{P} \leqslant \int_A \mathsf{M}(\eta \mid \mathscr{G}) d\mathsf{P}, \quad A \in \mathscr{G}.$$

Тогда требуемое неравенство следует из свойства **I** (§ 6).

C*. Это свойство вытекает из предыдущего, если учесть, что $-|\xi| \leqslant \xi \leqslant |\xi|$.

D*. Если множество $A \in \mathscr{G}$, то, согласно задаче 2 из § 6,

$$\int_A (a\xi + b\eta) d\mathsf{P} = \int_A a\xi \, d\mathsf{P} + \int_A b\eta \, d\mathsf{P} = \int_A a\mathsf{M}(\xi \mid \mathscr{G}) \, d\mathsf{P} + \\ + \int_A b\mathsf{M}(\eta \mid \mathscr{G}) \, d\mathsf{P} = \int_A [a\mathsf{M}(\xi \mid \mathscr{G}) + b\mathsf{M}(\eta \mid \mathscr{G})] \, d\mathsf{P},$$

что и доказывает свойство **D***.

E*. Это свойство следует из замечания, что $\mathsf{M}\xi$ является $\mathscr{F}_*$ — измеримой функцией и того факта, что если $A=\Omega$ или $A=\varnothing$, то очевидным образом

$$\int_A \xi\, d\mathsf{P} = \int_A \mathsf{M}\xi\, d\mathsf{P}.$$

F*. Поскольку ξ — $\mathscr{F}$-измерима и

$$\int_A \xi\, d\mathsf{P} = \int_A \xi\, d\mathsf{P},\quad A\in\mathscr{F},$$

то $\mathsf{M}(\xi\,|\,\mathscr{F})=\xi$ (п. н.).

G*. Это свойство вытекает из **E*** и **H***, если взять $\mathscr{G}_1=\{\varnothing,\ \Omega\}$ и $\mathscr{G}_2=\mathscr{G}$.

H*. Пусть $A\in\mathscr{G}_1$; тогда

$$\int_A \mathsf{M}(\xi\,|\,\mathscr{G}_1)\, d\mathsf{P} = \int_A \xi\, d\mathsf{P}.$$

Так как $\mathscr{G}_1\subseteq\mathscr{G}_2$, то $A\in\mathscr{G}_2$ и, значит,

$$\int_A \mathsf{M}[\mathsf{M}(\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]\, d\mathsf{P} = \int_A \mathsf{M}(\xi\,|\,\mathscr{G}_2)\, d\mathsf{P} = \int_A \xi\, d\mathsf{P}.$$

Следовательно, для $A\in\mathscr{G}_1$

$$\int_A \mathsf{M}(\xi\,|\,\mathscr{G}_1)\, d\mathsf{P} = \int_A \mathsf{M}[\mathsf{M}(\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]\, d\mathsf{P}$$

и по свойству **I** (§ 6) и задаче 5 (§ 6)

$$\mathsf{M}(\xi\,|\,\mathscr{G}_1) = \mathsf{M}[\mathsf{M}(\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]\ \text{(п. н.)}.$$

I*. Если $A\in\mathscr{G}_1$, то по определению $\mathsf{M}[\mathsf{M}(\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]$

$$\int_A \mathsf{M}[\mathsf{M}(\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]\, d\mathsf{P} = \int_A \mathsf{M}(\xi\,|\,\mathscr{G}_2)\, d\mathsf{P}.$$

Функция $\mathsf{M}(\xi\,|\,\mathscr{G}_2)$ является $\mathscr{G}_2$-измеримой и, поскольку $\mathscr{G}_2\subseteq\mathscr{G}_1$, то и $\mathscr{G}_1$-измеримой. Отсюда следует, что $\mathsf{M}(\xi\,|\,\mathscr{G}_2)$ есть один из вариантов условного математического ожидания $\mathsf{M}[\mathsf{M}(\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]$, что и доказывает свойство **I***.

J*. Поскольку $\mathsf{M}\xi$ является $\mathscr{G}$-измеримой функцией, то остается проверить, что для любого $B\in\mathscr{G}$

$$\int_A \xi\, d\mathsf{P} = \int_A \mathsf{M}\xi\, d\mathsf{P},$$

т. е. что $\mathsf{M}[\xi\cdot I_B]=\mathsf{M}\xi\cdot\mathsf{M}I_B$. Если $\mathsf{M}|\xi|<\infty$, то это сразу следует из теоремы 6 § 6. Общий случай сводится к этому с применением результата задачи 6 из § 6.

Доказательство свойства **K***, опирающееся на утверждение а) следующей далее теоремы 2, будет дано несколько позднее.

Теорема 2 (о сходимости под знаком условных математических ожиданий). *Пусть* $\{\xi_n\}_{n\geqslant 1}$ — *последовательность расширенных случайных величин.*

а) *Если* $|\xi_n|\leqslant\eta$, $\mathsf{M}\eta<\infty$ *и* $\xi_n\to\xi$ (п. н.), *то*

$$\mathsf{M}(\xi_n\,|\,\mathscr{G})\to\mathsf{M}(\xi\,|\,\mathscr{G})\quad(\text{п. н.})$$

и

$$\mathsf{M}(|\xi_n-\xi|\,|\,\mathscr{G})\to 0\quad(\text{п. н.})$$

b) *Если* $\xi_n\geqslant\eta$, $\mathsf{M}\eta>-\infty$ *и* $\xi_n\uparrow\xi$ (п. н.) *то*

$$\mathsf{M}(\xi_n\,|\,\mathscr{G})\uparrow\mathsf{M}(\xi\,|\,\mathscr{G})\quad(\text{п. н.}).$$

c) *Если* $\xi_n\leqslant\eta$, $\mathsf{M}\eta<\infty$ *и* $\xi_n\downarrow\xi$ (п. н.), то

$$\mathsf{M}(\xi_n\,|\,\mathscr{G})\downarrow\mathsf{M}(\xi\,|\,\mathscr{G})\quad(\text{п. н.})$$

d) *Если* $\xi_n\geqslant\eta$, $\mathsf{M}\eta>-\infty$, *то*

$$\mathsf{M}(\underline{\lim}\,\xi_n\,|\,\mathscr{G})\leqslant\underline{\lim}\,\mathsf{M}(\xi_n\,|\,\mathscr{G})\quad(\text{п. н.}).$$

e) *Если* $\xi_n\leqslant\eta$, $\mathsf{M}\eta<\infty$, *то*

$$\overline{\lim}\,\mathsf{M}(\xi_n\,|\,\mathscr{G})\leqslant\mathsf{M}(\overline{\lim}\,\xi_n\,|\,\mathscr{G})\quad(\text{п. н.}).$$

f) *Если* $\xi_n\geqslant 0$, *то*

$$\mathsf{M}\left(\sum\xi_n\,|\,\mathscr{G}\right)=\sum\mathsf{M}(\xi_n\,|\,\mathscr{G})\quad(\text{п. н.}).$$

Доказательство. а) Пусть $\zeta_n=\sup\limits_{m\geqslant n}|\xi_m-\xi|$. Поскольку $\xi_n\to\xi$ (п. н.), то $\zeta_n\downarrow 0$ (п. н.). Математические ожидания $\mathsf{M}\xi_n$ и $\mathsf{M}\xi$ конечны, поэтому в силу свойств **D***- и **C*** (п. н.)

$$|\mathsf{M}(\xi_n\,|\,\mathscr{G})-\mathsf{M}(\xi\,|\,\mathscr{G})|=|\mathsf{M}(\xi_n-\xi\,|\,\mathscr{G})|\leqslant$$
$$\leqslant\mathsf{M}(|\xi_n-\xi|\,|\,\mathscr{G})\leqslant\mathsf{M}(\zeta_n\,|\,\mathscr{G}).$$

Поскольку $\mathsf{M}(\zeta_{n+1}\,|\,\mathscr{G})\leqslant\mathsf{M}(\zeta_n\,|\,\mathscr{G})$ (п. н.), то (п. н.) существует предел $h=\lim\limits_n\mathsf{M}(\zeta_n\,|\,\mathscr{G})$. Тогда

$$0\leqslant\int\limits_\Omega h\,d\mathsf{P}\leqslant\int\limits_\Omega\mathsf{M}(\zeta_n\,|\,\mathscr{G})\,d\mathsf{P}=\int\limits_\Omega\zeta_n\,d\mathsf{P}\to 0,\quad n\to\infty,$$

где последнее утверждение следует из теоремы о мажорируемой сходимости, поскольку $0\leqslant\zeta_n\leqslant 2\eta$, $\mathsf{M}\eta<\infty$. Следовательно, $\int\limits_\Omega h\,d\mathsf{P}=0$ и по свойству **H** $h=0$ (п. н.).

b) Пусть сначала $\eta \equiv 0$. Поскольку $\mathsf{M}(\xi_n \mid \mathscr{G}) \leqslant \mathsf{M}(\xi_{n+1} \mid \mathscr{G})$ (п. н.), то существует (п. н.) предел $\zeta(\omega) = \lim\limits_n \mathsf{M}(\xi_n \mid \mathscr{G})$. Тогда из равенства

$$\int\limits_\Omega \xi_n \, d\mathsf{P} = \int\limits_\Omega \mathsf{M}(\xi_n \mid \mathscr{G}) \, d\mathsf{P}, \qquad A \in \mathscr{G},$$

и теоремы о монотонной сходимости

$$\int\limits_A \xi \, d\mathsf{P} = \int\limits_A \zeta \, d\mathsf{P}, \qquad A \in \mathscr{G}.$$

Следовательно, по свойству I и задаче 5 § 6 $\xi = \zeta$ (п. н.)

Для доказательства в общем случае заметим, что $0 \leqslant \xi_n^+ \uparrow \xi^+$, и по доказанному

$$\mathsf{M}(\xi_n^+ \mid \mathscr{G}) \uparrow \mathsf{M}(\xi^+ \mid \mathscr{G}) \text{ (п. н.).} \tag{7}$$

Но $0 \leqslant \xi_n^- \leqslant \xi^-$, $\mathsf{M}\xi^- < \infty$, поэтому в силу а)

$$\mathsf{M}(\xi_n^- \mid \mathscr{G}) \to \mathsf{M}(\xi^- \mid \mathscr{G}),$$

что вместе с (7) доказывает b).

Утверждение c) вытекает из b).

d) Пусть $\zeta_n = \inf\limits_{m \geqslant n} \xi_m$, тогда $\zeta_n \uparrow \zeta$, где $\zeta = \varliminf \xi_n$. Согласно b) $\mathsf{M}(\zeta_n \mid \mathscr{G}) \uparrow \mathsf{M}(\zeta \mid \mathscr{G})$ (п. н.). Поэтому (п. н.) $\mathsf{M}(\varliminf \xi_n \mid \mathscr{G}) = \mathsf{M}(\zeta \mid \mathscr{G}) = \lim\limits_n \mathsf{M}(\zeta_n \mid \mathscr{G}) = \varliminf \mathsf{M}(\zeta_n \mid \mathscr{G}) \leqslant \varliminf \mathsf{M}(\xi_n \mid \mathscr{G})$.

Утверждение e) вытекает из d).

f) Если $\xi_n \geqslant 0$, то по свойству **D***

$$\mathsf{M}\left(\sum_{k=1}^n \xi_k \,\Big|\, \mathscr{G}\right) = \sum_{k=1}^n \mathsf{M}(\xi_k \mid \mathscr{G}) \text{ (п. н.),}$$

что вместе с b) и доказывает требуемый результат.

Теорема доказана.

Приведем теперь доказательство свойства **K***. Пусть $\eta = I_B$, $B \in \mathscr{G}$. Тогда для всякого $A \in \mathscr{G}$

$$\int\limits_A \xi\eta \, d\mathsf{P} = \int\limits_{A \cap B} \xi \, d\mathsf{P} = \int\limits_{A \cap B} \mathsf{M}(\xi \mid \mathscr{G}) \, d\mathsf{P} = \int\limits_A I_B \mathsf{M}(\xi \mid \mathscr{G}) \, d\mathsf{P} = \int\limits_A \eta \mathsf{M}(\xi \mid \mathscr{G}) \, d\mathsf{P}.$$

В силу аддитивности интеграла Лебега равенство

$$\int\limits_A \xi\eta \, d\mathsf{P} = \int\limits_A \eta \mathsf{M}(\xi \mid \mathscr{G}) \, d\mathsf{P}, \qquad A \in \mathscr{G} \tag{8}$$

останется справедливым и для простых случайных величин $\eta = \sum_{k=1}^{n} y_k I_{B_k}$, $B_k \in \mathscr{G}$. Поэтому по свойству **I** (§ 6) для таких случайных величин

$$\mathsf{M}(\xi\eta \mid \mathscr{G}) = \eta \mathsf{M}(\xi \mid \mathscr{G}) \text{ (п.н.).} \tag{9}$$

Пусть теперь η — произвольная $\mathscr{G}$-измеримая случайная величина с $\mathsf{M}|\eta| < \infty$ и $\{\eta_n\}_{n \geqslant 1}$ — последовательность простых $\mathscr{G}$-измеримых случайных величин таких, что $|\eta_n| \leqslant \eta$ и $\eta_n \to \eta$. Тогда в силу (9)

$$\mathsf{M}(\xi\eta_n \mid \mathscr{G}) = \eta_n \mathsf{M}(\xi \mid \mathscr{G}) \text{ (п. н.).}$$

Ясно, что $|\xi\eta_n| \leqslant |\xi\eta|$, где $\mathsf{M}|\xi\eta| < \infty$. Поэтому по свойству a) $\mathsf{M}(\xi\eta_n \mid \mathscr{G}) \to \mathsf{M}(\xi\eta \mid \mathscr{G})$ (п. н.). Далее, так как $\mathsf{M}|\xi| < \infty$, то $\mathsf{M}(\xi \mid \mathscr{G})$ конечно (п. н.) (см. свойство **C*** и свойство **J** (§ 6)). Поэтому $\eta_n \mathsf{M}(\xi \mid \mathscr{G}) \to \eta \mathsf{M}(\xi \mid \mathscr{G})$ (п. н.) (Предположение о конечности почти наверное $\mathsf{M}(\xi \mid \mathscr{G})$ существенно, поскольку, согласно сноске на стр. 190, $0 \cdot \infty = 0$, но если $\eta_n = 1/n$, $\eta \equiv 0$, то $\frac{1}{n} \cdot \infty \not\to 0 \cdot \infty = 0$.)

5. Рассмотрим подробнее структуру условных математических ожиданий $\mathsf{M}(\xi \mid \mathscr{G}_\eta)$, обозначаемых, как было условлено выше, также через $\mathsf{M}(\xi \mid \eta)$.

Поскольку $\mathsf{M}(\xi \mid \eta)$ является $\mathscr{G}_\eta$-измеримой функцией, то, согласно теореме 3 из § 4 (точнее — очевидной ее модификации для расширенных случайных величин), найдется такая борелевская функция $m = m(y)$, определенная на $\bar{R}$ и со значениями в $\bar{R}$, что для всех $\omega \in \Omega$

$$m(\eta(\omega)) = \mathsf{M}(\xi \mid \eta)(\omega). \tag{10}$$

Эту функцию $m(y)$ будем обозначать через $\mathsf{M}(\xi \mid \eta = y)$ и называть *условным математическим ожиданием* ξ *относительно события* $\{\eta = y\}$, или *условным математическим ожиданием* ξ *при условии, что* $\eta = y$.

В соответствии с определением

$$\int_A \xi \, d\mathsf{P} = \int_A \mathsf{M}(\xi \mid \eta) \, d\mathsf{P} = \int_A m(\eta) \, d\mathsf{P}, \qquad A \in \mathscr{G}_\eta. \tag{11}$$

Поэтому по теореме 7 § 6 (о замене переменных под знаком интеграла Лебега)

$$\int_{\{\omega:\ \eta \in B\}} m(\eta) \, d\mathsf{P} = \int_B m(y) \, P_\eta(dy), \qquad B \in \mathscr{B}(\bar{R}), \tag{12}$$

где P_η — распределение вероятностей η. Следовательно, $m = m(y)$ есть борелевская функция такая, что для всякого $B \in \mathscr{B}(R)$

$$\int_{\{\omega:\ \eta \in B\}} \xi \, d\mathsf{P} = \int_B m(y) \, dP_\eta. \tag{13}$$

Это замечание подсказывает, что к определению условного математического ожидания $\mathsf{M}(\xi \mid \eta = y)$ можно прийти и иначе.

Определение 3. Пусть ξ и η — случайные величины (быть может, и расширенные) и $\mathsf{M}\xi$ определено. *Условным математическим ожиданием случайной величины ξ при условии, что $\eta = y$,* назовем всякую $\mathscr{B}(\bar{R})$-измеримую функцию $m = m(y)$, для которой

$$\int_{\{\omega:\, \eta \in B\}} \xi \, d\mathsf{P} = \int_B m(y) \, P_\eta(dy), \qquad B \in \mathscr{B}(\bar{R}). \tag{14}$$

Тот факт, что такая функция существует, следует из той же теоремы Радона — Никодима, если заметить, что функция множеств

$$Q(B) = \int_{\{\omega:\, \eta \in B\}} \xi \, d\mathsf{P}$$

является мерой со знаком, которая абсолютно непрерывна относительно меры P_η.

Предположим теперь, что $m(y)$ есть условное математическое ожидание в смысле определения 3. Тогда, применяя снова теорему о замене переменных под знаком интеграла Лебега, находим, что

$$\int_{\{\omega:\, \eta \in B\}} \xi \, d\mathsf{P} = \int_B m(y) \, P_\eta(dy) = \int_{\{\omega:\, \eta \in B\}} m(\eta) \, P_\eta(dy), \qquad B \in \mathscr{B}(\bar{R}).$$

Функция $m(\eta)$ является $\mathscr{F}_\eta$-измеримой, и множествами $\{\omega: \eta \in B\}$, $B \in \mathscr{B}(\bar{R})$, исчерпываются все множества из $\mathscr{F}_\eta$.

Отсюда вытекает, что $m(\eta)$ есть математическое ожидание $\mathsf{M}(\xi \mid \eta)$. Тем самым, зная $\mathsf{M}(\xi \mid \eta = y)$, можно восстановить $\mathsf{M}(\xi \mid \eta)$ и, наоборот, по $\mathsf{M}(\xi \mid \eta)$ найти $\mathsf{M}(\xi \mid \eta = y)$.

С интуитивной точки зрения условное математическое ожидание $\mathsf{M}(\xi \mid \eta = y)$ является более простым и понятным объектом, нежели $\mathsf{M}(\xi \mid \eta)$. Однако математическое ожидание $\mathsf{M}(\xi \mid \eta)$, рассматриваемое как $\mathscr{F}_\eta$-измеримая случайная величина, более удобно в работе.

Отметим, что приведенные выше свойства **A*** — **K*** и утверждения теоремы 2 легко переносятся на условные математические ожидания $\mathsf{M}(\xi \mid \eta = y)$ (с заменой «почти наверное» на «P_η-почти наверное»). Так, например, свойство **K*** переформулируется следующим образом: если $\mathsf{M}|\xi| < \infty$, $\mathsf{M}|\xi f(\eta)| < \infty$, где $f = f(y)$ — $\mathscr{B}(\bar{R})$-измеримая функция, то

$$\mathsf{M}(\xi f(\eta) \mid \eta = y) = f(y) \, \mathsf{M}(\xi \mid \eta = y) \qquad (P_\eta\text{-п. н.}). \tag{15}$$

Далее (ср. со свойством **J***), если ξ и η независимы, то

$$\mathsf{M}(\xi \mid \eta = y) = \mathsf{M}\xi \qquad (P_\eta\text{-п. н.}).$$

Отметим также, что если $B \in \mathscr{B}(R^2)$ и ξ и η независимы, то

$$\mathsf{M}[I_B(\xi, \eta) \mid \eta = y] = \mathsf{M} I_B(\xi, y) \qquad (P_\eta\text{-п. н.}), \tag{16}$$

и если $\varphi=\varphi(x, y)$ — $\mathscr{B}(R^2)$-измеримая функция такая, что $\mathsf{M}|\varphi(\xi, \eta)|<\infty$, то

$$\mathsf{M}[\varphi(\xi, \eta)\,|\,\eta=y]=\mathsf{M}[\varphi(\xi, y)] \qquad (P_\eta\text{-п. н.}).$$

Для доказательства (16) заметим следующее. Если $B=B_1\times B_2$, то для справедливости (16) надо лишь проверить, что

$$\int\limits_{\{\omega:\ \eta\in A\}} I_{B_1\times B_2}(\xi, \eta)\,\mathsf{P}(d\omega)=\int\limits_{(y\in A)} \mathsf{M}I_{B_1\times B_2}(\xi, y)\,P_\eta(dy).$$

Но левая часть есть $\mathsf{P}\{\xi\in B_1, \eta\in A\cap B_2\}$, а правая — $\mathsf{P}(\xi\in B_1)\times\times\mathsf{P}(\eta\in A\cap B_2)$, равенство которых следует из независимости ξ и η. В общем случае доказательство проводится с применением теоремы 1 из § 2 о монотонных классах (ср. с соответствующим местом в доказательстве теоремы Фубини).

Определение 4. *Условной вероятностью события* $A\in\mathscr{F}$ *при условии, что* $\eta=y$ (обозначение: $\mathsf{P}(A\,|\,\eta=y)$), *будем называть* $\mathsf{M}(I_A\,|\,\eta=y)$.

Понятно, что $\mathsf{P}(A\,|\,\eta=y)$ можно было бы определить как такую $\mathscr{B}(\bar{R})$-измеримую функцию, что

$$\mathsf{P}(A\cap\{\eta\in B\})=\int\limits_B \mathsf{P}(A\,|\,\eta=y)\,\mathsf{P}_\eta(dy), \qquad B\in\mathscr{B}(\bar{R}). \tag{17}$$

6. Приведем некоторые примеры вычисления условных вероятностей и условных математических ожиданий.

Пример 1. Пусть η — дискретная случайная величина с $\mathsf{P}(\eta=y_k)>0$, $\sum\limits_{k=1}^{\infty}\mathsf{P}(\eta=y_k)=1$. Тогда

$$\mathsf{P}(A\,|\,\eta=y_k)=\frac{\mathsf{P}(A\cap\{\eta=y_k\})}{\mathsf{P}(\eta=y_k)}, \qquad k\geqslant 1.$$

Для $y\notin\{y_1, y_2, \ldots\}$ условную вероятность $\mathsf{P}(A\,|\,\eta=y)$ можно определить произвольным образом, например положить равной нулю.

Если ξ — случайная величина, для которой существует $\mathsf{M}\xi$, то

$$\mathsf{M}(\xi\,|\,\eta=y_k)=\frac{1}{\mathsf{P}(\eta=y_k)}\int\limits_{\{\omega:\ \eta=y_k\}}\xi\,d\mathsf{P}.$$

Условное математическое ожидание $\mathsf{M}(\xi\,|\,\eta=y)$ для $y\notin\{y_1, y_2, \ldots\}$ определяется произвольно (например, полагается равным нулю).

Пример 2. Пусть (ξ, η) — пара случайных величин, распределение которых обладает плотностью $f_{\xi\eta}(x, y)$:

$$\mathsf{P}\{(\xi, \eta)\in B\}=\int\limits_B f_{\xi\eta}(x, y)\,dx\,dy, \qquad B\in\mathscr{B}(R^2).$$

Пусть $f_\xi(x)$ и $f_\eta(y)$ — плотности распределения вероятностей случайных величин ξ и η (см. (6.46), (6.55), (6.56)).

Обозначим

$$f_{\xi\mid\eta}(x\mid y)=\frac{f_{\xi\eta}(x,\ y)}{f_\eta(y)}, \tag{18}$$

полагая $f_{\xi\mid\eta}(x\mid y)=0$, если $f_\eta(y)=0$.

Тогда

$$\mathsf{P}(\xi\in C\mid\eta=y)=\int\limits_C f_{\xi\mid\eta}(x\mid y)\,dx,\qquad C\in\mathscr{B}(R), \tag{19}$$

т. е. $f_{\xi\mid y}(x\mid y)$ есть плотность условного распределения вероятностей.

В самом деле, для доказательства (19) достаточно убедиться в справедливости формулы (17) для $B\in\mathscr{B}(R)$, $A=\{\xi\in C\}$. В силу (6.43), (6.45) и теоремы Фубини

$$\int\limits_B\left[\int\limits_C f_{\xi\mid\eta}(x\mid y)\,dx\right]P_\eta(dy)=\int\limits_B\left[\int\limits_C f_{\xi\mid\eta}(x\mid y)\,dx\right]f_\eta(y)\,dy=$$
$$=\int\limits_{C\times B} f_{\xi\mid\eta}(x\mid y)\,f_\eta(y)\,dx\,dy=\int\limits_{C\times B} f_{\xi\eta}(x,\ y)\,dx\,dy=$$
$$=\mathsf{P}\{(\xi,\ \eta)\in C\times B\}=\mathsf{P}\{(\xi\in C)\cap(\eta\in B)\},$$

что и доказывает (17).

Аналогичным образом устанавливается следующий результат: если $\mathsf{M}\xi$ существует, то

$$\mathsf{M}(\xi\mid\eta=y)=\int\limits_{-\infty}^{\infty} xf_{\xi\mid\eta}(x\mid y)\,dx. \tag{20}$$

Пример 3. Пусть длительность работы некоторого прибора описывается неотрицательной случайной величиной $\eta=\eta(\omega)$, функция распределения которой $F_\eta(y)$ имеет плотность $f_\eta(y)$ (естественно, что $F_\eta(y)=f_\eta(y)=0$ для $y<0$). Найдем условное математическое ожидание $\mathsf{M}(\eta-a\mid\eta\geqslant a)$, т. е. среднее время, которое прибор еще проработает в предположении, что он уже проработал время a.

Пусть $\mathsf{P}(\eta\geqslant a)>0$. Тогда, согласно определению (см. п. 1) и (6.45),

$\mathsf{M}(\eta-a\mid\eta\geqslant a)=$

Интересно отметить, что если случайная величина η экспоненциально распределена, т. е.

$$f_\eta(y)=\begin{cases}\lambda e^{-\lambda y}, & y\geqslant 0,\\ 0, & y<0,\end{cases}\tag{21}$$

то $\mathsf{M}\eta=\mathsf{M}(\eta\mid\eta\geqslant 0)=1/\lambda$ и для любого $a>0$ $\mathsf{M}(\eta-a\mid\eta\geqslant a)=1/\lambda$. Иначе говоря, в этом случае среднее время, которое прибор еще проработает, в предположении, что он уже проработал время a, не зависит от значения a и совпадает просто со средним временем $\mathsf{M}\eta$.

В предположении (21) найдем условное распределение $\mathsf{P}(\eta-a\leqslant x\mid\eta\geqslant a)$.

Имеем

$$\mathsf{P}(\eta-a\leqslant x\mid\eta\geqslant a)=\frac{\mathsf{P}(a\leqslant\eta\leqslant a+x)}{\mathsf{P}(\eta\geqslant a)}=$$
$$=\frac{F_\eta(a+x)-F_\eta(a)+\mathsf{P}(\eta=a)}{1-F_\eta(a)+\mathsf{P}(\eta=a)}=\frac{[1-e^{-\lambda(a+x)}]-[1-e^{-\lambda a}]}{1-[1-e^{-\lambda a}]}=$$
$$=\frac{e^{-\lambda a}[1-e^{-\lambda x}]}{e^{-\lambda a}}=1-e^{-\lambda x}.$$

Таким образом, условное распределение $\mathsf{P}(\eta-a\leqslant x\mid\eta\geqslant a)$ совпадает с безусловным распределением $\mathsf{P}(\eta\leqslant x)$. Это замечательное свойство экспоненциального распределения является характеристическим: не существует других распределений с плотностями, обладающими свойством $\mathsf{P}(\eta-a\leqslant x\mid\eta\geqslant a)=\mathsf{P}(\eta\leqslant x)$, $a\geqslant 0$, $0\leqslant x<\infty$.

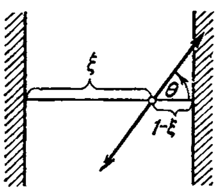


Рис. 29.

Пример 4 (игла Бюффона). Пусть на «коридор» бесконечной длины и единичной ширины (рис. 29) на плоскости «случайным» образом бросается игла единичной длины. Спрашивается, какова вероятность того, что игла пересечет (по крайней мере одну) стенку коридора?

Чтобы решить эту задачу, определим прежде всего, что означает, что игла бросается «случайным» образом. Пусть ξ — расстояние от центра иглы до левой стенки. Будем предполагать, что ξ равномерно распределено на отрезке $[0, 1]$, а (см. рис. 29) угол θ равномерно распределен на $[-\pi/2, \pi/2]$. Кроме того, будем предполагать ξ и θ независимыми.

Пусть A — событие, состоящее в том, что игла пересечет стенку коридора. Легко видеть, что если

$$B=\left\{(a, x)\colon |a|\leqslant\frac{\pi}{2},\ x\in[0, 1/2\cos a]\cup\left[1-\frac{1}{2}\cos a, 1\right]\right\},$$

то $A=\{\omega\colon (\theta, \xi)\in B\}$, и значит, интересующая нас вероятность

$$\mathsf{P}(A)=\mathsf{M}I_A(\omega)=\mathsf{M}I_B(\theta(\omega), \xi(\omega)).$$

В силу свойства G^* и формулы (16)

$$
\begin{aligned}
\mathsf{M} I_B(\theta(\omega),\ \xi(\omega)) &= \mathsf{M}(\mathsf{M}[I_B(\theta(\omega),\ \xi(\omega)) \mid \theta(\omega)]) = \\
&= \int_\Omega \mathsf{M}[I_B(\theta(\omega),\ \xi(\omega)) \mid \theta(\omega)]\, \mathsf{P}(d\omega) = \\
&= \int_{-\pi/2}^{\pi/2} \mathsf{M}[I_B(\theta(\omega),\ \xi(\omega)) \mid \theta(\omega) = a]\, P_\theta(da) = \\
&= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \mathsf{M} I_B(a,\ \xi(\omega))\, da = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos a\, da = \frac{2}{\pi},
\end{aligned}
$$

где мы воспользовались также тем, что

$$
\mathsf{M} I_B(a,\ \xi(\omega)) = \mathsf{P}\{\xi \in [0,\ 1/2\cos a] \cup [1 - 1/2\cos a]\} = \cos a.
$$

Итак, вероятность того, что «случайным» образом брошенная на коридор игла пересечет его стенки, равна $2/\pi$. Этот результат может быть положен в основу экспериментального определения значения числа π. В самом деле, пусть игла бросается независимым образом N раз. Определим ξ_i равным 1, если при i бросании игла пересекает коридор, и равным 0 в противном случае. Тогда в силу закона больших чисел (см., например, (1.5.6)) для всякого $\varepsilon > 0$

$$
\mathsf{P}\left\{\left|\frac{\xi_1 + \ldots + \xi_N}{N} - \mathsf{P}(A)\right| > \varepsilon\right\} \to 0, \qquad N \to \infty.
$$

В этом смысле частота

$$
\frac{\xi_1 + \ldots + \xi_N}{N} \approx \mathsf{P}(A) = \frac{2}{\pi}
$$

и, значит,

$$
\frac{2N}{\xi_1 + \ldots + \xi_N} \approx \pi.
$$

Именно эта формула и послужила основой для статистического определения значения числа π. В 1850 г. Р. Вольф (Цюрих) бросал иглу 5000 раз и получил для π значение 3,1596. По-видимому, этот способ явился одним из первых методов (известных теперь под названием «метода Монте — Карло») использования вероятностно-статистических закономерностей в численном анализе.

7. Если $\{\xi_n\}_{n \geqslant 1}$ — последовательность неотрицательных случайных величин, то, согласно утверждению f) теоремы 2,

$$
\mathsf{M}\left(\sum \xi_n \mid \mathscr{G}\right) = \sum \mathsf{M}(\xi_n \mid \mathscr{G}) \quad \text{(п. н.)}.
$$

В частности, если B_1, $B_2, \ldots$ — последовательность попарно непересекающихся множеств, то

$$\mathsf{P}\left(\sum B_n \,|\, \mathcal{G}\right) = \sum \mathsf{P}(B_n \,|\, \mathcal{G}) \quad (\text{п. н.}). \tag{22}$$

Важно подчеркнуть, что это равенство выполнено лишь почти наверное и, следовательно, условную вероятность $\mathsf{P}(B \,|\, \mathcal{G})(\omega)$ нельзя рассматривать при фиксированном ω как меру по B. Можно было бы подумать, что, за исключением некоторого множества $\mathcal{N}$ меры нуль, $\mathsf{P}(\cdot \,|\, \mathcal{G})(\omega)$ является все же мерой для $\omega \in \overline{\mathcal{N}}$. Однако это, вообще говоря, не так в силу следующего обстоятельства. Обозначим $\mathcal{N}(B_1, B_2, \ldots)$ то множество исходов ω, где для заданных B_1, $B_2, \ldots$ не выполнено свойство счетной аддитивности (22). Тогда исключительное множество $\mathcal{N}$ есть

$$\mathcal{N} = \bigcup \mathcal{N}(B_1, B_2, \ldots), \tag{23}$$

где объединение берется по всем непересекающимся множествам B_1, $B_2, \ldots$ из $\mathcal{F}$. Хотя P-мера каждого множества $\mathcal{N}(B_1, B_2, \ldots)$ равна нулю, P-мера множества $\mathcal{N}$ может оказаться (в силу несчетности объединения в (23)) ненулевой (Вспомним, что лебеговская мера отдельной точки равна нулю, а мера множества $\mathcal{N} = [0, 1)$, являющегося несчетной суммой одноточечных множеств $\{x\}$, $0 \leqslant x < 1$, равна единице).

В то же время было бы удобно, чтобы условная вероятность $\mathsf{P}(\cdot \,|\, \mathcal{G})(\omega)$ являлась мерой для каждого $\omega \in \Omega$, поскольку тогда, например, подсчет условных вероятностей $\mathsf{M}(\xi \,|\, \mathcal{G})$ можно было бы осуществлять (см. далее теорему 3) просто с помощью усреднения по мере $\mathsf{P}(\cdot \,|\, \mathcal{G})(\omega)$:

$$\mathsf{M}(\xi \,|\, \mathcal{G}) = \int_\Omega \xi(\omega)\, \mathsf{P}(d\omega \,|\, \mathcal{G}) \quad (\text{п. н.})$$

(ср. с (I.8.10)).

Введем такое

Определение 4. Функцию $P(\omega; B)$, определенную для всех $\omega \in \Omega$ и $B \in \mathcal{F}$, назовем *регулярной* условной вероятностью относительно $\mathcal{G}$, если:

a) для каждого $\omega \in \Omega$ $P(\omega; \cdot)$ есть вероятностная мера на $\mathcal{F}$;

b) для каждого $B \in \mathcal{F}$ $P(\omega; B)$ как функция от ω есть один из вариантов условной вероятности $\mathsf{P}(B \,|\, \mathcal{G})(\omega)$, т. е. $P(\omega; B) = \mathsf{P}(B \,|\, \mathcal{G})(\omega)$ (п. н.).

Теорема 3. *Пусть $P(\omega; B)$ — регулярная условная вероятность относительно $\mathcal{G}$ и ξ — интегрируемая случайная величина. Тогда*

$$\mathsf{M}(\xi \,|\, \mathcal{G})(\omega) = \int_\Omega \xi(\tilde{\omega})\, P(\omega; d\tilde{\omega}) \quad (\text{п. н.}). \tag{24}$$

Доказательство. Если $\xi = I_B$, $B \in \mathscr{F}$, то требуемая формула (24) превращается в равенство

$$\mathsf{P}(B \mid \mathscr{G})(\omega) = P(\omega;\ B) \quad (\text{п. н.}),$$

выполнимое в силу определения 4 b). Следовательно, (24) выполнено для простых функций.

Пусть теперь $\xi \geqslant 0$ и $\xi_n \uparrow \xi$, где ξ_r — простые функции. Тогда по свойству b) теоремы 2 $\mathsf{M}(\xi \mid \mathscr{G})(\omega) = \lim\limits_n \mathsf{M}(\xi_n \mid \mathscr{G})(\omega)$ (п. н.). Но поскольку для каждого $\omega \in \Omega$ $P(\omega;\ \cdot)$ есть мера, то по теореме о монотонной сходимости

$$\lim_n \mathsf{M}(\xi_n \mid \mathscr{G})(\omega) = \lim_n \int_\Omega \xi_n(\tilde{\omega})\, P(\omega; d\tilde{\omega}) = \lim_n \int_\Omega \xi(\tilde{\omega})\, P(\omega;\ d\tilde{\omega}).$$

Общий случай сводится к рассмотренному с помощью представления $\xi = \xi^+ - \xi^-$.

Теорема доказана.

Следствие. Пусть $\mathscr{G} = \mathscr{G}_\eta$, где η — случайная величина, причем пара (ξ, η) имеет плотность распределения вероятностей $f_{\xi\eta}(x, y)$. Пусть $\mathsf{M}\,|g(\xi)| < \infty$, тогда

$$\mathsf{M}(g(\xi) \mid \eta = y) = \int_{-\infty}^{\infty} g(x)\, f_{\xi \mid \eta}(x \mid y)\, dx,$$

где $f_{\xi \mid \eta}(x \mid y)$ — плотность условного распределения (см. (18)).

Чтобы сформулировать основной результат о существовании регулярных условных вероятностей, нам понадобятся следующие определения.

Определение 5. Пусть $(E, \mathscr{E})$ — измеримое пространство и $X = X(\omega)$ — случайный элемент со значениями в E и $\mathscr{G}$ — σ-подалгебра $\mathscr{F}$. Функция $Q(\omega;\ B)$, определенная для $\omega \in \Omega$ и $B \in \mathscr{E}$, называется *регулярным условным распределением X относительно $\mathscr{G}$*, если:

a) для каждого $\omega \in \Omega$ $Q(\omega;\ B)$ есть вероятностная мера на $(E, \mathscr{E})$;

b) для каждого $B \in \mathscr{E}$ $Q(\omega;\ B)$ как функция от ω есть один из вариантов условной вероятности $\mathsf{P}(X \in B \mid \mathscr{G})(\omega)$, т. е.

$$Q(\omega;\ B) = \mathsf{P}(X \in B \mid \mathscr{G})(\omega) \quad (\text{п. н.}).$$

Определение 6. Пусть ξ — случайная величина. Функция $F = F(\omega;\ x)$, $\omega \in \Omega$, $x \in R$, называется *регулярной функцией распределения для ξ относительно $\mathscr{G}$*, если:

a) для каждого $\omega \in \Omega$ $F(\omega;\ x)$ есть функция распределения на R;

b) для каждого $x \in R$ $F(\omega;\ x) = \mathsf{P}(\xi \leqslant x \mid \mathscr{G})(\omega)$ (п. н.).

Теорема 4. *Всегда существуют регулярная функция распределения и регулярное условное распределение случайной величины ξ относительно $\mathscr{G}$.*

Доказательство. Для каждого рационального $r \in R$ обозначим $F_r(\omega) = \mathsf{P}(\xi \leqslant r \mid \mathscr{G})(\omega)$, где $\mathsf{P}(\xi \leqslant r \mid \mathscr{G})(\omega) = \mathsf{M}(I_{\{\xi \leqslant r\}} \mid \mathscr{G})(\omega)$ — какой-нибудь вариант условной вероятности события $\{\xi \leqslant r\}$ относительно $\mathscr{G}$. Пусть $\{r_i\}$ — множество всех рациональных чисел на R. Если $r_i < r_j$, то в силу свойства B* $\mathsf{P}(\xi \leqslant r_i \mid \mathscr{G}) \leqslant \mathsf{P}(\xi \leqslant r_j \mid \mathscr{G})$ (п. н.) и, значит, если $A_{ij} = \{\omega\colon F_{r_j}(\omega) < F_{r_i}(\omega)\}$, $A = \bigcup A_{ij}$, то $\mathsf{P}(A) = 0$. Иначе говоря, множество тех ω, где у функций распределений $F_r(\omega)$, $r \in \{r_i\}$, нарушается монотонность, имеет меру нуль.

Пусть теперь $B_i = \left\{\omega\colon \lim\limits_{n\to\infty} F_{r_i + \frac{1}{n}}(\omega) \neq F_{r_i}(\omega)\right\}$, $B = \bigcup\limits_{i=1}^{\infty} B_i$. Ясно, что $I_{\left\{\xi \leqslant r_i + \frac{1}{n}\right\}} \downarrow I_{\{\xi \leqslant r_i\}}$, $n \to \infty$. Поэтому, согласно утверждению а) теоремы 2, $F_{r_i + \frac{1}{n}}(\omega) \to F_{r_i}(\omega)$ (п. н.) и, значит, множество B, где нарушается непрерывность справа (по рациональным числам), также имеет меру нуль, $\mathsf{P}(B) = 0$.

Далее, пусть $C = \left\{\omega\colon \lim\limits_{n\to\infty} F_n(\omega) \neq 1\right\} \cup \left\{\omega\colon \lim\limits_{n\to-\infty} F_n(\omega) > 0\right\}$. Тогда, поскольку $\{\xi \leqslant n\} \uparrow \Omega$, $n \to \infty$, а $\{\xi \leqslant n\} \downarrow \varnothing$, $n \to -\infty$, то $\mathsf{P}(C) = 0$.

Положим теперь

$$F(\omega;\ x) = \begin{cases} \lim\limits_{r \downarrow x} F_r(\omega), & \omega \notin A \cup B \cup C, \\ G(x), & \omega \in A \cup B \cup C, \end{cases}$$

где $G(x)$ — произвольная функция распределения на R, и покажем, что функция $F(\omega;\ x)$ удовлетворяет определению 6.

Пусть $\omega \notin A \cup B \cup C$. Тогда ясно, что $F(\omega;\ x)$ является неубывающей функцией от x. Если $x < x' \leqslant r$, то $F(\omega;\ x) \leqslant F(\omega;\ x') \leqslant F(\omega;\ r) = F_r(\omega) \downarrow F(\omega,\ x)$, когда $r \downarrow x$. Поэтому $F(\omega;\ x)$ непрерывна справа. Аналогично $\lim\limits_{x\to\infty} F(\omega;\ x) = 1$, $\lim\limits_{x\to-\infty} F(\omega;\ x) = 0$. Поскольку для $\omega \in A \cup B \cup C$ $F(\omega;\ x) = G(x)$, то для каждого $\omega \in \Omega$ $F(\omega;\ x)$ является функцией распределения на R, т. е. выполнено условие а) в определении 6.

Согласно конструкции $\mathsf{P}(\xi \leqslant r) \mid \mathscr{G})(\omega) = F_r(\omega) = F(\omega;\ r)$. Если $r \downarrow x$, то для всех $\omega \in \Omega$ $F(\omega;\ r) \downarrow F(\omega;\ x)$ в силу установленной непрерывности справа. Но из утверждения а) теоремы 2 $\mathsf{P}(\xi \leqslant r \mid \mathscr{G})(\omega) \to \mathsf{P}(\xi \leqslant x \mid \mathscr{G})(\omega)$ (п. н.). Поэтому $F(\omega;\ x) = \mathsf{P}(\xi \leqslant x \mid \mathscr{G})(\omega)$ (п. н.), что и доказывает свойство b) определения 6.

Обратимся теперь к доказательству существования регулярного условного распределения ξ относительно $\mathscr{G}$.

Пусть $F(\omega; \cdot\, x)$ — построенная выше функция. Положим

$$Q(\omega;\ B)=\int_B F(\omega;\ dx),$$

где интеграл понимается как интеграл Лебега — Стилтьеса, из свойств которого (см. п. 7 § 6) вытекает, что $Q(\omega;\ B)$ является мерой по B для каждого фиксированного $\omega \in \Omega$. Для установления того, что $Q(\omega;\ B)$ есть вариант условной вероятности $\mathsf{P}(\xi \in B \mid \mathscr{G})(\omega)$, воспользуемся принципом подходящих множеств.

Пусть $\mathscr{C}$ — совокупность множеств B из $\mathscr{B}(R)$, для которых $Q(\omega;\ B)=\mathsf{P}(\xi \in B \mid \mathscr{G})(\omega)$ (п. н.). Поскольку $F(\omega;\ x)=\mathsf{P}(\xi \leqslant x \mid \mathscr{G})(\omega)$ (п. н.), то в систему $\mathscr{C}$ входят множества B вида $B=(-\infty,\ x]$, $x \in R$. Значит, в $\mathscr{C}$ входят также все интервалы вида $(a,\ b]$ и алгебра $\mathscr{A}$, состоящая из конечных сумм непересекающихся множеств вида $(a,\ b]$. Тогда из свойства непрерывности меры $Q(\omega;\ B)$ (ω — фиксировано) и утверждения b) теоремы 2 следует, что $\mathscr{C}$ является монотонным классом, и поскольку $\mathscr{A} \subseteq \mathscr{C} \subseteq \mathscr{B}(R)$, то из теоремы 1 § 2

$$\mathscr{B}(R)=\sigma(\mathscr{A}) \subseteq \sigma(\mathscr{C})=\mu(\mathscr{C})=\mathscr{C} \subseteq \mathscr{B}(R),$$

откуда $\mathscr{C}=\mathscr{B}(R)$.

Теорема доказана

С помощью несложных топологических рассмотрений утверждение теоремы 4 о существовании регулярного условного распределения можно распространить на случайные элементы со значениями в так называемых борелевских пространствах. Дадим соответствующее

Определение 7. Измеримое пространство $(E,\ \mathscr{E})$ называется *борелевским пространством*, если оно борелевски эквивалентно некоторому борелевскому подмножеству числовой прямой, т. е. существует взаимно однозначное отображение $\varphi=\varphi(e)$: $(E,\ \mathscr{E}) \to (R,\ \mathscr{B}(R))$ такое, что:

1) $\varphi(E) \equiv \{\varphi(e):\ e \in E\}$ есть некоторое множество из $\mathscr{B}(R)$;
2) φ — $\mathscr{E}$-измеримо ($\varphi^{-1}(A) \in \mathscr{E}$, $A \in \varphi(E) \cap \mathscr{B}(R)$),
3) φ^{-1} — $\mathscr{B}(R)/\mathscr{E}$-измеримо ($\varphi(B) \in \varphi(E) \cap \mathscr{B}(R)$, $B \in \mathscr{E}$).

Теорема 5. *Пусть $X=X(\omega)$ — случайный элемент со значениями в борелевском пространстве $(E,\ \mathscr{E})$. Тогда существует регулярное условное распределение X относительно $\mathscr{G}$.*

Доказательство. Пусть $\varphi=\varphi(e)$ — функция из определения 7. В силу 2) из этого определения $\varphi(X(\omega))$ является случайной величиной. Поэтому по теореме 4 определено условное распределение $Q(\omega;\ A)$ случайной величины $\varphi(X(\omega))$ относительно $\mathscr{G}$, $A \in \varphi(E) \cap \mathscr{B}(R)$.

Введем функцию $\tilde{Q}(\omega;\ B)=Q(\omega;\ \varphi(B))$, $B \in \mathscr{E}$. В силу 3) определения 7 $\varphi(B) \in \varphi(E) \cap \mathscr{B}(R)$ и, следовательно, $\tilde{Q}(\omega;\ B)$

определено. Понятно, что при каждом ω $\tilde{Q}(\omega;\ B)$ является мерой по $B\in\mathscr{E}$. Зафиксируем теперь $B\in\mathscr{E}$. В силу взаимной однозначности отображения $\varphi=\varphi(e)$

$$\tilde{Q}(\omega;\ B)=Q(\omega;\ \varphi(B))=\mathsf{P}\{\varphi(X)\in\varphi(B)\mid\mathscr{G}\}=\mathsf{P}\{X\in B\mid\mathscr{G}\}\ \text{(п. н.)}.$$

Таким образом, $\tilde{Q}(\omega;\ B)$ является регулярным условным распределением X относительно $\mathscr{G}$.

Теорема доказана.

Следствие. Пусть $X=X(\omega)$ — случайный элемент со значениями в *полном сепарабельном метрическом пространстве* $(E,\ \mathscr{E})$. Тогда существует регулярное условное распределение X относительно $\mathscr{G}$. В частности, такое распределение существует в случае пространств $(R^n,\ \mathscr{B}(R^n))$, $(R^\infty,\ \mathscr{B}(R^\infty))$.

Доказательство следует из теоремы 5 и известного результата из топологии о том, что такие пространства являются борелевскими.

8. Развитая выше теория условных математических ожиданий позволяет дать обобщение теоремы Байеса, находящей применения в статистике.

Напомним, что если $\mathscr{D}=\{A_1,\ \ldots,\ A_n\}$ — некоторое разбиение пространства Ω с $\mathsf{P}(A_i)>0$, то теорема Байеса (I.3.9) утверждает, что для всякого B с $\mathsf{P}(B)>0$

$$\mathsf{P}(A_i\mid B)=\frac{\mathsf{P}(A_i)\,\mathsf{P}(B\mid A_i)}{\sum\limits_{j=1}^{n}\mathsf{P}(A_j)\,\mathsf{P}(B\mid A_j)}.\tag{25}$$

Поэтому, если $\theta=\sum\limits_{i=1}^{n}a_iI_{A_i}$ — дискретная случайная величина, то, согласно (I.8.10),

$$\mathsf{M}[g(\theta)\mid B]=\frac{\sum\limits_{i=1}^{n}g(a_i)\,\mathsf{P}(A_i)\,\mathsf{P}(B\mid A_i)}{\sum\limits_{j=1}^{n}\mathsf{P}(A_j)\,\mathsf{P}(B\mid A_j)},\tag{26}$$

или

$$\mathsf{M}[g(\theta)\mid B]=\frac{\int\limits_{-\infty}^{\infty}g(a)\,\mathsf{P}(B\mid\theta=a)\,P_\theta(da)}{\int\limits_{-\infty}^{\infty}\mathsf{P}(B\mid\theta=a)\,P_\theta(da)}.\tag{27}$$

Основываясь на данном в начале этого параграфа определении $\mathsf{M}[g(\theta)\mid B]$, нетрудно установить, что формула (27) остается справедливой для любого события B с $\mathsf{P}(B)>0$, случайных величин θ и функций $g=g(a)$ с $\mathsf{M}\,|g(\theta)|<\infty$.

Рассмотрим теперь аналог формулы (27) для условных математических ожиданий $\mathsf{M}[g(\theta)\,|\,\mathcal{G}]$ относительно некоторой σ-алгебры $\mathcal{G}$, $\mathcal{G} \subseteq \mathcal{F}$.

Пусть

$$Q(B) = \int_B g(\theta)\,\mathsf{P}(d\omega), \quad B \in \mathcal{G}. \tag{28}$$

Тогда в силу (4)

$$\mathsf{M}[g(\theta)\,|\,\mathcal{G}] = \frac{dQ}{d\mathsf{P}}(\omega). \tag{29}$$

Наряду с σ-алгеброй $\mathcal{G}$ рассмотрим σ-алгебру $\mathcal{G}_\theta$. Тогда, согласно (5),

$$\mathsf{P}(B) = \int_\Omega \mathsf{P}(B\,|\,\mathcal{G}_\theta)\,d\mathsf{P} \tag{30}$$

или по формуле замены переменных под знаком интеграла Лебега

$$\mathsf{P}(B) = \int_{-\infty}^{\infty} \mathsf{P}(B\,|\,\theta = a)\,\mathsf{P}_\theta(da). \tag{31}$$

Поскольку

$$Q(B) = \mathsf{M}[g(\theta)\,I_B] = \mathsf{M}[g(\theta)\cdot\mathsf{M}(I_B\,|\,\mathcal{G}_\theta)],$$

то

$$Q(B) = \int_{-\infty}^{\infty} g(a)\,\mathsf{P}(B\,|\,\theta = a)\,\mathsf{P}_\theta(da). \tag{32}$$

Предположим теперь, что условные вероятности $\mathsf{P}(B\,|\,\theta = a)$ являются регулярными и допускают представление

$$\mathsf{P}(B\,|\,\theta = a) = \int_B \rho(\omega;\ a)\,\lambda(d\omega), \tag{33}$$

где $\rho = \rho(x;\ a)$ — неотрицательная измеримая по паре переменных функция, а λ — некоторая σ-конечная мера на $(\Omega,\ \mathcal{G})$.

Пусть $\mathsf{M}\,|g(\theta)| < \infty$. Покажем, что ($\mathsf{P}$-п. н.)

$$\mathsf{M}[g(\theta)\,|\,\mathcal{G}] = \frac{\int_{-\infty}^{\infty} g(a)\,\rho(\omega,\ a)\,\mathsf{P}_\theta(da)}{\int_{-\infty}^{\infty} \rho(\omega,\ a)\,P_\theta(da)} \tag{34}$$

(*обобщенная теорема Байеса*).

Для доказательства (34) нам понадобится следующая

Лемма. *Пусть $(\Omega,\ \mathcal{F})$ — некоторое измеримое пространство.*

а) *Пусть μ и λ — σ-конечные меры, $\mu \ll \lambda$ и $f = f(\omega)$ — $\mathcal{F}$-измеримая функция. Тогда*

$$\int_\Omega f\,d\mu = \int_\Omega f\,\frac{d\mu}{d\lambda}\cdot d\lambda \tag{35}$$

(в том смысле, что если существует один из интегралов, то существует и второй, и они совпадают).

b) *Если* ν — *мера со знаком и* μ, λ — σ-*конечные меры*, $\nu \ll \mu$, $\mu \ll \lambda$, *то*

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda} \ (\lambda\text{-п. н.}) \tag{36}$$

и

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\lambda} \Big/ \frac{d\mu}{d\lambda} \ (\mu\text{-п. н.}). \tag{37}$$

Доказательство. a) Поскольку

$$\mu(A) = \int_A \left(\frac{d\mu}{d\lambda}\right) d\lambda, \quad A \in \mathscr{F},$$

то (35) очевидным образом выполнено для всякой простой функции $f = \sum f_i I_{A_i}$. Общий случай следует из представления $f = f^+ - f^-$ и теоремы о монотонной сходимости.

b) Из утверждения a) с $f = \frac{d\nu}{d\mu}$ находим

$$\nu(A) = \int_A \left(\frac{d\nu}{d\mu}\right) d\mu = \int_A \left(\frac{d\nu}{d\mu}\right) \cdot \left(\frac{d\mu}{d\lambda}\right) \cdot d\lambda.$$

Тогда $\nu \ll \lambda$ и, значит,

$$\nu(A) = \int_A \frac{d\nu}{d\lambda} \, d\lambda,$$

откуда в силу произвольности множества A и свойства I (§ 6) следует (36).

Свойство (37) вытекает из (36) и того замечания, что $\mu\left\{\omega\colon \frac{d\mu}{d\lambda} = 0\right\} = \int\limits_{\left\{\omega\colon \frac{d\mu}{d\lambda} = 0\right\}} \frac{d\mu}{d\lambda} \, d\lambda = 0$ (на множестве $\left\{\omega\colon \frac{d\mu}{d\lambda} = 0\right\}$ правую часть в (37) можно определить произвольно, например, положить равной нулю). Лемма доказана.

Для доказательства (34) заметим, что в силу теоремы Фубини и предположения (33)

$$\mathsf{Q}(B) = \int_B \left[\int_{-\infty}^{\infty} g(a)\, \rho(x;\ a)\, P_\theta(da)\right] \lambda(dx), \tag{38}$$

$$\mathsf{P}(B) = \int_B \left[\int_{-\infty}^{\infty} \rho(x;\ a)\, P_\theta(da)\right] \lambda(dx). \tag{39}$$

Тогда в силу леммы

$$\frac{dQ}{dP} = \frac{dQ/d\lambda}{dP/d\lambda} \quad (\text{P-п. н.}),$$

что с учетом (38), (39) и (29) дает формулу (34).

Замечание. Формула (34) остается справедливой, если вместо случайной величины θ рассмотреть случайный элемент со значениями в некотором измеримом пространстве $(E, \mathscr{E})$ (с заменой интеграла по R интегралом по E).

Остановимся на некоторых частных случаях формулы (34).

Пусть σ-алгебра $\mathscr{G}$ порождается случайной величиной ξ, $\mathscr{G} = \mathscr{G}_\xi$. Предположим, что

$$\mathsf{P}(\xi \in A \mid \theta = a) = \int_A q(x;\ a)\, \lambda(dx), \qquad A \in \mathscr{B}(R), \tag{40}$$

где $q = q(x;\ a)$ — некоторая неотрицательная измеримая по паре переменных функция, а λ — некоторая σ-конечная мера на $(R, \mathscr{B}(R))$. Тогда из формулы замены переменных под знаком интеграла Лебега и (34) находим, что

$$\mathsf{M}[g(\theta) \mid \xi = x] = \frac{\int_{-\infty}^{\infty} g(a)\, q(x;\ a)\, P_\theta(da)}{\int_{-\infty}^{\infty} q(x;\ a)\, P_\theta(da)}. \tag{41}$$

Пусть, в частности, (θ, ξ) — пара дискретных случайных величин, $\theta = \sum a_i I_{A_i}$, $\xi = \sum x_j I_{B_j}$. Тогда, выбирая в качестве λ *считающую* меру ($\lambda(\{x_i\}) = 1$, $i = 1, 2, \ldots$) из (40), получим

$$\mathsf{M}[g(\theta) \mid \xi = x_j] = \frac{\sum_i g(a_i)\, \mathsf{P}(\xi = x_j \mid \theta = a_i)\, \mathsf{P}(\theta = a_i)}{\sum_i \mathsf{P}(\xi = x_j \mid \theta = a_i)\, \mathsf{P}(\theta = a_i)}. \tag{42}$$

(Ср. с (26).)

Пусть теперь (θ, ξ) — пара абсолютно непрерывных величин с плотностью $f_{\theta, \xi}(a, x)$. Тогда в силу (19) представление (40) выполнено с $q(x;\ a) = f_{\xi \mid \theta}(x \mid a)$ и мерой Лебега λ. Поэтому

$$\mathsf{M}[g(\theta) \mid \xi = x] = \frac{\int_{-\infty}^{\infty} g(a)\, f_{\xi \mid \theta}(x \mid a)\, f_\theta(a)\, da}{\int_{-\infty}^{\infty} f_{\xi \mid \theta}(x \mid a)\, f_\theta(a)\, da}. \tag{43}$$

9. Задачи.

1. Пусть ξ и η — независимые одинаково распределенные случайные величины и $\mathsf{M}\xi$ определено. Показать, что

$$\mathsf{M}(\xi \mid \xi + \eta) = \mathsf{M}(\eta \mid \xi + \eta) = \frac{\xi + \eta}{2} \quad (\text{п. н.}).$$

2. Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $\mathsf{M}|\xi_i|<\infty$. Показать, что

$$\mathsf{M}(\xi_1 \mid S_n, S_{n+1}, \ldots) = \frac{S_n}{n} \text{ (п. н.)},$$

где $S_n = \xi_1 + \ldots + \xi_n$.

3. Предположим, что случайные элементы (X, Y) таковы, что существует регулярное распределение $P_x(B) = \mathsf{P}(Y \in B \mid X = x)$. Показать, что если $\mathsf{M}|g(X, Y)| < \infty$, то P_x-п. н.

$$\mathsf{M}[g(X, Y) \mid X = x] = \int g(x, y)\, P_x(dy).$$

4. Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$. Показать, что

$$\mathsf{M}(\xi \mid a < \xi \leqslant b) = \frac{\int_a^b x\, dF_\xi(x)}{F_\xi(b) - F_\xi(a)}$$

(предполагается, что $F_\xi(b) - F_\xi(a) > 0$).

5. Пусть $g = g(x)$ — выпуклая книзу борелевская функция и $\mathsf{M}|g(\xi)| < \infty$. Показать, что для условных математических ожиданий справедливо *неравенство Иенсена*

$$g(\mathsf{M}(\xi \mid \mathscr{G})) \leqslant \mathsf{M}(g(\xi) \mid \mathscr{G}).$$

6. Показать, что случайная величина ξ и σ-алгебра $\mathscr{G}$ независимы (т. е. для любого $B \in \mathscr{G}$ случайные величины ξ и $I_B(\omega)$ независимы) тогда и только тогда, когда для каждой борелевской функции $g(x)$ с $\mathsf{M}|g(\xi)| < \infty$ $\mathsf{M}(g(\xi) \mid \mathscr{G}) = \mathsf{M}g(\xi)$.

§ 8. Случайные величины. II

1. В первой главе были введены такие характеристики простых случайных величин, как дисперсия, ковариация и коэффициент корреляции. Соответствующим образом эти понятия вводятся и в общем случае. А именно, пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство и $\xi = \xi(\omega)$ — случайная величина, для которой определено математическое ожидание $\mathsf{M}\xi$.

Дисперсией случайной величины ξ называется величина

$$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2.$$

Величина $\sigma = +\sqrt{\mathsf{D}\xi}$ называется *стандартным отклонением*.

Если ξ — случайная величина с гауссовской (нормальной) плотностью то смысл параметров m и σ, входящих в (1), оказывается очень простым:

$$m = \mathsf{M}\xi, \quad \sigma^2 = \mathsf{D}\xi.$$

Таким образом, распределение вероятностей этой случайной величины ξ, называемой *гауссовской*, или *нормально распределенной*, полностью определяется ее средним значением m и дисперсией σ^2. (В этой связи понятна часто используемая для этого запись: $\xi \sim \mathcal{N}(m, \sigma^2)$.)

Пусть теперь (ξ, η) — пара случайных величин. Их ковариацией называется величина

$$\operatorname{cov}(\xi, \eta) = \mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) \tag{2}$$

(предполагается, что математическое ожидание определено).

Если $\operatorname{cov}(\xi, \eta) = 0$, то говорят, что случайные величины ξ и η *не коррелированы*.

Если $\mathsf{D}\xi > 0$, $\mathsf{D}\eta > 0$, то величина

$$\rho(\xi, \eta) \equiv \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\mathsf{D}\xi \cdot \mathsf{D}\eta}} \tag{3}$$

называется коэффициентом корреляции случайных величин ξ и η.

Свойства дисперсии, ковариации и коэффициента корреляции для простых случайных величин были изложены в § 4 гл. I. В общем случае эти свойства формулируются совершенно аналогичным образом.

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор, компоненты которого имеют конечный второй момент. Назовем *матрицей ковариации* (ковариационной матрицей) вектора ξ матрицу (порядка $n \times n$) $\mathbb{R} = \|R_{ij}\|$, где $R_{ij} = \operatorname{cov}(\xi_i, \xi_j)$. Ясно, что матрица $\mathbb{R}$ является *симметрической*. Кроме того, она *неотрицательно определена*, т. е.

$$\sum_{i, j=1}^{n} R_{ij}\lambda_i\lambda_j \geqslant 0$$

для любых $\lambda_i \in R$, $i = 1, \ldots, n$, поскольку

$$\sum_{i, j}^{n} R_{ij}\lambda_i\lambda_j = \mathsf{M}\left[\sum_{i=1}^{n} (\xi_i - \mathsf{M}\xi_i)\lambda_i\right]^2 \geqslant 0.$$

Следующая лемма показывает, что справедлив и обратный результат.

Лемма. *Для того чтобы матрица $\mathbb{R}$ порядка $n \times n$ была ковариационной матрицей некоторого вектора $\xi = (\xi_1, \ldots, \xi_n)$, необходимо и достаточно, чтобы эта матрица была симметрической и неотрицательно определенной, или, что эквивалентно,* *существовала бы матрица A (порядка $n\times k$, $1\leqslant k\leqslant n$) такая, что*

$$\mathbb{R}=AA^*,$$

где $*$ — символ транспонирования.

Доказательство. Как показано выше, всякая ковариационная матрица является симметрической и неотрицательно определенной.

Обратно, пусть $\mathbb{R}$ — такая матрица. Из теории матриц известно, что для всякой симметрической неотрицательно определенной матрицы $\mathbb{R}$ можно найти такую ортогональную матрицу $\mathcal{O}$ (т. е. $\mathcal{O}\mathcal{O}^*=E$ — единичная матрица), что

$$\mathcal{O}^*\mathbb{R}\mathcal{O}=D,$$

где

$$D=\begin{pmatrix} d_1 & \ddots & 0 \\ 0 & & d_n \end{pmatrix}$$

— диагональная матрица с неотрицательными элементами d_i, $i=1, \ldots, n$.

Отсюда следует, что

$$\mathbb{R}=\mathcal{O}D\mathcal{O}^*=(\mathcal{O}B)(B^*\mathcal{O}^*),$$

где B — диагональная матрица с элементами $b_i=+\sqrt{d_i}$, $i=1, \ldots, n$. Поэтому, если положить $A=\mathcal{O}B$, то для матрицы $\mathbb{R}$ получим требуемое представление $\mathbb{R}=AA^*$.

Ясно, что всякая матрица AA^* является симметрической и неотрицательно определенной. Поэтому осталось лишь показать, что $\mathbb{R}$ является ковариационной матрицей некоторого случайного вектора.

Пусть $\eta_1, \eta_2, \ldots, \eta_n$ — последовательность независимых нормально распределенных случайных величин, $\mathcal{N}(0, 1)$. (Существование такой последовательности вытекает, например, из следствия 1 к теореме 1 § 9 и, в сущности, может быть легко выведено из теоремы 2 § 3). Тогда случайный вектор $\xi=A\eta$ (векторы рассматриваются как векторы-столбцы) обладает требуемым свойством. Действительно,

$$\mathsf{M}\xi\xi^*=\mathsf{M}(A\eta)(A\eta)^*=A\cdot\mathsf{M}\eta\eta^*\cdot A^*=AEA^*=AA^*.$$

(Если $\zeta=\|\zeta_{ij}\|$ — матрица, элементами которой являются случайные величины, то под $\mathsf{M}\zeta$ понимается матрица $\|\mathsf{M}\zeta_{ij}\|$).

Лемма доказана.

Обратимся теперь к двумерной гауссовской (нормальной) плотности

$$f_{\xi\eta}(x, y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2}-\right.\right.$$
$$\left.\left.-2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}, \quad (4)$$

характеризуемой пятью параметрами m_1, m_2, σ_1, σ_2 и ρ (ср. с (3.14)), где $|m_1|<\infty$, $|m_2|<\infty$, $\sigma_1>0$, $\sigma_2>0$, $|\rho|<1$. Простой подсчет раскрывает смысл этих параметров:

$$m_1 = \mathsf{M}\xi, \quad \sigma_1^2 = \mathsf{D}\xi,$$

$$m_2 = \mathsf{M}\eta, \quad \sigma_2^2 = \mathsf{D}\eta,$$

$$\rho = \rho(\xi, \eta).$$

В § 4 гл. I было объяснено, что если величины ξ и η не коррелированы ($\rho(\xi, \eta) = 0$), то отсюда еще не вытекает, что они независимы. Однако если пара (ξ, η) — гауссовская, то из некоррелированности ξ и η следует, что они независимы.

В самом деле, если в (4) $\rho = 0$, то

$$f_{\xi\eta}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} \cdot e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}.$$

Но в силу (6.55) и (4)

$$f_\xi(x) = \int_{-\infty}^{\infty} f_{\xi\eta}(x, y)\, dy = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}},$$

$$f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi\eta}(x, y)\, dx = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}.$$

Поэтому

$$f_{\xi\eta}(x, y) = f_\xi(x) \cdot f_\eta(y),$$

откуда следует, что величины ξ и η независимы (см. конец п. 8 § 6).

2. Убедительной иллюстрацией полезности введенного выше в § 7 понятия условного математического ожидания является его применение к решению следующей задачи, относящейся к *теории оценивания* (ср. с п. 8 § 4 гл. I).

Пусть (ξ, η) — пара случайных величин, из которых ξ наблюдаема, а η наблюдению не подлежит. Спрашивается, как по значениям наблюдений над ξ «оценить» ненаблюдаемую компоненту η?

Чтобы сделать эту задачу более определенной, введем понятие оценки. Пусть $\varphi = \varphi(x)$ — борелевская функция. Случайную величину $\varphi(\xi)$ будем называть *оценкой* η по ξ, а величину $\mathsf{M}[\eta - \varphi(\xi)]^2$ (среднеквадратической) ошибкой этой оценки. Оценку $\varphi^*(\xi)$ назовем *оптимальной* (в среднеквадратическом смысле), если

$$\Delta \equiv \mathsf{M}[\eta - \varphi^*(\xi)]^2 = \inf_\varphi \mathsf{M}[\eta - \varphi(\xi)]^2, \tag{5}$$

где inf берется по классу всех борелевских функций $\varphi = \varphi(x)$.

Теорема 1. *Пусть* $\mathsf{M}\eta^2 < \infty$. *Тогда оптимальная оценка* $\varphi^* = \varphi^*(\xi)$ *существует и в качестве* $\varphi^*(x)$ *может быть взята функция*

$$\varphi^*(x) = \mathsf{M}(\eta \mid \xi = x). \tag{6}$$

Доказательство. Без ограничения общности можно рассматривать только те оценки $\varphi(\xi)$, для которых $\mathsf{M}\varphi^2(\xi) < \infty$. Тогда, если $\varphi(\xi)$ — такая оценка, а $\varphi^*(\xi) = \mathsf{M}(\eta \mid \xi)$, то

$$\begin{aligned}\mathsf{M}[\eta - \varphi(\xi)]^2 &= \mathsf{M}[(\eta - \varphi^*(\xi)) + (\varphi^*(\xi) - \varphi(\xi))]^2 = \\ &= \mathsf{M}[\eta - \varphi^*(\xi)]^2 + \mathsf{M}[\varphi^*(\xi) - \varphi(\xi)]^2 + \\ &\quad + 2\mathsf{M}[(\eta - \varphi^*(\xi))(\varphi^*(\xi) - \varphi(\xi))] \geqslant \mathsf{M}[\eta - \varphi^*(\xi)]^2,\end{aligned}$$

поскольку $\mathsf{M}[\varphi^*(\xi) - \varphi(\xi)]^2 \geqslant 0$ и по свойствам условных математических ожиданий

$$\begin{aligned}\mathsf{M}[(\eta - \varphi^*(\xi))(\varphi^*(\xi) - \varphi(\xi))] &= \mathsf{M}\{\mathsf{M}[(\eta - \varphi^*(\xi))(\varphi^*(\xi) - \varphi(\xi)] \mid \xi]\} = \\ &= \mathsf{M}\{(\varphi^*(\xi) - \varphi(\xi))\,\mathsf{M}(\eta - \varphi^*(\xi) \mid \xi)\} = 0.\end{aligned}$$

Теорема доказана.

Замечание. Из доказательства теоремы видно, что ее утверждение справедливо и в том случае, когда ξ не только случайная величина, но и произвольный случайный элемент со значениями в некотором измеримом пространстве $(E, \mathscr{E})$. Под оценками $\varphi = \varphi(x)$ тогда следует понимать $\mathscr{E}/\mathscr{B}(R)$-измеримые функции.

Рассмотрим структуру функции $\varphi^*(x)$ в предположении, что (ξ, η) — гауссовская пара с плотностью, задаваемой формулой (4).

Из (1), (4) и (7.10) находим, что плотность $f_{\eta \mid \xi}(y \mid x)$ условного распределения вероятностей задается формулой

$$f_{\eta \mid \xi}(y \mid x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\,\sigma_2}\, e^{-\frac{(y - m(x))^2}{2\sigma_2^2(1-\rho^2)}}, \tag{7}$$

где

$$m(x) = m_2 + \frac{\sigma_2}{\sigma_1}\rho \cdot (x - m_1). \tag{8}$$

Тогда из следствия к теореме 3 § 7

$$\mathsf{M}(\eta \mid \xi = x) = \int_{-\infty}^{\infty} y f_{\eta \mid \xi}(y \mid x)\, dy = m(x) \tag{9}$$

и

$$\begin{aligned}\mathsf{D}(\eta \mid \xi = x) &\equiv \mathsf{M}[(\eta - \mathsf{M}(\eta \mid \xi = x))^2 \mid \xi = x] = \\ &= \int_{-\infty}^{\infty} (y - m(x))^2 f_{\eta \mid \xi}(y \mid x)\, dy = \\ &= \sigma_2^2(1 - \rho^2).\end{aligned} \tag{10}$$

Заметим, что условная дисперсия $\mathsf{D}(\eta \mid \xi = x)$ не зависит от x и, значит,

$$\Delta = \mathsf{M}[\eta - \mathsf{M}(\eta \mid \xi = x)]^2 = \sigma_2^2(1 - \rho^2). \tag{11}$$

Формулы (9), (11) получены в предположении $\mathsf{D}\xi > 0$, $\mathsf{D}\eta > 0$. Если же $\mathsf{D}\xi > 0$, а $\mathsf{D}\eta = 0$, то они выполняются очевидным образом.

Итак, справедлив следующий результат (ср. с (I.4.16), (I.4.17)).

Теорема 2. *Пусть (ξ, η) — гауссовский вектор с $\mathsf{D}\xi > 0$. Тогда оптимальная оценка η по ξ есть*

$$\mathsf{M}(\eta \mid \xi) = \mathsf{M}\eta + \frac{\operatorname{cov}(\xi, \eta)}{\mathsf{D}\xi}(\xi - \mathsf{M}\xi), \tag{12}$$

а ее ошибка

$$\Delta \equiv \mathsf{M}[\eta - \mathsf{M}(\eta \mid \xi)]^2 = \mathsf{D}\eta - \frac{\operatorname{cov}^2(\xi, \eta)}{\mathsf{D}\xi}. \tag{13}$$

Замечание. Кривая $y(x) = \mathsf{M}(\eta \mid \xi = x)$ называется *кривой регрессии η на ξ или η по отношению к ξ*. В гауссовском случае $\mathsf{M}(\eta \mid \xi = x) = a + bx$ и, следовательно, регрессия η на ξ является *линейной*. Поэтому нет ничего удивительного в том, что правые части формул (12) и (13) совпадают с соответствующими частями формул (I.4.16) и (I.4.17) для оптимальной линейной оценки и ее ошибки.

Следствие. Пусть ε_1 и ε_2 — независимые гауссовские случайные величины с нулевыми средними и единичной дисперсией и

$$\xi = a_1\varepsilon_1 + a_2\varepsilon_2, \quad \eta = b_1\varepsilon_1 + b_2\varepsilon_2.$$

Тогда $\mathsf{M}\xi = \mathsf{M}\eta = 0$, $\mathsf{D}\xi = a_1^2 + a_2^2$, $\mathsf{D}\eta = b_1^2 + b_2^2$, $\operatorname{cov}(\xi, \eta) = a_2 b_1 + a_2 b_2$, и если $a_1^2 + a_2^2 > 0$, то

$$\mathsf{M}(\eta \mid \xi) = \frac{a_1 b_1 + a_2 b_2}{a_1^2 + a_2^2}\,\xi, \tag{14}$$

$$\Delta = \frac{(a_1 b_2 - a_2 b_1)^2}{a_1^2 + a_2^2}. \tag{15}$$

3. Рассмотрим вопросы отыскания функций распределения для случайных величин, являющихся функциями от других случайных величин.

Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$ (и плотностью $f_\xi(x)$, если таковая существует), $\varphi = \varphi(x)$ — некоторая борелевская функция и $\eta = \varphi(\xi)$. Обозначая $I_y = (-\infty, y)$, находим

$$F_\eta(y) = \mathsf{P}(\eta \leqslant y) = \mathsf{P}(\varphi(\xi) \in I_y) = \mathsf{P}(\xi \in \varphi^{-1}(I_y)) = \int_{\varphi^{-1}(I_y)} F_\xi(dx), \tag{16}$$

что дает выражение для функции распределения $F_\eta(y)$ через функцию распределения $F_\xi(x)$ и функцию φ.

Так, если $\eta = a\xi + b$, $a > 0$, то

$$F_\eta(y) = \mathsf{P}\left(\xi \leqslant \frac{y-b}{a}\right) = F_\xi\left(\frac{y-b}{a}\right). \tag{17}$$

Если $\eta = \xi^2$, то, очевидно, $F_\eta(y) = 0$ для $y < 0$, а для $y \geqslant 0$

$$F_\eta(y) = \mathsf{P}(\xi^2 \leqslant y) = \mathsf{P}(-\sqrt{y} \leqslant \xi \leqslant \sqrt{y}) = \\ = F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y}) + \mathsf{P}(\xi = -\sqrt{y}). \tag{18}$$

Обратимся теперь к вопросу отыскания плотности $f_\eta(y)$.

Предположим, что область значений случайной величины ξ есть (конечный или бесконечный) открытый интервал $I = (a, b)$, а функция $\varphi = \varphi(x)$, определенная для $x \in I$, является непрерывно дифференцируемой и либо строго возрастающей, либо строго убывающей. Будем предполагать также, что $\varphi'(x) \neq 0$, $x \in I$.

Обозначим $h(y) = \varphi^{-1}(y)$ и предположим для определенности, что $\varphi(x)$ строго возрастает. Тогда для $y \in \varphi(I)$

$$F_\eta(y) = \mathsf{P}(\eta \leqslant y) = \mathsf{P}(\varphi(\xi) \leqslant y) = \mathsf{P}(\xi \leqslant \varphi^{-1}(y)) = \\ = \mathsf{P}(\xi \leqslant h(y)) = \int_{-\infty}^{h(y)} f_\xi(x)\,dx. \tag{19}$$

Согласно задаче (15) из § 6

$$\int_{-\infty}^{h(y)} f_\xi(x)\,dx = \int_{-\infty}^{y} f_\xi(h(z))\,h'(z)\,dz \tag{20}$$

и, значит,

$$f_\eta(y) = f_\xi(h(y))\,h'(y). \tag{21}$$

Аналогично, если функция $\varphi(x)$ является строго убывающей, то

$$f_\eta(y) = f_\xi(h(y))\,(-h'(y)).$$

Таким образом, в обоих случаях

$$f_\eta(y) = f_\xi(h(y))\,|h'(y)|. \tag{22}$$

Например, если $\eta = a\xi + b$, $a \neq 0$, то $h(y) = \frac{y-b}{a}$ и $f_\eta(y) = \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right)$.

Если $\xi \sim \mathcal{N}(m, \sigma^2)$, а $\eta = e^\xi$, то из (22) находим, что

$$f_\eta(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma y} \exp\left[-\dfrac{\ln\left(\frac{y}{M}\right)^2}{2\sigma^2}\right], & y > 0, \\ 0, & y \leqslant 0, \end{cases} \tag{23}$$

где $M = e^m$.

Распределение вероятностей с плотностью (23) называется *логарифмически нормальным.*

Если функция $\varphi=\varphi(x)$ не является строго возрастающей или строго убывающей, то формула (22) неприменима. Однако для многих приложений вполне достаточно следующее ее обобщение.

Пусть функция $\varphi=\varphi(x)$ определена на множестве $\sum_{k=1}^{n}[a_k, b_k]$, причем на каждом открытом интервале $I_k=(a_k, b_k)$ является непрерывно дифференцируемой либо строго возрастающей, либо строго убывающей, $\varphi'(x)\neq 0$ при $x\in I_k$. Пусть $h_k=h_k(y)$ — обратная функция к $\varphi(x)$, $x\in I_k$. Тогда имеет место следующее обобщение формулы (22):

$$f_\eta(y)=\sum_{k=1}^{n} f_\xi(h_k(y))\,|h'_k(y)|\cdot I_{D_k}(y), \tag{24}$$

где D_k — область определения функции $h_k(y)$.

Так, например, если $\eta=\xi^2$, то, беря $I_1=(-\infty, 0)$, $I_2=(0, \infty)$, находим, что $h_1(y)=-\sqrt{y}$, $h_2(y)=\sqrt{y}$, и, значит,

$$f_\eta(y)=\begin{cases}\dfrac{1}{2\sqrt{y}}[f_\xi(\sqrt{y})+f_\xi(-\sqrt{y})], & y>0,\\ 0, & y\leqslant 0.\end{cases} \tag{25'}$$

Заметим, что этот результат следует также из (18), поскольку $\mathsf{P}(\xi=-\sqrt{y})=0$. В частности, если $\xi\sim\mathcal{N}(0, 1)$, то

$$f_{\xi^2}(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi y}}e^{-y/2}, & y>0,\\ 0, & y\leqslant 0.\end{cases} \tag{26}$$

Несложный подсчет показывает также, что

$$f_{|\xi|}(y)=\begin{cases} f_\xi(y)+f_\xi(-y), & y>0,\\ 0, & y\leqslant 0,\end{cases} \tag{27}$$

$$f_{+\sqrt{|\xi|}}(y)=\begin{cases} 2y\,(f_\xi(y^2)+f_\xi(-y^2)), & y>0,\\ 0, & y\leqslant 0.\end{cases} \tag{28}$$

4. Обратимся теперь к функциям от многих случайных величин.

Если ξ и η — случайные величины с совместным распределением $F_{\xi\eta}(x, y)$, а $\varphi=\varphi(x, y)$ — некоторая борелевская функция, то для $\zeta=\varphi(\xi, \eta)$ сразу получаем, что

$$F_\zeta(z)=\int\limits_{\{x,\,y:\ \varphi(x,\,y)\leqslant z\}} dF_{\xi\eta}(x, y). \tag{29}$$

Например, если $\varphi(x, y) = x + y$, а ξ и η независимы (и, значит, $F_{\xi\eta}(x, y) = F_\xi(x) \cdot F_\eta(y)$), то, применяя теорему Фубини, получим

$$F_\zeta(z) = \int\limits_{\{x,\, y\colon\, x+y \leqslant z\}} dF_\xi(x) \cdot dF_\eta(y) = \int\limits_{R^2} I_{\{x+y \leqslant z\}}(x, y)\, dF_\xi(x)\, dF_\eta(y) =$$

$$= \int\limits_{-\infty}^{\infty} dF_\xi(x) \left\{ \int\limits_{-\infty}^{\infty} I_{\{x+y \leqslant z\}}(x, y)\, dF_\eta(y) \right\} = \int\limits_{-\infty}^{\infty} F_\eta(z - x)\, dF_\xi(x) \qquad (30)$$

и аналогично

$$F_\zeta(z) = \int\limits_{-\infty}^{\infty} F_\xi(z - y)\, dF_\eta(y). \qquad (31)$$

Если F и G — две функции распределения, то функцию

$$H(z) = \int\limits_{-\infty}^{\infty} F(z - x)\, dG(x)$$

принято обозначать $F * G$ и называть *сверткой* F и G.

Таким образом, *функция распределения F_ζ суммы двух независимых случайных величин ξ и η есть свертка их функций распределения F_ξ и F_η*:

$$F_\zeta = F_\xi * F_\eta.$$

Ясно при этом, что $F_\xi * F_\eta = F_\eta * F_\xi$.

Предположим теперь, что независимые случайные величины ξ и η имеют плотности f_ξ и f_η. Тогда из (31), снова применяя теорему Фубини, найдем, что

$$F_\zeta(z) = \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{z-y} f_\xi(u)\, du \right] f_\eta(y)\, dy =$$

$$= \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{z} f_\xi(u - y)\, du \right] f_\eta(y)\, dy = \int\limits_{-\infty}^{z} \left[\int\limits_{-\infty}^{\infty} f_\xi(u - y) f_\eta(y)\, dy \right] du,$$

откуда

$$f_\zeta(z) = \int\limits_{-\infty}^{\infty} f_\xi(z - y) f_\eta(y)\, dy, \qquad (32)$$

и аналогично

$$f_\zeta(z) = \int\limits_{-\infty}^{\infty} f_\eta(z - x) f_\xi(x)\, dx. \qquad (33)$$

Рассмотрим несколько примеров на применение этих формул.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — последовательность независимых одинаково распределенных случайных величин с равномерной на $[-1, 1]$ плотностью

$$f(x)=\begin{cases} 1/2, & |x|\leqslant 1, \\ 0, & |x|>1. \end{cases}$$

Тогда из (32) находим

$$f_{\xi_1+\xi_2}(x)=\begin{cases} \dfrac{2-|x|}{4}, & |x|\leqslant 2, \\ 0, & |x|>2, \end{cases}$$

$$f_{\xi_1+\xi_2+\xi_3}(x)=\begin{cases} \dfrac{(3-|x|)^2}{16}, & 1\leqslant|x|\leqslant 3, \\ \dfrac{3-x^2}{8}, & 0\leqslant|x|\leqslant 1, \\ 0, & |x|>3, \end{cases}$$

и вообще (по индукции)

$$f_{\xi_1+\ldots+\xi_n}(x)=\begin{cases} \dfrac{1}{2^n(n-1)!}\displaystyle\sum_{k=0}^{\left[\frac{n+x}{2}\right]}(-1)^k C_n^k(n+x-2k)^{n-1}, & |x|\leqslant n, \\ 0, & |x|>n. \end{cases}$$

Пусть теперь $\xi\sim\mathcal{N}(m_1, \sigma_1^2)$, $\eta\sim\mathcal{N}(m_2, \sigma_2^2)$. Если обозначить

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

то

$$f_\xi(x)=\frac{1}{\sigma_1}\varphi\left(\frac{x-m_1}{\sigma_1}\right), \quad f_\eta(x)=\frac{1}{\sigma_2}\varphi\left(\frac{x-m_2}{\sigma_2}\right),$$

и из (32) легко находим, что

$$f_{\xi+\eta}(x)=\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\varphi\left(\frac{x-(m_1+m_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right).$$

Таким образом, *сумма двух независимых гауссовских случайных величин снова есть гауссовская случайная величина* со средним m_1+m_2 и дисперсией $\sigma_1^2+\sigma_2^2$.

Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины, каждая из которых нормально распределена с нулевым средним и единичной дисперсией. Тогда, используя (26), нетрудно (по индукции) найти, что

$$f_{\xi_1^2+\ldots+\xi_n^2}(x)=\begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)}x^{(n/2)-1}e^{-x/2}, & x>0, \\ 0, & x\leqslant 0. \end{cases} \tag{34}$$

Обычно величина $\xi_1^2+\ldots+\xi_n^2$ обозначается χ_n^2, а ее распределение (с плотностью (32)) называется *χ^2-распределением* (хи-квадрат распределением) с n степенями свободы (ср. с табл. 2 в § 3).

Если обозначить $\chi_n=+\sqrt{\chi_n^2}$, то из (28) и (34) следует, что

$$f_{\chi_n}(x)=\begin{cases}\dfrac{2x^{n-1}e^{-x^2/2}}{2^{n/2}\Gamma(n/2)}, & x\geqslant 0,\\ 0, & x<0.\end{cases} \tag{35}$$

Распределение вероятностей с такой плотностью принято называть *χ-распределением* (хи-распределением) с n степенями свободы.

Пусть снова ξ и η — независимые случайные величины с плотностями f_ξ и f_η. Тогда

$$F_{\xi\eta}(z)=\iint\limits_{\{x,\, y:\ xy\leqslant z\}} f_\xi(x)f_\eta(y)\,dx\,dy,$$

$$F_{\xi/\eta}(z)=\iint\limits_{\left\{x,\, y:\ \frac{x}{y}\leqslant z\right\}} f_\xi(x)f_\eta(y)\,dx\,dy.$$

Отсюда нетрудно получить, что

$$f_{\xi\eta}(z)=\int_{-\infty}^{\infty} f_\xi\left(\frac{z}{y}\right)f_\eta(y)\frac{dy}{|y|}=\int_{-\infty}^{\infty} f_\eta\left(\frac{z}{x}\right)f_\xi(x)\frac{dx}{|x|} \tag{36}$$

и

$$f_{\frac{\xi}{\eta}}(z)=\int_{-\infty}^{\infty} f_\xi(zy)f_\eta(y)\,|y|\,dy. \tag{37}$$

Полагая в (37) $\xi=\xi_0$ и $\eta=\sqrt{\dfrac{\xi_1^2+\ldots+\xi_n^2}{n}}$, где $\xi_0, \xi_1, \ldots, \xi_n$ — независимые гауссовские случайные величины с нулевыми средними и дисперсиями $\sigma^2>0$, и используя (35), найдем, что

$$f_{\frac{\xi_0}{\sqrt{\frac{1}{n}(\xi_1^2+\ldots+\xi_n^2)}}}(x)=\frac{1}{\sqrt{\pi n}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\,\frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}}. \tag{38}$$

Величина $\dfrac{\xi_0}{\sqrt{\frac{1}{n}(\xi_1^2+\ldots+\xi_n^2)}}$ обычно обозначается через t, а ее распределение называется *t-распределением* или *распределением Стьюдента* с n степенями свободы (ср. с табл. 2 в § 3). Заметим, что это распределение не зависит от σ.

5. Задачи.

1. Проверить справедливость формул (9), (10), (24), (27), (28), (34)—(38).

2. Пусть $\xi_1, \ldots, \xi_n$, $n \geqslant 2$, — независимые одинаково распределенные случайные величины с функцией распределения $F(x)$ (и плотностью $f(x)$, если таковая существует) и $\bar{\xi} = \max(\xi_1, \ldots, \xi_n)$, $\underline{\xi} = \min(\xi_1, \ldots, \xi_n)$, $\rho = \bar{\xi} - \underline{\xi}$. Показать, что

$$F_{\bar{\xi}, \underline{\xi}}(x, y) = \begin{cases} (F(y))^n - (F(y) - F(x))^n, & y > x, \\ (F(y))^n, & y \leqslant x, \end{cases}$$

$$f_{\bar{\xi}, \underline{\xi}}(x, y) = \begin{cases} n(n-1)[F(y) - F(x)]^{n-2} f(x) f(y), & y > x, \\ 0, & y < x, \end{cases}$$

$$F_\rho(x) = \begin{cases} n \int\limits_{-\infty}^{\infty} [F(y) - F(y-x)]^{n-1} f(y)\, dy, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

$$f_\rho(x) = \begin{cases} n(n-1) \int\limits_{-\infty}^{\infty} [F(y) - F(y-x)]^{n-2} f(y-x) f(y)\, dy, & x > 0, \\ 0, & x < 0. \end{cases}$$

3. Пусть ξ_1 и ξ_2 — независимые пуассоновские случайные величины с параметрами λ_1 и λ_2 соответственно. Показать, что $\xi_1 + \xi_2$ также имеет пуассоновское распределение с параметром $\lambda_1 + \lambda_2$.

4. Пусть в (4) $m_1 = m_2 = 0$ Показать, что

$$f_{\xi/\eta}(z) = \frac{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}}{\pi(\sigma_2^2 z^2 - 2\rho\sigma_1\sigma_2 z + \sigma_1^2)}.$$

5. Величина $\rho^*(\xi, \eta) = \sup\limits_{u, v} \rho(u(\xi), v(\eta))$, где супремум берется по всем борелевским функциям $u = u(x)$ и $v = v(x)$, для которых коэффициент корреляции $\rho(u(\xi), v(\eta))$ определен, называется *максимальным коэффициентом корреляции* ξ *и* η. Показать, что сcучайные величины ξ и η независимы тогда и только тогда, когда $\rho^*(\xi, \eta) = 0$.

6. Пусть $\tau_1, \tau_2, \ldots, \tau_n$ — независимые неотрицательные одинаково распределенные случайные величины с экспоненциальной плотностью распределения

$$f(t) = \lambda e^{-\lambda t}, \qquad t \geqslant 0.$$

Показать, что распределение случайной величины $\tau_1 + \ldots + \tau_k$ имеет плотность

$$\frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!}, \qquad t \geqslant 0, \qquad 1 \leqslant k \leqslant n,$$

и

7. Пусть $\xi \sim \mathscr{N}(0, \sigma^2)$. Показать, что для всякого $p \geqslant 1$

$$\mathsf{M}|\xi|^p = C_p \sigma^p,$$

где

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

и $\Gamma(s) = \int\limits_0^\infty e^{-x} x^{s-1}\, dx$ — гамма-функция Эйлера. В частности, для любого целого $n \geqslant 1$

$$\mathsf{M}\xi^{2n} = (2n-1)!!\, \sigma^{2n}.$$

§ 9. Построение процесса с заданными конечномерными распределениями

1. Пусть $\xi = \xi(\omega)$ — случайная величина, заданная на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ и

$$F_\xi(x) = \mathsf{P}\{\omega: \xi(\omega) \leqslant x\} -$$

ее функция распределения. Понятно, что $F_\xi(x)$ является функцией распределения на числовой прямой в смысле определения 1 § 3.

Поставим сейчас следующий вопрос. Пусть $F = F(x)$ — некоторая функция распределения на R. Спрашивается, существует ли случайная величина, имеющая функцию $F(x)$ своей функцией распределения?

Одна из причин, оправдывающая эту постановку вопроса, состоит в следующем. Многие утверждения теории вероятностей начинаются словами: «Пусть ξ — случайная величина с функцией распределения $F(x)$, тогда...». Поэтому, чтобы утверждения подобного типа были содержательными, надо иметь уверенность, что рассматриваемый объект действительно существует. Поскольку для задания случайной величины нужно прежде всего задать область ее определения $(\Omega, \mathscr{F})$, а для того, чтобы говорить о ее распределении надо иметь вероятностную меру P на $(\Omega, \mathscr{F})$, то правильная постановка вопроса о существовании случайной величины с заданной функцией распределения $F(x)$ такова:

Существуют ли вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ *и случайная величина* $\xi = \xi(\omega)$ *на нем такие, что*

$$\mathsf{P}\{\omega: \xi(\omega) \leqslant x\} = F(x)?$$

Покажем, что ответ на этот вопрос положительный и, в сущности, он содержится в теореме 1 § 1.

Действительно, положим

$$\Omega = R, \quad \mathscr{F} = \mathscr{B}(R).$$

Тогда из теоремы 1 § 1 следует, что на $(R, \mathscr{B}(R))$ существует (и притом единственная) вероятностная мера P, для которой $\mathsf{P}(a, b] = F(b) - F(a)$, $a < b$.

Положим $\xi(\omega) \equiv \omega$. Тогда

$$\mathsf{P}\{\omega: \xi(\omega) \leqslant x\} = \mathsf{P}\{\omega: \omega \leqslant x\} = \mathsf{P}(-\infty, x] = F(x).$$

Таким образом, требуемое вероятностное пространство и искомая случайная величина построены.

2. Поставим теперь аналогичный вопрос для случайных процессов.

Пусть $X = (\xi_t)_{t \in T}$ — случайный процесс (в смысле определения 3 § 5), заданный на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ для $t \in T \subseteq R$.

С физической точки зрения наиболее важной вероятностной характеристикой случайного процесса является набор $\{F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)\}$ его *конечномерных функций распределения*

$$F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = P\{\omega: \xi_{t_1} \leqslant x_1, \ldots, \xi_{t_n} \leqslant x_n\}, \tag{1}$$

заданных для всех наборов $t_1, \ldots, t_n$ с $t_1 < t_2 < \ldots < t_n$.

Из (1) видно, что для каждого набора $t_1, \ldots, t_n$ с $t_1 < t_2 < \ldots \ldots < t_n$ функции $F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)$ являются n-мерными функциями распределения (в смысле определения 2 § 3) и что набор $\{F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)\}$ удовлетворяет следующим условиям *согласованности*:

$$F_{t_1, \ldots, t_k}(x_1, \ldots, x_k) = F_{t_1, \ldots, t_n}(x_1, \ldots, x_k, +\infty, \ldots, +\infty), \tag{2}$$

где $k < n$.

Естественно теперь поставить такой вопрос: при каких условиях заданное семейство $\{F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)\}$ функций распределения $F_{t_1 \ldots t_n}(x_1, \ldots, x_n)$ (в смысле определения 2 § 3) может быть семейством конечномерных функций распределения некоторого случайного процесса? Весьма примечательно, что все такие дополнительные условия исчерпываются условиями согласованности (2).

Теорема 1 (теорема Колмогорова о существовании процесса). *Пусть $\{F_{t_1 \ldots t_n}(x_1 \ldots x_n)\}$, где $t_i \in T \subseteq R$, $t_1 < t_2 < \ldots < t_n$, $n \geqslant 1$, заданное семейство конечномерных функций распределения, удовлетворяющих условиям согласованности (2). Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и случайный процесс $X = (\xi_t)_{t \in T}$, такие что*

Доказательство. Положим

$$\Omega = R^T, \quad \mathscr{F} = \mathscr{B}(R^T),$$

т. е. возьмем в качестве пространства Ω пространство действительных функций $\omega = (\omega_t)_{t \in T}$ с σ-алгеброй, порожденной цилиндрическими множествами.

Пусть $\tau = [t_1, \ldots, t_n]$, $t_1 < t_2 < \ldots < t_n$. Тогда, согласно теореме 2 из § 3, в пространстве $(R^n, \mathscr{B}(R^n))$ можно построить и и притом единственную вероятностную меру P_τ такую, что

$$P_\tau\{(\omega_{t_1}, \ldots, \omega_{t_n}): \omega_{t_1} \leqslant x_1, \ldots, \omega_{t_n} \leqslant x_n\} = F_{t_1 \ldots t_n}(x_1, \ldots, x_n). \quad (4)$$

Из условий согласованности (2) вытекает, что семейство $\{P_\tau\}$ также является согласованным (см. (3.20)). Согласно теореме 4 из § 3 на пространстве $(R^T, \mathscr{B}(R^T))$ существует вероятностная мера P такая, что

$$\mathsf{P}\{\omega: (\omega_{t_1}, \ldots, \omega_{t_n}) \in B\} = P_\tau(B)$$

для всякого набора $\tau = [t_1, \ldots, t_n]$, $t_1 < \ldots < t_n$.

Отсюда следует также, что выполнено условие (4). Таким образом, в качестве искомого случайного процесса $X = \{\xi_t(\omega)\}_{t \in T}$ можно взять процесс, определенный следующим образом:

$$\xi_t(\omega) = \omega_t, \quad t \in T. \quad (5)$$

Теорема доказана.

Замечание 1. Построенное вероятностное пространство $(R^T, \mathscr{B}(R^T), \mathsf{P})$ часто называют *каноническим*, а задание случайного процесса равенством (5) — *координатным способом* построения процесса.

Замечание 2. Пусть $(E_\alpha, \mathscr{E}_\alpha)$ — полные сепарабельные метрические пространства, α принадлежит произвольному множеству индексов $\mathfrak{A}$. Пусть $\{P_\tau\}$ — набор согласованных конечномерных функций распределения P_τ, $\tau = [\alpha_1, \ldots, \alpha_n]$ на $(E_{\alpha_1} \times \ldots \times E_{\alpha_n}, \mathscr{E}_{\alpha_1} \otimes \ldots \otimes \mathscr{E}_{\alpha_n})$. Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и семейство $\mathscr{F}/\mathscr{E}_\alpha$-измеримых функций $(X_\alpha(\omega))_{\alpha \in \mathfrak{A}}$ такие, что

$$\mathsf{P}\{(X_{\alpha_1}, \ldots, X_{\alpha_n}) \in B\} = P_\tau(B)$$

для любых $\tau = [\alpha_1, \ldots, \alpha_n]$ и $B \in \mathscr{E}_{\alpha_1} \otimes \ldots \otimes \mathscr{E}_{\alpha_n}$.

Этот результат, обобщающий утверждение теоремы 1, следует из теоремы 4 § 3, если положить $\Omega = \prod_\alpha E_\alpha$, $\mathscr{F} = \bigotimes_\alpha \mathscr{E}_\alpha$ и $X_\alpha(\omega) = \omega_\alpha$ для каждого $\omega = (\omega_\alpha)$, $\alpha \in \mathfrak{A}$.

Следствие 1. Пусть $F_1(x)$, $F_2(x)$, … — последовательность одномерных функций распределения. Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и последовательность независимых случайных величин $\xi_1, \xi_2, \ldots$ такие, что

$$P\{\omega: \xi_i(\omega) \leqslant x\} = F_i(x). \tag{6}$$

В частности, существует вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$, на котором определена бесконечная последовательность бернуллиевских случайных величин (в этой связи см. п. 2 § 5 гл. I). Отметим, что в качестве Ω можно здесь взять пространство

$$\Omega = \{\omega: \omega = (a_1, a_2, \ldots), a_i = 0,1\}$$

(ср. также с теоремой 2).

Для доказательства следствия достаточно положить $F_{1, \ldots, n}(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n)$ и применить теорему 1.

Следствие 2. Пусть $T = [0, \infty)$ и $\{p(s, x; t, B)\}$ — семейство неотрицательных функций, определенных для $s, t \in T$, $t > s$, $x \in R$, $B \in \mathscr{B}(R)$ и удовлетворяющих следующим условиям:

a) $p(s, x; t, B)$ является при фиксированных s, x и t *вероятностной мерой по* B;

b) при фиксированных s, t и B $p(s, x; t, B)$ является *борелевской функцией по* x;

c) для всех $0 \leqslant s < t < \tau$ и $B \in \mathscr{B}(R)$ выполняется *уравнение Колмогорова — Чэпмена*

$$p(s, x; \tau, B) = \int_R p(s, x; t, dy)\, p(t, y; \tau, B). \tag{7}$$

И пусть $\pi = \pi(B)$ — вероятностная мера на $(R, \mathscr{B}(R))$. Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и случайный процесс $X = \{\xi_t\}_{t \geqslant 0}$ на нем такие, что для $0 = t_0 < t_1 < \ldots < t_n$,

$$\mathsf{P}\{\xi_{t_0} \leqslant x_0, \xi_{t_1} \leqslant x_1, \ldots, \xi_{t_n} \leqslant x_n\} =$$

$$= \int_{-\infty}^{x_0} \pi(dy_0) \int_{-\infty}^{x_1} p(0, y_0; t_1, dy_1) \ldots \int_{-\infty}^{x_n} p(t_{n-1}, y_{n-1}; t_n, dy_n). \tag{8}$$

Так построенный процесс X называется *марковским процессом* с начальным распределением π и системой переходных вероятностей $\{p(s, x; t, B)\}$.

Следствие 3. Пусть $T = \{0, 1, 2, \ldots\}$ и $\{P_k(x; B)\}$ — семейство неотрицательных функций, определенных для $k \geqslant 1$, $x \in R$, $B \in \mathscr{B}(R)$ и таких, что функция $p_k(x; B)$ есть вероятностная мера по B (при фиксированных k и x) и измерима по x (при фиксированных k и B). Пусть, кроме того, $\pi = \pi(B)$ — вероятностная мера на $(R, \mathscr{B}(R))$.

Тогда можно построить вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ с семейством случайных величин $X = \{\xi_0, \xi_1, \ldots\}$ на нем таких, что

$$\mathsf{P}\{\xi_0 \leqslant x_0,\ \xi_1 \leqslant x_1,\ \ldots\} = \\ = \int_{-\infty}^{x_0} \pi(dy_0) \int_{-\infty}^{x_1} p_1(y_0;\ dy_1) \ldots \int_{-\infty}^{x_n} p_n(y_{n-1};\ dy_n).$$

3. В соответствии со следствием 1 существует последовательность независимых случайных величин $\xi_1, \xi_2, \ldots$, одномерные функции распределения которых есть соответственно $F_1, F_2, \ldots$

Пусть теперь $(E_1, \mathscr{E}_1)$, $(E_2, \mathscr{E}_2)$, ... — полные сепарабельные метрические пространства и $P_1, P_2, \ldots$ — вероятностные меры на них. Тогда из замечания 2 следует, что существует вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и последовательность *независимых* элементов $X_1, X_2, \ldots$ таких, что X_n — $\mathscr{F}/\mathscr{E}_n$-измеримы, и $\mathsf{P}(X_n \in B) = P_n(B)$, $B \in \mathscr{E}_n$.

Оказывается, что этот результат остается справедливым и в том случае, когда пространства $(E_n, \mathscr{E}_n)$ являются *произвольными измеримыми пространствами.*

Теорема 2 (теорема Ионеску-Тулчи о продолжении меры и существовании случайной последовательности). *Пусть $(\Omega_n, \mathscr{F}_n)$, $n = 1, 2, \ldots,$ — произвольные измеримые пространства и $\Omega = \prod \Omega_n$, $\mathscr{F} = \bigotimes \mathscr{F}_n$. Предположим, что на $(\Omega_1, \mathscr{F}_1)$ задана вероятностная мера P_1 и для каждого набора $(\omega_1, \ldots, \omega_n) \in \Omega_1 \times \ldots \times \Omega_n$, $n \geqslant 1$, на $(\Omega_{n+1}, \mathscr{F}_{n+1})$ заданы вероятностные меры $P(\omega_1, \ldots, \omega_n; \cdot)$. Будем предполагать, что $P(\omega_1, \ldots, \omega_n; B)$ для каждого $B \in \mathscr{F}_{n+1}$ являются борелевскими функциями от $(\omega_1, \ldots, \omega_n)$, и пусть*

$$P_n(A_1 \times \ldots \times A_n) = \\ = \int_{A_1} P_1(d\omega_1) \int_{A_2} P(\omega_1;\ d\omega_2) \ldots \int_{A_n} P(\omega_1, \ldots, \omega_{n-1};\ d\omega_n), \tag{9}$$

$$A_i \in \mathscr{F}_i, \quad n \geqslant 1.$$

Тогда на $(\Omega, \mathscr{F})$ существуют единственная вероятностная мера P такая, что для любого $n \geqslant 1$

$$\mathsf{P}\{\omega\colon \omega_1 \in A_1, \ldots, \omega_n \in A_n\} = P_n(A_1 \times \ldots \times A_n), \tag{10}$$

и случайная последовательность $X = (X_1(\omega), X_2(\omega), \ldots)$ такая, что

$$\mathsf{P}\{\omega\colon X_1(\omega) \in A_1 \ldots, X_n(\omega) \in A_n\} = P_n(A_1 \times \ldots \times A_n), \tag{11}$$

где $A_i \in \mathscr{E}_i$.

Доказательство. Первый шаг в доказательстве состоит в установлении того, что для каждого $n > 1$ функцию множеств P_n, заданную на прямоугольниках $A_1 \times \ldots \times A_n$ с помощью равенства (9), можно продолжить на σ-алгебру $\mathscr{F}_1 \otimes \ldots \otimes \mathscr{F}_n$.

С этой целью для каждого $n \geqslant 2$ и $B \in \mathscr{F}_1 \otimes \ldots \otimes \mathscr{F}_n$ положим

$$P_n(B) = \int_{\Omega_1} P_1(d\omega_1) \int_{\Omega_2} P(\omega_1; d\omega_2) \int_{\Omega_{n-1}} P(\omega_1, \ldots, \omega_{n-2}; d\omega_{n-1}) \times \\ \times \int_{\Omega_n} I_B(\omega_1, \ldots, \omega_n) P(\omega_1, \ldots, \omega_{n-1}; d\omega_n). \quad (12)$$

Нетрудно видеть, что для $B = A_1 \times \ldots \times A_n$ правая часть в (12) совпадает с правой частью в (9). Кроме того, для $n = 2$, так же как и в теореме 8 § 6, устанавливается, что P_2 является мерой. Отсюда по индукции легко устанавливается, что P_n являются ме́рами для произвольного $n \geqslant 2$.

Следующий шаг в доказательстве такой же, как и в теореме Колмогорова о продолжении меры в $(R^\infty, \mathscr{B}(R^\infty))$ (теорема 3 § 3). А именно, для всякого цилиндрического множества $J_n(B) = \{\omega \in \Omega: (\omega_1, \ldots, \omega_n) \in B\}$, $B \in \mathscr{F}_1 \otimes \ldots \otimes \mathscr{F}_n$, определим функцию множеств P с помощью равенства

$$\mathsf{P}(J_n(B)) = P_n(B). \quad (13)$$

Используя (12) и то обстоятельство, что $P(\omega_1, \ldots, \omega_k; \cdot)$ являются мерами, нетрудно установить, что определение (13) корректно в том смысле, что значение $\mathsf{P}(J_n(B))$ не зависит от способа представления цилиндрического множества.

Отсюда вытекает, что функция множеств P, определенная в (13) для цилиндрических множеств и, очевидным образом, на алгебре, содержащей все цилиндрические множества, является на этой алгебре конечно-аддитивной мерой. Остается проверить ее счетную аддитивность на этой алгебре и затем воспользоваться теоремой Каратеодори.

В теореме 3 § 3 осуществление указанной проверки основывалось на том свойстве пространств $(R^n, \mathscr{B}(R^n))$, что для каждого борелевского множества B можно найти компакт $A \subseteq B$, вероятностная мера которого сколь угодно близка к мере множества B. В рассматриваемом случае этот момент доказательства видоизменяется следующим образом.

Пусть, как и в теореме 3 § 3, $\{\hat{B}_n\}_{n \geqslant 1}$ — последовательность цилиндрических множеств

$$\hat{B}_n = \{\omega: (\omega_1, \ldots, \omega_n) \in B_n\},$$

убывающих к пустому множеству $\varnothing$, но

$$\lim_{n \to \infty} \mathsf{P}(\hat{B}_n) > 0. \quad (14)$$

Из (12) для $n > 1$

$$\mathsf{P}(\hat{B}_n) = \int_{\Omega_1} f_n^{(1)}(\omega_1) P_1(d\omega_1),$$

где

$$f_n^{(1)}(\omega_1)=\int_{\Omega_2} P(\omega_1;\ d\omega_2)\ldots\int_{\Omega_n} I_{B_n}(\omega_1,\ \ldots,\ \omega_n)\,P(\omega_2,\ \ldots,\ \omega_{n-1};\ d\omega_n).$$

Поскольку $\hat{B}_{n+1}\subseteq\hat{B}_n$, то $B_{n+1}\subseteq B_n\times\Omega_{n+1}$ и, значит, $I_{B_{n+1}}(\omega_1,\ \ldots,\ \omega_{n+1})\leqslant I_{B_n}(\omega_1,\ \ldots,\ \omega_n)\,I_{\Omega_{n+1}}(\omega_{n+1})$. Поэтому последовательность функций $\{f_n^{(1)}(\omega_1)\}_{n\geqslant 1}$ является убывающей. Пусть $f^{(1)}(\omega_1)=\lim_n f_n^{(1)}(\omega_1)$. Тогда по теореме о мажорируемой сходимости

$$\lim_n \mathsf{P}(\hat{B}_n)=\lim_n\int_{\Omega_1} f_n^{(1)}(\omega_1)\,P_1(d\omega_1)=\int_{\Omega_1} f^{(1)}(\omega_1)\,P_1(d\omega_1).$$

По предположению $\lim_n \mathsf{P}(\hat{B}_n)>0$. Отсюда следует, что найдется такое $\omega_1^0\in B$, что $f^{(1)}(\omega_1^0)>0$, поскольку, если точка $\omega_1\notin B_1$, то $f_n^{(1)}(\omega_1)=0$ для всех $n\geqslant 1$.

Далее, для $n>2$

$$f_n^{(1)}(\omega_1^0)=\int_{\Omega_2} f_n^{(2)}(\omega_2)\,P(\omega_1^0;\ d\omega_2), \tag{15}$$

где

$$f_n^{(2)}(\omega_2)=\int_{\Omega_3} P(\omega_1^0,\ \omega_2;\ d\omega_3)\ldots$$
$$\ldots\int_{\Omega_n} I_{B_n}(\omega_1^0,\ \omega_2,\ \ldots,\ \omega_n)\,P(\omega_1^0,\ \omega_2,\ \ldots,\ \omega_{n-1},\ d\omega_n).$$

Как и в случае последовательности $\{f_n^{(1)}(\omega_1)\}$, устанавливается, что последовательность $\{f_n^{(2)}(\omega_2)\}$ является убывающей. Пусть $f^{(2)}(\omega_2)=\lim_{n\to\infty} f_n^{(2)}(\omega_2)$. Тогда из (15) следует, что

$$0<f^{(1)}(\omega_1^0)=\int_{\Omega_2} f^{(2)}(\omega_2)\,P(\omega_1^0;\ d\omega_2),$$

и найдется такая точка $\omega_2^0\in\Omega_2$, что $f^{(2)}(\omega_2^0)>0$. При этом $(\omega_1^0,\ \omega_2^0)\in B_2$. Продолжая указанный процесс, получим, что для любого n найдется точка $(\omega_1^0,\ \ldots,\ \omega_n^0)\in B_n$. Следовательно, точка $(\omega_1^0,\ \ldots,\ \omega_n^0,\ \ldots)\in\bigcap\hat{B}_n$, но в то же время, по предположению, $\bigcap\hat{B}_n=\varnothing$. Полученное противоречие показывает, что $\lim_n \mathsf{P}(\hat{B}_n)=0$.

Итак, утверждение теоремы в части, касающейся существования вероятностной меры P, доказано. Заключительная часть очевидным образом следует из предыдущей, если положить $X_n(\omega)=\omega_n$, $n\geqslant 1$.

Следствие 1. Пусть $(E_n,\ \mathscr{E}_n)_{n\geqslant 1}$ — произвольные измеримые пространства и $(P_n)_{n\geqslant 1}$ — вероятностные меры на них. Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и семейство независимых случайных элементов $X_1, X_2, \ldots$ со значениями в $(E_1, \mathscr{E}_1)$, $(E_2, \mathscr{E}_2), \ldots$, соответственно такие, что

$$\mathsf{P}\{\omega\colon X_n(\omega) \in B\} = P_n(B), \quad B \in \mathscr{E}_n, \ n \geqslant 1.$$

Следствие 2. Пусть $E = \{1, 2, \ldots\}$, $\{p_k(x; y)\}$ — семейство неотрицательных функций, $k \geqslant 1$, $x, y \in E$, таких, что $\sum\limits_{y \in E} p_k(x; y) = 1$, $x \in E$, $k \geqslant 1$. Пусть, кроме того, $\pi = \pi(x)$ распределение вероятностей на E $(\pi(x) \geqslant 0, \ \sum\limits_{x \in E} \pi(x) = 1)$.

Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и семейство случайных величин $X = \{\xi_0, \xi_1, \ldots\}$ на нем такие, что

$$\mathsf{P}\{\xi_0 = x_0, \xi_1 = x_1, \ldots, \xi_n = x_n\} = \\ = \pi(x_0)\, p_1(x_0, x_1) \ldots p_n(x_{n-1}, x_n) \tag{16}$$

(ср. с (I.12.4)) для всех $x_i \in E$ и $n \geqslant 1$. В качестве Ω можно взять пространство

$$\Omega = \{\omega\colon \omega = (x_0, x_1, \ldots), \ x_i \in E\}.$$

Последовательность случайных величин $X = \{\xi_0, \xi_1, \ldots\}$, удовлетворяющих условию (16), называют *марковской цепью* со счетным множеством состояний E, с матрицами переходных вероятностей $\{p_k(x, y)\}$ и начальным распределением вероятностей π. (Ср. с определением в § 12 гл. I.)

4. Задачи.

1. Пусть $\Omega = [0, 1]$, $\mathscr{F}$ — класс борелевских множеств на $[0, 1]$, P — мера Лебега на $[0, 1]$. Показать, что пространство $(\Omega, \mathscr{F}, \mathsf{P})$ является универсальным в том смысле, что для любой функции распределения $F(x)$ на $(\Omega, \mathscr{F}, \mathsf{P})$ можно так определить случайную величину $\xi = \xi(\omega)$, что ее функция распределения $F_\xi(x) = \mathsf{P}(\xi \leqslant x)$ совпадает с функцией $F(x)$. (Указание. $\xi(\omega) = F^{-1}(\omega)$, $0 < \omega < 1$, где $F^{-1}(\omega) = \sup\{x\colon F(x) < \omega\}$, когда $0 < \omega < 1$, а $\xi(0)$, $\xi(1)$ могут быть взяты произвольными.)

2. Проверить согласованность семейств распределений в следствиях к теоремам 1 и 2.

3. Вывести утверждение следствия 2 к теореме 2 из теоремы 1.

§ 10. Разные виды сходимости последовательностей случайных величин

1. Как и в математическом анализе, в теории вероятностей приходится иметь дело с разными видами сходимости случайных величин. Ниже будут рассмотрены следующие основные виды сходимости: *по вероятности, с вероятностью единица, в среднем порядка p, по распределению.*

Начнем с определений. Пусть $\xi, \xi_1, \xi_2, \ldots$ — случайные величины, заданные на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$.

Определение 1. Последовательность случайных величин $\xi_1, \xi_2, \ldots$ называется сходящейся *по вероятности* к случайной величине ξ (обозначение: $\xi_n \xrightarrow{\mathsf{P}} \xi$), если для любого $\varepsilon > 0$

$$\mathsf{P}\{|\xi_n - \xi| > \varepsilon\} \to 0, \quad n \to \infty. \tag{1}$$

С этим видом сходимости мы уже встречались в связи с законом больших чисел в схеме Бернулли, утверждающему, что

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \to 0, \quad n \to \infty$$

(см. обозначения в § 5 гл. I). В анализе этот вид сходимости принято называть сходимостью по мере.

Определение 2. Последовательность случайных величин $\xi_1, \xi_2, \ldots$ называется сходящейся *с вероятностью единица* (*почти наверное, почти всюду*) к случайной величине ξ, если

$$\mathsf{P}\{\omega\colon \xi_n \not\to \xi\} = 0, \tag{2}$$

т. е. если множество исходов ω, для которых $\xi_n(\omega)$ не сходятся к $\xi(\omega)$, имеет нулевую вероятность.

Этот вид сходимости обозначают следующим образом: $\xi_n \to \xi$ (P-п. н.), или $\xi_n \xrightarrow{\text{п. н.}} \xi$, или $\xi_n \xrightarrow{\text{п.в.}} \xi$.

Определение 3. Последовательность случайных величин $\xi_1, \xi_2, \ldots$ называется сходящейся *в среднем порядка* p, $0 < p < \infty$, к случайной величине ξ, если

$$\mathsf{M}|\xi_n - \xi|^p \to 0, \quad n \to \infty. \tag{3}$$

В анализе этот вид сходимости называют *сходимостью в смысле* L^p. В этой связи (3) обычно записывают в виде $\xi_n \xrightarrow{L^p} \xi$. В частном случае $p = 2$ эту сходимость называют также сходимостью *в среднем квадратическом* и пишут $\xi = \text{l. i. m.}\, \xi_n$ (l. i. m. — сокращение от limit in mean — сходимость в среднем).

Определение 4. Последовательность случайных величин $\xi_1, \xi_2, \ldots$ называется сходящейся *по распределению* к случайной величине ξ (обозначение: $\xi_n \xrightarrow{d} \xi$), если для любой ограниченной непрерывной функции $f = f(x)$

$$\mathsf{M}f(\xi_n) \to \mathsf{M}f(\xi), \quad n \to \infty. \tag{4}$$

Наименование этого вида сходимости объясняется тем, что, как будет показано в § 1 гл. III условие (4) эквивалентно сходимости функций распределения $F_{\xi_n}(x)$ к функции распределения

$F_\xi(x)$ в каждой точке x, где функция $F_\xi(x)$ *непрерывна.* Эту сходимость обозначают $F_{\xi_n} \Rightarrow F_\xi$.

Подчеркнем, что сходимость по распределению случайных величин определяется только в терминах сходимости их функций распределения. Поэтому об этом виде сходимости имеет смысл говорить и тогда, когда случайные величины заданы на разных вероятностных пространствах. Этот вид сходимости будет подробно изучаться в гл. III, где, в частности, будет объяснено, почему в определении сходимости $F_{\xi_n} \Rightarrow F_\xi$ требуется сходимость лишь в точках непрерывности функции $F_\xi(x)$, а не для всех x.

2. В математическом анализе для решения вопроса о сходимости (в том или ином смысле) заданной последовательности функций оказывается полезным понятие фундаментальной последовательности, или последовательности Коши. Введем аналогичные понятия для первых трех рассмотренных видов сходимости последовательностей случайных величин.

Будем говорить, что последовательность случайных величин $\{\xi_n\}_{n\geqslant 1}$ *фундаментальна по вероятности, с вероятностью единица* и *в среднем порядка* p, $0<p<\infty$, если выполнены соответственно следующие условия: для любого $\varepsilon>0$ $\mathsf{P}\{|\xi_n-\xi_m|<\varepsilon\}\to 0$, $n, m\to\infty$, последовательность $\{\xi_n(\omega)\}_{n\geqslant 1}$ фундаментальна для почти всех $\omega\in\Omega$, последовательность функций $\{\xi_n(\omega)\}_{n\geqslant 1}$ фундаментальна в смысле L^p, т. е. $\mathsf{M}|\xi_n-\xi_m|^p\to 0$, $n, m\to\infty$.

3. Теорема 1. а) *Для того чтобы $\xi_n\to\xi$ (P-п. н.), необходимо и достаточно, чтобы для любого $\varepsilon>0$*

$$\mathsf{P}\left\{\sup_{k\geqslant n}|\xi_k-\xi|\geqslant\varepsilon\right\}\to 0, \qquad n\to\infty. \tag{5}$$

b) *Последовательность $\{\xi_n\}_{n\geqslant 1}$ фундаментальна с вероятностью единица тогда и только тогда, когда для любого $\varepsilon>0$*

$$\mathsf{P}\left\{\sup_{\substack{k\geqslant n\\ l\geqslant n}}|\xi_k-\xi_l|\leqslant\varepsilon\right\}\to 0, \qquad n\to\infty, \tag{6}$$

или, что эквивалентно,

$$\mathsf{P}\left\{\sup_{k\geqslant 0}|\xi_{n+k}-\xi_n|\geqslant\varepsilon\right\}\to 0, \qquad n\to\infty. \tag{7}$$

Доказательство. а) Пусть $A_n^\varepsilon=\{\omega\colon |\xi_n-\xi|\geqslant\varepsilon\}$, $A^\varepsilon = \overline{\lim} A_n^\varepsilon \equiv \bigcap_{n=1}^\infty \bigcup_{k\geqslant n} A_k^\varepsilon$. Тогда

$$\{\omega\colon \xi_n \not\to \xi\} = \bigcup_{\varepsilon>0} A^\varepsilon = \bigcup_{m=1}^\infty A^{1/m}.$$

Но

$$\mathsf{P}(A^\varepsilon) = \lim_n \mathsf{P}\left(\bigcup_{k \geqslant n} A_k^\varepsilon\right),$$

поэтому утверждение а) является результатом следующей цепочки импликаций:

$$0 = \mathsf{P}\{\omega\colon \xi_n \not\to \xi\} = \mathsf{P}\left(\bigcup_{\varepsilon > 0} A^\varepsilon\right) \Leftrightarrow \mathsf{P}\left(\bigcup_{m=1}^{\infty} A^{1/m}\right) = 0 \Leftrightarrow$$
$$\Leftrightarrow \mathsf{P}(A^{1/m}) = 0,\ m \geqslant 1 \Leftrightarrow \mathsf{P}(A^\varepsilon) = 0,\ \varepsilon > 0 \Leftrightarrow$$
$$\Leftrightarrow \mathsf{P}\left(\bigcup_{k \geqslant n} A_k^\varepsilon\right) \to 0,\ n \to \infty \Leftrightarrow \mathsf{P}\left(\sup_{k \geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right) \to 0,\ n \to \infty.$$

b) Обозначим $B_{k,l}^\varepsilon = \{\omega\colon\ |\xi_k - \xi_l| \geqslant \varepsilon\}$, $B^\varepsilon = \bigcap_{n=1}^{\infty} \bigcup_{\substack{k \geqslant n \\ l \geqslant n}} B_{k,l}^\varepsilon$.

Тогда $\{\omega\colon\ \{\xi_n(\omega)\}_{n \geqslant 1}$ *не фундаментальна*$\} = \bigcup_{\varepsilon > 0} B^\varepsilon$, и так же, как в а), показывается, что $\mathsf{P}\{\omega\colon\ \{\xi_n(\omega)\}_{n \geqslant 1}$ *не фундаментальна*$\} = 0 \Leftrightarrow (6)$. Эквивалентность же утверждений (6) и (7) следует из очевидных неравенств

$$\sup_{k \geqslant 0} |\xi_{n+k} - \xi_n| \leqslant \sup_{\substack{k \geqslant 0 \\ l \geqslant 0}} |\xi_{n+k} - \xi_{n+l}| \leqslant 2 \sup_{k \geqslant 0} |\xi_{n+k} - \xi_n|.$$

Теорема доказана.

Следствие. Поскольку

$$\mathsf{P}\left\{\sup_{k \geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right\} = \mathsf{P}\left\{\bigcup_{k \geqslant n} (|\xi_k - \xi| \geqslant \varepsilon)\right\} \leqslant \sum_{k \geqslant n} \mathsf{P}\{|\xi_k - \xi| \geqslant \varepsilon\},$$

то выполнение для каждого $\varepsilon > 0$ условия

$$\sum_{k=1}^{\infty} \mathsf{P}\{|\xi_k - \xi| \geqslant \varepsilon\} < \infty \tag{8}$$

достаточно для сходимости $\xi_n \xrightarrow{\text{п. н.}} \xi$.

В связи с условием (8) уместно сейчас отметить, что положенные при его выводе рассуждения позволяют установить следующий простой, но важный результат, являющийся основным средством при исследовании свойств, выполняющихся с вероятностью единица.

Пусть $A_1, A_2, \ldots$ — некоторая последовательность событий из $\mathscr{F}$. Напомним (см. табл. в § 1), что через $\{A_n \text{б.ч.}\}$ обозначается событие $\overline{\lim}\, A_n$, состоящее в том, что произойдет бесконечно много событий из $A_1, A_2, \ldots$,

Лемма Бореля—Кантелли.

a) *Если* $\sum \mathsf{P}(A_n) < \infty$, *то* $\mathsf{P}\{A_n \text{ б.ч.}\} = 0$.

b) *Если* $\sum \mathsf{P}(A_n) = \infty$ *и события* $A_1, A_2, \ldots$ *независимы, то* $\mathsf{P}\{A_n \text{ б.ч.}\} = 1$.

Доказательство. a) По определению

$$\{A_n \text{ б. ч}\} = \overline{\lim} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k.$$

Поэтому

$$\mathsf{P}\{A_n \text{ б. ч.}\} = \mathsf{P}\left\{\bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k\right\} = \lim \mathsf{P}\left(\bigcup_{k \geqslant n} A_k\right) \leqslant \lim \sum_{k \geqslant n} \mathsf{P}(A_k),$$

откуда и следует утверждение a).

b) Если события $A_1, A_2, \ldots$ независимы, то таковыми же будут и события $\bar{A}_1, \bar{A}_2, \ldots$. Тогда для любого $N \geqslant n$

$$\mathsf{P}\left(\bigcap_{k=n}^{N} \bar{A}_k\right) = \prod_{k=n}^{N} \mathsf{P}(\bar{A}_k),$$

откуда нетрудно вывести, что

$$\mathsf{P}\left(\bigcap_{k=n}^{\infty} \bar{A}_k\right) = \prod_{k=n}^{\infty} \mathsf{P}(\bar{A}_k). \tag{9}$$

В силу неравенства $\log(1-x) \leqslant -x$, $0 \leqslant x < 1$,

$$\log \prod_{k=n}^{\infty} [1 - \mathsf{P}(A_k)] = \sum_{k=n}^{\infty} \log[1 - \mathsf{P}(A_k)] \leqslant -\sum_{k=n}^{\infty} \mathsf{P}(A_k) = -\infty.$$

Следовательно, для любого n

$$\mathsf{P}\left(\bigcap_{k=n}^{\infty} \bar{A}_k\right) = 0$$

и, значит, $\mathsf{P}(A_n \text{ б. ч.}) = 1$.

Лемма доказана.

Следствие 1. Если $A_n^\varepsilon = \{\omega\colon |\xi_n - \xi| \leqslant \varepsilon\}$, то условие (8) означает, что $\sum_{n=1}^{\infty} \mathsf{P}(A_n^\varepsilon) < \infty$, $\varepsilon > 0$, и по лемме Бореля — Кантелли $\mathsf{P}(A^\varepsilon) = 0$, $\varepsilon > 0$, где $A^\varepsilon = \overline{\lim} A_n^\varepsilon$. Тем самым

$$\sum \mathsf{P}\{|\xi_k - \xi| \geqslant \varepsilon\} < \infty,\ \varepsilon > 0 \Rightarrow \mathsf{P}(A^\varepsilon) = 0,$$
$$\varepsilon > 0 \Leftrightarrow \mathsf{P}\{\omega\colon \xi_n \nrightarrow \xi) = 0,$$

что уже отмечалось выше.

Следствие 2. Пусть $\{\varepsilon_n\}_{n\geqslant 1}$ — последовательность положительных чисел таких, что $\varepsilon_n \downarrow 0$, $n\to\infty$. Тогда, если

$$\sum_{n=1}^{\infty} \mathsf{P}\{|\xi_n-\xi|\geqslant\varepsilon_n\}<\infty, \tag{10}$$

то $\xi_n \xrightarrow{\text{п. н.}} \xi$.

В самом деле, пусть $A_n=\{|\xi_n-\xi|\geqslant\varepsilon_n\}$. Тогда по лемме Бореля — Кантелли $\mathsf{P}(A_n \text{ б. ч.})=0$. А это означает, что для почти каждого исхода $\omega\in\Omega$ найдется такое $N=N(\omega)$, что для $n\geqslant N(\omega)$ $|\xi_n(\omega)-\xi(\omega)|\leqslant\varepsilon_n$. Но $\varepsilon_n\downarrow 0$, поэтому $\xi_n(\omega)\to\xi(\omega)$ для почти всех $\omega\in\Omega$.

4. Теорема 2. *Имеют место следующие импликации:*

$$\xi_n \xrightarrow{\text{п. н.}} \xi \Rightarrow \xi_n \xrightarrow{\mathsf{P}} \xi, \tag{11}$$

$$\xi_n \xrightarrow{L^p} \xi \Rightarrow \xi_n \xrightarrow{\mathsf{P}} \xi, \quad p>0, \tag{12}$$

$$\xi_n \xrightarrow{\mathsf{P}} \xi \Rightarrow \xi_n \xrightarrow{d} \xi. \tag{13}$$

Доказательство. Утверждение (11) следует из сравнения определения сходимости по вероятности с критерием (5), а импликация (12) — из неравенства Чебышева.

Для доказательства (13) пусть $|f(x)|\leqslant c$, $\varepsilon>0$ и N таково, что $\mathsf{P}(|\xi|>N)\leqslant\varepsilon/4c$. Выберем δ таким, чтобы для всех $|x|\leqslant N$ и $|x-y|\leqslant\delta$ было выполнено неравенство $|f(x)-f(y)|\leqslant\varepsilon/4c$. Тогда (ср. с доказательством теоремы Вейерштрасса в п. 5 § 5 гл. I)

$$\begin{aligned}\mathsf{M}|f(\xi_n)-f(\xi)| &= \mathsf{M}(|f(\xi_n)-f(\xi)|;\ |\xi_n-\xi|\leqslant\delta,\ |\xi|\leqslant N)+\\ &+\mathsf{M}(|f(\xi_n)-f(\xi)|;\ |\xi_n-\xi|\leqslant\delta,\ |\xi|>N)+\\ &+\mathsf{M}(|f(\xi_n)-f(\xi)|;\ |\xi_n-\xi|>\delta)\leqslant\\ &\leqslant\varepsilon/2+\varepsilon/2+2c\mathsf{P}\{|\xi_n-\xi|>\delta\}=\varepsilon+2c\mathsf{P}\{|\xi_n-\xi|>\delta\}.\end{aligned}$$

Но $\mathsf{P}\{|\xi_n-\xi|>\delta\}\to 0$, поэтому для достаточно больших n $\mathsf{M}|f(\xi_n)-f(\xi)|\leqslant 2\varepsilon$, что в силу произвольности $\varepsilon>0$ доказывает импликацию (13).

Теорема доказана.

Приведем ряд примеров, показывающих, в частности, что в (11), (12) обратные импликации, вообще говоря, несправедливы.

Пример 1. ($\xi_n \xrightarrow{\mathsf{P}} \xi \not\Rightarrow \xi_n \xrightarrow{\text{п. н}} \xi$; $\xi_n \xrightarrow{L^p} \xi \not\Rightarrow \xi_n \xrightarrow{\text{п. н}} \xi$). Пусть $\Omega=[0, 1]$, $\mathscr{F}=\mathscr{B}([0, 1])$, P — мера Лебега. Положим

$$A_n^i=\left[\frac{i-1}{n}, \frac{i}{n}\right],\quad \xi_n^i=I_{A_n^i}(\omega),\qquad i=1, 2, \ldots, n;\quad n\geqslant 1.$$

Тогда последовательность случайных величин

$$\{\xi_1^1;\ \xi_2^1,\ \xi_2^2;\ \xi_3^1,\ \xi_3^2,\ \xi_3^3;\ \ldots\}$$

сходится и по вероятности, и в среднем порядка $p>0$, но не сходится ни в одной точке $\omega\in[0, 1]$.

Пример 2. $\left(\xi_n \xrightarrow{\text{п. н.}} \xi \Rightarrow \xi_n \xrightarrow{\mathsf{P}} \xi \nRightarrow \xi_n \xrightarrow{L^p} \xi,\ p>0\right)$. Снова пусть $\Omega=[0, 1]$, $\mathscr{F}=\mathscr{B}([0, 1])$, P — мера Лебега и

$$\xi_n(\omega)=\begin{cases} e^n, & 0\leqslant\omega\leqslant 1/n,\\ 0, & \omega>1/n.\end{cases}$$

Тогда последовательность $\{\xi_n\}$ сходится с вероятностью единица (и, следовательно, по вероятности) к нулю, однако для любого $p>0$

$$\mathsf{M}|\xi_n|^p=\frac{e^{np}}{n}\to\infty,\quad n\to\infty.$$

Пример 3. $\left(\xi_n \xrightarrow{L^p} \xi \nRightarrow \xi_n \xrightarrow{\text{п. н.}} \xi\right)$. Пусть $\{\xi_n\}$ — последовательность независимых случайных величин с

$$\mathsf{P}(\xi_n=1)=p_n,\quad \mathsf{P}(\xi_n=0)=1-p_n.$$

Тогда нетрудно установить, что

$$\xi_n \xrightarrow{\mathsf{P}} 0 \Leftrightarrow p_n\to 0,\quad n\to\infty, \tag{14}$$

$$\xi_n \xrightarrow{L^p} 0 \Leftrightarrow p_n\to 0,\quad n\to\infty, \tag{15}$$

$$\xi_n \xrightarrow{\text{п. н.}} 0 \Leftrightarrow \sum_{n=1}^{\infty} p_n<\infty. \tag{16}$$

В частности, при $p_n=1/n$ $\xi_n \xrightarrow{L^p} 0$ для любого $p>0$, но $\xi_n \overset{\text{п. н.}}{\not\to} 0$.

В следующей теореме выделяется один интересный случай, когда из сходимости почти наверное следует сходимость в смысле L^1.

Теорема 3. *Пусть $\{\xi_n\}$ — последовательность неотрицательных случайных величин таких, что $\xi_n \xrightarrow{\text{п. н.}} \xi$ и $\mathsf{M}\xi_n\to\mathsf{M}\xi<\infty$. Тогда*

$$\mathsf{M}|\xi_n-\xi|\to 0,\qquad n\to\infty. \tag{17}$$

Доказательство. Для достаточно больших n $\mathsf{M}\xi_n<\infty$, поэтому для них

$$\mathsf{M}|\xi-\xi_n|=\mathsf{M}(\xi-\xi_n)I_{\{\xi\geqslant\xi_n\}}+\mathsf{M}(\xi_n-\xi)I_{\{\xi_n>\xi\}}=$$
$$=2\mathsf{M}(\xi-\xi_n)I_{\{\xi\geqslant\xi_n\}}+\mathsf{M}(\xi_n-\xi).$$

Но $0\leqslant(\xi-\xi_n)I_{\{\xi\geqslant\xi_n\}}\leqslant\xi$. Поэтому по теореме о мажорируемой сходимости $\lim\limits_n(\xi-\xi_n)I_{\{\xi\geqslant\xi_n\}}=0$, что вместе с предположением $\mathsf{M}\xi_n\to\mathsf{M}\xi$ доказывает (17).

Замечание. Теорема о мажорируемой сходимости справедлива и тогда, когда в ней сходимость почти наверное заменяется на сходимость по вероятности (см. задачу 1). Поэтому в теореме 3 сходимость «$\xi_n \xrightarrow{\text{п. н.}} \xi$» можно заменить на сходимость «$\xi_n \xrightarrow{\mathsf{P}} \xi$».

5. Из математического анализа известно, что всякая фундаментальная числовая последовательность $\{x_n\}$, $x_n \in R$, является сходящейся (критерий Коши). Приведем аналогичные результаты для сходимости последовательности случайных величин.

Теорема. 4 (критерий Коши сходимости почти наверное). *Для того чтобы последовательность случайных величин $\{\xi_n\}_{n\geqslant 1}$ была сходящейся с вероятностью единица (к некоторой случайной величине ξ), необходимо и достаточно, чтобы она была фундаментальна с вероятностью единица.*

Доказательство. Если $\xi_n \xrightarrow{\text{п. н.}} \xi$, то

$$\sup_{\substack{k\geqslant n\\ l\geqslant n}} |\xi_k - \xi_l| \leqslant \sup_{k\geqslant n} |\xi_k - \xi| + \sup_{l\geqslant n} |\xi_l - \xi|,$$

откуда вытекает необходимость условия теоремы.

Пусть теперь последовательность $\{\xi_n\}_{n\geqslant 1}$ фундаментальна с вероятностью единица. Обозначим $\mathscr{N} = \{\omega\colon \{\xi_n(\omega)\}$ *не фундаментальная*$\}$. Тогда для всех $\omega \in \Omega \setminus \mathscr{N}$ числовая последовательность $\{\xi_n(\omega)\}_{n\geqslant 1}$ является фундаментальной и, согласно критерию Коши для числовых последовательностей, существует $\lim \xi_n(\omega)$. Положим

$$\xi(\omega) = \begin{cases} \lim \xi_n(\omega), & \omega \in \Omega \setminus \mathscr{N}, \\ 0, & \omega \in \mathscr{N}. \end{cases} \tag{18}$$

Так определенная функция является случайной величиной и, очевидно, $\xi_n \xrightarrow{\text{п. н.}} \xi$.

Теорема доказана.

Прежде чем переходить к случаю сходимости по вероятности, установим следующий полезный результат.

Теорема 5. *Если последовательность $\{\xi_n\}$ фундаментальна (сходится) по вероятности, то из нее можно извлечь подпоследовательность $\{\xi_{n_k}\}$, фундаментальную (сходящуюся) с вероятностью единица.*

Доказательство. Пусть последовательность $\{\xi_n\}$ фундаментальна по вероятности. В силу теоремы 4 достаточно доказать, что из нее можно извлечь подпоследовательность, сходящуюся почти наверное.

Положим $n_1 = 1$ и по индукции определим n_k, как то наименьшее $n > n_{k-1}$, для которого при всех $s \geqslant n$, $t \geqslant n$

$$\mathsf{P}\{|\xi_t - \xi_s| > 2^{-k}\} < 2^{-k}.$$

Тогда

$$\sum_k \mathsf{P}\{|\xi_{n_{k+1}} - \xi_{n_k}| > 2^{-k}\} < \sum 2^{-k} < \infty$$

и по лемме Бореля — Кантелли

$$\mathsf{P}\{|\xi_{n_{k+1}}-\xi_{n_k}|>2^{-k}\text{ б.ч.}\}=0.$$

Поэтому с вероятностью единица

$$\sum_{k=1}^{\infty}|\xi_{n_{k+1}}-\xi_{n_k}|<\infty.$$

Пусть $\mathscr{N}=\left\{\omega:\ \sum|\xi_{n_{k+1}}-\xi_{n_k}|=\infty\right\}$. Тогда, если положить

$$\xi(\omega)=\begin{cases}\xi_{n_1}(\omega)+\sum\limits_{k=1}^{\infty}\left(\xi_{n_{k+1}}^{(\omega)}-\xi_{n_k}(\omega)\right), & \omega\in\Omega\setminus\mathscr{N},\\ 0, & \omega\in\mathscr{N},\end{cases}$$

то получим $\xi_{n_k}\xrightarrow{\text{п. н.}}\xi$.

Если же исходная последовательность сходится по вероятности, то она и фундаментальна по вероятности (см. далее (19)) и, следовательно, этот случай сводится к уже разобранному.

Теорема доказана.

Теорема 6 (критерий Коши сходимости по вероятности). *Для того чтобы последовательность случайных величин* $\{\xi_n\}_{n\geqslant 1}$ *была сходящейся по вероятности, необходимо и достаточно, чтобы она была фундаментальна по вероятности.*

Доказательство. Если $\xi_n\xrightarrow{\mathsf{P}}\xi$, то

$$\mathsf{P}\{|\xi_n-\xi_m|\geqslant\varepsilon\}\leqslant\mathsf{P}\{|\xi_n-\xi|\geqslant\varepsilon/2\}+\mathsf{P}\{|\xi_m-\xi|\geqslant\varepsilon/2\}\qquad(19)$$

и, следовательно, последовательность $\{\xi_n\}$ фундаментальна по вероятности.

Обратно, если $\{\xi_n\}$ фундаментальна по вероятности, то тогда, согласно теореме 5, найдутся подпоследовательность $\{\xi_{n_k}\}$ и случайная величина ξ такие, что $\xi_{n_k}\xrightarrow{\text{п. н.}}\xi$. Но тогда

$$\mathsf{P}\{|\xi_n-\xi|\geqslant\varepsilon\}\leqslant\mathsf{P}\{|\xi_n-\xi_{n_k}|\geqslant\varepsilon/2\}+\mathsf{P}\{|\xi_{n_k}-\xi|\geqslant\varepsilon/2\},$$

откуда ясно, что $\xi_n\xrightarrow{\mathsf{P}}\xi$. Теорема доказана.

В связи со сходимостью в среднем порядка $p>0$ сделаем прежде всего несколько замечаний о пространствах L^p.

Будем обозначать через $L^p=L^p(\Omega,\ \mathscr{F},\ \mathsf{P})$ — пространство случайных величин $\xi=\xi(\omega)$ с $\mathsf{M}|\xi|^p\equiv\int_\Omega|\xi|^p\,d\mathsf{P}<\infty$. Предположим, что $p\geqslant 1$ и положим

$$\|\xi\|_p=(\mathsf{M}|\xi|^p)^{1/p}.$$

Ясно, что

$$\|\xi\|_p\geqslant 0,\qquad(20)$$

$$\|c\xi\|_p=|c|\,\|\xi\|_p,\ c\text{ — постоянная},\qquad(21)$$

и в силу неравенства Минковского (6.31)

$$\|\xi+\eta\|_p \leqslant \|\xi\|_p + \|\eta\|_p. \tag{22}$$

Таким образом, в соответствии с известными определениями функционального анализа функция $\|\cdot\|_p$, определенная на L^p и удовлетворяющая условиям (20) — (22), является (для $p \geqslant 1$) *полунормой*.

Чтобы она была и *нормой*, нужно еще выполнение свойства

$$\|\xi\|_p = 0 \Rightarrow \xi = 0. \tag{23}$$

Это свойство, конечно, не выполнено, поскольку, согласно свойству H (§ 6), можно лишь утверждать, что $\xi = 0$ почти наверное. Однако если под L^p понимать пространство, элементами которого являются не случайные величины ξ с $\mathsf{M}|\xi|^p < \infty$, а классы эквивалентных случайных величин (ξ эквивалентно η, если $\xi = \eta$ почти наверное), то $\|\cdot\|_p$ становится нормой, а L^p — нормированным линейным пространством. Если в каждом классе эквивалентных случайных величин выбрать по одному элементу, беря функцию, тождественно равную нулю, в качестве представителя в классе функций, ей эквивалентных, то полученное пространство (которое также обозначается L^p) будет уже линейным нормированным пространством функций (а не классом эквивалентности).

Один из важных результатов функционального анализа состоит в доказательстве того, что пространства L^p, $p \geqslant 1$, являются *полными*, т. е. всякая фундаментальная последовательность является сходящейся. Сформулируем и докажем этот результат на вероятностном языке.

Теорема 7 (критерий Коши сходимости в среднем порядка $p \geqslant 1$). *Для того чтобы последовательность случайных величин $\{\xi_n\}_{n\geqslant 1}$ из L^p сходилась в среднем порядка $p \geqslant 1$ к случайной величине, принадлежащей L^p, необходимо и достаточно, чтобы эта последовательность была фундаментальной в среднем порядка p.*

Доказательство. Необходимость следует из неравенства Минковского. Пусть $\{\xi_n\}$ — фундаментальна ($\|\xi_n - \xi_m\|_p \to 0$, $n, m \to \infty$). Как и в доказательстве теоремы 5, выберем подпоследовательность $\{\xi_{n_k}\}$ такую, что $\xi_{n_k} \xrightarrow{\text{п. н.}} \xi$, где ξ — некоторая случайная величина с $\|\xi\|_p < \infty$.

Положим $n_1 = 1$ и по индукции выберем n_k, как то наименьшее $n > n_{k-1}$, для которого при всех $s \geqslant n$, $t \geqslant n$

$$\|\xi_t - \xi_s\|_p < 2^{-2k}.$$

Обозначим

Тогда в силу неравенства Чебышева

$$\mathsf{P}(A_k) \leqslant \frac{\mathsf{M}\,|\xi_{n_{k+1}} - \xi_{n_k}|^r}{2^{-kr}} \leqslant \frac{2^{-2kr}}{2^{-kr}} = 2^{-kr} \leqslant 2^{-k}.$$

Так же как в теореме 5, отсюда выводится, что существует такая случайная величина ξ, что $\xi_{n_k} \xrightarrow{\text{п. н.}} \xi$.

Выведем отсюда, что $\|\xi_n - \xi\|_p \to 0$, $n \to \infty$. С этой целью зафиксируем $\varepsilon > 0$ и выберем $N = N(\varepsilon)$ таким, что $\|\xi_n - \xi_m\|_p^p < \varepsilon$ для всех $n \geqslant N$, $m \geqslant N$. Тогда для любого фиксированного $n \geqslant N$ в силу леммы Фату

$$\mathsf{M}\,|\xi_n - \xi|^p = \mathsf{M}\left\{\lim_{n_k\to\infty} |\xi_n - \xi_{n_k}|^p\right\} = \mathsf{M}\left\{\underline{\lim}_{n_k\to\infty} |\xi_n - \xi_{n_k}|^p\right\} \leqslant$$

$$\leqslant \underline{\lim}_{n_k\to\infty} \mathsf{M}\,|\xi_n - \xi_{n_k}|^p = \underline{\lim}_{n_k\to\infty} \|\xi_n - \xi_{n_k}\|_p^p \leqslant \varepsilon.$$

Следовательно, $\mathsf{M}\,|\xi_n - \xi|^p \to 0$, $n \to \infty$. Ясно также, что поскольку $\xi = (\xi - \xi_n) + \xi_n$, то в силу неравенства Минковского $\mathsf{M}\,|\xi|^p < \infty$.

Теорема доказана.

Замечание 1. В соответствии с терминологией функционального анализа полные нормированные линейные пространства называются *банаховскими* пространствами. Таким образом, пространства L^p, $p \geqslant 1$, являются банаховскими.

Замечание 2. Если $0 < p < 1$, то $\|\xi\|_p = (\mathsf{M}\,|\xi|^p)^{1/p}$ не удовлетворяет неравенству треугольника (22) и, следовательно, не является нормой. Тем не менее пространства (классов эквивалентности) L^p, $0 < p < 1$, являются *полными* относительно метрики $d(\xi, \eta) \equiv \mathsf{M}\,|\xi - \eta|^p$.

Замечание 3. Обозначим $L^\infty = L^\infty(\Omega, \mathscr{F}, \mathsf{P})$ пространство (классов эквивалентности) случайных величин $\xi = \xi(\omega)$, для которых $\|\xi\|_\infty < \infty$, где величина $\|\xi\|_\infty$, называемая существенным супремумом ξ, определяется формулой

$$\|\xi\|_\infty \equiv \operatorname{ess\,sup} |\xi| \equiv \inf\{0 \leqslant c \leqslant \infty\colon \mathsf{P}(|\xi| > c) = 0\}.$$

Функция $\|\cdot\|_\infty$ является нормой, и относительно этой нормы пространство L^∞ является полным.

6. Задачи.

1. Используя теорему 5, показать, что в теоремах 3 и 4 из § 6 сходимость почти наверное может быть заменена сходимостью по вероятности.

2. Доказать, что пространство L^∞ полно.

3. Показать, что если $\xi_n \xrightarrow{\mathsf{P}} \xi$ и в то же время $\xi_n \xrightarrow{\mathsf{P}} \eta$, то ξ и η эквивалентны ($\mathsf{P}(\xi \neq \eta) = 0$).

4. Пусть $\xi_n \xrightarrow{\mathsf{P}} \xi$, $\eta_n \xrightarrow{\mathsf{P}} \eta$ и случайные величины ξ и η эквивалентны. Показать, что для любого $\varepsilon > 0$

$$\mathsf{P}\{|\xi_n - \eta_n| \geqslant \varepsilon\} \to 0, \qquad n \to \infty.$$

5. Пусть $\xi_n \xrightarrow{\mathsf{P}} \xi$, $\eta_n \xrightarrow{\mathsf{P}} \eta$. Показать, что $a\xi_n + b\eta_n \xrightarrow{\mathsf{P}} a\xi + b\eta$ (a, b — постоянные), $|\xi_n| \xrightarrow{\mathsf{P}} |\xi|$, $\xi_n\eta_n \xrightarrow{\mathsf{P}} \xi\eta$.

6. Пусть $(\xi_n - \xi)^2 \xrightarrow{\mathsf{P}} 0$. Показать, что $\xi_n^2 \xrightarrow{\mathsf{P}} \xi^2$.

7. Показать, что если $\xi_n \xrightarrow{d} C$, где C — постоянная, то имеет место и сходимость по вероятности:

$$\xi_n \xrightarrow{d} C \Rightarrow \xi_n \xrightarrow{\mathsf{P}} C.$$

8. Пусть последовательность $\{\xi_n\}_{n \geqslant 1}$ такова, что для некоторого $p > 0$ $\sum_{n=1}^{\infty} \mathsf{M}|\xi_n|^p < \infty$. Показать, что $\xi_n \to 0$ (P-п. н.).

9. Пусть $\{\xi_n\}_{n \geqslant 1}$ — последовательность независимых одинаково распределенных случайных величин. Доказать, что

$$\mathsf{M}|\xi_1| < \infty \Leftrightarrow \sum_{n=1}^{\infty} \mathsf{P}\{|\xi_1| > \varepsilon \cdot n\} < \infty \Leftrightarrow \sum_{n=1}^{\infty} \mathsf{P}\left\{\left|\frac{\xi_n}{n}\right| > \varepsilon\right\} < \infty \Rightarrow$$

$$\Rightarrow \frac{\xi_n}{n} \to 0 \ (\mathsf{P}\text{-п. н.}).$$

10. Пусть $\{\xi_n\}_{n \geqslant 1}$ — некоторая последовательность случайных величин. Предположим, что существуют случайная величина ξ и подпоследовательность $\{n_k\}$ такие, что $\xi_{n_k} \to \xi$ (P-п. н.) и $\max\limits_{n_{k-1} < l \leqslant n_k} |\xi_l - \xi_{n_{k-1}}| \to 0$ (P-п. н.) при $k \to \infty$. Показать, что тогда $\xi_n \to \xi$ (P-п. н.).

11. Определим d-метрику во множестве случайных величин, полагая

$$d(\xi, \eta) = \mathsf{M}\frac{|\xi - \eta|}{1 + |\xi - \eta|}$$

и отождествляя случайные величины, совпадающие почти наверное. Показать, что сходимость по вероятности эквивалентна сходимости в d-метрике.

12. Показать, что не существует метрики во множестве случайных величин такой, что сходимость в ней эквивалентна сходимости почти наверное.

§ 11. Гильбертово пространство случайных величин с конечным вторым моментом

1. Среди банаховских пространств L^p, $p \geqslant 1$, рассмотренных выше, особо важную роль играет пространство $L^2 = L^2(\Omega, \mathscr{F}, \mathsf{P})$ — пространство (классов эквивалентных) случайных величин с конечным вторым моментом.

Если ξ, $\eta \in L^2$, то положим

$$(\xi, \eta) \equiv \mathsf{M}\,\xi\eta. \tag{1}$$

Ясно, что для ξ, η, $\zeta \in L^2$

$$(a\xi + b\eta, \zeta) = a(\xi, \zeta) + b(\eta, \zeta), \qquad a, b \in R,$$

$$(\xi, \xi) \geqslant 0$$

и

$$(\xi, \xi) = 0 \Rightarrow \xi = 0.$$

Тем самым (ξ, η) является *скалярным произведением*. Относительно нормы

$$\|\xi\| = (\xi, \xi)^{1/2}, \tag{2}$$

индуцируемой этим скалярным произведением, пространство L^2 (как было показано в § 10) является *полным*. Поэтому в соответствии с терминологией функционального анализа пространство с введенным скалярным произведением (1) является *гильбертовым пространством* случайных величин (с конечным вторым моментом).

Методы гильбертова пространства широко используются в теории вероятностей при исследовании свойств, определяемых лишь первыми двумя моментами рассматриваемых случайных величин («L^2-теория»). В этой связи остановимся на основных понятиях и фактах, необходимых для изложения L^2-теории (гл. VI).

2. Две случайные величины ξ и η из L^2 будем называть ортогональными ($\xi \perp \eta$), если их скалярное произведение $(\xi, \eta) \equiv \mathsf{M}\xi\eta = 0$. Согласно § 8 величины ξ и η назывались некоррелированными, если $\operatorname{cov}(\xi, \eta) = 0$, т. е. если

$$\mathsf{M}\xi\eta = \mathsf{M}\xi \cdot \mathsf{M}\eta.$$

Отсюда следует, что для случайных величин с нулевыми средними значениями понятия их ортогональности и некоррелированности совпадают.

Система $M \subseteq L^2$ будет называться *системой ортогональных* случайных величин, если $\xi \perp \eta$ для любых ξ, $\eta \in M$ $(\xi \neq \eta)$.

Если к тому же для всех $\xi \in M$ их норма $\|\xi\| = 1$, то M называется *ортонормированной системой* случайных величин.

3. Пусть $M = \{\eta_1, \ldots, \eta_n\}$ — ортонормированная система и ξ — какая-то случайная величина из L^2. В классе линейных оценок вида $\sum_{i=1}^{n} a_i \eta_i$ найдем наилучшую (в среднеквадратическом смысле) оценку случайной величины ξ (ср. с п. 2 § 8).

Простой подсчет показывает, что

$$\begin{aligned}
\mathsf{M}\left|\xi - \sum_{i=1}^{n} a_i\eta_i\right|^2 &= \left\|\xi - \sum_{i=1}^{n} a_i\eta_i\right\|^2 = \left(\xi - \sum_{i=1}^{n} a_i\eta_i,\ \xi - \sum_{i=1}^{n} a_i\eta_i\right) = \\
&= \|\xi\|^2 - 2\sum_{i=1}^{n} a_i(\xi, \eta_i) + \left(\sum_{i=1}^{n} a_i\eta_i,\ \sum_{i=1}^{n} a_i\eta_i\right) = \\
&= \|\xi\|^2 - 2\sum_{i=1}^{n} a_i(\xi, \eta_i) + \sum_{i=1}^{n} a_i^2 = \\
&= \|\xi\|^2 - \sum_{i=1}^{n} |(\xi, \eta_i)|^2 + \sum_{i=1}^{n} |a_i - (\xi, \eta_i)|^2 \geqslant \\
&\geqslant \|\xi\|^2 - \sum_{i=1}^{n} |(\xi, \eta_i)|^2, \qquad (3)
\end{aligned}$$

где мы воспользовались тем, что

$$a_i^2 - 2a_i(\xi, \eta_i) = |a_i - (\xi, \eta_i)|^2 - |(\xi, \eta_i)|^2.$$

Отсюда ясно, что инфинум $\mathsf{M}\left|\xi - \sum_{i=1}^{n} a_i\eta_i\right|^2$ по всем действительным $a_1, \ldots, a_n$ достигается при $a_i = (\xi, \eta_i)$, $i = 1, \ldots, n$.

Таким образом, оптимальной (в среднеквадратическом смысле) линейной оценкой ξ по $\eta_1, \ldots, \eta_n$ является оценка

$$\hat{\xi} = \sum_{i=1}^{n} (\xi, \eta_i)\eta_i. \qquad (4)$$

При этом

$$\Delta \equiv \inf \mathsf{M}\left|\xi - \sum_{i=1}^{n} a_i\eta_i\right|^2 = \mathsf{M}|\xi - \hat{\xi}|^2 = \|\xi\|^2 - \sum_{i=1}^{n} |(\xi, \eta_i)|^2 \qquad (5)$$

(ср. с (I.4.17) и (8.13)).

Из (3) вытекает также следующее *неравенство Бесселя*: если $M=\{\eta_1, \eta_2, \dots\}$ — некоторая ортонормированная система и $\xi \in L^2$, то

$$\sum_{i=1}^{\infty} |(\xi, \eta_i)|^2 \leqslant \|\xi\|^2; \tag{6}$$

при этом равенство достигается тогда и только тогда, когда

$$\xi = \text{l. i. m.} \sum_{i=1}^{\infty} (\xi, \eta_i)\eta_i. \tag{7}$$

Оценку $\hat{\xi}$, являющуюся *оптимальной линейной оценкой*, часто обозначают $\hat{\mathsf{M}}(\xi \mid \eta_1, \dots, \eta_n)$ и называют *условным математическим ожиданием* (ξ относительно $\eta_1, \dots, \eta_n$) *в широком смысле*.

Это название объясняется следующим. Если рассматривать всевозможные оценки $\varphi=\varphi(\eta_1, \dots, \eta_n)$ случайной величины ξ по $\eta_1, \dots, \eta_n$ (φ — борелевская функция), то оптимальной оценкой будет оценка $\varphi^*=\mathsf{M}(\xi \mid \eta_1, \dots, \eta_n)$, т. е. условное математическое ожидание ξ относительно $\eta_1, \dots, \eta_n$ (ср. с теоремой 1 § 8). Поэтому оптимальную линейную оценку по аналогии обозначают $\hat{\mathsf{M}}(\xi \mid \eta_1, \dots, \eta_n)$ и называют условным математическим ожиданием в широком смысле. В этой связи отметим, что если $\eta_1, \dots, \eta_n$ образуют гауссовскую систему (см. далее § 13), то $\mathsf{M}(\xi \mid \eta_1, \dots, \eta_n)$ и $\hat{\mathsf{M}}(\xi \mid \eta_1, \dots, \eta_n)$ совпадают.

Остановимся на *геометрическом смысле* оценки $\hat{\xi}=\hat{\mathsf{M}}(\xi \mid \eta_1, \dots, \eta_n)$.

Обозначим через $\mathscr{L}=\mathscr{L}(\eta_1, \dots, \eta_n)$ *линейное многообразие*, порожденное ортонормированной системой случайных величин $\eta_1, \dots, \eta_n$ (т. е. совокупность случайных величин вида $\sum_{i=1}^{n} a_i\eta_i$, $a_i \in R$).

Тогда из вышеизложенного вытекает, что ξ допускает «ортогональное разложение»

$$\xi = \hat{\xi} + (\xi - \hat{\xi}), \tag{8}$$

где $\hat{\xi} \in \mathscr{L}$, а $\xi - \hat{\xi} \perp \mathscr{L}$ в том смысле, что $\xi - \hat{\xi} \perp \lambda$ для любого $\lambda \in \mathscr{L}$. Естественно поэтому $\hat{\xi}$ назвать *проекцией* ξ на $\mathscr{L}$ («ближайшим» к ξ элементом из $\mathscr{L}$), а $\xi - \hat{\xi}$ — *перпендикуляром* к $\mathscr{L}$.

4. Предположение ортонормированности случайных величин $\eta_1, \dots, \eta_n$ позволило просто найти оптимальную линейную оценку (проекцию) $\hat{\xi}$ для ξ по $\eta_1, \dots, \eta_n$. Сложнее обстоит дело, если отказаться от предположения ортонормированности. Однако случай произвольных величин $\eta_1, \dots, \eta_n$ в определенном смысле может быть, как будет ниже показано, сведен к уже рассмотренному случаю ортонормированных величин. Для простоты дальнейшего изложения будем предполагать, что все рассматриваемые случайные величины имеют нулевые средние.

Будем говорить, что случайные величины $\eta_1, \ldots, \eta_n$ *линейно независимы*, если равенство

$$\sum_{i=1}^{n} a_i \eta_i = 0 \quad (\mathsf{P}\text{-п. н.})$$

выполнено лишь тогда, когда все a_i равны нулю.

Рассмотрим матрицу ковариаций

$$\mathbb{R} \equiv \mathsf{M}\eta\eta^*$$

вектора $\eta = (\eta_1, \ldots, \eta_n)$. Она является симметрической и неотрицательно определенной и, как отмечалось в § 8, найдется ортогональная матрица $\mathcal{O}$, приводящая ее к диагональному виду

$$\mathcal{O}^* \mathbb{R} \mathcal{O} = D,$$

где

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

— матрица с неотрицательными элементами d_i, являющимися характеристическими числами матрицы $\mathbb{R}$, т. е. корнями λ характеристического уравнения $\det(\mathbb{R} - \lambda E) = 0$.

Если величины $\eta_1, \ldots, \eta_n$ линейно независимы, то детерминант Грама (т. е. $\det \mathbb{R}$) не равен нулю и, значит, все $d_i > 0$. Пусть

$$B = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$$

и

$$\beta = B^{-1} \mathcal{O}^* \eta. \tag{9}$$

Тогда матрица ковариаций вектора β

$$\mathsf{M}\beta\beta^* = B^{-1}\mathcal{O}^* \mathsf{M}\eta\eta^* \mathcal{O} B^{-1} = B^{-1}\mathcal{O}^* \mathbb{R} \mathcal{O} B^{-1} = E,$$

и, следовательно, вектор $\beta = (\beta_1, \ldots, \beta_n)$ состоит из некоррелированных случайных величин. Ясно также, что

$$\eta = (\mathcal{O} B)\beta. \tag{10}$$

Таким образом, если $\eta_1, \ldots, \eta_n$ линейно независимы, то найдется такая ортонормированная система $\beta_1, \ldots, \beta_n$, что выполнены соотношения (9) и (10). При этом

$$\mathscr{L}\{\eta_1, \ldots, \eta_n\} = \mathscr{L}\{\beta_1, \ldots, \beta_n\}.$$

Изложенный способ получения ортонормированной системы $\beta_1, \ldots, \beta_n$ в ряде задач сказывается не очень удобным. Дело в том, что если трактовать η_i как значение случайной последовательности $(\eta_1, \ldots, \eta_n)$ в момент времени i, то построенное выше значение β_i сказывается зависящим не только от «прошлого» $(\eta_1, \ldots, \eta_i)$, но и от «будущего» $(\eta_{i+1}, \ldots, \eta_n)$. Приводимый ниже *процесс ортогонализации Грама — Шмидта* не страдает этим недостатком, более того, он обладает тем преимуществом, что может быть применен к бесконечным последовательностям *линейно независимых* случайных величин (т. е. последовательностям, у которых любое конечное число величин является линейно независимыми).

Пусть $\eta_1, \eta_2, \ldots$ — последовательность линейно независимых случайных величин из L^2. Построим по индукции последовательность $\varepsilon_1, \varepsilon_2, \ldots$ следующим образом. Пусть $\varepsilon_1 = \frac{\eta_1}{\|\eta_1\|}$. Если $\varepsilon_1, \ldots, \varepsilon_{k-1}$ уже выбраны так, что они ортонормированы, то положим

$$\varepsilon_n = \frac{\eta_n - \hat{\eta}_n}{\|\eta_n - \hat{\eta}_n\|}, \tag{11}$$

где $\hat{\eta}_n$ есть проекция η_n на линейное многообразие $\mathscr{L}(\varepsilon_1, \ldots, \varepsilon_{n-1})$, порожденное величинами $\eta_1, \ldots, \eta_{n-1}$

$$\hat{\eta}_n = \sum_{k=1}^{n-1} (\eta_n, \varepsilon_k)\varepsilon_k. \tag{12}$$

Поскольку величины $\eta_1, \ldots, \eta_n$ линейно независимы и $\mathscr{L}(\eta_1, \ldots, \eta_{n-1}) = \mathscr{L}(\varepsilon_1, \ldots, \varepsilon_{n-1})$, то $\|\eta_n - \hat{\eta}_n\| > 0$ и, следовательно, ε_n определено.

По построению $\|\varepsilon_n\| = 1$, $n \geq 1$, и ясно, что $(\varepsilon_n, \varepsilon_k) = 0$, $k < n$. Тем самым последовательность $\varepsilon_1, \varepsilon_2, \ldots$ является ортонормированной. При этом, согласно (11),

$$\eta_n = \hat{\eta}_n + b_n\varepsilon_n,$$

где $b_n = \|\eta_n - \hat{\eta}_n\|$, а $\hat{\eta}_n$ определяется формулой (12).

Пусть теперь $\eta_1, \ldots, \eta_n$ — произвольная система случайных величин (не обязательно линейно независимых). Пусть $\det \mathbb{R} = 0$, где $\mathbb{R} \equiv \|r_{ij}\|$ — матрица ковариаций вектора $(\eta_1, \ldots, \eta_n)$, и пусть

$$\operatorname{rang} \mathbb{R} = r < n.$$

Тогда, как известно из алгебры, квадратичная форма

$$Q(a) = \sum_{i,j=1}^{n} r_{ij}a_i a_j, \qquad a = (a_1, \ldots, a_n),$$

такова, что существует ровно $n-r$ линейно независимых векторов $a^{(1)}, \ldots, a^{(n-r)}$ таких, что $Q(a^{(i)})=0$, $i=1, \ldots, n-r$.

Но

$$Q(a)=\mathsf{M}\left(\sum_{k=1}^{n} a_k \eta_k\right)^2.$$

Следовательно, с вероятностью единица

$$\sum_{k=1}^{n} a_k^{(i)} \eta_k = 0, \quad i=1, \ldots, n-r.$$

Иначе говоря, существует ровно $n-r$ линейных соотношений между величинами $\eta_1, \ldots, \eta_n$. Поэтому, если, скажем, $\eta_1, \ldots, \eta_r$ линейно независимы, то все остальные величины $\eta_{r+1}, \ldots, \eta_n$ линейно через них выражаются и, значит, $\mathscr{L}(\eta_1, \ldots, \eta_n) = \mathscr{L}(\eta_1, \ldots, \eta_r)$. Отсюда ясно, что с помощью процесса ортогонализации можно найти r ортонормированных случайных величин $\varepsilon_1, \ldots, \varepsilon_r$ таких, что все $\eta_1, \ldots, \eta_n$ линейно через них выражаются и $\mathscr{L}(\eta_1, \ldots, \eta_n) = \mathscr{L}(\varepsilon_1, \ldots, \varepsilon_r)$.

5. Пусть $\eta_1, \eta_2, \ldots$, — последовательность случайных величин из L^2. Будем обозначать через $\mathscr{L} = \mathscr{L}(\eta_1, \eta_2, \ldots)$ *линейное многообразие*, порожденное величинами $\eta_1, \eta_2, \ldots$, т. е. совокупность случайных величин вида $\sum_{i=1}^{n} a_i \eta_i$, $n \geqslant 1$, $a_i \in R$. Через $\overline{\mathscr{L}} = \overline{\mathscr{L}}(\eta_1, \eta_2, \ldots)$ обозначим *замкнутое линейное многообразие*, порожденное $\eta_1, \eta_2, \ldots$, т. е. совокупность случайных величин из $\mathscr{L}$ и их пределов в среднеквадратическом смысле.

Говорят, что система случайных величин $\eta_1, \eta_2, \ldots$ образует *счетный ортонормированный базис* (иначе — *полную ортонормированную систему*) в L^2, если:

a) $\eta_1, \eta_2, \ldots$ — ортонормированная система,

b) $\overline{\mathscr{L}}(\eta_1, \eta_2, \ldots) = L^2$.

Гильбертово пространство со счетным ортонормированным базисом называют *сепарабельным*.

В силу условия b) для любого $\xi \in L^2$ и заданного $\varepsilon > 0$ найдутся такие $a_1, \ldots, a_n$, что

$$\left\| \xi - \sum_{i=1}^{n} a_i \eta_i \right\| \leqslant \varepsilon.$$

Тогда, согласно (3),

$$\left\| \xi - \sum_{i=1}^{n} (\xi, \eta_i) \eta_i \right\| \leqslant \varepsilon$$

и, следовательно, для сепарабельных гильбертовых пространств L^2 любой элемент ξ представим в виде

$$\xi = \sum_{i=1}^{\infty} (\xi, \eta_i) \cdot \eta_i, \tag{13}$$

точнее,

$$\xi = \underset{n}{\text{l. i. m.}} \sum_{i=1}^{n} (\xi, \eta_i)\, \eta_i.$$

Отсюда и из (3) тогда заключаем, что имеет место следующее *равенство Парсеваля*:

$$\|\xi\|^2 = \sum_{i=1}^{\infty} |(\xi, \eta_i)|^2, \qquad \xi \in L^2. \tag{14}$$

Нетрудно доказать, что верно и обратное: если $\eta_1, \eta_2, \ldots$ — некоторая ортонормированная система и выполнено любое из условий (13) или (14), то эта система является базисом.

Приведем примеры сепарабельных гильбертовых пространств и их базисов.

Пример 1. Пусть $\Omega = R$, $\mathscr{F} = \mathscr{B}(R)$ и P — гауссовская мера

$$\mathsf{P}(-\infty, a] = \int_{-\infty}^{a} \varphi(x)\, dx, \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Обозначим $D = \frac{d}{dx}$ и введем функции

$$H_n(x) = \frac{(-1)^n D^n \varphi(x)}{\varphi(x)}, \qquad n \geqslant 0. \tag{15}$$

Нетрудно найти, что

$$\begin{aligned} D\varphi(x) &= -x\varphi(x), \\ D^2\varphi(x) &= (x^2-1)\,\varphi(x), \\ D^3\varphi(x) &= (3x - x^3)\,\varphi(x), \\ &\dots\dots\dots\dots \end{aligned} \tag{16}$$

Отсюда следует, что $H_n(x)$ являются полиномами (называемыми *полиномами Эрмита*). Из (15), (16) находим, что

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \end{aligned}$$

Простой подсчет показывает, что

$$(H_m, H_n) = \int_{-\infty}^{\infty} H_m(x) H_n(x)\, d\mathsf{P} = \\ = \int_{-\infty}^{\infty} H_m(x) H_n(x) \varphi(x)\, dx = n!\delta_{mn},$$

где δ_{mn} — символ Кронекера (0, если $m \neq n$, и 1, если $m = n$). Поэтому, если положить

$$h_n(x) = \frac{H_n(x)}{\sqrt{n}},$$

то система этих *нормированных полиномов Эрмита* $\{h_n(x)\}_{n \geqslant 0}$ будет ортонормированной системой. Из функционального анализа известно, что если

$$\lim_{c \downarrow 0} \int_{-\infty}^{\infty} e^{c|x|} \mathsf{P}(dx) < \infty, \tag{17}$$

то система функций $\{1, x, x^2, \ldots\}$ является плотной в L^2, т. е. любая функция $\xi = \xi(x)$ из L^2 может быть представлена или в виде $\sum_{i=1}^{n} a_i \eta_i(x)$, где $\eta_i(x) = x^i$, или в виде их пределов (в среднеквадратическом смысле). Если применить процесс ортогонализации Грамма — Шмидта к последовательности функций $\eta_1(x)$, $\eta_2(x)$, ... с $\eta_i(x) = x^i$, то полученная ортонормированная система будет в точности совпадать с системой нормированных полиномов Эрмита. В рассматриваемом нами случае условие (17) выполнено. Следовательно, полиномы $\{h_n(x)\}_{n \geqslant 0}$ образуют базис и, значит, любая случайная величина $\xi = \xi(x)$ на рассматриваемом вероятностном пространстве представима в виде

$$\xi(x) = \text{l. i. m.} \sum_{i=0}^{\infty} (\xi, h_i) h_i(x). \tag{18}$$

Пример 2. Пусть $\Omega = \{0, 1, 2, \ldots,\}$ и $P = \{P_1, P_2, \ldots\}$ — пуассоновское распределение:

$$P_x = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots; \ \lambda > 0.$$

Положим $\Delta f(x) = f(x) - f(x-1)$ $(f(x) = 0,\ x < 0)$ и по аналогии с (15) определим *полиномы Пуассона — Шарлье*

Поскольку

$$(\Pi_m, \Pi_n) = \sum_{x=0}^{\infty} \Pi_m(x) \Pi_n(x) P_x = c_n \delta_{mn},$$

где c_n — положительные константы, то система *нормированных полиномов Пуассона — Шарлье* $\{\pi_n(x)\}_{n \geq 0}$, $\pi_n(x) = \frac{\Pi_n(x)}{\sqrt{c_n}}$, образует ортонормированную систему, которая в силу выполнимости условия (17) является базисом.

Пример 3. Приводимые в этом примере ортонормированные системы функций Радемахера и Хаара интересны как для теории функций, так и для теории вероятностей.

Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$ и P — мера Лебега. Как упоминалось в § 1, каждое число $x \in [0, 1)$ может быть однозначно разложено в двоичную дробь

$$x = \frac{x_1}{2} + \frac{x_2}{2^2} + \ldots,$$

где $x_i = 0$ или 1. (Для однозначности разложения мы уславливаемся рассматривать только те разложения, которые содержат *бесконечное* число нулей. Так, из двух разложений

$$\frac{1}{2} = \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \ldots = \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots,$$

мы берем первое.)

Образуем случайные величины $\xi_1(x)$, $\xi_2(x)$, ..., положив

$$\xi_n(x) = x_n.$$

Тогда для любых a_i, принимающих значения 0 или 1,

$$\mathsf{P}\{x\colon \xi_1 = a_1, \ldots, \xi_n = a_n\} = \mathsf{P}\left\{x\colon \frac{a_1}{2} + \frac{a_2}{2^2} + \ldots + \frac{a_n}{2^n} \leq x < \right.$$
$$\left. < \frac{a_1}{2} + \frac{a_2}{2^2} + \ldots + \frac{a_n}{2^n} + \frac{1}{2^n}\right\} =$$
$$= \mathsf{P}\left\{x\colon x \in \left[\frac{a_1}{2} + \ldots + \frac{a_n}{2^n}, \; \frac{a_1}{2} + \ldots + \frac{a_n}{2^n} + \frac{1}{2^n}\right)\right\} = \frac{1}{2^n}.$$

Отсюда непосредственно следует, что $\xi_1, \xi_2, \ldots$ образует *последовательность независимых бернуллиевских случайных величин* (рис. 30 показывает, как устроены $\xi_1 = \xi_1(x)$ и $\xi_2 = \xi_2(x)$).

Если теперь положить $R_n(x) = 1 - 2\xi_n(x)$, $n \geq 1$, то нетрудно проверить, что система $\{R_n\}$ (*функций Радемахера*, рис. 31) является ортонормированной:

$$\mathsf{M} R_n R_m = \int_0^1 R_n(x) R_m(x)\, dx = \delta_{nm}.$$

Заметим, что $(1,\ R_n) \equiv \mathsf{M}R_n = 0$. Отсюда следует, что эта система *не является* полной.

Однако систему Радемахера можно использовать для построения так называемой *системы Хаара*, которая и проще устроена, и к тому же является как *ортонормированной*, так и *полной*.

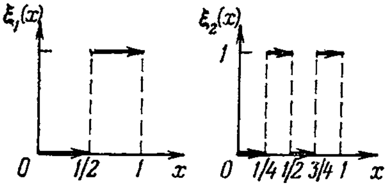


Рис 30.

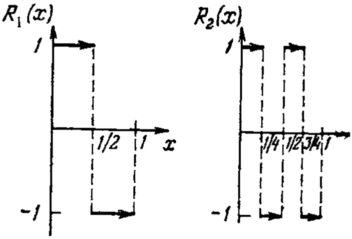


Рис 31 Функции Радемахера

Снова пусть $\Omega = [0,\ 1)$, $\mathscr{F} = \mathscr{B}([0,\ 1))$. Положим

$$H_1(x) = 1,$$
$$H_2(x) = R_1(x),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$H_n(x) = \begin{cases} 2^{j/2} R_j(x), & \text{если } \frac{k-1}{2^j} \leqslant x < \frac{k}{2^j},\ n = 2^j + k, \\ & 1 \leqslant k \leqslant 2^j,\ j \geqslant 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Нетрудно проверить, что $H_n(x)$ можно записать и в таком виде:

$$H_{2^m+1}(x) = \begin{cases} 2^{m/2}, & 0 \leqslant x < 2^{-(m+1)}, \\ -2^{m/2}, & 2^{-(m+1)} \leqslant x < 2^{-m}, \quad m = 1,\ 2,\ \dots, \\ 0 & \text{в других случаях,} \end{cases}$$

$$H_{2^m+j}(x) = H_{2^m+1}\left(x - \frac{j-1}{2^m}\right), \quad j = 1, \dots, 2^m.$$

На рис. 32 приведены графики первых восьми функций, дающих представление о структуре образования и поведении функций Хаара.

Система функций Хаара является, как нетрудно проверить, ортонормированной. Более того, она полна и в L^1, и в L^2, т. е. если функция $f = f(x) \in L^p$ для $p = 1$ или $p = 2$, то

$$\int_0^1 \left| f(x) - \sum_{k=1}^{n} (f,\ H_k) H_k(x) \right|^p dx \to 0, \qquad n \to \infty,$$

и обладает к тому же тем свойством, что с вероятностью единица (по лебеговской мере)

$$\sum_{k=1}^{n} (f, H_k) H_k(x) \to f(x), \quad n \to \infty.$$

Мы докажем эти факты в § 4 гл. VII, выведя их из общих теорем о сходимости мартингалов, что, в частности, будет служить хорошей иллюстрацией применения мартингальных методов к теории функций.

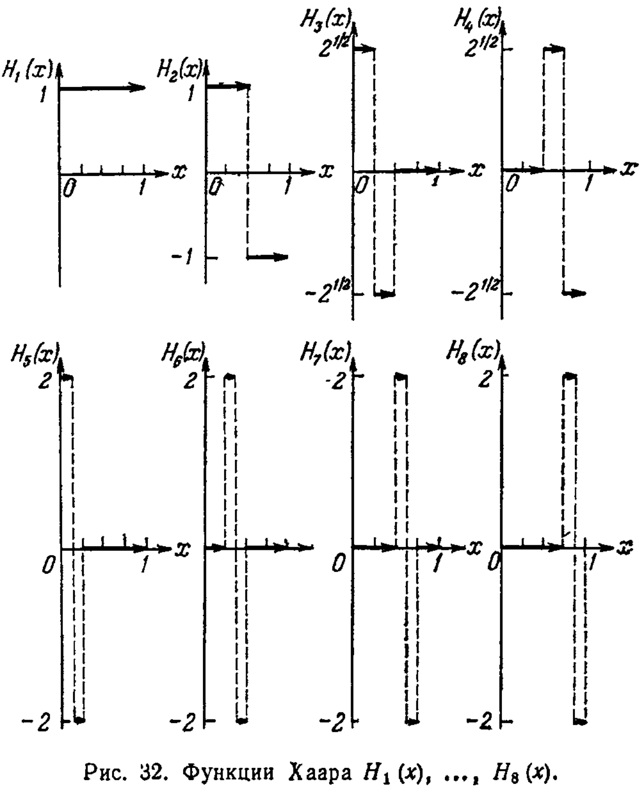


Рис. 32. Функции Хаара $H_1(x), \ldots, H_8(x)$.

6. Если $\eta_1, \ldots, \eta_n$ — некоторая конечная ортонормированная система, то, как было показано выше, для всякой случайной величины $\xi \in L^2$ в линейном многообразии $\mathscr{L} = \mathscr{L}(\eta_1, \ldots, \eta_n)$ можно найти случайную величину $\hat{\xi}$ (проекцию ξ на $\mathscr{L}$) такую, что

$$\|\xi - \hat{\xi}\| = \inf\{\|\xi - \zeta\| : \zeta \in \mathscr{L}(\eta_1, \ldots, \eta_n)\}.$$

При этом $\hat{\xi} = \sum_{i=1}^{n} (\xi, \eta_i)\eta_i$. Этот результат допускает естественное обобщение на тот случай, когда $\eta_1, \eta_2, \ldots$ — счетная ортонормированная система (не обязательно являющаяся базисом). А именно, справедлив следующий результат.

Теорема. *Пусть $\eta_1, \eta_2, \ldots$ — ортонормированная система случайных величин, $\overline{\mathscr{L}} = \overline{\mathscr{L}}(\eta_1, \eta_2, \ldots)$ — замкнутое линейное многообразие, порожденное ими. Тогда существует и притом единственный элемент $\hat{\xi} \in \overline{\mathscr{L}}$ такой, что*

$$\|\xi - \hat{\xi}\| = \inf\{\|\xi - \zeta\|: \zeta \in \overline{\mathscr{L}}\}. \tag{20}$$

При этом

$$\hat{\xi} = \underset{n}{\mathrm{l.i.m.}} \sum_{i=1}^{n} (\xi, \eta_i)\eta_i \tag{21}$$

и $\xi - \hat{\xi} \perp \zeta$, $\zeta \in \overline{\mathscr{L}}$.

Доказательство. Обозначим $d = \inf\{\|\xi - \zeta\|: \zeta \in \overline{\mathscr{L}}\}$ и выберем последовательность $\zeta_1, \zeta_2, \ldots$ так, что $\|\xi - \zeta_n\| \to d$. Покажем, что эта последовательность является фундаментальной. Простой подсчет показывает, что

$$\|\zeta_n - \zeta_m\|^2 = 2\|\zeta_n - \xi\|^2 + 2\|\zeta_m - \xi\|^2 - 4\left\|\frac{\zeta_n + \zeta_m}{2} - \xi\right\|^2.$$

Ясно, что $\frac{\zeta_n + \zeta_m}{2} \in \overline{\mathscr{L}}$, поэтому $\left\|\frac{\zeta_n + \zeta_m}{2} - \xi\right\|^2 \geq d^2$ и, следовательно, $\|\zeta_n - \zeta_m\|^2 \to 0$, $n, m \to \infty$.

Пространство L^2 является полным (теорема 7 § 10). Поэтому найдется такой элемент $\hat{\xi}$, что $\|\zeta_n - \hat{\xi}\| \to 0$. Множество $\overline{\mathscr{L}}$ замкнуто, поэтому $\hat{\xi} \in \overline{\mathscr{L}}$. Далее, $\|\zeta_n - \xi\| \to d$, следовательно, $\|\xi - \hat{\xi}\| = d$, что и доказывает существование требуемого элемента.

Покажем, что $\hat{\xi}$ — единственный элемент в $\overline{\mathscr{L}}$ с требуемым свойством. Пусть $\tilde{\xi} \in \overline{\mathscr{L}}$ и

$$\|\xi - \hat{\xi}\| = \|\xi - \tilde{\xi}\| = d.$$

Тогда (в силу задачи 3)

$$\|\hat{\xi} + \tilde{\xi} - 2\xi\|^2 + \|\hat{\xi} - \tilde{\xi}\|^2 = 2\|\hat{\xi} - \xi\|^2 + 2\|\tilde{\xi} - \xi\|^2 = 4d^2.$$

Но

$$\|\hat{\xi} + \tilde{\xi} - 2\xi\|^2 = 4\left\|\frac{1}{2}(\hat{\xi} + \tilde{\xi}) - \xi\right\|^2 \geq 4d^2.$$

Следовательно, $\|\hat{\xi} - \tilde{\xi}\|^2 = 0$, что и доказывает единственность «ближайшего» к ξ элемента из $\overline{\mathscr{L}}$.

Докажем теперь, что $\xi-\hat{\xi}\perp\zeta$, $\zeta\in\overline{\mathscr{L}}$. В силу (20) для любого $c\in R$

$$\|\xi-\hat{\xi}-c\zeta\|\geqslant\|\xi-\hat{\xi}\|.$$

Но

$$\|\xi-\hat{\xi}-c\zeta\|^2=\|\xi-\hat{\xi}\|^2+c^2\|\zeta\|^2-2(\xi-\hat{\xi},\ c\zeta).$$

Поэтому

$$c^2\|\zeta\|^2\geqslant 2(\xi-\hat{\xi},\ c\zeta). \tag{22}$$

Возьмем $c=\lambda(\xi-\hat{\xi},\ \zeta)$, $\lambda\in R$. Тогда из (22) получим, что

$$(\xi-\hat{\xi},\ \zeta)^2[\lambda^2\|\zeta\|^2-2\lambda]\geqslant 0.$$

При достаточно малых положительных λ $\lambda^2\|\zeta\|^2-2\lambda<0$. Поэтому $(\xi-\hat{\xi},\ \zeta)=0$, $\zeta\in\overline{\mathscr{L}}$.

Осталось доказать представление (21).

Множество $\overline{\mathscr{L}}=\overline{\mathscr{L}}(\eta_1,\ \eta_2,\ldots)$ является замкнутым подпространством в L^2 и, следовательно, само является гильбертовым пространством (с тем же самым скалярным произведением). Для этого гильбертова пространства $\overline{\mathscr{L}}$ система $\eta_1,\ \eta_2,\ldots$ является базисом (задача 4) и, следовательно,

$$\hat{\xi}=\text{l.i.m.}\sum_{k=1}^{\infty}(\hat{\xi},\ \eta_k)\,\eta_k. \tag{23}$$

Но $\xi-\hat{\xi}\perp\eta_k$, $k\geqslant 1$, а значит, $(\hat{\xi},\ \eta_k)=(\xi,\ \eta_k)$, $k\geqslant 0$, что вместе с (23) доказывает (21).

Теорема доказана.

Замечание. Как и в конечномерном случае, $\hat{\xi}$ будем называть проекцией ξ на $\overline{\mathscr{L}}=\overline{\mathscr{L}}(\eta_1,\ \eta_2,\ldots)$, $\xi-\hat{\xi}$ — перпендикуляром, а представление

$$\xi=\hat{\xi}+(\xi-\hat{\xi})$$

— ортогональным разложением.

Величину $\hat{\xi}$ обозначают также $\hat{\mathsf{M}}(\xi\mid\eta_1,\ \eta_2,\ldots)$ и называют условным математическим ожиданием в широком смысле (ξ относительно $\eta_1,\ \eta_2,\ldots$). С точки зрения оценивания ξ по $\eta_1,\ \eta_2,\ldots$ величина $\hat{\xi}$ является оптимальной линейной оценкой, ошибка которой

$$\Delta\equiv\mathsf{M}\,|\xi-\hat{\xi}|^2\equiv\|\xi-\hat{\xi}\|^2=\|\xi\|^2-\sum_{i=1}^{\infty}|(\xi,\ \eta_i)|^2,$$

что следует из (5) и (23).

7. Задачи.

1. Показать, что если $\xi = \text{l.i.m.}\, \xi_n$, то $\|\xi_n\| \to \|\xi\|$.

2. Показать, что если $\xi = \text{l.i.m.}\, \xi_n$ и $\eta = \text{l.i.m.}\, \eta_n$, то $(\xi_n, \eta_n) \to (\xi, \eta)$.

3. Показать, что норма $\|\cdot\|$ удовлетворяет свойству «параллелограмма»

$$\|\xi+\eta\|^2 + \|\xi-\eta\|^2 = 2(\|\xi\|^2 + \|\eta\|^2).$$

4. Пусть $\{\xi_1, \ldots, \xi_n\}$ — семейство ортогональных случайных величин. Показать, что для них справедлива «теорема Пифагора»:

$$\left\| \sum_{i=1}^{n} \xi_i \right\|^2 = \sum_{i=1}^{n} \|\xi_i\|^2.$$

5. Пусть $\eta_1, \eta_2, \ldots$ — ортонормированная система и $\overline{\mathscr{L}} = \overline{\mathscr{L}}(\eta_1, \eta_2, \ldots)$ — замкнутое линейное многообразие, порожденное $\eta_1, \eta_2, \ldots$. Доказать, что эта система является базисом для (гильбертова) пространства $\overline{\mathscr{L}}$.

6. Пусть $\xi_1, \xi_2, \ldots$ — последовательность ортогональных случайных величин, $S_n = \xi_1 + \ldots + \xi_n$. Показать, что если $\sum_{n=1}^{\infty} \mathsf{M}\xi_n^2 < \infty$, то найдется такая случайная величина S с $\mathsf{M}S^2 < \infty$, что $\text{l.i.m.}\, S_n = S$, т. е. $\|S_n - S\|^2 = \mathsf{M}|S_n - S|^2 \to 0$, $n \to \infty$.

7. Показать, что в пространстве $L^2 = L^2([-\pi, \pi], \mathscr{B}([-\pi, \pi]))$ с мерой Лебега μ система функций $\left\{\frac{1}{\sqrt{2\pi}} e^{i\lambda n}, n = 0, \pm 1, \ldots\right\}$ образует ортонормированный базис.

§ 12. Характеристические функции

1. Метод характеристических функций является одним из основных средств аналитического аппарата теории вероятностей. Наиболее ярко это будет продемонстрировано в гл. III при доказательстве предельных теорем и, в частности, при доказательстве центральной предельной теоремы, обобщающей теорему Муавра — Лапласа. Здесь же мы ограничимся определениями и изложением основных свойств характеристических функций.

Прежде всего сделаем одно замечание общего характера.

Наряду со случайными величинами (принимающими действительные значения) теория характеристических функций требует привлечения комплекснозначных случайных величин (см. п. 1 § 5).

Многие из определений и свойств, относящихся к случайным величинам, легко переносятся и на комплексный случай. Так, математическое ожидание $\mathsf{M}\zeta$ комплекснозначной случайной величины $\zeta = \xi + i\eta$ считается определенным, если определены математические ожидания $\mathsf{M}\xi$ и $\mathsf{M}\eta$. В этом случае по определению полагаем $\mathsf{M}\zeta = \mathsf{M}\xi + i\mathsf{M}\eta$. Из определения 5 (§ 5) независимости случайных элементов нетрудно вывести, что комплекснозначные величины $\zeta_1 = \xi_1 + i\eta_1$, $\zeta_2 = \xi_2 + i\eta_2$ независимы тогда и только тогда, когда независимы пары случайных величин (ξ_1, η_1) и (ξ_2, η_2), или, что то же самое, независимы σ-алгебры $\mathscr{F}_{\xi_1, \eta_1}$ и $\mathscr{F}_{\xi_2, \eta_2}$.

Наряду с пространством L^2 действительных случайных величин с конечным вторым моментом можно ввести в рассмотрение гильбертово пространство комплекснозначных случайных величин $\zeta = \xi + i\eta$ с $\mathsf{M}|\zeta|^2 < \infty$, где $|\zeta|^2 = \xi^2 + \eta^2$, и скалярным произведением $(\zeta_1, \zeta_2) = \mathsf{M}\zeta_1\bar{\zeta}_2$, где $\bar{\zeta}_2$ — комплексно-сопряженная случайная величина. В дальнейшем как действительнозначные, так и комплекснозначные случайные величины будем называть просто случайными величинами, отмечая, если это необходимо, о каком конкретно случае идет речь.

Условимся также о следующих обозначениях.

При алгебраических операциях векторы $a \in R^n$ будут рассматриваться как вектор-столбцы,

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

а a^* — как вектор-строки, $a^* = (a_1, \ldots, a_n)$. Если $a, b \in R^n$, то под их скалярным произведением (a, b) будет пониматься величина $\sum_{i=1}^{n} a_i b_i$. Ясно, что $(a, b) = a^* b$.

Если $a \in R^n$ и $\mathbb{R} = \| r_{ij} \|$ — матрица порядка $n \times n$, то

$$(\mathbb{R}a, a) = a^* \mathbb{R} a = \sum_{i, j=1}^{n} r_{ij} a_i a_j. \qquad (1)$$

2. Определение 1. Пусть $F = F(x_1, \ldots, x_n)$ — n-мерная функция распределения в $(R^n, \mathscr{B}(R^n))$. Ее *характеристической функцией* называется функция

$$\varphi(t) = \int_{R^n} e^{i(t, x)}\, dF(x), \qquad t \in R^n. \qquad (2)$$

Определение 2. Если $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор, определенный на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ со значениями в R^n, то его *характеристической функцией* называется функция

$$\varphi_\xi(t) = \int_{R^n} e^{i(t, x)}\, dF_\xi(x), \qquad t \in R^n, \qquad (3)$$

где $F_\xi = F_\xi(x_1, \ldots, x_n)$ — функция распределения вектора $\xi = (\xi_1, \ldots, \xi_n)$.

Если функция $F(x)$ имеет плотность $f=f(x)$, то тогда

$$\varphi(t)=\int_{R^n} e^{i(t,x)}f(x)\,dx.$$

Иначе говоря, в этом случае характеристическая функция $\varphi(t)$ есть не что иное, как преобразование Фурье функции $f(x)$.

Из (3) и теоремы 6.7 (о замене переменных под знаком интеграла Лебега) вытекает, что характеристическую функцию $\varphi_\xi(t)$ случайного вектора можно определить также равенством

$$\varphi_\xi(t)=\mathsf{M}e^{i(t,\xi)}, \qquad t\in R^n. \tag{4}$$

Приведем теперь основные свойства характеристических функций, формулируя и доказывая их лишь в случае $n=1$. Некоторые наиболее важные результаты, относящиеся к общему случаю, даются в виде задач.

Пусть $\xi=\xi(\omega)$ — случайная величина, $F_\xi=F_\xi(x)$ — ее функция распределения и

$$\varphi_\xi(t)=\mathsf{M}e^{it\xi}$$

— характеристическая функция.

Сразу отметим, что если $\eta=a\xi+b$, то

$$\varphi_\eta(t)=\mathsf{M}e^{it\eta}=\mathsf{M}e^{it(a\xi+b)}=e^{itb}\mathsf{M}e^{iat\xi}.$$

Поэтому

$$\varphi_\eta(t)=e^{itb}\varphi_\xi(at). \tag{5}$$

Далее, если $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины и $S_n=\xi_1+\ldots+\xi_n$, то

$$\varphi_{S_n}(t)=\prod_{j=1}^{n}\varphi_{\xi_j}(t). \tag{6}$$

В самом деле,

$$\begin{aligned}\varphi_{S_n}&=\mathsf{M}e^{it(\xi_1+\cdots+\xi_n)}=\mathsf{M}e^{it\xi_1}\ldots e^{it\xi_n}=\\&=\mathsf{M}e^{it\xi_1}\ldots\mathsf{M}e^{it\xi_n}=\prod_{j=1}^{n}\varphi_{\xi_j}(t),\end{aligned}$$

где мы воспользовались тем, что математическое ожидание произведения независимых (ограниченных) случайных величин (как действительных так и комплексных, см. теорему 6 в § 6 и задачу 1) равно произведению их математических ожиданий.

Свойство (6) является ключевым при доказательстве предельных теорем для сумм независимых случайных величин методом характеристических функций (см. § 3 гл. III). В этой связи отметим, что функция распределения F_{S_n} выражается через функции распределения отдельных слагаемых уже значительно более сложным образом, а именно, $F_{S_n}=F_{\xi_1}*\ldots*F_{\xi_n}$, где знак $*$ означает свертку распределений (см. п. 4 § 8).

Приведем примеры характеристических функций.

Пример 1. Пусть ξ — бернуллиевская случайная величина с $\mathsf{P}(\xi=1)=p$, $\mathsf{P}(\xi=0)=q$, $p+q=1$, $1>p>0$, тогда

$$\varphi_\xi(t)=pe^{it}+q.$$

Если $\xi_1, \ldots, \xi_n$ — независимые одинаково распределенные (как ξ) случайные величины, то

$$\varphi_{\frac{S_n-np}{\sqrt{npq}}}(t)=\mathsf{M}e^{it\frac{S_n-np}{\sqrt{npq}}}=e^{-it\sqrt{\frac{np}{q}}}\left[pe^{i\frac{t}{\sqrt{npq}}}+q\right]^n=$$

$$=\left[pe^{it\sqrt{\frac{q}{np}}}+qe^{-it\sqrt{\frac{p}{nq}}}\right]^n. \tag{7}$$

Заметим, что при $n\to\infty$ отсюда следует, что

$$\varphi_{\frac{S_n-np}{\sqrt{npq}}}(t)\to e^{-t^2/2}. \tag{8}$$

Пример 2. Пусть $\xi\sim\mathscr{N}(m, \sigma^2)$, $|m|<\infty$, $\sigma^2>0$. Покажем, что

$$\varphi_\xi(t)=e^{itm-\frac{t^2\sigma^2}{2}}. \tag{9}$$

Положим $\eta=\frac{\xi-m}{\sigma}$. Тогда $\eta\sim\mathscr{N}(0, 1)$ и, так как в силу (5)

$$\varphi_\xi(t)=e^{itm}\varphi_\eta(\sigma t),$$

то достаточно лишь показать, что

$$\varphi_\eta(t)=e^{-t^2/2}. \tag{10}$$

Имеем

$$\varphi_\eta(t)=\mathsf{M}e^{it\eta}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{itx}e^{-x^2/2}\,dx=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sum_{n=0}^{\infty}\frac{(itx)^n}{n!}e^{-x^2/2}\,dx=\sum_{n=0}^{\infty}\frac{(it)^n}{n!}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^ne^{-x^2/2}\,dx=$$

$$=\sum_{n=0}^{\infty}\frac{(it)^{2n}}{(2n)!}(2n-1)!!=\sum_{n=0}^{\infty}\frac{(it)^{2n}}{(2n)!}\frac{(2n)!}{2^nn!}=\sum_{n=0}^{\infty}\left(-\frac{t^2}{2}\right)\cdot\frac{1}{n!}=e^{-t^2/2},$$

где мы воспользовались тем (см. задачу 7 в § 8), что

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{2n}e^{-x^2/2}\,dx=\mathsf{M}\eta^{2n}=(2n-1)!!$$

Пример 3. Пусть ξ — пуассоновская случайная величина,

$$\mathsf{P}(\xi=k)=\frac{e^{-\lambda}\lambda^k}{k!},\qquad k=0,\ 1,\ldots$$

Тогда

$$\mathsf{M}e^{it\xi}=\sum_{k=0}^{\infty}e^{itk}\frac{e^{-\lambda}\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!}=\exp\{\lambda(e^{it}-1)\}. \tag{11}$$

3. Как отмечалось в п. 1 § 9, с каждой функцией распределения в $(R,\ \mathscr{B}(R))$ можно связать случайную величину, имеющую эту функцию в качестве своей функции распределения. Поэтому при изложении свойств характеристических функций (в смысле как определения 1, так и определения 2) можно ограничиться рассмотрением характеристических функций $\varphi(t)=\varphi_\xi(t)$ случайных величин $\xi=\xi(\omega)$.

Теорема 1. *Пусть ξ — случайная величина с функцией распределения $F=F(x)$ и*

$$\varphi(t)=\mathsf{M}e^{it\xi}$$

— ее характеристическая функция.

Имеют место следующие свойства:

1) $|\varphi(t)|\leqslant\varphi(0)=1$;

2) *$\varphi(t)$ равномерно непрерывна по $t\in R$;*

3) $\varphi(t)=\overline{\varphi(-t)}$;

4) *$\varphi(t)$ является действительнозначной функцией тогда и только тогда, когда распределение F симметрично* $\left(\int_B dF(x)=\right.$ $\left.=\int_{-B}dF(x),\ B\in\mathscr{B}(R),\ -B=\{-x\colon\ x\in B\}\right)$;

5) *если для некоторого $n\geqslant 1$ $\mathsf{M}|\xi|^n<\infty$, то при всех $r\leqslant n$ существуют производные $\varphi^{(r)}(t)$ и*

$$\varphi^{(r)}(t)=\int_R (ix)^r e^{itx}\,dF(x), \tag{12}$$

$$\mathsf{M}\xi^r=\frac{\varphi^{(r)}(0)}{i^r}, \tag{13}$$

$$\varphi(t)=\sum_{r=0}^{n}\frac{(it)^r}{r!}\mathsf{M}\xi^r+\frac{(it)^n}{n!}\varepsilon_n(t), \tag{14}$$

где $|\varepsilon_n(t)|\leqslant 3\mathsf{M}|\xi|^n$ *и* $\varepsilon_n(t)\to 0,\ t\to 0$.

6) *Если существует и является конечной $\varphi^{(2n)}(0)$, то $\mathsf{M}\xi^{2n}<\infty$.*

7) *Пусть $\mathsf{M}|\xi|^n<\infty$ для всех $n\geqslant 1$ и*

$$\overline{\lim_n}\frac{(\mathsf{M}|\xi|^n)^{1/n}}{n}=\frac{1}{R}<\infty,$$

тогда при всех $|t|<R$

$$\varphi(t)=\sum_{n=0}^{\infty}\frac{(it)^n}{n!}\mathsf{M}\xi^n. \qquad (15)$$

Доказательство. Свойства 1) и 3) очевидны. Свойство 2) следует из оценки

$$|\varphi(t+h)-\varphi(t)|=|\mathsf{M}e^{it\xi}(e^{ih\xi}-1)|\leqslant \mathsf{M}|e^{ih\xi}-1|$$

и теоремы о мажорируемой сходимости, согласно которой $\mathsf{M}|e^{ih\xi}-1|\to 0$, $h\to 0$.

Свойство 4). Пусть F симметрична. Тогда, если $g(x)$ — ограниченная борелевская нечетная функция, то $\int_R g(x)\,dF(x)=0$ (заметим, что для простых нечетных функций это следует сразу из определения симметричности F). Поэтому $\int_R \sin tx\,dF(x)=0$ и, значит,

$$\varphi(t)=\mathsf{M}\cos t\xi.$$

Обратно, пусть $\varphi_\xi(t)$ является действительной функцией. Тогда в силу 3)

$$\varphi_{-\xi}(t)=\varphi_\xi(-t)=\overline{\varphi_\xi(t)}=\varphi_\xi(t),\qquad t\in R.$$

Отсюда (как это будет доказано ниже в теореме 2) следует, что функции распределения $F_{-\xi}$ и F_ξ случайных величин $-\xi$ и ξ совпадают, а значит (по теореме 3.1),

$$\mathsf{P}(\xi\in B)=\mathsf{P}(-\xi\in B)=\mathsf{P}(\xi\in -B)$$

для любого $B\in\mathscr{B}(R)$.

Свойство 5). Если $\mathsf{M}|\xi|^n<\infty$, то в силу неравенств Ляпунова (6.28) $\mathsf{M}|\xi|^r<\infty$, $r\leqslant n$.

Рассмотрим отношение

$$\frac{\varphi(t+h)-\varphi(t)}{h}=\mathsf{M}e^{it\xi}\left(\frac{e^{ih\xi}-1}{h}\right).$$

Поскольку

$$\frac{|e^{ihx}-1|}{h}\leqslant x,$$

и $\mathsf{M}|\xi|<\infty$, то по теореме о мажорируемой сходимости существует

$$\lim_{h\to 0}\mathsf{M}e^{it\xi}\left(\frac{e^{ih\xi}-1}{h}\right),$$

равный

$$\mathsf{M}e^{it\xi}\lim_{h\to 0}\left(\frac{e^{ih\xi}-1}{h}\right)=i\mathsf{M}(\xi e^{it\xi})=i\int_{-\infty}^{\infty}xe^{itx}\,dF(x). \qquad (16)$$

Поэтому существует производная $\varphi'(t)$ и

$$\varphi'(t) = i\,(\mathsf{M}\xi e^{it\xi}) = i\int_{-\infty}^{\infty} x e^{itx}\,dF(x).$$

Существование производных $\varphi^{(r)}(t)$, $1 < r \leqslant n$, и справедливость формул (12) устанавливаются по индукции.

Формулы (13) следуют непосредственно из (12). Установим справедливость представления (14).

Поскольку для действительных y

$$e^{iy} = \cos y + i\sin y = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}[\cos\theta_1 y + i\sin\theta_2 y],$$

где $|\theta_1| \leqslant 1$, $|\theta_2| \leqslant 1$, то

$$e^{it\xi} = \sum_{k=0}^{n-1} \frac{(it\xi)^k}{k!} + \frac{(it\xi)^n}{n!}[\cos\theta_1(\omega)\,t\xi + i\sin\theta_2(\omega)\,t\xi] \tag{17}$$

и

$$\mathsf{M}e^{it\xi} = \sum_{k=0}^{n-1} \frac{(it)^k}{k!}\mathsf{M}\xi^k + \frac{(it)^n}{n!}[\mathsf{M}\xi^n + \varepsilon_n(t)], \tag{18}$$

где

$$\varepsilon_n(t) = \mathsf{M}\,[\xi^n(\cos\theta_1(\omega)\,t\xi + i\sin\theta_2(\omega)\,t\xi - 1)].$$

Ясно, что $|\varepsilon_n(t)| \leqslant 3\mathsf{M}|\xi^n|$, причем по теореме о мажорируемой сходимости $\varepsilon_n(t) \to 0$, $t \to 0$.

Свойство 6). Доказательство будем вести по индукции. Предположим сначала, что производная $\varphi''(0)$ существует и конечна. Покажем, что тогда $\mathsf{M}\xi^2 < \infty$. По правилу Лопиталя и лемме Фату

$$\varphi''(0) = \lim_{h\to 0}\frac{1}{2}\left[\frac{\varphi'(2h)-\varphi'(0)}{2h} + \frac{\varphi'(0)-\varphi'(-2h)}{2h}\right] =$$

$$= \lim_{h\to 0}\frac{2\varphi'(2h)-2\varphi'(-2h)}{8h} = \lim_{h\to 0}\frac{1}{4h^2}[\varphi(2h) - 2\varphi(0) + \varphi(-2h)] =$$

$$= \lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{e^{ihx}-e^{-ihx}}{2h}\right)^2 dF(x) =$$

$$= -\lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{\sin hx}{hx}\right)^2 x^2\,dF(x) \leqslant -\int_{-\infty}^{\infty}\lim_{h\to 0}\left(\frac{\sin hx}{hx}\right)^2 x^2\,dF(x) =$$

$$= -\int_{-\infty}^{\infty} x^2\,dF(x).$$

Поэтому

$$\int_{-\infty}^{\infty} x^2\,dF(x) \leqslant -\varphi''(0) < \infty.$$

Пусть теперь $\varphi^{(2k+2)}(0)$ существует, конечна и $\int_{-\infty}^{+\infty} x^{2k}\,dF(x) < \infty$. Если $\int_{-\infty}^{\infty} x^{2k}\,dF(x) = 0$, то и $\int_{-\infty}^{\infty} x^{2k+2}\,dF(x) = 0$. Так что будем предполагать, что $\int_{-\infty}^{\infty} x^{2k}\,dF(x) > 0$. Тогда, согласно свойству 5),

$$\varphi^{(2k)}(t) = \int_{-\infty}^{\infty} (ix)^{2k} e^{itx}\,dF(x)$$

и, значит,

$$(-1)^k \varphi^{(2k)}(t) = \int_{-\infty}^{\infty} e^{itx}\,dG(x),$$

где $G(x) = \int_{-\infty}^{x} u^{2k}\,dF(u)$.

Следовательно, функция $(-1)^k \varphi^{(2k)}(t)\,G^{-1}(\infty)$ является характеристической функцией вероятностного распределения $G(x)\cdot G^{-1}(\infty)$ и по доказанному

$$G^{-1}(\infty)\int_{-\infty}^{\infty} x^2\,dG(x) < \infty.$$

Но $G^{-1}(\infty) > 0$, значит,

$$\int_{-\infty}^{\infty} x^{2k+2}\,dF(x) = \int_{-\infty}^{\infty} x^2\,dG(x) < \infty.$$

Свойство 7). Пусть $0 < t_0 < R$. Тогда, используя формулу Стирлинга, находим, что

$$\overline{\lim}\,\frac{(\mathsf{M}\,|\xi|^n)^{1/n}}{n} < \frac{1}{t_0} \Rightarrow \overline{\lim}\,\frac{(\mathsf{M}\,|\xi|^n t_0^n)^{1/n}}{n} < 1 \Rightarrow \overline{\lim}\left(\frac{\mathsf{M}\,|\xi|^n t_0^n}{n!}\right)^{1/n} < 1.$$

Следовательно, по признаку Коши ряд $\sum \frac{\mathsf{M}\,|\xi|^n t_0^n}{n!}$ сходится, а значит, сходится и ряд $\sum_{r=0}^{\infty} \frac{(it)^r}{r!}\mathsf{M}\xi^r$ для любого $|t| \leqslant t_0$. Но в силу (14) для любого $n \geqslant 1$

$$\varphi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!}\mathsf{M}\xi^r + R_n(t),$$

где $|R_n(t)| \leqslant 3\frac{|t|^n}{n!}\mathsf{M}|\xi|^n$. Поэтому для всех $|t|<R$

$$\varphi(t)=\sum_{r=0}^{\infty}\frac{(it)^r}{r!}\mathsf{M}\xi^r.$$

Теорема доказана.

Замечание 1. Аналогично доказательству (14) устанавливается, что если для некоторого $n\geqslant 1$ $\mathsf{M}|\xi|^n<\infty$, то

$$\varphi(t)=\sum_{k=0}^{n}\frac{i^k(t-s)^k}{k!}\int_{-\infty}^{\infty}x^k e^{isx}\,dF(x)+\frac{i^n(t-s)^n}{n!}\varepsilon_n(t-s), \tag{19}$$

где $|\varepsilon_n(t-s)|\leqslant 3\mathsf{M}|\xi^n|$ и $\varepsilon_n(t-s)\to 0$, $t-s\to 0$.

Замечание 2. По поводу условия, фигурирующего в свойстве (7), см. также далее п. 9, посвященный вопросу об «единственности проблемы моментов».

4. Следующая теорема показывает, что характеристическая функция однозначно определяет функцию распределения.

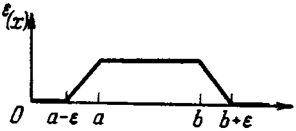


Рис. 33.

Теорема 2 (единственности). *Пусть F и G — две функции распределения, имеющие одну и ту же характеристическую функцию, т. е. для всех $t\in R$*

$$\int_{-\infty}^{\infty}e^{itx}\,dF(x)=\int_{-\infty}^{\infty}e^{itx}\,dG(x). \tag{20}$$

Тогда $F(x)\equiv G(x)$.

Доказательство. Зафиксируем a, $b\in R$, $\varepsilon>0$ и рассмотрим функцию $f^\varepsilon=f^\varepsilon(x)$, изображенную на рис. 33. Покажем, что

$$\int_{-\infty}^{\infty}f^\varepsilon(x)\,dF(x)=\int_{-\infty}^{\infty}f^\varepsilon(x)\,dG(x). \tag{21}$$

Пусть $n\geqslant 0$ таково, что $[a-\varepsilon,\ b+\varepsilon]\subseteq[-n,\ n]$, и последовательность $\{\delta_n\}$ такая, что $1\geqslant\delta_n\downarrow 0$, $n\to\infty$. Как всякая непрерывная на $[-n,\ n]$ функция с равными значениями в концевых точках, функция $f^\varepsilon=f^\varepsilon(x)$ может быть равномерно аппроксимирована (теорема Вейерштрасса — Стоуна) тригонометрическими полиномами, т. е. существует *конечная* сумма

$$f_n^\varepsilon(x)=\sum_k a_k\exp\left(i\pi x\frac{k}{n}\right) \tag{22}$$

такая, что

$$\sup_{-n\leqslant x\leqslant n}|f^\varepsilon(x)-f_n^\varepsilon(x)|\leqslant\delta_n. \tag{23}$$

Продолжим периодически функцию $f_n^\varepsilon(x)$ для всех $x \in R$ и заметим, что

$$\sup_x |f_n^\varepsilon(x)| \leqslant 2.$$

Тогда, поскольку в силу (20)

$$\int_{-\infty}^{\infty} f_n^\varepsilon(x)\, dF(x) = \int_{-\infty}^{\infty} f_n^\varepsilon(x)\, dG(x),$$

то

$$\left| \int_{-\infty}^{\infty} f^\varepsilon(x)\, dF(x) - \int_{-\infty}^{\infty} f^\varepsilon(x)\, dG(x) \right| = \left| \int_{-n}^{n} f^\varepsilon\, dF - \int_{-n}^{n} f^\varepsilon\, dG \right| \leqslant$$

$$\leqslant \left| \int_{-n}^{n} f_n^\varepsilon\, dF - \int_{-n}^{n} f_n^\varepsilon\, dG \right| + 2\delta_n \leqslant$$

$$\leqslant \left| \int_{-\infty}^{\infty} f_n^\varepsilon\, dF - \int_{-\infty}^{\infty} f_n^\varepsilon\, dG \right| + 2\delta_n + 2F(\overline{[-n, n]}) + 2G(\overline{[-n, n]}), \quad (24)$$

где $F(A) = \int_A dF(x)$, $G(A) = \int_A dG(x)$. При $n \to \infty$ правая часть в (24) стремится к нулю, что и доказывает равенство (21).

При $\varepsilon \to 0$ $f^\varepsilon(x) \to I_{(a, b]}(x)$. Поэтому по теореме о мажорируемой сходимости из (21) следует, что

$$\int_{-\infty}^{\infty} I_{(a, b]}(x)\, dF(x) = \int_{-\infty}^{\infty} I_{(a, b]}(x)\, dG(x),$$

т. е. $F(b) - F(a) = G(b) - G(a)$, откуда в силу произвольности a и b следует, что $F(x) = G(x)$ для всех $x \in R$.

Теорема доказана.

5. Предыдущая теорема говорит о том, что функция распределения $F = F(x)$ однозначно восстанавливается по своей характеристической функции $\varphi = \varphi(t)$. Следующая теорема дает явное представление функции F через φ.

Теорема 3 (формула обращения). *Пусть $F = F(x)$ — функция распределения и*

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

— ее характеристическая функция.

а) *Для любых двух точек a, b $(a < b)$, где функция $F = F(x)$ непрерывна,*

$$F(b) - F(a) = \lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt; \quad (25)$$

b) *Если* $\int_{-\infty}^{\infty}|\varphi(t)|\,dt<\infty$, *то функция распределения* $F(x)$ *имеет плотность* $f(x)$,

$$F(x)=\int_{-\infty}^{x} f(y)\,dy \tag{26}$$

и

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt. \tag{27}$$

Доказательство. Прежде всего отметим, что если функция $F(x)$ имеет плотность $f(x)$, то

$$\varphi(t)=\int_{-\infty}^{\infty} e^{itx}f(x)\,dx, \tag{28}$$

и поэтому формула (27) есть не что иное, как преобразование Фурье от (интегрируемой) функции $\varphi(t)$. Интегрируя левую и правые части (27) и применяя теорему Фубини, получим

$$F(b)-F(a)=\int_a^b f(x)\,dx=\frac{1}{2\pi}\int_a^b\left[\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt\right]dx=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\left[\int_a^b e^{-itx}\,dx\right]dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\,\frac{e^{-ita}-e^{-itb}}{it}\,dt.$$

После этих рассмотрений, объясняющих до некоторой степени формулу (25), перейдем к ее доказательству.

a) Имеем

$$\Phi_c\equiv\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\,\varphi(t)\,dt=$$

$$=\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\left[\int_{-\infty}^{\infty} e^{itx}\,dF(x)\right]dt=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\,e^{itx}\,dt\right]dF(x)=$$

$$=\int_{-\infty}^{\infty}\Psi_c(x)\,dF(x), \tag{29}$$

где мы положили

и воспользовались теоремой Фубини, справедливость которой в данном случае следует из того, что

$$\left|\frac{e^{-ita}-e^{-itb}}{it}\cdot e^{itx}\right| = \left|\frac{e^{-ita}-e^{-itb}}{it}\right| = \left|\int_a^b e^{-itx}\,dx\right| \leqslant b-a$$

и

$$\int_{-c}^{c}\int_{-\infty}^{\infty}(b-a)\,dF(x) \leqslant 2c(b-a) < \infty.$$

Далее

$$\Psi_c(x) = \frac{1}{2\pi}\int_{-c}^{c}\frac{\sin t(x-a)-\sin t(x-b)}{t}\,dt =$$

$$= \frac{1}{2\pi}\int_{-c(x-a)}^{c(x-a)}\frac{\sin v}{v}\,dv - \frac{1}{2\pi}\int_{-c(x-b)}^{c(x-b)}\frac{\sin u}{u}\,du. \tag{30}$$

Функция

$$g(s,\ t) = \int_s^t \frac{\sin v}{v}\,dv$$

равномерно непрерывна по s и t и

$$g(s,\ t) \to \pi \tag{31}$$

при $s\downarrow-\infty$ и $t\uparrow\infty$. Поэтому существует такая константа C, что для всех c и x $|\Psi_c(x)| < C < \infty$. Кроме того, из (30) и (31) следует, что

$$\Psi_c(x) \to \Psi(x), \quad c \to \infty,$$

где

$$\Psi(x) = \begin{cases} 0, & x<a, \quad x>b, \\ 1/2, & x=a, \quad x=b, \\ 1, & a<x<b. \end{cases}$$

Пусть μ — мера на $(R,\ \mathscr{B}(R))$ такая, что $\mu(a,\ b] = F(b) - F(a)$. Тогда, применяя теорему о мажорируемой сходимости и пользуясь формулами задачи 1 в § 3, находим, что при $c \to \infty$

$$\Phi_c = \int_{-\infty}^{\infty}\Psi_c(x)\,dF(x) \to \int_{-\infty}^{\infty}\Psi(x)\,dF(x) =$$

$$= \mu(a,\ b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\} =$$

$$= F(b-) - F(a) + \frac{1}{2}[F(a) - F(a-) + F(b) - F(b-)] =$$

$$= \frac{F(b)+F(b-)}{2} - \frac{F(a)+F(a-)}{2} = F(b) - F(a),$$

где последнее равенство справедливо для любых точек a и b, являющихся точками непрерывности функции $F(x)$.

Итак, формула (25) доказана.

b) Пусть $\int\limits_{-\infty}^{\infty} |\varphi(t)|\,dt < \infty$. Обозначим

$$f(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} \varphi(t)\,dt.$$

Из теоремы о мажорируемой сходимости следует, что эта функция непрерывна по x и, следовательно, она интегрируема на интервале $[a, b]$. Поэтому, снова применяя теорему Фубини, находим, что

$$\int\limits_a^b f(x)\,dx = \int\limits_a^b \frac{1}{2\pi}\left(\int\limits_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt\right)dx =$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \varphi(t)\left[\int\limits_a^b e^{-itx}\,dx\right]dt = \lim_{c\to\infty}\frac{1}{2\pi}\int\limits_{-c}^{c}\varphi(t)\left[\int\limits_a^b e^{-itx}\,dx\right]dt =$$

$$= \lim_{c\to\infty}\frac{1}{2\pi}\int\limits_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,dt = F(b) - F(a)$$

для всех точек a и b, являющихся точками непрерывности функции $F(x)$.

Отсюда вытекает, что

$$F(x) = \int\limits_{-\infty}^{x} f(y)\,dy, \qquad x \in R,$$

и так как $f(x)$ — непрерывная, а $F(x)$ — неубывающая функции, то $f(x)$ есть плотность $F(x)$.

Теорема доказана.

Следствие. Формула обращения (25) дает другое доказательство утверждения теоремы 2.

Теорема 4. *Для того чтобы компоненты случайного вектора $\xi = (\xi_1, \ldots, \xi_n)$ были независимы, необходимо и достаточно, чтобы его характеристическая функция была произведением характеристических функций компонент:*

$$\mathsf{M}e^{i(t_1\xi_1 + \ldots + t_n\xi_n)} = \prod_{k=1}^{n} \mathsf{M}e^{it_k\xi_k}, \qquad (t_1, \ldots, t_n) \in R^n.$$

Доказательство. Необходимость следует из задачи 1. Для доказательства достаточности обозначим $F = F(x_1, \ldots, x_n)$ — функцию распределения вектора $\xi = (\xi_1, \ldots, \xi_n)$ и $F_k(x)$ — функцию распределения ξ_k, $1 \leqslant k \leqslant n$. Положим $G = G(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n)$. Тогда по теореме Фубини для всех $(t_1, \ldots, t_n) \in R^n$

$$\int_{R^n} e^{i(t_1x_1+\ldots+t_nx_n)}\, dG(x_1 \ldots x_n) = \prod_{k=1}^{n} \int_R e^{it_kx_k}\, dF_k(x) =$$

$$= \prod_{k=1}^{n} \mathsf{M} e^{it_k\xi_k} = \mathsf{M} e^{i(t_1\xi_1+\ldots+t_k\xi_k)} = \int_{R^n} e^{i(t_1x_1+\ldots+t_nx_n)}\, dF(x_1 \ldots x_n).$$

Поэтому по теореме 2 (точнее по ее многомерному аналогу; см. задачу 3) $F = G$, и, следовательно, согласно теореме из § 5, величины $\xi_1, \ldots, \xi_n$ независимы.

6. В теореме 1 сформулированы некоторые необходимые условия, которым удовлетворяет характеристическая функция. Таким образом, если для функции $\varphi = \varphi(t)$ не выполняется, скажем, одно из первых трех утверждений этой теоремы, то это означает, что рассматриваемая функция не является характеристической.

Сложнее обстоит дело с проверкой того, является ли интересующая нас функция $\varphi = \varphi(t)$ характеристической. Сформулируем (без доказательства) ряд результатов в этом направлении.

Теорема Бохнера—Хинчина. *Пусть $\varphi(t)$ — непрерывная функция, $t \in R$, и $\varphi(0) = 1$. Для того чтобы $\varphi(t)$ была характеристической, необходимо и достаточно, чтобы она была неотрицательно-определенной, т. е. для любых действительных $t_1, \ldots, t_n$ и любых комплексных чисел $\lambda_1, \ldots, \lambda_n$, $n = 1, 2, \ldots,$*

$$\sum_{i,j=1}^{n} \varphi(t_i - t_j)\lambda_i\bar{\lambda}_j \geqslant 0. \tag{32}$$

Необходимость условия (32) очевидна, поскольку, если

$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$, то

$$\sum_{i,j=1}^{n} \varphi(t_i - t_j)\lambda_i\bar{\lambda}_j = \int_{-\infty}^{\infty} \left| \sum_{k=1}^{n} \lambda_k e^{it_kx} \right|^2 dF(x) \geqslant 0.$$

Труднее доказывается достаточность условия (32).

Теорема Пойа. *Пусть непрерывная, четная и выпуклая книзу функция $\varphi(t)$ такова, что $\varphi(t) \geqslant 0$, $\varphi(0) = 1$, $\varphi(t) \to 0$ при $t \to \infty$. Тогда $\varphi(t)$ является характеристической функцией.*

Эта теорема дает весьма удобный способ конструирования функций, являющихся характеристическими. Таковыми будут, например, функции

$$\varphi_1(t) = e^{-|t|},$$

$$\varphi_2(t) = \begin{cases} 1 - |t|, & |t| \leqslant 1, \\ 0, & |t| > 1. \end{cases}$$

Таковой будет и функция $\varphi_3(t)$, изображенная на рис. 34. На интервале $[-a, a]$ функция $\varphi_3(t)$ совпадает с функцией $\varphi_2(t)$. Однако отвечающие им функции распределения F_2 и F_3, очевидно, различны. Этот пример показывает, что для совпадения функций распределения недостаточно, вообще говоря, совпадения их характеристических функций на конечном интервале.

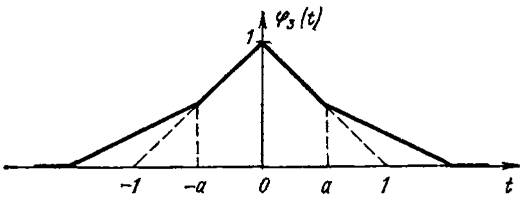


Рис. 34.

Теорема Марцинкевича. *Если характеристическая функция $\varphi(t)$ имеет вид* $\exp \mathscr{P}(t)$, *где* $\mathscr{P}(t)$ — *полином, то степень этого полинома не может быть больше двух.*

Из этой теоремы вытекает, например, что функция e^{-t^4} *не является* характеристической функцией.

7. Следующая теорема является примером результата, показывающего, как по свойствам характеристической функции случайной величины могут быть сделаны нетривиальные заключения о структуре этой величины.

Теорема 5. *Пусть $\varphi_\xi(t)$ — характеристическая функция случайной величины ξ.*

a) *Если $|\varphi_\xi(t_0)| = 1$ для некоторого $t_0 \neq 0$, то случайная величина ξ является решетчатой с шагом $h = \frac{2\pi}{t_0}$, т. е.*

$$\sum_{n=-\infty}^{\infty} \mathsf{P}\{\xi = a + nh\} = 1, \tag{33}$$

где a — некоторая константа.

b) *Если $|\varphi_\xi(t)| = |\varphi_\xi(\alpha t)| = 1$ для двух различных точек t и αt, где α — иррациональное число, то случайная величина ξ является вырожденной:*

$$\mathsf{P}\{\xi = a\} = 1,$$

где a — некоторая константа.

c) *Если* $|\varphi_\xi(t)| \equiv 1$, *то случайная величина* ξ *вырождена.*

Доказательство. a) Если $|\varphi_\xi(t_0)| = 1$, $t_0 \neq 0$, то найдется число a такое, что для этого t_0 $\varphi(t_0) = e^{it_0 a}$. Тогда

$$e^{it_0 a} = \int_{-\infty}^{\infty} e^{it_0 x}\, dF(x) \Rightarrow 1 = \int_{-\infty}^{\infty} e^{it_0(x-a)}\, dF(x) \Rightarrow 1 =$$

$$= \int_{-\infty}^{\infty} \cos t_0(x-a)\, dF(x) \Rightarrow \int_{-\infty}^{\infty} [1 - \cos t_0(x-a)]\, dF(x) = 0.$$

Поскольку $1 - \cos t_0(x-a) \geqslant 0$, то из свойства H (п. 2 § 6) следует, что (P-п. н.)

$$1 = \cos t_0(\xi - a),$$

что эквивалентно соотношению (33).

b) Из предположения $|\varphi_\xi(t)| = |\varphi_\xi(\alpha t)| = 1$ и (33) следует, что

$$\sum_{n=-\infty}^{\infty} \mathsf{P}\left\{\xi = a + \frac{2\pi}{t} n\right\} = \sum_{m=-\infty}^{\infty} \mathsf{P}\left\{\xi = b + \frac{2\pi}{\alpha t} m\right\} = 1.$$

Если ξ не является вырожденной, то тогда во множествах

$$\left\{a + \frac{2\pi}{t} n,\ n = 0, \pm 1, \ldots\right\} \quad \text{и} \quad \left\{b + \frac{2\pi}{\alpha t} m,\ m = 0, \pm 1, \ldots\right\}$$

найдутся по крайней мере по две совпадающие точки:

$$a + \frac{2\pi}{t} n_1 = b + \frac{2\pi}{\alpha t} m_1, \quad a + \frac{2\pi}{t} n_2 = b + \frac{2\pi}{\alpha t} m_2,$$

откуда

$$\frac{2\pi}{t}(n_1 - n_2) = \frac{2\pi}{\alpha t}(m_1 - m_2),$$

что противоречит предположению об иррациональности числа α. Утверждение c) следует из b).

Теорема доказана.

8. Пусть $\xi = (\xi_1, \ldots, \xi_k)$ — случайный вектор,

$$\varphi_\xi(t) = \mathsf{M} e^{i(t,\xi)}, \qquad t = (t_1, \ldots, t_k),$$

— его характеристическая функция. Будем предполагать, что для некоторого $n \geqslant 1$ $\mathsf{M}|\xi_i|^n < \infty$, $i = 1, \ldots, k$. Из неравенства Гёльдера (6.29) и неравенства Ляпунова (6.27) отсюда следует, что существуют (смешанные) моменты $\mathsf{M}\left(\xi_1^{\nu_1} \ldots \xi_k^{\nu_k}\right)$ для всех неотрицательных $\nu_1, \ldots, \nu_k$ таких, что $\nu_1 + \ldots + \nu_k \leqslant n$.

Как и в теореме 1, из этого выводится существование и непрерывность частных производных

$$\frac{\partial^{\nu_1 + \ldots + \nu_k}}{\partial t_1^{\nu_1} \ldots \partial t_k^{\nu_k}} \varphi_\xi(t_1, \ldots, t_k)$$

для $\nu_1+\ldots+\nu_k \leqslant n$. Тогда, разлагая $\varphi_\xi(t_1, \ldots, t_k)$ в ряд Тейлора, найдем, что

$$\varphi_\xi(t_1, \ldots, t_k) = \sum_{\nu_1+\ldots+\nu_k \leqslant n} \frac{i^{\nu_1+\ldots+\nu_k}}{\nu_1! \ldots \nu_k!} m_\xi^{(\nu_1 \ldots \nu_k)} t_1^{\nu_1} \ldots t_k^{\nu_k} + o(|t|^n), \tag{34}$$

где $|t| = |t_1| + \ldots + |t_k|$ и

$$m_\xi^{(\nu_1 \ldots \nu_k)} = \mathsf{M}\xi_1^{\nu_1} \ldots \xi_k^{\nu_k}$$

— *смешанный момент порядка* $\nu = (\nu_1, \ldots, \nu_k)$.

Функция $\varphi_\xi(t_1, \ldots, t_k)$ непрерывна, $\varphi_\xi(0, \ldots, 0) = 1$, и поэтому в некоторой окрестности нуля ($|t| < \delta$) она не обращается в нуль. В этой окрестности существуют и являются непрерывными частные производные

$$\frac{\partial^{\nu_1+\ldots+\nu_k}}{\partial t_1^{\nu_1} \ldots \partial t_k^{\nu_k}} \ln \varphi_\xi(t_1, \ldots, t_k),$$

где под $\ln z$ понимается главное значение логарифма (если $z = re^{i\theta}$, то $\ln z$ полагается равным $\ln r + i\theta$). Поэтому $\ln \varphi_\xi(t_1, \ldots, t_k)$ может быть представлен по формуле Тейлора

$$\ln \varphi_\xi(t_1 \ldots t_k) = \sum_{\nu_1+\ldots+\nu_k \leqslant n} \frac{i^{\nu_1+\ldots+\nu_k}}{\nu_1! \ldots \nu_k!} s_\xi^{(\nu_1 \ldots \nu_k)} t_1^{\nu_1} \ldots t_k^{\nu_k} + o(|t|^n), \tag{35}$$

где коэффициенты $s_\xi^{(\nu_1 \ldots \nu_k)}$ называют (*смешанными*) *семиинвариантами* или *кумулянтами* порядка $\nu = (\nu_1 \ldots \nu_k)$ вектора $\xi = (\xi_1, \ldots, \xi_k)$.

Заметим, что если ξ и η — два *независимых* вектора, то

$$\ln \varphi_{\xi+\eta}(t) = \ln \varphi_\xi(t) + \ln \varphi_\eta(t), \tag{36}$$

и поэтому

$$s_{\xi+\eta}^{(\nu_1 \ldots \nu_k)} = s_\xi^{(\nu_1 \ldots \nu_k)} + s_\eta^{(\nu_1 \ldots \nu_k)}. \tag{37}$$

(Именно это свойство и оправдывает название «семиинварианты» для $s_\xi^{(\nu_1 \ldots \nu_k)}$.)

Чтобы упростить запись и придать формулам (34), (35) «одномерный» вид, введем следующие обозначения.

Если $\nu = (\nu_1, \ldots, \nu_k)$ — вектор с неотрицательными целочисленными компонентами, то положим

$$\nu! = \nu_1! \ldots \nu_k!, \quad |\nu| = \nu_1 + \ldots + \nu_k, \quad t^\nu = t_1^{\nu_1} \ldots t_k^{\nu_k}.$$

Пусть также $s_\xi^{(\nu)} = s_\xi^{(\nu_1 \ldots \nu_k)}$, $m_\xi^{(\nu)} = m_\xi^{(\nu_1 \ldots \nu_k)}$.

Тогда представления (34), (35) примут следующий вид.

$$\varphi_\xi(t) = \sum_{|\nu| \leqslant n} \frac{i^{|\nu|}}{\nu!} m_\xi^{(\nu)} t^\nu + o(|t|^n), \tag{38}$$

$$\ln \varphi_\xi(t) = \sum_{|\nu| \leqslant n} \frac{i^{|\nu|}}{\nu!} s_\xi^{(\nu)} t^\nu + o(|t|^n). \tag{39}$$

Следующая теорема и ее следствия дают *формулы связи моментов и семиинвариантов.*

Теорема 6. *Пусть* $\xi = (\xi_1, \ldots, \xi_k)$ — *случайный вектор с* $\mathsf{M}|\xi_i|^n < \infty$, $i = 1, \ldots, k$, $n \geqslant 1$. *Тогда для всех* $\nu = (\nu_1, \ldots, \nu_k)$ *с* $|\nu| \leqslant n$

$$m_\xi^{(\nu)} = \sum_{\lambda^{(1)} + \ldots + \lambda^{(q)} = \nu} \frac{1}{q!} \frac{\nu!}{\lambda^{(1)}! \ldots \lambda^{(q)}!} \prod_{p=1}^{q} s_\xi^{(\lambda^{(p)})}, \tag{40}$$

$$s_\xi^{(\nu)} = \sum_{\lambda^{(1)} + \ldots + \lambda^{(q)} = \nu} \frac{(-1)^{q-1}}{q} \frac{\nu!}{\lambda^{(1)} \ldots \lambda^{(q)}} \prod_{p=1}^{q} m_\xi^{(\lambda^{(p)})}, \tag{41}$$

где $\sum\limits_{\lambda^{(1)} + \ldots + \lambda^{(q)} = \nu}$ *означает суммирование по всем упорядоченным наборам целых неотрицательных векторов* $\lambda^{(p)}$, $|\lambda^{(p)}| > 0$, *дающих в сумме вектор* ν.

Доказательство. Поскольку

$$\varphi_\xi(t) = \exp(\ln \varphi_\xi(t)),$$

то, разлагая exp по формуле Тейлора и учитывая (39), получим

$$\varphi_\xi(t) = 1 + \sum_{q=1}^{n} \frac{1}{q!} \left(\sum_{1 \leqslant |\lambda| \leqslant n} \frac{i^{|\lambda|}}{\lambda!} s_\xi^{(\lambda)} t^\lambda \right)^q + o(|t|^n). \tag{42}$$

Сравнивая члены при t^λ в правых частях (38) и (42) и учитывая, что $|\lambda^{(1)}| + \ldots + |\lambda^{(q)}| = |\lambda^{(1)} + \ldots + \lambda^{(q)}|$, получаем формулу (40).

Далее,

$$\ln \varphi_\xi(t) = \ln \left[1 + \sum_{1 \leqslant |\lambda| \leqslant n} \frac{i^{|\lambda|}}{\lambda!} m_\xi^{(\lambda)} t^\lambda + o(|t|^n) \right]. \tag{43}$$

При малых z справедливо разложение

$$\ln(1 + z) = \sum_{q=1}^{n} \frac{(-1)^{q-1}}{q} z^q + o(z^q).$$

Применяя это разложение к (43) и приравнивая затем коэффициенты при t^{λ} с соответствующими коэффициентами в правой части (38), получим формулу (41).

Следствие 1. Справедливы следующие формулы, связывающие моменты и семиинварианты:

$$m_{\xi}^{(\nu)} = \sum_{\{r_1\lambda^{(1)}+\ldots+r_x\lambda^{(x)}=\nu\}} \frac{1}{r_1!\ldots r_x!} \frac{\nu!}{(\lambda^{(1)}!)^{r_1}\ldots(\lambda^{(x)}!)^{r_x}} \prod_{j=1}^{x} \left[s_{\xi}^{(\lambda^{(j)})}\right]^{r_j}, \tag{44}$$

$$s_{\xi}^{(\nu)} = \sum_{\{r_1\lambda^{(1)}+\ldots+r_x\lambda^{(x)}=\nu\}} \frac{(-1)^{q-1}(q-1)!}{r_1!\ldots r_x!} \frac{\nu!}{(\lambda^{(1)}!)^{r_1}\ldots(\lambda^{(x)}!)^{r_x}} \times \prod_{j=1}^{x} \left[m_{\xi}^{(\lambda^{(j)})}\right]^{r_j}, \tag{45}$$

где $\sum\limits_{\{r_1\lambda^{(1)}+\ldots+r_x\lambda^{(x)}=\nu\}}$ означает суммирование по всем неупорядоченным наборам различных целых неотрицательных векторов $\lambda^{(j)}$, $|\lambda^{(j)}|>0$ и по всем упорядоченным наборам целых положительных чисел r_j таким, что $r_1\lambda^{(1)}+\ldots+r_x\lambda^{(x)}=\nu$.

Для доказательства (44) предположим, что среди векторов $\lambda^{(1)}, \ldots, \lambda^{(q)}$, участвующих в формуле (40), r_1 векторов равны $\lambda^{(i_1)}, \ldots, r_x$ векторов равны $\lambda^{(i_x)}$ ($r_j>0$, $r_1+\ldots+r_x=q$), причем все векторы $\lambda^{(i_s)}$ различны. Существует ровно $\frac{q!}{r_1!\ldots r_x!}$ различных наборов векторов, совпадающих с точностью до порядка с набором $\{\lambda^{(1)}, \ldots, \lambda^{(q)}\}$. Но если два набора, скажем, $\{\lambda^{(1)}, \ldots, \lambda^{(q)}\}$ и $\{\bar{\lambda}^{(1)}, \ldots, \bar{\lambda}^{(q)}\}$, отличаются лишь порядком, то $\prod\limits_{p=1}^{q} s_{\xi}^{(\lambda^{(p)})} = \prod\limits_{p=1}^{q} s_{\xi}^{(\bar{\lambda}^{(p)})}$. Поэтому, отождествляя наборы, совпадающие с точностью до порядка, из (40) получаем (44).

Аналогичным образом из (41) выводится формула (45).

Следствие 2. Рассмотрим тот частный случай, когда $\nu = (1, \ldots, 1)$. В этом случае моменты $m_{\xi}^{(\nu)} \equiv \mathsf{M}\xi_1 \ldots \xi_k$ и соответствующие семиинварианты будем называть *простыми*.

Формулы связи простых моментов и семиинвариантов получаются из приведенных формул. Однако их удобнее записать по-другому.

Для этого введем следующие обозначения.

Пусть $\xi = (\xi_1, \ldots, \xi_k)$ — рассматриваемый вектор, $I_{\xi} = \{1, 2, \ldots, k\}$ — множество индексов компонент этого вектора. Если $I \subseteq I_{\xi}$, то через ξ_I будем обозначать вектор, состоящий из тех компонент вектора ξ, индексы которых принадлежат I. Пусть $\chi(I)$ — вектор $\{\chi_1, \ldots, \chi_n\}$, у которого $\chi_i = 1$, если $i \in I$, и $\chi_i = 0$, если $i \notin I$. Эти векторы находятся во взаимнооднозначном соответствии с множествами $I \subseteq I_\xi$. Поэтому обозначим

$$m_\xi(I) = m_\xi^{(\chi(I))}, \quad s_\xi(I) = s_\xi^{(\chi(I))}.$$

Иначе говоря, $m_\xi(I)$ и $s_\xi(I)$ являются простыми моментами и семиинвариантами подвектора ξ_I вектора ξ.

Далее, назовем *разбиением* множества I неупорядоченный набор непересекающихся непустых множеств I_p, такой, что $\sum_p I_p = I$.

С учетом этих обозначений имеют место формулы

$$m_\xi(I) = \sum_{\sum_{p=1}^{q} I_p = I} \prod_{p=1}^{q} s_\xi(I_p), \tag{46}$$

$$s_\xi(I) = \sum_{\sum_{p=1}^{q} I_p = I} (-1)^{q-1}(q-1)! \prod_{p=1}^{q} m_\xi(I_p). \tag{47}$$

Для доказательства представления (46) обратимся к формуле (44). Если $\nu = \chi(I)$ и $\lambda^{(1)} + \ldots + \lambda^{(q)} = \nu$, то $\lambda^{(p)} = \chi(I_p)$, $I_p \subseteq I$, все $\lambda^{(p)}$ различны, $\lambda^{(p)}! = \nu! = 1$ и каждому неупорядоченному набору $\{\chi(I_1), \ldots, \chi(I_q)\}$ взаимнооднозначно соответствует разбиение $I = \sum_{p=1}^{q} I_p$. Следовательно, из (44) следует (46).

Аналогичным образом из (45) выводится справедливость представления (47).

Пример 1. Пусть ξ — случайная величина ($k = 1$) и $m_n = m_\xi^{(n)} = \mathsf{M}\xi^n$, $s_n = s_\xi^{(n)}$. Тогда из (40) и (41) получаем следующие формулы:

$$\begin{aligned} m_1 &= s_1, \\ m_2 &= s_2 + s_1^2, \\ m_3 &= s_3 + 3s_1s_2 + s_1^3, \\ m_4 &= s_4 + 3s_2^2 + 4s_1s_3 + 6s_1^2s_2 + s_1^4, \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{48}$$

и

$$\begin{aligned} s_1 &= m_1 = \mathsf{M}\xi, \\ s_2 &= m_2 - m_1^2 = \mathsf{D}\xi, \\ s_3 &= m_3 - 3m_1m_2 + 2m_1^3, \\ s_4 &= m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4, \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{49}$$

Пример 2. Пусть $\xi \sim \mathcal{N}(m, \sigma^2)$. Поскольку, согласно (9),

$$\ln \varphi_\xi(t) = itm - \frac{t^2\sigma^2}{2},$$

то в силу (39) $s_1 = m$, $s_2 = \sigma^2$ и все семиинварианты, начиная с третьего, равны нулю, т. е. $s_n = 0$, $n \geqslant 3$.

Заметим, что в силу теоремы Марцинкевича функция вида $\exp \mathscr{P}(t)$, где $\mathscr{P}(t)$ — полином, может быть характеристической только в том случае, когда степень этого полинома не больше двух. Отсюда, в частности, вытекает, что гауссовское распределение является *единственным* распределением, обладающим тем свойством, что все его семиинварианты s_n, начиная с некоторого номера, обращаются в нуль.

Пример 3. Если ξ — пуассоновская случайная величина с параметром $\lambda > 0$, то, согласно (11),

$$\ln \varphi_\xi(t) = \lambda(e^{it} - 1).$$

Отсюда следует, что для всех $n \geqslant 1$

$$s_n = \lambda. \tag{50}$$

Пример 4. Пусть $\xi = (\xi_1, \ldots, \xi_k)$ — случайный вектор. Тогда

$$\begin{aligned} m_\xi(1) &= s_\xi(1), \\ m_\xi(1, 2) &= s_\xi(1, 2) + s_\xi(1)\, s_\xi(2), \\ m_\xi(1, 2, 3) &= s_\xi(1, 2, 3) + s_\xi(1, 2)\, s_\xi(3) + \\ &\quad + s_\xi(1, 3)\, {}_\xi(2) + \\ &\quad + s_\xi(2, 3)\, s_\xi(1) + s_\xi(1)\, s_\xi(2)\, s_\xi(3) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{51}$$

Эти формулы показывают, что простые моменты выражаются через простые семиинварианты весьма *симметричным* образом. Если положить $\xi_1 \equiv \xi_2 \equiv \ldots \equiv \xi_k$, то из них получатся, конечно, формулы (48).

Из (51) становится понятным «групповое» происхождение коэффициентов в формулах (48). Из (51) следует также, что

$$s_\xi(1, 2) = m_\xi(1, 2) - m_\xi(1)\, m_\xi(2) = \mathsf{M}\xi_1\xi_2 - \mathsf{M}\xi_1 \mathsf{M}\xi_2, \tag{52}$$

т. е. $s_\xi(1, 2)$ есть не что иное, как *ковариация* случайных величин ξ_1, ξ_2.

9. Пусть ξ — случайная величина с функцией распределения $F = F(x)$ и характеристической функцией $\varphi(t)$. Предположим, что существуют все моменты $m_n = \mathsf{M}\xi^n$, $n \geqslant 1$.

Из теоремы 2 следует, что характеристическая функция однозначно определяет распределение вероятностей. Поставим сейчас следующий вопрос (единственность проблемы моментов): *однозначно ли определяют моменты $\{m_n\}_{n\geq 1}$ распределение вероятностей?*

Точнее, пусть F и G — две функции распределения, у которых все моменты совпадают, т. е. для всех целых $n \geqslant 0$

$$\int_{-\infty}^{\infty} x^n dF(x) = \int_{-\infty}^{\infty} x^n dG(x). \tag{53}$$

Спрашивается, вытекает ли отсюда совпадение функций F и G?

Вообще говоря, ответ на этот вопрос отрицательный. Чтобы в этом убедиться, рассмотрим распределение F с плотностью

$$f(x) = \begin{cases} ke^{-\alpha x^\lambda}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

где $\alpha > 0$, $0 < \lambda < 1/2$, а константа k выбрана из соображений нормировки $\int_0^\infty f(x)\,dx = 1$.

Обозначим $\beta = \alpha \operatorname{tg} \lambda\pi$, и пусть $g(x) = 0$ для $x \leqslant 0$ и

$$g(x) = ke^{-\alpha x^\lambda}[1 + \varepsilon \sin(\beta x^\lambda)], \quad |\varepsilon| < 1, \quad x > 0.$$

Ясно, что $g(x) \geqslant 0$. Покажем, что при всех целых $n \geqslant 0$

$$\int_0^\infty x^n e^{-\alpha x^\lambda} \sin \beta x^\lambda \, dx = 0. \tag{54}$$

Известно, что для $p > 0$ и комплексных q с $\operatorname{Re} q > 0$

$$\int_0^\infty t^{p-1} e^{-qt}\,dt = \frac{\Gamma(p)}{q^p}.$$

Положим здесь $p = \frac{n+1}{\lambda}$, $q = \alpha + i\beta$, $t = x^\lambda$. Тогда

$$\int_0^\infty x^{\lambda\left(\frac{n+1}{\lambda} - 1\right)} e^{-(\alpha+i\beta)\,x^\lambda} \lambda x^{\lambda-1}\,dx =$$

$$= \lambda \int_0^\infty x^n e^{-(\alpha+i\beta)x^\lambda} dx = \lambda \int_0^\infty x^n e^{-\alpha x^\lambda} \cos \beta x^\alpha dx -$$

$$- i\lambda \int_0^\infty x^n e^{-\alpha x^\lambda} \sin \beta x^\lambda \, dx = \frac{\Gamma\left(\frac{n+1}{\lambda}\right)}{\alpha^{\frac{n+1}{\lambda}} (1 + i \operatorname{tg} \lambda\pi)^{\frac{n+1}{\lambda}}}. \tag{55}$$

Но

$$(1+i\operatorname{tg}\lambda\pi)^{\frac{n+1}{\lambda}}=(\cos\lambda\pi+i\sin\lambda\pi)^{\frac{n+1}{\lambda}}(\cos\lambda\pi)^{-\frac{n+1}{\lambda}}=$$

$$=e^{i\pi(n+1)}(\cos\lambda\pi)^{-\frac{n+1}{\lambda}}=\cos\pi(n+1)\cdot\cos(\lambda\pi)^{-\frac{n+1}{\lambda}},$$

поскольку $\sin\pi(n+1)=0$.

Тем самым правая часть в (55) является действительной и, значит, при всех целых $n\geqslant 0$ справедлива формула (54). Возьмем теперь в качестве $G(x)$ функцию распределения с плотностью $g(x)$. Тогда из (54) следует, что у функций распределения F и G все моменты совпадают, т. е. для всех целых $n\geqslant 0$ справедливы равенства (53).

Приведем теперь некоторые достаточные условия, обеспечивающие единственность проблемы моментов.

Теорема 7. *Пусть $F=F(x)$ — функция распределения и $\mu_n=\int_{-\infty}^{\infty}|x|^n\,dF(x)$. Если*

$$\varlimsup_{n\to\infty}\frac{\mu_n^{1/n}}{n}<\infty, \tag{56}$$

то моменты $\{m_n\}_{n\geqslant 1}$, где $m_n=\int_{-\infty}^{\infty}x^n d\,F(x)$, однозначно определяют функцию распределения $F=F(x)$.

Доказательство. Из (56) и утверждения 7) теоремы 1 следует, что найдется такое $t_0>0$, что для всех $|t|\leqslant t_0$ характеристическая функция $\varphi(t)=\int_{-\infty}^{\infty}e^{itx}\,dF(x)$ представима в виде

$$\varphi(t)=\sum_{k=0}^{\infty}\frac{(it)^k}{k!}m_k$$

и, следовательно, моменты $\{m_n\}_{n\geqslant 1}$ однозначно определяют значение характеристической функции $\varphi(t)$ для всех $|t|\leqslant t_0$.

Возьмем точку s с $|s|\leqslant t_0/2$. Тогда из (56), так же как и при доказательстве (15), показывается, что для всех $|t-s|\leqslant t_0$

$$\varphi(t)=\sum_{k=0}^{\infty}\frac{i^k(t-s)^k}{k!}\varphi^{(k)}(s),$$

где

$$\varphi^{(k)}(s)=i^k\int_{-\infty}^{\infty}x^k e^{isx}\,dF(x)$$

однозначно определяется по моментам $\{m_n\}_{n\geqslant 1}$. Следовательно, эти моменты определяют однозначно $\varphi(t)$ для всех $|t|\leqslant\frac{3}{2}t_0$.

Продолжая этот процесс, убеждаемся в том, что $\{m_n\}_{n\geq 1}$ определяют однозначно $\varphi(t)$ при всех t, а значит, и функцию распределения $F(x)$.

Теорема доказана.

Следствие 1. Моменты однозначно определяют распределение вероятностей, *сосредоточенное на конечном интервале.*

Следствие 2. Для единственности проблемы моментов достаточно, чтобы

$$\overline{\lim_{n\to\infty}} \frac{(m_{2n})^{1/2n}}{2n} < \infty. \tag{57}$$

Для доказательства достаточно заметить, что нечетные моменты оцениваются по четным, и затем воспользоваться условием (56).

Пример. Пусть $F(x)$ — функция нормального распределения,

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{t^2}{2\sigma^2}}\, dt.$$

Тогда $m_{2n+1} = 0$, $m_{2n} = \frac{(2n)!}{2^n n!}\sigma^{2n}$ и из (57) следует, что эти моменты являются моментами только нормального распределения.

Приведем в заключение (без доказательства)

Критерий Карлемана единственности проблемы моментов.

a) *Пусть* $\{m_n\}_{n\geq 1}$ — *моменты некоторого распределения вероятностей, причем*

$$\sum_{n=0}^{\infty} \frac{1}{(m_{2n})^{1/2n}} = \infty.$$

Тогда они определяют распределение вероятностей однозначно.

b) *Если* $\{m_n\}_{n\geq 1}$ — *моменты распределения, сосредоточенного на* $[0, \infty)$, *то для однозначности достаточно потребовать, чтобы*

$$\sum_{n=0}^{\infty} \frac{1}{(m_n)^{1/2n}} = \infty.$$

10. Задачи.

1. Пусть ξ и η — независимые случайные величины, $f(x) = f_1(x) + if_2(x)$, $g(x) = g_1(x) + ig_2(x)$, где $f_k(x)$, $g_k(x)$ — борелевские функции, $k = 1,2$. Показать, что если $\mathsf{M}|f(\xi)| < \infty$ $\mathsf{M}|g(\xi)| < \infty$, то

$$\mathsf{M}|f(\xi)g(\eta)| < \infty$$

и

$$\mathsf{M}f(\xi)g(\eta) = \mathsf{M}f(\xi) \cdot \mathsf{M}g(\eta).$$

2. Пусть $\xi=(\xi_1,\ldots,\xi_n)$ и $\mathsf{M}\|\xi\|^n<\infty$, где $\|\xi\|=+\sqrt{\sum\xi_i^2}$. Показать, что

$$\varphi_\xi(t)=\sum_{k=0}^{n}\frac{i^k}{k!}\mathsf{M}(t,\xi)^m+\varepsilon_n(t)\|t\|^n,$$

где $t=(t_1,\ldots,t_n)$ и $\varepsilon_n(t)\to 0$, $t\to 0$.

3. Доказать теорему 2 для n-мерных функций распределения $F=F_n(x_1,\ldots,x_n)$ и $G=G_n(x_1,\ldots,x_n)$.

4. Пусть $F=F(x_1,\ldots,x_n)$ — n-мерная функция распределения, $\varphi=\varphi(t_1,\ldots,t_n)$ — ее характеристическая функция. Используя обозначение (3.12), установить справедливость формулы обращения

$$\mathsf{P}(a,\ b]=\lim_{c\to\infty}\frac{1}{(2\pi)^n}\int_{-c}^{c}\prod_{k=1}^{n}\frac{e^{-it_ka_k}-e^{-it_kb_k}}{it_k}\varphi(t_1\ldots t_k)\,dt_1\ldots dt_k.$$

(Предполагается, что $(a,\ b]$ является интервалом непрерывности функции $\mathsf{P}(a,\ b]$, т. е. при всех $k=1,\ldots,n$ точки a_k, b_k являются точками непрерывности маргинальных функций распределения $F_k(x_k)$, полученных из $F(x_1,\ldots,x_n)$, если положить все переменные, за исключением x_k, равными $+\infty$.)

5. Пусть $\varphi_k(t)$, $k\geqslant 1$, — характеристические функции, а неотрицательные числа λ_k, $k\geqslant 1$, таковы, что $\sum\lambda_k=1$. Показать, что функция $\sum\lambda_k\varphi_k(t)$ является характеристической.

6. Если $\varphi(t)$ — характеристическая функция, то будут ли $\operatorname{Re}\varphi(t)$ и $\operatorname{Im}\varphi(t)$ — характеристическими функциями?

7. Пусть φ_1, φ_2, φ_3 — характеристические функции и $\varphi_1\varphi_2=\varphi_1\varphi_3$. Следует ли отсюда, что $\varphi_2=\varphi_3$?

8. Составить таблицу характеристических функций для распределений, приведенных в табл. 1 и 2 § 3.

9. Пусть ξ — целочисленная случайная величина и $\varphi_\xi(t)$ — ее характеристическая функция. Показать, что

$$\mathsf{P}(\xi=k)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-ikt}\varphi_\xi(t)\,dt,\qquad k=0,\ \pm 1,\ \pm 2\ldots$$

§ 13. Гауссовские системы

1. Гауссовские, или нормально распределенные, случайные величины, гауссовские процессы и системы играют исключительно важную роль в теории вероятностей и математической статистике. Объясняется это прежде всего справедливостью центральной предельной теоремы (§ 4 гл. III), частным случаем которой яв-

ляется теорема Муавра — Лапласа (§ 6 гл.I). Согласно этой теореме нормальное распределение носит универсальный характер в том смысле, что распределение суммы большого числа независимых случайных величин или случайных векторов, подчиняющихся не слишком стеснительным условиям, хорошо аппроксимируется этим распределением.

Именно это обстоятельство дает теоретическое объяснение распространенному в статистической практике «закону ошибок», выражающемуся в том, что ошибка измерения, слагающаяся из большого числа независимых «элементарных» ошибок, подчиняется нормальному распределению.

Многомерное гауссовское распределение описывается небольшим числом параметров, что является несомненным его достоинством при построении простых вероятностных моделей. Гауссовские случайные величины имеют конечный второй момент, и, следовательно, их свойства могут изучаться методами гильбертова пространства. Важным при этом оказывается то обстоятельство, что в гауссовском случае некоррелированность превращается в независимость, что дает возможность значительно усилить результаты «L^2-теории».

2. Напомним, что (согласно § 8) случайная величина $\xi=\xi(\omega)$ называлась гауссовской или нормально распределенной с параметрами m и σ^2 $(\xi\sim\mathscr{N}(m,\ \sigma^2))$, $|m|<\infty$, $\sigma^2>0$, если ее плотность $f_\xi(x)$ имеет следующий вид:

$$f_\xi(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{1}$$

где $\sigma=+\sqrt{\sigma^2}$.

При $\sigma\downarrow 0$ плотности $f_\xi(x)$ «сходятся к δ-функции, сосредоточенной в точке $x=m$». Поэтому естественно сказать, что случайная величина ξ нормально распределена с параметрами m и $\sigma^2=0$ $(\xi\sim\mathscr{N}(m,\ 0))$, если ξ такова, что $\mathsf{P}(\xi=m)=1$.

Можно дать, однако, такое определение, которое сразу будет охватывать как *невырожденный* $(\sigma^2>0)$, так и *вырожденный* $(\sigma^2=0)$ случаи. С этой целью рассмотрим характеристическую функцию $\varphi_\xi(t)\equiv\mathsf{M}e^{it\xi}$, $t\in R$.

Если $\mathsf{P}(\xi=m)=1$, то очевидно, что

$$\varphi_\xi(t)=e^{itm}, \tag{2}$$

а если $\xi\sim\mathscr{N}(m,\ \sigma^2)$, $\sigma^2>0$, то, согласно (12.9),

$$\varphi_\xi(t)=e^{itm-\frac{t^2\sigma^2}{2}} \tag{3}$$

Легко видеть, что при $\sigma^2=0$ правая часть (3) совпадает с правой частью (2). Отсюда и из теоремы 1 § 12 следует, что гауссовскую случайную величину ξ с параметрами m и σ^2 $(|m|<\infty,\ \sigma^2\geqslant 0)$ можно определить как такую величину, для которой характеристическая функция $\varphi_\xi(t)$ задается формулой (3). Подход, основанный на привлечении характеристических функций, особенно удобен в многомерном случае.

Пусть $\xi=(\xi_1, \ldots, \xi_n)$ — случайный вектор и

$$\varphi_\xi(t)=\mathsf{M}e^{i(t, \xi)},\ t=(t_1, \ldots, t_n)\in R^n, \tag{4}$$

— его характеристическая функция (см. определение 2 в § 12).

Определение 1. Случайный вектор $\xi=(\xi_1, \ldots, \xi_n)$ называется *гауссовским* или нормально распределенным, если его характеристическая функция $\varphi_\xi(t)$ имеет следующий вид:

$$\varphi_\xi(t)=e^{i(t, m)-\frac{1}{2}(\mathbb{R}t, t)}, \tag{5}$$

где $m=(m_1, \ldots, m_n)$, $|m_k|<\infty$ и $\mathbb{R}=\|r_{kl}\|$ — симметрическая неотрицательно определенная матрица порядка $n\times n$ (для краткости будем использовать обозначение: $\xi\sim\mathcal{N}(m, \mathbb{R})$).

В связи с данным определением возникает прежде всего вопрос о том, а является ли функция (5) характеристической? Покажем, что это действительно так.

С этой целью предположим сначала, что матрица $\mathbb{R}$ является невырожденной. Тогда определены обратная матрица $A=\mathbb{R}^{-1}$ и функция

$$f(x)=\frac{|A|^{1/2}}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}(A(x-m), (x-m))\right\}, \tag{6}$$

где $x=(x_1, \ldots, x_n)$, $|A|=\det A$. Эта функция является неотрицательной. Покажем, что

$$\int_{R^n} e^{i(t, x)} f(x)\,dx=e^{i(t, m)-\frac{1}{2}(\mathbb{R}t, t)},$$

или, что то же,

$$I_n\equiv\int_{R^n} e^{i(t, x-m)}\frac{|A|^{1/2}}{(2\pi)^{n/2}}e^{-\frac{1}{2}(A(x-m), (x-m))}dx=e^{-\frac{1}{2}(\mathbb{R}t, t)}. \tag{7}$$

Сделаем в интеграле замену переменных

$$x-m=\mathcal{O}u,\quad t=\mathcal{O}v,$$

где $\mathcal{O}$ — ортогональная матрица такая, что

$$\mathcal{O}\mathbb{R}\mathcal{O}=D,$$

— диагональная матрица с $d_i \geqslant 0$ (см. доказательство леммы в § 8). Поскольку $|\mathbb{R}| = \det \mathbb{R} \neq 0$, то $d_i > 0$, $i = 1, \ldots, n$. Поэтому

$$|A| = |\mathbb{R}^{-1}| = d_1^{-1} \ldots d_n^{-1}. \tag{8}$$

Далее (см. обозначения п.1 § 12)

$$\begin{aligned} i(t,\ x-m) - \frac{1}{2}(A(x-m),\ (x-m)) &= \\ &= i(\mathcal{O}v,\ \mathcal{O}u) - \frac{1}{2}(A\mathcal{O}u,\ \mathcal{O}u) = \\ &= i(\mathcal{O}v)^*\mathcal{O}u - \frac{1}{2}(\mathcal{O}u)^*A(\mathcal{O}u) = \\ &= iv^*u - \frac{1}{2}u^*\mathcal{O}^*A\mathcal{O}u = iv^*u - \frac{1}{2}u^*D^{-1}u. \end{aligned}$$

Вместе с (8) и (12.9) это дает

$$\begin{aligned} I_n &= \frac{1}{(2\pi)^{n/2}(d_1 \ldots d_n)^{1/2}} \int_{R^n} e^{iv^*u - \frac{1}{2}u^*D^{-1}u}\,du = \\ &= \prod_{k=1}^{n} \frac{1}{(2\pi d_k)^{1/2}} \int_{-\infty}^{\infty} e^{iv_k u_k - \frac{u_k^2}{2d_k}}\,du_k = \prod_{k=1}^{n} e^{-\frac{v_k^2 d_k}{2}} = \\ &= e^{-\frac{1}{2}v^*Dv} = e^{-\frac{1}{2}v^*\mathcal{O}^*\mathbb{R}\mathcal{O}v} = e^{-\frac{1}{2}t^*\mathbb{R}t} = e^{-\frac{1}{2}(\mathbb{R}t,\,t)}. \end{aligned}$$

Из (6) следует также, что

$$\int_{R^n} f(x)\,dx = 1. \tag{9}$$

Таким образом, функция (5) является характеристической функцией n-мерного (невырожденного) гауссовского распределения (см. п. 3 § 3).

Пусть теперь матрица $\mathbb{R}$ вырожденная. Возьмем $\varepsilon > 0$ и рассмотрим положительно определенную симметрическую матрицу $\mathbb{R}^\varepsilon \equiv \mathbb{R} + \varepsilon E$. Тогда по доказанному функция

$$\varphi^\varepsilon(t) = e^{itm - \frac{1}{2}(\mathbb{R}^\varepsilon t,\ t)}$$

является характеристической:

$$\varphi^\varepsilon(t) = \int_{R^n} e^{i(t,\ x)}\,dF_\varepsilon(x),$$

где $F_\varepsilon(x) = F_\varepsilon(x_1, \ldots, x_n)$ — n-мерная функция распределения.

При $\varepsilon \to 0$

$$\varphi^\varepsilon(t) \to \varphi(t) = e^{itm - \frac{1}{2}(\mathbb{R}t,t)}.$$

Предельная функция $\varphi(t)$ непрерывна в нулевой точке $(0, \ldots, 0)$. Поэтому, согласно теореме 1 и задаче 1 из § 3 гл. III, она является характеристической.

Итак, корректность определения 1 установлена.

3. Выясним смысл вектора m и матрицы $\mathbb{R} = \| r_{kl} \|$, входящих в характеристическую функцию (5).

Поскольку

$$\ln \varphi_\xi(t) = i(t, m) - \frac{1}{2}(\mathbb{R}t, t) = i \sum_{k=1}^{n} t_k m_k - \frac{1}{2} \sum_{k, l=1}^{n} r_{kl} t_k t_l, \tag{10}$$

то из (12.35) и формул связи моментов и семиинвариантов находим, что

$$m_1 = s_\xi^{(1, 0, \ldots, 0)} = \mathsf{M}\xi_1, \ldots, m_k = s_\xi^{(0, \ldots, 0, 1)} = \mathsf{M}\xi_k.$$

Аналогично

$$r_{11} = s_\xi^{(2, 0, \ldots, 0)} = \mathsf{D}\xi_1, \quad r_{12} = s_\xi^{(1, 1, 0 \ldots)} = \operatorname{cov}(\xi_1, \xi_2),$$

и вообще

$$r_{kl} = \operatorname{cov}(\xi_k, \xi_l).$$

Таким образом, m есть *вектор средних значений* ξ, а $\mathbb{R}$ — *матрица ковариаций*.

Если матрица $\mathbb{R}$ невырожденная, то к этому результату можно было бы прийти и иначе. Именно, в этом случае вектор ξ имеет плотность $f(x)$, задаваемую формулой (6). Тогда прямой подсчет показывает, что

$$\mathsf{M}\xi_k \equiv \int x_k f(x)\, dx = m_k, \tag{11}$$

$$\operatorname{cov}(\xi_k, \xi_l) = \int (x_k - m_k)(x_l - m_l) f(x)\, dx = r_{kl}.$$

4. Обратимся к рассмотрению некоторых свойств гауссовских векторов.

Теорема 1. a) *У гауссовского вектора некоррелированность его компонент эквивалентна их независимости;*

b) *вектор* $\xi = (\xi_1, \ldots, \xi_n)$ *является гауссовским тогда и только тогда, когда для любого вектора* $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\lambda_k \in R$, *случайные величины* $(\xi, \lambda) = \lambda_1 \xi_1 + \ldots + \lambda_n \xi_n$ *имеет гауссовское распределение.*

Доказательство. a) Если компоненты вектора $\xi = (\xi_1, \ldots, \xi_n)$ некоррелированы, то из вида характеристической функции $\varphi_\xi(t)$ следует, что она является произведением характеристических функций

$$\varphi_\xi(t) = \prod_{k=1}^{n} \varphi_{\xi_k}(t_k).$$

Поэтому в силу теоремы 4 § 12 компоненты $\xi_1, \ldots, \xi_n$ независимы.

Обратное утверждение очевидно, поскольку из независимости всегда следует некоррелированность.

b) Если ξ — гауссовский вектор, то из (5) следует, что

$$\mathsf{M}\exp\left\{it\left(\xi_1\lambda_1+\ldots+\xi_n\lambda_n\right)\right\}=\exp\left\{it\left(\sum\lambda_k m_k\right)-\right.$$
$$\left.-\frac{t^2}{2}\left(\sum r_{kl}\lambda_k\lambda_l\right)\right\},\qquad t\in R,$$

и, следовательно,

$$(\xi,\ \lambda)\sim\mathscr{N}\left(\sum\lambda_k m_k,\ \sum r_{kl}\lambda_k\lambda_l\right).$$

Обратно, гауссовость случайной величины $(\xi,\ \lambda)=\xi_1\lambda_1+\ldots+\xi_n\lambda_n$ означает, в частности, что

$$\mathsf{M}e^{i(\xi,\lambda)}=e^{i\mathsf{M}(\xi,\lambda)-\frac{\mathsf{D}(\xi,\lambda)}{2}}=e^{i\sum\lambda_k\mathsf{M}\xi_k-\frac{1}{2}\sum\lambda_k\lambda_l\operatorname{cov}(\xi_k,\ \xi_l)}.$$

В силу произвольности $\lambda_1,\ldots\lambda_n$ и из определения 1 отсюда следует, что вектор $\xi=(\xi_1,\ \ldots,\ \xi_n)$ — гауссовский.

Теорема доказана.

Замечание. Пусть $(\theta,\ \xi)$ — гауссовский вектор с $\theta=(\theta_1,\ldots,\theta_k)$, $\xi=(\xi_1,\ \ldots,\ \xi_k)$. Если векторы θ и ξ некоррелированы, т. е. $\operatorname{cov}(\theta_i,\ \xi_j)=0$, $i=1,\ \ldots,\ k$, $j=1,\ \ldots,\ l$, то они и независимы.

Доказательство — то же, что и для утверждения а) теоремы.

Пусть $\xi=(\xi_1,\ldots,\ \xi_n)$ — гауссовский вектор, и для простоты будем предполагать, что вектор средних значений является нулевым. Если $\operatorname{rang}\mathbb{R}=r<n$, то, как было показано в § 11, существует ровно $n-r$ линейных соотношений между величинами $\xi_1,\ \ldots,\ \xi_n$. В этом случае можно считать, что, скажем, величины $\xi_1,\ \ldots,\ \xi_r$ линейно независимы, а все остальные через них линейно выражаются. Поэтому все основные свойства вектора $\xi=(\xi_1,\ \ldots,\ \xi_n)$ определяются первыми r компонентами $(\xi_1,\ \ldots,\ \xi_r)$, для которых соответствующая матрица ковариаций уже является невырожденной.

Итак, можно считать, что исходный вектор $\xi=(\xi_1,\ldots,\ \xi_n)$ уже таков, что его компоненты линейно независимы и, значит, $|\mathbb{R}|>0$.

Пусть $\mathscr{O}$ — ортогональная матрица, приводящая $\mathbb{R}$ к диагональному виду

$$\mathscr{O}^*\mathbb{R}\mathscr{O}=D.$$

Все диагональные элементы матрицы D положительны, и, следовательно, определена обратная матрица. Положим $B^2=D$ и

$$\beta=B^{-1}\mathscr{O}^*\xi.$$

Тогда легко убедиться, что

$$\mathsf{M}e^{i(t,\ \beta)}=\mathsf{M}e^{i\beta^* t}=e^{-\frac{1}{2}(Et,\ t)},$$

т. е. вектор $\beta = (\beta_1, \ldots, \beta_n)$ — это гауссовский вектор с некоррелированными, а значит (теорема 1), и независимыми компонентами. Тогда, обозначая $A = \mathscr{O}B$, получаем, что исходный гауссовский вектор $\xi = (\xi_1, \ldots, \xi_n)$ представляется в виде

$$\xi = A\beta, \tag{12}$$

где $\beta = (\beta_1, \ldots, \beta_n)$ — гауссовский вектор с независимыми компонентами, $\beta_k \sim \mathscr{N}(0,1)$. Отсюда вытекает следующий результат. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — вектор с линейно независимыми компонентами, $\mathsf{M}\xi_k = 0$, $k = 1, \ldots, n$. Этот вектор является гауссовским тогда и только тогда, когда существуют независимые гауссовские величины $\beta_1, \ldots, \beta_n$, $\beta_k \sim \mathscr{N}(0,1)$, и невырожденная матрица A порядка n такие, что $\xi = A\beta$. При этом $\mathbb{R} = AA^*$ — матрица ковариаций вектора ξ.

Если $|\mathbb{R}| \neq 0$, то, согласно методу ортогонализации Грама — Шмидта (см. § 11),

$$\xi_k = \hat{\xi}_k + b_k\varepsilon_k, \qquad k = 1, \ldots, n, \tag{13}$$

где в силу гауссовости вектор $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k) \sim \mathscr{N}(0, E)$,

$$\hat{\xi}_k = \sum_{l=1}^{k-1} (\xi_k, \varepsilon_l)\,\varepsilon_l, \tag{14}$$

$$b_k = \|\xi_k - \hat{\xi}_k\| \tag{15}$$

и

$$\mathscr{L}\{\xi_1, \ldots, \xi_k\} = \mathscr{L}\{\varepsilon_1, \ldots, \varepsilon_k\}. \tag{16}$$

Из ортогонального разложения (13) сразу получаем, что

$$\hat{\xi}_k = \mathsf{M}(\xi_k \mid \xi_{k-1}, \ldots, \xi_1). \tag{17}$$

Вместе с (16) и (14) отсюда следует, что в гауссовском случае условное математическое ожидание $\mathsf{M}(\xi_k \mid \xi_{k-1}, \ldots, \xi_1)$ является линейной функцией от $(\xi_1, \ldots, \xi_{k-1})$:

$$\mathsf{M}(\xi_k \mid \xi_{k-1}, \ldots, \xi_1) = \sum_{i=1}^{k-1} a_i\xi_i. \tag{18}$$

(В случае $k = 2$ этот результат был установлен в § 8.)

Поскольку, согласно замечанию к теореме 1 § 8, $\mathsf{M}(\xi_k|\xi_{k-1}, \ldots, \xi_1)$ является оптимальной (в среднеквадратическом смысле) оценкой ξ_k по $\xi_1, \ldots, \xi_{k-1}$, то из (18) следует, что в гауссовском случае оптимальная оценка оказывается *линейной*.

Используем эти результаты для отыскания оптимальной оценки вектора $\theta = (\theta_1, \ldots, \theta_k)$ по вектору $\xi = (\xi_1, \ldots, \xi_l)$ в предположении, что (θ, ξ) — гауссовский вектор. Обозначим

$$m_\theta = \mathsf{M}\theta, \quad m_\xi = \mathsf{M}\xi$$

— векторы-столбцы средних значений и

$$D_{\theta\theta} \equiv \operatorname{cov}(\theta, \theta) \equiv \| \operatorname{cov}(\theta_i, \theta_j) \|, \quad 1 \leqslant i, j \leqslant k,$$
$$D_{\theta\xi} \equiv \operatorname{cov}(\theta, \xi) \equiv \| \operatorname{cov}(\theta_i, \xi_j) \|, \quad 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant l,$$
$$D_{\xi\xi} \equiv \operatorname{cov}(\xi, \xi) \equiv \| \operatorname{cov}(\xi_i, \xi_j) \|, \quad 1 \leqslant i, j \leqslant l,$$

— матрицы ковариаций. Предположим, что матрица $D_{\xi\xi}$ имеет обратную матрицу. Тогда справедлива следующая

Теорема 2 (теорема о нормальной корреляции). *Для гауссовского вектора* (θ, ξ) *оптимальная оценка* $\mathsf{M}(\theta \mid \xi)$ *вектора* θ *по* ξ *и ее матрица ошибок*

$$\Delta = \mathsf{M}[\theta - \mathsf{M}(\theta \mid \xi)][\theta - \mathsf{M}(\theta \mid \xi)]^*$$

задаются следующими формулами:

$$\mathsf{M}(\theta \mid \xi) = m_\theta + D_{\theta\xi} D_{\xi\xi}^{-1} (\xi - m_\xi), \tag{19}$$

$$\Delta = D_{\theta\theta} - D_{\theta\xi} D_{\xi\xi}^{-1} (D_{\theta\xi})^*. \tag{20}$$

Доказательство. Образуем вектор

$$\eta = (\theta - m_\theta) - D_{\theta\xi} D_{\xi\xi}^{-1} (\xi - m_\xi). \tag{21}$$

Тогда непосредственно проверяется, что $\mathsf{M}\eta(\xi - m_\xi)^* = 0$, т. е. вектор η не коррелирован с вектором $(\xi - m_\xi)$. Но в силу гауссовости (θ, ξ) вектор (η, ξ) также будет гауссовским. Отсюда в силу замечания к теореме 1 векторы η и $\xi - m_\xi$ независимы. Значит, независимы η и ξ и, следовательно, $\mathsf{M}(\eta \mid \xi) = \mathsf{M}\eta = 0$. Поэтому

$$\mathsf{M}[\theta - m_\theta \mid \xi] - D_{\theta\xi} D_{\xi\xi}^{-1} (\xi - m_\xi) = 0,$$

что и доказывает представление (19).

Для доказательства (20) рассмотрим условную ковариацию

$$\operatorname{cov}(\theta, \theta \mid \xi) \equiv \mathsf{M}[(\theta - \mathsf{M}(\theta \mid \xi))(\theta - \mathsf{M}(\theta \mid \xi))^* \mid \xi]. \tag{22}$$

Поскольку $\theta - \mathsf{M}(\theta \mid \xi) = \eta$, то в силу независимости η и ξ находим, что

$$\operatorname{cov}(\theta, \theta \mid \xi) = \mathsf{M}(\eta\eta^* \mid \xi) = \mathsf{M}\eta\eta^* =$$
$$= D_{\theta\theta} + D_{\theta\xi} D_{\xi\xi}^{-1} D_{\xi\xi} D_{\xi\xi}^{-1} D_{\theta\xi}^* - 2 D_{\theta\xi} D_{\xi\xi}^{-1} D_{\xi\xi} D_{\xi\xi}^{-1} D_{\theta\xi}^* = D_{\theta\theta} - D_{\theta\xi} D_{\xi\xi}^{-1} D_{\theta\xi}^*.$$

Поскольку $\operatorname{cov}(\theta, \theta \mid \xi)$ не зависит от «случая», то

$$\Delta = \mathsf{M} \operatorname{cov}(\theta, \theta \mid \xi) = \operatorname{cov}(\theta, \theta \mid \xi),$$

что и доказывает представление (20).

Следствие. Пусть $(\theta; \xi_1, \ldots, \xi_n)$ — $(n+1)$-мерный гауссовский вектор, причем $\xi_1, \ldots, \xi_n$ независимы. Тогда

$$\mathsf{M}(\theta \mid \xi_1, \ldots, \xi_n) = \mathsf{M}\theta + \sum_{i=1}^{n} \frac{\operatorname{cov}(\theta, \xi_i)}{\mathsf{D}\xi_i}(\xi_i - \mathsf{M}\xi_i),$$

$$\Delta = \mathsf{D}\theta - \sum_{i=1}^{n} \frac{\operatorname{cov}^2(\theta, \xi_i)}{\mathsf{D}\xi_i}$$

(ср. с формулами (8.12), (8.13)).

5. Пусть $\xi_1, \xi_2, \ldots$ — последовательность гауссовских случайных векторов, сходящаяся по вероятности к вектору ξ. Покажем, что вектор ξ также является гауссовским.

В соответствии с утверждением а) теоремы 1 достаточно показать это лишь для случайных величин.

Пусть $m_n = \mathsf{M}\xi_n$, $\sigma_n^2 = \mathsf{D}\xi_n$. Тогда по теореме Лебега о мажорируемой сходимости

$$\lim_{n\to\infty} e^{itm_n - \frac{1}{2}\sigma_n^2 t^2} = \lim_{n\to\infty} \mathsf{M}e^{it\xi_n} = \mathsf{M}e^{it\xi}.$$

Из существования предела в левой части вытекает, что найдутся такие m и σ^2, что

$$m = \lim_{n\to\infty} m_n, \quad \sigma^2 = \lim_{n\to\infty} \sigma_n^2.$$

Следовательно,

$$\mathsf{M}e^{it\xi} = e^{itm - \frac{1}{2}\sigma^2 t^2},$$

т. е. $\xi \sim \mathscr{N}(m, \sigma^2)$.

Отсюда, в частности, вытекает, что замкнутое линейное многообразие $\overline{\mathscr{L}}\{\xi_1, \xi_2, \ldots\}$, порожденное гауссовскими величинами $\xi_1, \xi_2, \ldots$ (см. п. 5 § 11), состоит из гауссовских величин.

6. Перейдем теперь к определению общих гауссовских систем.

Определение 2. Совокупность случайных величин $\xi = \{\xi_\alpha\}$, где α принадлежит некоторому множеству индексов $\mathfrak{A}$, называется *гауссовской системой*, если для любого $n \geqslant 1$ и любых $\alpha_1, \ldots, \alpha_n$ из $\mathfrak{A}$ случайный вектор $(\xi_{\alpha_1}, \ldots, \xi_{\alpha_n})$ является гауссовским.

Отметим некоторые свойства гауссовских систем.

a) Если $\xi = (\xi_\alpha)$, $\alpha \in \mathfrak{A}$, — гауссовская система, то всякая ее подсистема $\xi' = (\xi_{\alpha'})$, $\alpha' \in \mathfrak{A}' \subseteq \mathfrak{A}$, также является гауссовской.

b) Если ξ_α, $\alpha \in \mathfrak{A}$, — независимые гауссовские величины, то система $\xi = (\xi_\alpha)$, $\alpha \in \mathfrak{A}$, является гауссовской.

c) Если $\xi = (\xi_\alpha)$, $\alpha \in \mathfrak{A}$, — гауссовская система, то замкнутое линейное многообразие $\overline{\mathscr{L}}(\xi)$, состоящее из величин вида $\sum_{i=1}^{n} c_{\alpha_i}\xi_{\alpha_i}$

и их пределов в среднеквадратическом смысле, образует гауссовскую систему.

Заметим, что утверждение, обратное к свойству а), вообще говоря, неверно. Например, пусть ξ_1 и η_1 независимы и $\xi_1 \sim \mathscr{N}(0, 1)$, $\eta_1 \sim \mathscr{N}(0, 1)$. Определим систему

$$(\xi, \eta) = \begin{cases} (\xi_1, |\eta_1|), & \text{если } \xi_1 \geqslant 0, \\ (\xi_1, -|\eta_1|), & \text{если } \xi_1 < 0. \end{cases} \tag{23}$$

Тогда нетрудно проверить, что каждая из величин ξ и η гауссовская, а вектор (ξ, η) гауссовским не является.

Пусть $\xi = \xi_\alpha$, $\alpha \in \mathfrak{A}$, — некоторая гауссовская система с вектором средних значений $m = (m_\alpha)$, $\alpha \in \mathfrak{A}$, и матрицей ковариаций $\mathbb{R} = (r_{\alpha\beta})_{\alpha, \beta \in \mathfrak{A}}$, где $m_\alpha = \mathsf{M}\xi_\alpha$. Матрица $\mathbb{R}$ является, очевидно, симметрической ($r_{\alpha\beta} = r_{\beta\alpha}$) и неотрицательно-определенной в том смысле, что для любого вектора $c = (c_\alpha)_{\alpha \in \mathfrak{A}}$ со значениями в $R^{\mathfrak{A}}$, у которого лишь конечное число координат c_α отлично от нуля,

$$(\mathbb{R}c, c) \equiv \sum_{\alpha, \beta} r_{\alpha\beta} c_\alpha c_\beta \geqslant 0. \tag{24}$$

Поставим сейчас обратный вопрос. Пусть задано некоторое параметрическое множество $\mathfrak{A} = \{\alpha\}$, вектор $m = (m_\alpha)_{\alpha \in \mathfrak{A}}$ и симметрическая неотрицательно-определенная матрица $\mathbb{R} = (r_{\alpha\beta})_{\alpha, \beta \in \mathfrak{A}}$. Спрашивается, существует ли вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ и на нем гауссовская система случайных величин $\xi = (\xi_\alpha)_{\alpha \in \mathfrak{A}}$ такие, что

$$\mathsf{M}\xi_\alpha = m_\alpha,$$
$$\operatorname{cov}(\xi_\alpha, \xi_\beta) = r_{\alpha\beta}, \quad \alpha, \beta \in \mathfrak{A}?$$

Если взять конечный набор $\alpha_1, \ldots, \alpha_n$, то по вектору $\overline{m} = (m_{\alpha_1}, \ldots, m_{\alpha_n})$ и матрице $\overline{R} = (r_{\alpha\beta})$, $\alpha, \beta = \alpha_1, \ldots, \alpha_n$, в R^n можно построить гауссовское распределение $F_{\alpha_1 \ldots \alpha_n}(x_1 \ldots x_n)$ с характеристической функцией

$$\varphi(t) = e^{i(t, \overline{m}) - \frac{1}{2}(\overline{\mathbb{R}}t, t)}, \quad t = (t_{\alpha_1}, \ldots, t_{\alpha_n}).$$

Нетрудно проверить, что семейство

$$\{F_{\alpha_1, \ldots \alpha_n}(x_1, \ldots, x_n);\ \alpha_i \in \mathfrak{A}\}$$

является согласованным. Следовательно, по теореме Колмогорова (теорема 1 § 9 и замечание 2 к ней) ответ на поставленный выше вопрос является положительным.

7. Если $\mathfrak{A}=\{1, 2, \ldots\}$, то в соответствии с терминологией, принятой в § 5, систему случайных величин $\xi=(\xi_\alpha)_{\alpha\in\mathfrak{A}}$, будем называть *случайной последовательностью* и обозначать $\xi=(\xi_1, \xi_2, \ldots)$. Гауссовская последовательность полностью описывается вектором средних значений $m=(m_1, m_2, \ldots)$ и матрицей ковариаций $\mathbb{R}=\|r_{ij}\|$, $r_{ij}=\operatorname{cov}(\xi_i, \xi_j)$. В частности, если $r_{ij}=\sigma_i^2\delta_{ij}$, то $\xi=(\xi_1, \xi_2, \ldots)$ есть гауссовская последовательность независимых случайных величин с $\xi_i \sim \mathcal{N}(m_i, \sigma_i^2)$, $i\geqslant 1$.

В том случае, когда $\mathfrak{A}=[0, 1]$, $[0, \infty)$, $(-\infty, \infty)$ ..., систему величин $\xi=(\xi_t)$, $t\in\mathfrak{A}$, называют *случайным процессом с непрерывным временем*.

Остановимся на некоторых примерах гауссовских случайных процессов. Если считать их средние значения равными нулю, то полное описание вероятностных свойств таких процессов определяется видом матрицы ковариации $\|r_{st}\|$. Будем обозначать r_{st} через $r(s, t)$ и называть эту функцию от s и t *ковариационной функцией*.

Пример 1. Если $T=[0, \infty)$ и

$$r(s, t)=\min(s, t), \tag{25}$$

то гауссовский процесс $\xi=(\xi_t)_{t\geqslant 0}$ с такой функцией ковариации (см. задачу 2) и $\xi_0\equiv 0$ называется *процессом броуновского движения (винеровским процессом)*.

Отметим, что этот процесс имеет *независимые приращения*, т. е. для любых $t_1<t_2<\ldots<t_n$ случайные величины

$$\xi_{t_2}-\xi_{t_1}, \ldots, \xi_{t_n}-\xi_{t_{n-1}}$$

являются независимыми. В самом деле, в силу гауссовости достаточно проверить лишь попарную некоррелированность приращений. Но если $s<t<u<v$, то

$$\mathsf{M}[\xi_t-\xi_s][\xi_v-\xi_u]=$$
$$=[r(t, v)-r(t, u)]-[r(s, v)-r(s, u)]=(t-t)-(s-s)=0.$$

Пример 2. Процесс $\xi=(\xi_t)$, $0\leqslant t\leqslant 1$, с $\xi_0\equiv 0$ и

$$r(s, t)=\min(s, t)-st \tag{26}$$

называется *условным винеровским процессом* (заметим, что поскольку $r(1, 1)=0$, то $\mathsf{P}(\xi_1=0)=1$).

Пример 3. Процесс $\xi=(\xi_t)$, $-\infty<t<\infty$, с

$$r(s, t)=e^{-|t-s|} \tag{27}$$

называют *гауссовско-марковским*.

8. Задачи.

1. Пусть ξ_1, ξ_2, ξ_3 — независимые гауссовские случайные величины, $\xi_i \sim \mathcal{N}(0, 1)$. Показать, что

$$\frac{\xi_1+\xi_2\xi_3}{\sqrt{1+\xi_3^2}} \sim \mathcal{N}(0, 1).$$

(В этой связи возникает интересная задача описания всех *нелинейных* преобразований от независимых гауссовских величин $\xi_1, \ldots, \xi_n$, распределение которых также является гауссовским.)

2. Доказать, что функции (25), (26), (27) являются неотрицательно-определенными (и, следовательно, действительно являются ковариационными функциями).

3. Пусть A — некоторая матрица порядка $m\times n$. Назовем матрицу $A^{\oplus}$ порядка $n\times m$ *псевдообратной* к матрице A, если найдутся такие матрицы U и V, что

$$AA^{\oplus}A = A, \quad A^{\oplus} = UA^* = A^*V.$$

Показать, что матрица $A^{\oplus}$, определяемая этими условиями, существует и единственна.

4. Показать, что формулы (19) и (20) в теореме о нормальной корреляции остаются справедливыми и в случае вырождения матрицы $\mathsf{D}_{\xi\xi}$, если в них вместо $\mathsf{D}_{\xi\xi}^{-1}$ рассматривать псевдообратную матрицу $\mathsf{D}_{\xi\xi}^{\oplus}$.

5. Пусть $(\theta, \xi) = (\theta_1 \ldots \theta_k; \xi_1, \ldots, \xi_l)$ — гауссовский вектор с невырожденной матрицей $\Delta \equiv \mathsf{D}_{\theta\theta} - \mathsf{D}_{\xi\xi}^{\oplus}\mathsf{D}_{\theta\xi}^*$. Показать, что у функции распределения $\mathsf{P}(\theta \leqslant a \mid \xi) = \mathsf{P}(\theta_1 \leqslant a_1, \ldots, \theta_k \leqslant a_k \mid \xi)$ существует (P-п. н.) плотность $p(a_1, \ldots, a_k \mid \xi)$, определяемая формулой $p(a_1, \ldots, a_k \mid \xi) =$

$$= \frac{|\Delta|^{-1/2}}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}(a - \mathsf{M}(\theta \mid \xi))^* \Delta^{-1} (a - \mathsf{M}(\theta \mid \xi))\right\}.$$

6. (С. Н. Бернштейн). Пусть ξ и η — независимые одинаково распределенные случайные величины с конечной дисперсией. Доказать, что если $\xi+\eta$ и $\xi-\eta$ независимы, то ξ и η являются гауссовскими величинами.

ГЛАВА III

СХОДИМОСТЬ ВЕРОЯТНОСТНЫХ МЕР. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

§ 1. Слабая сходимость вероятностных мер и распределений

1. Многие из фундаментальных результатов теории вероятностей формулируются в виде предельных теорем. В форме предельной теоремы было сформулировано утверждение, названное законом больших чисел Я. Бернулли, эту форму имела теорема Муавра — Лапласа, с которых, собственно говоря, и началась истинная теория вероятностей.

В настоящей главе мы остановимся на двух основных аспектах теории предельных теорем: на понятии слабой сходимости и на методе характеристических функций, являющимся одним из самых мощных средств доказательства предельных теорем и их уточнений.

Напомним для начала формулировку закона больших чисел (гл. I, § 5) в схеме Бернулли.

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{P}(\xi_i = 1) = p$, $\mathsf{P}(\xi_i = 0) = q$, $p + q = 1$. Используя введенное в § 10 гл. II понятие сходимости по вероятности закон больших чисел Я. Бернулли можно сформулировать в следующей форме:

$$\frac{S_n}{n} \xrightarrow{\mathsf{P}} p, \qquad n \to \infty, \tag{1}$$

где $S_n = \xi_1 + \ldots + \xi_n$. (В гл. IV будет показано, что на самом деле здесь имеет место и сходимость с вероятностью единица.)

Обозначим

$$\begin{gathered} F_n(x) = \mathsf{P}\left\{\frac{S_n}{n} \leqslant x\right\}, \\ F(x) = \begin{cases} 1, & x \geqslant p, \\ 0, & x < p \end{cases} \end{gathered} \tag{2}$$

где $F(x)$ — функция распределения вырожденной случайной величины $\xi \equiv p$. Пусть также P_n и P — вероятностные меры на $(R, \mathscr{B}(R))$, отвечающие функциям распределения F_n и F.

В соответствии с теоремой 2 из § 10 главы II сходимость по вероятности $\frac{S_n}{n} \xrightarrow{\mathsf{P}} p$ влечет за собой сходимость по распределению $\frac{S_n}{n} \xrightarrow{d} p$, означающую, что

$$\mathsf{M} f\left(\frac{S_n}{n}\right) \to \mathsf{M} f(p), \quad n \to \infty, \tag{3}$$

для любой функции $f = f(x)$ из класса $\mathbb{C}(R)$ непрерывных ограниченных функций на R.

Поскольку

$$\mathsf{M} f\left(\frac{S_n}{n}\right) = \int_R f(x)\, \mathsf{P}_n(dx), \quad \mathsf{M} f(p) = \int_R f(x)\, \mathsf{P}(dx),$$

то (3) можно переписать в форме

$$\int_R f(x)\, \mathsf{P}_n(dx) \to \int_R f(x)\, \mathsf{P}(dx), \quad f \in \mathbb{C}(R), \tag{4}$$

или (в соответствии с обозначениями § 6 гл. II) — в форме

$$\int_R f(x)\, dF_n(x) \to \int_R f(x)\, dF(x), \quad f \in \mathbb{C}(R). \tag{5}$$

В анализе сходимость (4) называют *слабой сходимостью* (мер P_n к мере P, $n \to \infty$) и записывают в виде $\mathsf{P}_n \xrightarrow{w} \mathsf{P}$. Естественно и сходимость (5) также назвать слабой сходимостью функций распределений F_n к F и обозначить ее $F_n \xrightarrow{w} F$.

Итак, можно утверждать, что в схеме Бернулли

$$\frac{S_n}{n} \xrightarrow{\mathsf{P}} p \Rightarrow F_n \xrightarrow{w} F. \tag{6}$$

Из (1) нетрудно также вывести, что для функций распределения, введенных в (2),

$$F_n(x) \to F(x), \quad n \to \infty,$$

для всех точек $x \in R$ *за исключением одной точки* $x = p$, где функция $F(x)$ терпит разрыв.

Это обстоятельство показывает, что слабая сходимость $F_n \xrightarrow{w} F$ *не влечет* за собой поточечную сходимость функций $F_n(x)$ к $F(x)$, $n \to \infty$, для *всех* точек $x \in R$. Оказывается, однако, что как в случае схемы Бернулли так и в общем случае произвольных функций распределения, слабая сходимость эквивалентна (см. далее теорему 2) так называемой сходимости в основном в смысле следующего определения.

Определение 1. Последовательность функций распределения $\{F_n\}$, заданных на числовой прямой, называется сходящейся *в основном* к функции распределения F (обозначение: $F_n \Rightarrow F$), если при $n \to \infty$

$$F_n(x) \to F(x), \quad x \in \mathbb{C}(F),$$

где $\mathbb{C}(F)$ — множество точек непрерывности предельной функции $F = F(x)$.

В рассматриваемом случае схемы Бернулли функция $F = F(x)$ вырождена, и отсюда нетрудно вывести (см. задачу 7 к § 10 в гл. II), что

$$(F_n \Rightarrow F) \Rightarrow \left(\frac{S_n}{n} \xrightarrow{\mathsf{P}} p\right).$$

Таким образом, с учетом приводимой ниже теоремы 2

$$\left(\frac{S_n}{n} \xrightarrow{\mathsf{P}} p\right) \Rightarrow \left(F_n \xrightarrow{w} F\right) \Leftrightarrow (F_n \Rightarrow F) \Rightarrow \left(\frac{S_n}{n} \xrightarrow{\mathsf{P}} p\right) \tag{7}$$

и, следовательно, утверждение закона больших чисел можно рассматривать как одну из теорем о слабой сходимости функций распределений, определенных в (2).

Обозначим

$$F_n(x) = \mathsf{P}\left\{\frac{S_n - np}{\sqrt{npq}} \leqslant x\right\},$$

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du. \tag{8}$$

Теорема Муавра — Лапласа (§ 6 гл. I) утверждает, что $F_n(x) \to F(x)$ для всех $x \in R$ и, следовательно, $F_n \Rightarrow F$. В силу отмеченной эквивалентности слабой сходимости $F_n \xrightarrow{w} F$ и сходимости в основном $F_n \Rightarrow F$ можно, следовательно, сказать, что теорема Муавра — Лапласа есть также утверждение о слабой сходимости функций распределений, определенных в (8).

Эти два примера оправдывают концепцию слабой сходимости вероятностных мер, вводимую далее в определении 2. Хотя для случая числовой прямой слабая сходимость равносильна сходимости в основном соответствующих функций распределения, предпочтительнее, однако, в качестве исходной рассматривать именно слабую сходимость, во-первых, потому что она проще поддается анализу, и, во-вторых, по той причине, что она имеет смысл и для более общих пространств, нежели числовая прямая, в частности для метрических пространств, важнейшими примерами которых для нас являются пространства R^n, R^∞, C и D (см. § 3 гл. II).

2. Пусть $(E, \mathscr{E}, \rho)$ — метрическое пространство с метрикой $\rho = \rho(x, y)$, σ-алгеброй $\mathscr{E}$ борелевских подмножеств, порожденных

открытыми множествами, и пусть P, P_1, P_2, ...— вероятностные меры на $(E, \mathscr{E}, \rho)$.

Определение 2. Последовательность вероятностных мер $\{\mathsf{P}_n\}$ называется *слабо сходящейся* к вероятностной мере P (обозначение: $\mathsf{P}_n \xrightarrow{w} \mathsf{P}$), если

$$\int_E f(x)\,\mathsf{P}_n(dx) \to \int_E f(x)\,\mathsf{P}(dx) \tag{9}$$

для любой функции $f=f(x)$ из класса $\mathbb{C}(E)$ непрерывных ограниченных функций на E.

Определение 3. Последовательность вероятностных мер $\{\mathsf{P}_n\}$ называется *сходящейся в основном* к вероятностной мере P (обозначение: $\mathsf{P}_n \Rightarrow \mathsf{P}$), если

$$\mathsf{P}_n(A) \to \mathsf{P}(A) \tag{10}$$

для любого множества A из $\mathscr{E}$, для которого

$$\mathsf{P}(\partial A) = 0. \tag{11}$$

(Через ∂A обозначается граница множества A:

$$\partial A = [A] \cap [\bar{A}],$$

где $[A]$ — замыкание множества A.)

Следующая важная теорема показывает эквивалентность понятий слабой сходимости и сходимости в основном для вероятностных мер, а также содержит другие равносильные формулировки.

Теорема 1. *Следующие утверждения эквивалентны:*

I. $\mathsf{P}_n \xrightarrow{w} \mathsf{P}$.

II. $\overline{\lim}\, \mathsf{P}_n(A) \leqslant \mathsf{P}(A)$, A — *замкнутые множества.*

III. $\underline{\lim}\, \mathsf{P}_n(A) \geqslant \mathsf{P}(A)$, A — *открытые множества.*

IV. $\mathsf{P}_n \Rightarrow \mathsf{P}$.

Доказательство. (I) $\Rightarrow$ (II). Пусть A — замкнутое множество, $f(x) = I_A(x)$ и

$$f_\varepsilon(x) = g\left(\frac{1}{\varepsilon}\rho(x, A)\right), \quad \varepsilon > 0,$$

где

$$\rho(x, A) = \inf\{\rho(x, y) : y \in A\},$$

$$g(t) = \begin{cases} 1, & t \leqslant 0, \\ 1-t, & 0 \leqslant t \leqslant 1, \\ 0, & t \geqslant 1. \end{cases}$$

Обозначим также

$$A_\varepsilon=\{x:\ \rho(x,\ A)<\varepsilon\}$$

и заметим, что $A_\varepsilon\downarrow A$, $\varepsilon\downarrow 0$.

Поскольку функции $f_\varepsilon(x)$ ограничены, непрерывны и

$$\mathsf{P}_n(A)=\int_E I_A(x)\,\mathsf{P}_n(dx)\leqslant\int_E f_\varepsilon(x)\,\mathsf{P}_n(dx),$$

то

$$\overline{\lim_n}\,\mathsf{P}_n(A)\leqslant\overline{\lim_n}\int_E f_\varepsilon(x)\,\mathsf{P}_n(dx)=$$

$$=\int_E f_\varepsilon(x)\,\mathsf{P}(dx)\leqslant\mathsf{P}(A_\varepsilon)\downarrow\mathsf{P}(A),\quad\varepsilon\downarrow 0,$$

что и доказывает требуемую импликацию.

Импликации (II) $\Rightarrow$ (III) и (III) $\Rightarrow$ (II) становятся очевидными, если от множеств перейти к их дополнениям.

(III) $\Rightarrow$ (IV). Пусть $A^0=A\setminus\partial A$ — внутренность, а $[A]$ — замыкание множества A. Тогда в силу II, III и предположения $\mathsf{P}(\partial A)=0$

$$\overline{\lim_n}\,\mathsf{P}_n(A)\leqslant\overline{\lim_n}\,\mathsf{P}_n([A])\leqslant\mathsf{P}([A])=\mathsf{P}(A),$$

$$\overline{\lim_n}\,\mathsf{P}_n(A)\geqslant\underline{\lim_n}\,\mathsf{P}_n(A^0)\geqslant\mathsf{P}(A^0)=\mathsf{P}(A),$$

и, значит, $\mathsf{P}_n(A)\to\mathsf{P}(A)$ для всякого A с $\mathsf{P}(\partial A)=0$.

(IV) $\Rightarrow$ (I). Пусть $f=f(x)$ — непрерывная ограниченная функция с $|f(x)|<M$. Обозначим

$$D=\{t\in R:\ \mathsf{P}\{x:\ f(x)=t\}\neq 0\}$$

и рассмотрим разбиение $T_k=(t_0,\ t_1,\ \ldots,\ t_k)$ интервала $[-M,M]$:

$$-M=t_0<t_1<\ldots<t_k=M,\quad k\geqslant 1,$$

с $t_i\notin D$, $i=0,\ 1,\ \ldots,\ k$. (Заметим, что множество D не более чем счетно, поскольку множества $f^{-1}\{t\}$ не пересекаются, а мера P конечна.)

Пусть $B_i=\{x:\ t_i\leqslant f(x)<t_{i+1}\}$. Поскольку функция $f(x)$ непрерывная и, следовательно, множество $f^{-1}(t_i,\ t_{i+1})$ открыто, то $\partial B_i\subseteq f^{-1}\{t_i\}\cup f^{-1}\{t_{i+1}\}$. Точки $t_i,\ t_{i+1}\notin D$, поэтому $\mathsf{P}(\partial B_i)=0$ и в силу (IV)

$$\sum_{i=0}^{k-1}t_i\mathsf{P}_n(B_i)\to\sum_{i=0}^{k-1}t_i\mathsf{P}(B_i).\tag{12}$$

Но

$$\left|\int_E f(x)\,\mathsf{P}_n(dx)-\int_E f(x)\,\mathsf{P}(dx)\right|\leqslant\left|\int_E f(x)\,\mathsf{P}_n(dx)-\sum_{i=0}^{k-1}t_i\mathsf{P}_n(B_i)\right|+$$

$$+\left|\sum_{i=0}^{k-1}t_i\mathsf{P}_n(B_i)-\sum_{i=0}^{k-1}t_i\mathsf{P}(B_i)\right|+\left|\sum_{i=0}^{k-1}t_i\mathsf{P}(B_i)-\int_E f(x)\,\mathsf{P}(dx)\right|\leqslant$$

$$\leqslant 2\max_{0\leqslant i\leqslant k-1}(t_{i+1}-t_i)+\left|\sum_{i=0}^{k-1}t_i\mathsf{P}_n(B_i)-\sum_{i=0}^{k-1}t_i\mathsf{P}(B_i)\right|,$$

откуда в силу (12) и произвольности разбиений T_k, $k\geqslant 1$,

$$\lim_n\int_E f(x)\,\mathsf{P}_n(dx)=\int_E f(x)\,\mathsf{P}(dx).$$

Теорема доказана.

Замечание 1. Участвующие в доказательстве импликации $\mathrm{I}\Rightarrow\mathrm{II}$ функции $f(x)=I_A(x)$ и $f_\varepsilon(x)$ являются соответственно *полунепрерывными сверху* и *равномерно непрерывными*. Учитывая это обстоятельство, нетрудно показать, что каждое из условий теоремы эквивалентно одному из нижеследующих условий:

(V) $\int_E f(x)\mathsf{P}_n(dx)\to\int_E f(x)\mathsf{P}(dx)$ *для всех ограниченных равномерно непрерывных функций* $f(x)$;

(VI) $\overline{\lim_n}\int_E f(x)\mathsf{P}_n(dx)\leqslant\int_E f(x)\mathsf{P}(dx)$ *для всех ограниченных функций* $f(x)$, *являющихся полунепрерывными сверху* $\left(\overline{\lim_n} f(x_n)\leqslant f(x),\ x_n\to x\right)$;

(VII) $\underline{\lim_n}\int_E f(x)\mathsf{P}_n(dx)\geqslant\int_E f(x)\mathsf{P}(dx)$ *для всех ограниченных функций* $f(x)$, *являющихся полунепрерывными снизу* $\left(\underline{\lim_n} f(x_n)\geqslant f(x),\ x_n\to x\right)$.

Замечание 2. Теорема 1 допускает естественное обобщение на тот случай, когда вместо вероятностных мер P и P_n, заданных на $(E, \mathscr{E}, \rho)$, рассматриваются *произвольные* (не обязательно вероятностные) *конечные меры* μ и μ_n. Для таких мер совершенно аналогично вводятся понятия слабой сходимости $\mu_n\xrightarrow{w}\mu$, сходимости в основном $\mu_n\Rightarrow\mu$ и, так же как в теореме 1, устанавливается эквивалентность следующих условий:

I*. $\mu_n\xrightarrow{w}\mu$;

II*. $\overline{\lim_n}\,\mu_n(A)\leqslant\mu(A)$, A — *замкнутые множества, и* $\mu_n(E)\to\mu(E)$;

III*. $\varliminf_n \mu_n(A) \geqslant \mu(A)$, *$A$ — открытые множества, и* $\mu_n(E) \to \mu(E)$;

IV*. $\mu_n \Rightarrow \mu$.

Каждое из этих условий равносильно любому из условий V*, VI*, VII*, формулируемых как и V, VI, VII с заменой мер P_n и P на μ_n и μ соответственно.

3. Пусть $(R, \mathscr{B}(R))$ — числовая прямая с системой борелевских множеств $\mathscr{B}(R)$, порожденных евклидовой метрикой $\rho(x, y) = |x-y|$ (ср. с замечанием 2 в п. 2 § 2 гл. II). Обозначим P, P_n, $n \geqslant 1$, вероятностные меры на $(R, \mathscr{B}(R))$, и пусть F, F_n, $n \geqslant 1$, — соответствующие им функции распределения. Тогда справедлива

Теорема 2. *Следующие условия эквивалентны*:

(1) $\mathsf{P}_n \xrightarrow{w} \mathsf{P}$,

(2) $\mathsf{P}_n \Rightarrow \mathsf{P}$,

(3) $F_n \xrightarrow{w} F$,

(4) $F_n \Rightarrow F$.

Доказательство. Поскольку (2) $\Leftrightarrow$ (1) $\Leftrightarrow$ (3), то достаточно доказать, что (2) $\Leftrightarrow$ (4).

Если $\mathsf{P}_n \Rightarrow \mathsf{P}$, то, в частности,

$$\mathsf{P}_n(-\infty, x] \to \mathsf{P}(-\infty, x]$$

для всех $x \in R$ таких, что $\mathsf{P}\{x\}=0$. А это и означает, что $F_n \Rightarrow F$.

Пусть теперь $F_n \Rightarrow F$. Для доказательства сходимости $\mathsf{P}_n \Rightarrow \mathsf{P}$ достаточно (в силу теоремы 1) показать, что $\varliminf_n \mathsf{P}_n(A) \geqslant \mathsf{P}(A)$ для всякого открытого множества A.

Если A — открытое множество, то найдется счетная система непересекающихся открытых интервалов $I_1, I_2, \ldots$ (вида (a, b)) таких, что $A = \sum_{k=1}^{\infty} I_k$. Зафиксируем $\varepsilon > 0$ и выберем в каждом интервале $I_k = (a_k, b_k)$ подынтервал $I_k' = (a_k', b_k']$ такой, что a_k', $b_k' \in \mathbb{C}(F)$ и $\mathsf{P}(I_k) \leqslant \mathsf{P}(I_k') + \varepsilon \cdot 2^{-k}$. (Поскольку множество точек разрыва функции $F = F(x)$ не более чем счетно, такие интервалы I_k', $k \geqslant 1$, действительно существуют.) Тогда по лемме Фату

$$\varliminf_n \mathsf{P}_n(A) = \varliminf_n \sum_{k=1}^{\infty} \mathsf{P}_n(I_k) \geqslant \sum_{k=1}^{\infty} \varliminf_n \mathsf{P}_n(I_k) \geqslant$$

$$\geqslant \sum_{k=1}^{\infty} \varliminf_n \mathsf{P}_n(I_k').$$

Но

$$\mathsf{P}_n(I'_k) = F_n(b'_k) - F_n(a'_k) \to F(b'_k) - F(a'_k) = \mathsf{P}(I'_k).$$

Поэтому

$$\varliminf_n \mathsf{P}_n(A) \geqslant \sum_{k=1}^{\infty} \mathsf{P}(I'_k) \geqslant \sum_{k=1}^{\infty} (\mathsf{P}(I_k) - \varepsilon \cdot 2^{-k}) = \mathsf{P}(A) - \varepsilon,$$

что в силу произвольности $\varepsilon > 0$ доказывает, что $\varliminf_n \mathsf{P}_n(A) \geqslant \mathsf{P}(A)$, если A — открытое множество.

Теорема доказана.

4. Пусть $(E, \mathscr{E})$ — измеримое пространство. Систему подмножеств $\mathscr{K}_0(E) \subseteq \mathscr{E}$ назовем *определяющим классом*, если для любых двух вероятностных мер P и Q, заданных на $(E, \mathscr{E})$, из равенства

$$\mathsf{P}(A) = \mathsf{Q}(A) \quad \textit{для всех } A \in \mathscr{K}_0(E)$$

вытекает, что эти меры совпадают тождественно, т. е.

$$\mathsf{P}(A) = \mathsf{Q}(A) \quad \textit{для всех } A \in \mathscr{E}.$$

Если $(E, \mathscr{E}, \rho)$ — метрическое пространство, то систему подмножеств $\mathscr{K}_1(E) \subseteq \mathscr{E}$ назовем *классом, определяющим сходимость*, если для любых мер $\mathsf{P}, \mathsf{P}_1, \mathsf{P}_2, \ldots$ из того, что

$$\mathsf{P}_n(A) \to \mathsf{P}(A) \quad \textit{для всех } A \in \mathscr{K}_1(E) \textit{ с } \mathsf{P}(\partial A) = 0$$

вытекает, что

$$\mathsf{P}_n(A) \to \mathsf{P}(A) \quad \textit{для всех } A \in \mathscr{E} \textit{ с } \mathsf{P}(\partial A) = 0.$$

В случае $(E, \mathscr{E}) = (R, \mathscr{B}(R))$ в качестве определяющего класса $\mathscr{K}_0(R)$ можно взять класс «элементарных» множеств $\mathscr{K} = \{(-\infty, x], x \in R\}$ (теорема 1 из § 3 гл. II). Из эквивалентности условий (2) и (4) теоремы 2 вытекает, что класс $\mathscr{K}$ является также и классом, определяющим сходимость.

Естественно возникает вопрос о таких определяющих классах и для более общих пространств.

В случае пространств R^n, $n \geqslant 2$, класс $\mathscr{K}$ «элементарных» множеств вида $(-\infty, x] = (-\infty, x_1] \times \ldots \times (-\infty, x_n]$, $x = (x_1, \ldots, x_n) \in R^n$ является как определяющим классом (теорема 2 из § 3 гл. II), так и классом, определяющим сходимость (задача 2).

В случае пространства R^∞ цилиндрические множества $\mathscr{K}_0(R^\infty)$ являются теми «элементарными» множествами, по вероятностям которых однозначно определяется вероятность для всех борелевских множеств (теорема 3 из § 3 гл. II). Оказывается, что в этом случае класс цилиндрических множеств является тем классом, который определяет также и сходимость (задача 3). Таким образом, $\mathscr{K}_1(R^\infty) = \mathscr{K}_0(R^\infty)$.

Можно было бы ожидать, что и в случае более общих пространств класс цилиндрических множеств является классом, определяющим сходимость. Однако, вообще говоря, это не верно.

Так, например, рассмотрим пространство $(C, \mathscr{B}_0(C), \rho)$ с равномерной метрикой ρ (см. п. 6 § 2 гл. II). Пусть P — вероятностная мера, целиком сосредоточенная на функции $x_t \equiv 0$, $0 \leqslant t \leqslant 1$, а P_n — вероятностные меры, $n \geqslant 1$, каждая из которых сосредоточена на функции x_n, изображенной на рис. 35. Нетрудно убедиться, что $\mathsf{P}_n(A) \to \mathsf{P}(A)$ для всех цилиндрических множеств A с $\mathsf{P}(\partial A) = 0$. Но, если взять, например, множество

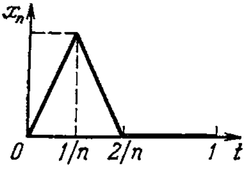


Рис. 35.

$$A = \{x \in C \colon |x_t| \leqslant \tfrac{1}{2},\ 0 \leqslant t \leqslant 1\} \in \mathscr{B}_0(C),$$

то $\mathsf{P}(\partial A) = 0$, $\mathsf{P}_n(A) = 0$, $\mathsf{P}(A) = 1$ и, следовательно, $\mathsf{P}_n \not\Rightarrow \mathsf{P}$.

Таким образом, $\mathscr{K}_0(C) = \mathscr{B}_0(C)$, но $\mathscr{K}_0(C) \subset \mathscr{K}_1(C)$ (включение строгое!).

5. Задачи.

1. Будем говорить, что функция $F = F(x)$, заданная на R^n, *непрерывна в точке* $x \in R^n$, если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что $|F(x) - F(y)| < \varepsilon$ для всех $y \in R^n$, удовлетворяющих неравенству

$$x - \delta e < y < x + \delta e,$$

где $e = (1, \ldots, 1) \in R^n$. Будем говорить также, что последовательность функций распределения $\{F_n\}$ *сходится в основном* к функции распределения F $(F_n \Rightarrow F)$, если $F_n(x) \to F(x)$ для всех точек $x \in R^n$, где функция $F = F(x)$ непрерывна.

Показать, что утверждение теоремы 2 остается справедливым для R^n, $n > 1$. (См. замечание к теореме 2.)

2. Показать, что в случае пространств R^n класс «элементарных» множеств $\mathscr{K}$ является классом, определяющим сходимость.

3. Пусть E — одно из пространств R^∞, C или D. Будем говорить, что последовательность вероятностных мер $\{\mathsf{P}_n\}$ (заданных на σ-алгебре $\mathscr{E}$ борелевских множеств, порожденных открытыми множествами) сходится *в основном в смысле конечномерных распределений* к вероятностной мере P (обозначение: $\mathsf{P}_n \overset{f}{\Rightarrow} \mathsf{P}$), если $\mathsf{P}_n(A) \to \mathsf{P}(A)$, $n \to \infty$, для всех *цилиндрических* множеств A с $\mathsf{P}(\partial A) = 0$.

Показать, что в случае пространства R^∞

$$\left(\mathsf{P}_n \overset{f}{\Rightarrow} \mathsf{P}\right) \Leftrightarrow (\mathsf{P}_n \Rightarrow \mathsf{P}).$$

4. Пусть F и G — функции распределения на числовой прямой и

$$L(F, G) = \inf\{h > 0:\ F(x-h) - h \leqslant G(x) \leqslant F(x+h) + h\}$$

— *расстояние Леви* (между F и G). Показать, что сходимость в основном эквивалентна сходимости в метрике Леви:

$$(F_n \Rightarrow F) \Leftrightarrow L(F_n, F) \to 0.$$

5. Пусть $F_n \Rightarrow F$ и функция распределения F является непрерывной. Показать, что тогда сходимость $F_n(x)$ к $F(x)$ равномерна:

$$\sup_x |F_n(x) - F(x)| \to 0, \qquad n \to \infty.$$

6. Доказать утверждение, сформулированное в замечании 1 к теореме 1.

7. Убедиться в справедливости эквивалентности условий (I*) — (IV*), сформулированных в замечании 2 к теореме 1.

8. Показать, что $\mathsf{P}_n \xrightarrow{w} \mathsf{P}$ тогда и только тогда, когда всякая подпоследовательность $\{\mathsf{P}_{n'}\}$ последовательности $\{\mathsf{P}_n\}$ содержит подпоследовательность $\{\mathsf{P}_{n''}\}$ такую, что $\mathsf{P}_{n''} \xrightarrow{w} \mathsf{P}$.

§ 2. Относительная компактность и плотность семейств вероятностных распределений

1. Если задана последовательность вероятностных мер, то прежде чем рассматривать вопрос о ее (слабой) сходимости к той или иной вероятностной мере, следует, конечно, выяснить, а сходится ли вообще эта последовательность к некоторой мере или имеет она хотя бы одну сходящуюся подпоследовательность.

Так, например, последовательность $\{\mathsf{P}_n\}$, где $\mathsf{P}_{2n} = \mathsf{P}$, $\mathsf{P}_{2n+1} = \mathsf{Q}$, а P и Q — различные вероятностные меры, не является, очевидно, сходящейся, но имеет две сходящиеся подпоследовательности $\{\mathsf{P}_{2n}\}$ и $\{\mathsf{P}_{2n+1}\}$.

Совсем просто устроенная последовательность $\{\mathsf{P}_n\}$ вероятностных мер P_n, $n \geqslant 1$, каждая из которых сосредоточена в точке $\{n\}$ ($\mathsf{P}_n\{n\} = 1$), не только не является сходящейся, но и не содержит ни одной сходящейся подпоследовательности. (Поскольку $\lim\limits_n \mathsf{P}_n(a, b] = 0$ для любых $a < b$, то предельная мера должна была бы быть тождественно равной нулю, а это противоречит тому, что $1 = \mathsf{P}_n(R) \not\to 0$, $n \to \infty$.) Интересно отметить, что в этом примере соответствующая последовательность функций распределения $\{F_n\}$, где

$$F_n(x) = \begin{cases} 1, & x \geqslant n, \\ 0, & x < n, \end{cases}$$

является, очевидно, сходящейся: для любого $x \in R$

$$F_n(x) \to G(x) \equiv 0.$$

Однако предельная функция $G = G(x)$ не является функцией распределения (в смысле определения 1 из § 3 гл. II).

Этот пример поучителен с той точки зрения, что, как он показывает, класс функций распределения не является компактным. Он подсказывает также, что для сходимости последовательности функций распределения к функции, которая являлась бы также функцией распределения, нужны некоторые условия, предотвращающие «утечку массы на бесконечность».

После этих вводных замечаний, поясняющих характер возникающих здесь трудностей, перейдем к основным определениям.

2. Будем предполагать, что все рассматриваемые меры определены на метрическом пространстве $(E, \mathscr{E}, \rho)$.

Определение 1. Семейство вероятностных мер $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$ назовем *относительно компактным*, если любая последовательность мер из $\mathscr{P}$ содержит подпоследовательность, слабо сходящуюся к некоторой вероятностной мере.

Подчеркнем, что в этом определении предельная мера предполагается *вероятностной*, хотя, быть может, и не принадлежащей исходному классу $\mathscr{P}$. (Именно с этим последним обстоятельством связано появление слова «относительно» в данном определении.)

Проверка того, что данное семейство вероятностных мер относительно компактно, является делом далеко не простым. Желательно поэтому иметь простые и удобные критерии, позволяющие осуществлять эту проверку. Этой цели служит

Определение 2. Семейство вероятностных мер $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$ называется *плотным*, если для каждого $\varepsilon > 0$ можно указать компакт $K \subseteq E$ такой, что

$$\sup_{\alpha \in \mathfrak{A}} \mathsf{P}_\alpha (E \setminus K) \leqslant \varepsilon. \qquad (1)$$

Определение 3. Семейство функций распределения $\mathscr{F} = \{F_\alpha;\ \alpha \in \mathfrak{A}\}$, определенных на R^n, $n \geqslant 1$, называется *относительно компактным* (*плотным*), если таковым является соответствующее семейство вероятностных мер $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$, где P_α — мера, построенная по F_α.

3. Следующий результат играет фундаментальную роль во всей проблематике слабой сходимости вероятностных мер.

Теорема 1 (теорема Прохорова). *Пусть $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$ — семейство вероятностных мер, заданных на полном сепарабельном метрическом пространстве $(E, \mathscr{E}, \rho)$. Семейство $\mathscr{P}$ является относительно компактным тогда и только тогда, когда оно является плотным.*

Доказательство этой теоремы будет приведено лишь для случая числовой прямой. (Почти без всяких изменений это доказательство переносится на случай произвольных евклидовых пространств R^n, $n \geqslant 2$. Затем справедливость теоремы устанавливается последовательно для R^∞, для σ-компактных пространств и, наконец, для общих полных сепарабельных метрических пространств путем сведения каждого из этих случаев к предыдущему.)

Необходимость. Пусть семейство вероятностных мер $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$, заданных на $(R, \mathscr{B}(R))$, относительно компактно, но не плотно. Тогда найдется такое $\varepsilon > 0$, что для любого компакта $K \subseteq R$

$$\sup_\alpha \mathsf{P}_\alpha(R \setminus K) > \varepsilon,$$

а значит, и для любого интервала $I = (a, b)$

$$\sup_\alpha \mathsf{P}_\alpha(R \setminus I) > \varepsilon.$$

Отсюда вытекает, что для каждого интервала $I_n = (-n, n)$, $n \geqslant 1$ найдется такая мера P_{α_n}, что

$$\mathsf{P}_{\alpha_n}(R \setminus I_n) > \varepsilon.$$

Раз исходное семейство $\mathscr{P}$ относительно компактно, то из последовательности $\{\mathsf{P}_{\alpha_n}\}_{n \geqslant 1}$ можно извлечь подпоследовательность, скажем, $\{\mathsf{P}_{\alpha_{n_k}}\}$ такую, что $\mathsf{P}_{\alpha_{n_k}} \xrightarrow{w} \mathsf{Q}$, где Q — некоторая вероятностная мера. Тогда в силу эквивалентности условий I и II в теореме 1 из § 1 для всякого $n \geqslant 1$

$$\varlimsup_{k \to \infty} \mathsf{P}_{\alpha_{n_k}}(R \setminus I_n) \leqslant \mathsf{Q}(R \setminus I_n). \tag{2}$$

Но $\mathsf{Q}(R \setminus I_n) \downarrow 0$, $n \to \infty$, а левая часть в (2) больше $\varepsilon > 0$. Это противоречие показывает, что относительная компактность влечет за собой плотность.

Для доказательства достаточности нам необходим один общий результат (называемый теоремой Хелли) о *секвенциальной компактности* семейства обобщенных функций распределения (п. 2 § 3 гл. II).

Обозначим через $\mathscr{J} = \{G\}$ совокупность функций $G = G(x)$ (обобщенных функций распределения), удовлетворяющих следующим свойствам:

1) $G(x)$ — *не убывают*;
2) $0 \leqslant G(-\infty)$, $G(+\infty) \leqslant 1$;
3) $G(x)$ — *непрерывны справа*.

Ясно, что $\mathscr{J}$ включает в себя класс функций распределения $\mathscr{F} = \{F\}$, для которых $F(-\infty) = 0$ и $F(+\infty) = 1$.

Теорема 2 (теорема Хелли). *Класс $\mathscr{J}=\{G\}$ обобщенных функций распределения является секвенциально компактным, т. е. для любой последовательности $\{G_n\}$ функций из $\mathscr{J}$ найдутся функции $G\in\mathscr{J}$ и подпоследовательность $\{n_k\}\subseteq\{n\}$ такие, что*

$$G_{n_k}(x)\to G(x),\qquad k\to\infty,$$

для любой точки x из множества $\mathbb{C}(G)$ точек непрерывности функции $G=G(x)$.

Доказательство. Обозначим через $T=\{x_1, x_2, \ldots\}$ счетное всюду плотное множество в R. Поскольку числовая последовательность $\{G_n(x_1)\}$ *ограничена*, то найдется подпоследовательность $N_1=\{n_1^{(1)}, n_2^{(1)}, \ldots\}$ такая, что при $i\to\infty$ $G_{n_i^{(1)}}(x_1)$ сходятся к некоторому числу g_1. В свою очередь из последовательности N_1 можно извлечь подпоследовательность $N_2=\{n_1^{(2)}, n_2^{(2)}, \ldots\}$ такую, что $G_{n_i^{(2)}}(x_2)$ сходятся при $i\to\infty$ к некоторому числу g_2 и т. д.

Определим на множестве $T\subseteq R$ функцию $G_T(x)$, полагая

$$G_T(x_i)=g_i,\quad x_i\in T,$$

и рассмотрим «канторовскую» диагональную последовательность $N=\{n_1^{(1)}, n_2^{(2)}, \ldots\}$. Тогда для любого $x_i\in T$ при $m\to\infty$

$$G_{n_m^{(m)}}(x_i)\to G_T(x_i).$$

Определим, наконец, функцию $G=G(x)$ для всех $x\in R$, полагая

$$G(x)=\inf\{G_T(y):\ y\in T,\ y>x\}. \tag{3}$$

Мы утверждаем, что $G=G(x)$ есть искомая функция и $G_{n_m^{(m)}}(x)\to G(x)$ для всех точек x, где $G(x)$ непрерывна.

Поскольку все рассматриваемые функции G_n являются неубывающими, то $G_{n_m^{(m)}}(x)\leqslant G_{n_m^{(m)}}(y)$ для всех x и y, принадлежащих множеству T и удовлетворяющих неравенству $x\leqslant y$. Поэтому для таких x и y

$$G_T(x)\leqslant G_T(y).$$

Отсюда и из определения (3) следует, что функция $G=G(x)$ является неубывающей.

Покажем теперь, что она непрерывна справа. Пусть $x_k\downarrow x$ и $d=\lim\limits_k G(x_k)$. Ясно, что $G(x)\leqslant d$, и надо установить, что на самом деле $G(x)=d$. Предположим противное, т. е. пусть $G(x)<d$. Из (3) следует, что тогда найдется такая точка $y\in T$, $x<y$, что $G_T(y)<d$. Для достаточно больших k $x<x_k<y$, а, значит, $G(x_k)\leqslant G_T(y)<d$ и $\lim\limits_k G(x_k)<d$, что противоречит равенству $d=\lim\limits_k G(x_k)$. Итак, построенная функция G принадлежит $\mathscr{J}$.

Установим теперь сходимость $G_{n_m^{(m)}}(x^0) \to G(x^0)$ для всякой точки $x^0 \in \mathbb{C}(G)$.

Если $x^0 < y \in T$, то

$$\overline{\lim_m}\, G_{n_m^{(m)}}(x^0) \leqslant \overline{\lim_m}\, G_{n_m^{(m)}}(y) = G_T(y),$$

откуда

$$\overline{\lim_m}\, G_{n_m^{(m)}}(x^0) \leqslant \inf\{G_T(y):\ y > x^0,\ y \in T\} = G(x^0). \tag{4}$$

С другой стороны, пусть $x^1 < y < x^0$, $y \in T$. Тогда

$$G(x^1) \leqslant G_T(y) = \lim_m G_{n_m^{(m)}}(y) = \underline{\lim_m}\, G_{n_m^{(m)}}(y) \leqslant \underline{\lim_m}\, G_{n_m^{(m)}}(x^0).$$

Поэтому, полагая $x^1 \uparrow x^0$, получим, что

$$G(x^0-) \leqslant \underline{\lim_m}\, G_{n_m^{(m)}}(x^0). \tag{5}$$

Но если $G(x^0-) = G(x^0)$, то тогда из (4) и (5) заключаем, что $G_{n_m^{(m)}}(x^0) \to G(x^0)$, $m \to \infty$.

Теорема доказана.

Завершим теперь доказательство теоремы 1.

Достаточность. Пусть семейство $\mathscr{P}$ плотно и $\{\mathsf{P}_n\}$ — некоторая последовательность вероятностных мер из $\mathscr{P}$. Обозначим через $\{F_n\}$ последовательность соответствующих функций распределения.

В силу теоремы Хелли найдутся подпоследовательность $\{F_{n_k}\} \subseteq \{F_n\}$ и обобщенная функция распределения $G \in \mathscr{J}$ такие, что $F_{n_k}(x) \to G(x)$ для $x \in \mathbb{C}(G)$. Покажем, что в силу предположения о плотности семейства $\mathscr{P}$ функция $G = G(x)$ является на самом деле «настоящей» функцией распределения ($G(-\infty) = 0$, $G(+\infty) = 1$).

Возьмем $\varepsilon > 0$, и пусть $I = (a, b]$ — тот интервал, для которого

$$\sup_n \mathsf{P}_n(R \setminus I) < \varepsilon,$$

или, что эквивалентно,

$$1 - \varepsilon \leqslant \mathsf{P}_n(a, b], \qquad n \geqslant 1.$$

Выберем точки a', $b' \in \mathbb{C}(G)$, такими, что $a' < a$, $b' > b$. Тогда

$$1 - \varepsilon \leqslant \mathsf{P}_{n_k}(a, b] \leqslant \mathsf{P}_{n_k}(a', b'] = F_{n_k}(b') - F_{n_k}(a') \to G(b') - G(a').$$

Отсюда следует, что $G(+\infty) - G(-\infty) = 1$, и поскольку $0 \leqslant G(-\infty) \leqslant G(+\infty) \leqslant 1$, то $G(-\infty) = 0$ и $G(+\infty) = 1$.

Таким образом, предельная функция $G = G(x)$ является функцией распределения и $F_{n_k} \Rightarrow G$, что вместе с теоремой 2 из § 1 доказывает, что $\mathsf{P}_{n_k} \xrightarrow{w} \mathsf{Q}$, где Q — вероятностная мера, построенная по функции распределения G.

Теорема 1 доказана.

4. Задачи.

1. Провести доказательство теорем 1 и 2 для пространств R^n, $n \geqslant 2$.

2. Пусть P_α — гауссовская мера на числовой прямой с параметрами m_α и σ_α^2, $\alpha \in \mathfrak{A}$. Показать, что семейство $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$ является плотным тогда и только тогда, когда существуют константы a и b такие, что

$$|m_\alpha| \leqslant a, \quad \sigma_\alpha^2 \leqslant b, \quad \alpha \in \mathfrak{A}.$$

3. Привести примеры плотных и неплотных семейств вероятностных мер $\mathscr{P} = \{\mathsf{P}_\alpha;\ \alpha \in \mathfrak{A}\}$, определенных на $(R^\infty, \mathscr{B}(R^\infty))$.

§ 3. Метод характеристических функций в доказательстве предельных теорем

1. Доказательство первых предельных теорем теории вероятностей — закона больших чисел и теорем Муавра — Лапласа и Пуассона для схемы Бернулли — основывалось на прямом анализе допредельных функций распределений F_n, которые довольно просто выражаются через биномиальные вероятности. (В схеме Бернулли суммируемые случайные величины принимают только два значения, что и дает, в сущности, возможность явно найти функции F_n.) Однако для случайных величин более сложной природы подобный метод прямого анализа функций F_n становится практически неосуществимым.

Первый шаг в доказательстве предельных теорем для сумм произвольно распределенных независимых случайных величин был сделан Чебышевым.

Предложенное им неравенство, известное теперь как «неравенство Чебышева», не только дало возможность элементарно доказать закон больших чисел Я. Бернулли, но и установить весьма общие условия справедливости этого закона для сумм $S_n = \xi_1 + \ldots + \xi_n$, $n \geqslant 1$, независимых случайных величин в форме утверждения, что для всякого $\varepsilon > 0$

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - \frac{\mathsf{M}S_n}{n}\right| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{1}$$

(См. задачу 2.)

Далее, Чебышевым был создан (и Марковым усовершенствован) так называемый «метод моментов», который позволил установить, что утверждение теоремы Муавра—Лапласа, записанное в виде

$$\mathsf{P}\left\{\frac{S_n-\mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}\leqslant x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}\,du, \tag{2}$$

носит универсальный характер в том смысле, что оно справедливо в очень общих предположениях относительно природы суммируемых случайных величин. Именно это дало основание называть утверждение (2) *центральной предельной теоремой* теории вероятностей.

Несколько позже Ляпунов предложил иной метод доказательства центральной предельной теоремы, в основе которого лежала (восходящая к Лапласу) идея «характеристической функции» распределения вероятностей. Последующее развитие показало, что «метод характеристических функций» Ляпунова является весьма эффективным при доказательстве самых разнообразных предельных теорем, что и обусловило его развитие и широкое применение.

Сущность этого метода состоит в следующем.

2. Мы уже знаем (§ 12 гл. II), что между функциями распределения и характеристическими функциями существует взаимно однозначное соответствие. Поэтому изучение свойств функций распределения можно проводить, изучая соответствующие характеристические функции. Замечательным сказывается то обстоятельство, что слабая сходимость $F_n\xrightarrow{w}F$ функций распределения эквивалентна поточечной сходимости $\varphi_n\to\varphi$ соответствующих характеристических функций. Более того, имеет место следующий результат, являющийся основным средством доказательства теорем о слабой сходимости распределений на числовой прямой.

Теорема 1 (теорема непрерывности). *Пусть $\{F_n\}$ — последовательность функций распределения $F_n=F_n(x)$, $x\in R$, и $\{\varphi_n\}$ — соответствующая последовательность характеристических функций,*

$$\varphi_n(t)=\int_{-\infty}^{\infty}e^{itx}\,dF_n(x),\qquad t\in R.$$

1) *Если $F_n\xrightarrow{w}F$, где $F=F(x)$ — некоторая функция распределения, то $\varphi_n(t)\to\varphi(t)$, $t\in R$, где $\varphi(t)$ — характеристическая функция $F=F(x)$.*

2) *Если при каждом $t\in R$ существует $\lim\limits_n\varphi_n(t)$ и функция $\varphi(t)=\lim\limits_n\varphi_n(t)$ непрерывна в точке $t=0$, то она является характеристической функцией некоторого распределения вероятностей $F=F(x)$ и*

$$F_n\xrightarrow{w}F.$$

Доказательство утверждения 1) сразу следует из определения слабой сходимости, примененного к функциям $\operatorname{Re} e^{itx}$ и $\operatorname{Im} e^{itx}$.

Доказательству утверждения 2) предпошлем несколько вспомогательных предложений.

Лемма 1. *Пусть $\{\mathsf{P}_n\}$ — плотное семейство вероятностных мер. Предположим, что каждая слабо сходящаяся подпоследовательность $\{\mathsf{P}_{n'}\}$ последовательности $\{\mathsf{P}_n\}$ сходится к одной и той же вероятностной мере P. Тогда и вся последовательность $\{\mathsf{P}_n\}$ слабо сходится к P.*

Доказательство. Допустим, что $\mathsf{P}_n \overset{w}{\nrightarrow} \mathsf{P}$. Тогда найдется такая ограниченная непрерывная функция $f = f(x)$, что

$$\int_R f(x)\, \mathsf{P}_n(dx) \nrightarrow \int_R f(x)\, \mathsf{P}(dx).$$

Отсюда следует, что существуют $\varepsilon > 0$ и бесконечная последовательность чисел $\{n'\} \subseteq \{n\}$ такие, что

$$\left| \int_R f(x)\, \mathsf{P}_{n'}(dx) - \int_R f(x)\, \mathsf{P}(dx) \right| \geqslant \varepsilon > 0. \tag{3}$$

По теореме Прохорова (§ 2) из последовательности $\{\mathsf{P}_{n'}\}$ можно выбрать подпоследовательность $\{\mathsf{P}_{n''}\}$ такую, что $\mathsf{P}_{n''} \xrightarrow{w} \mathsf{Q}$, где Q — некоторая вероятностная мера.

По предположению леммы $\mathsf{Q} = \mathsf{P}$, и, значит,

$$\int_R f(x)\, \mathsf{P}_{n''}(dx) \to \int_R f(x)\, \mathsf{P}(dx),$$

что находится в противоречии с (3). Лемма доказана.

Лемма 2. *Пусть $\{\mathsf{P}_n\}$ — плотное семейство вероятностных мер на $(R, \mathscr{B}(R))$. Последовательность $\{\mathsf{P}_n\}$ слабо сходится к некоторой вероятностной мере тогда и только тогда, когда для каждого $t \in R$ существует $\lim\limits_n \varphi_n(t)$, где $\varphi_n(t)$ — характеристическая функция меры P_n:*

$$\varphi_n(t) = \int_R e^{itx}\, \mathsf{P}_n(dx).$$

Доказательство. Если семейство $\{\mathsf{P}_n\}$ плотно, то по теореме Прохорова найдется подпоследовательность $\{\mathsf{P}_{n'}\}$ и вероятностная мера P такие, что $\mathsf{P}_{n'} \xrightarrow{w} \mathsf{P}$. Предположим, что вся последовательность $\{\mathsf{P}_n\}$ не сходится к P $\left(\mathsf{P}_n \overset{w}{\nrightarrow} \mathsf{P}\right)$. Тогда в силу леммы 1 найдется подпоследовательность $\{\mathsf{P}_{n''}\}$ и вероятностная мера Q такие, что $\mathsf{P}_{n''} \xrightarrow{w} \mathsf{Q}$, причем $\mathsf{P} \neq \mathsf{Q}$.

Воспользуемся теперь тем, что при каждом $t \in R$ существует $\lim_n \varphi_n(t)$. Тогда

$$\lim_{n'} \int_R e^{itx} \mathsf{P}_{n'}(dx) = \lim_{n''} \int_R e^{itx} \mathsf{P}_{n''}(dx)$$

и, значит,

$$\int_R e^{itx} \mathsf{P}(dx) = \int_R e^{itx} \mathsf{Q}(dx), \qquad t \in R.$$

Но характеристическая функция однозначно определяет распределение (теорема 2 § 12 гл. II). Поэтому $\mathsf{P} = \mathsf{Q}$, что противоречит предположению $\mathsf{P}_n \overset{w}{\nrightarrow} \mathsf{P}$.

Что же касается обратного утверждения леммы, то оно непосредственно следует из определения слабой сходимости.

Следующая лемма дает оценку «хвостов» функции распределения по поведению ее характеристической функции в окрестности нуля.

Лемма 3. *Пусть $F = F(x)$ — функция распределения на числовой прямой и $\varphi = \varphi(t)$ — ее характеристическая функция. Тогда существует такая константа $K > 0$, что для всякого $a > 0$*

$$\int_{|x| \geqslant 1/a} dF(x) \leqslant \frac{K}{a} \int_0^a [1 - \operatorname{Re} \varphi(t)]\, dt. \tag{4}$$

Доказательство. Поскольку $\operatorname{Re} \varphi(t) = \int_{-\infty}^{\infty} \cos tx\, dF(x)$, то, применяя теорему Фубини, находим, что

$$\frac{1}{a} \int_0^a [1 - \operatorname{Re} \varphi(t)]\, dt = \frac{1}{a} \int_0^a \left[\int_{-\infty}^{\infty} (1 - \cos tx)\, dF(x) \right] dt =$$

$$= \int_{-\infty}^{\infty} \left[\frac{1}{a} \int_0^a (1 - \cos tx)\, dt \right] dF(x) = \int_{-\infty}^{\infty} \left(1 - \frac{\sin ax}{ax}\right) dF(x) \geqslant$$

$$\geqslant \inf_{|y| \geqslant 1} \left(1 - \frac{\sin y}{y}\right) \cdot \int_{|ax| \geqslant 1} dF(x) = \frac{1}{K} \int_{|x| \geqslant 1/a} dF(x),$$

где

$$\frac{1}{K} = \inf_{|y| \geqslant 1} \left(1 - \frac{\sin y}{y}\right) = 1 - \sin 1 \geqslant \frac{1}{7},$$

так что (4) справедливо с константой $K = 7$. Лемма доказана.

Доказательство утверждения 2) теоремы 1. Пусть $\varphi_n(t) \to \varphi(t)$, $n \to \infty$, где функция $\varphi(t)$ непрерывна в нуле. Покажем, что отсюда следует плотность семейства вероятностных мер $\{\mathsf{P}_n\}$, где P_n — мера, соответствующая функции распределения F_n.

В силу (4) и теоремы о мажорируемой сходимости

$$\mathsf{P}_n\left\{R\diagdown\left(-\frac{1}{a},\ \frac{1}{a}\right)\right\} = \int\limits_{|x|\geqslant\frac{1}{a}} dF_n(x) \leqslant \frac{K}{a}\int\limits_0^a [1-\operatorname{Re}\varphi_n(t)]\,dt \to$$

$$\to \frac{K}{a}\int\limits_0^a [1-\operatorname{Re}\varphi(t)]\,dt$$

при $n\to\infty$.

Поскольку по предположению функция $\varphi(t)$ непрерывна в нуле и $\varphi(0)=1$, то для всякого $\varepsilon>0$ можно найти такое $a>0$, что для всех $n\geqslant 1$

$$\mathsf{P}_n\left\{R\diagdown\left(-\frac{1}{a},\ \frac{1}{a}\right)\right\}\leqslant\varepsilon.$$

Следовательно, семейство $\{\mathsf{P}_n\}$ плотно, и в силу леммы 2 существует вероятностная мера P такая, что

$$\mathsf{P}_n \xrightarrow{w} \mathsf{P}.$$

Отсюда

$$\varphi_n(t)=\int\limits_{-\infty}^{\infty} e^{itx}\mathsf{P}_n(dx)\to\int\limits_{-\infty}^{\infty} e^{itx}\mathsf{P}(dx),$$

и в то же самое время $\varphi_n(t)\to\varphi(t)$. Поэтому $\varphi(t)$ является характеристической функцией вероятностной меры P.

Теорема доказана.

Следствие. Пусть $\{F_n\}$ — последовательность функций распределения и $\{\varphi_n\}$ — соответствующая последовательность характеристических функций Пусть, кроме того, F — функция распределения, φ — ее характеристическая функция. Тогда $F_n\xrightarrow{w}F$, если и только если $\varphi_n(t)\to\varphi(t)$ для всех $t\in R$.

Замечание. Пусть $\eta,\ \eta_1,\ \eta_2,\ \ldots$ — случайные величины и $F_{\eta_n}\xrightarrow{w}F_\eta$. В соответствии с определением § 10 гл. II тогда говорят, что *случайные величины* $\eta_1,\ \eta_2,\ \ldots$ *сходятся по распределению* к η, и записывают это в виде $\eta_n\xrightarrow{w}\eta$. Эта запись наглядна, и поэтому часто в формулировках предельных теорем ее предпочитают записи $F_{\eta_n}\xrightarrow{w}F_\eta$.

3. В следующем параграфе теорема 1 будет применена для доказательства центральной предельной теоремы для независимых разнораспределенных случайных величин. Доказательство будет вестись при выполнении так называемого «условия Линдеберга». Затем будет показано, что «условие Ляпунова» обеспечивает выполнение «условия Линдеберга». Сейчас же мы остановимся на применении метода характеристических функций к доказательству некоторых простых предельных теорем.

Теорема 2 (закон больших чисел). *Пусть* $\xi_1, \xi_2, \ldots$ — *последовательность независимых одинаково распределенных случайных величин с* $\mathsf{M}|\xi_1| < \infty$, $S_n = \xi_1 + \ldots + \xi_n$ *и* $\mathsf{M}\xi_1 = m$. *Тогда* $\frac{S_n}{n} \xrightarrow{\mathsf{P}} m$, *т. е. для всякого* $\varepsilon > 0$

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right\} \to 0, \qquad n \to \infty.$$

Доказательство. Пусть $\varphi(t) = \mathsf{M}e^{it\xi_1}$ и $\varphi_{\frac{S_n}{n}}(t) = \mathsf{M}e^{it\frac{S_n}{n}}$.

Тогда в силу независимости случайных величин и формулы (II.12.6)

$$\varphi_{\frac{S_n}{n}}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n.$$

Но согласно (II.12.14)

$$\varphi(t) = 1 + itm + o(t), \qquad t \to 0.$$

Значит, для всякого фиксированного $t \in R$

$$\varphi\left(\frac{t}{n}\right) = 1 + i\frac{t}{n}m + o\left(\frac{1}{n}\right), \qquad n \to \infty.$$

и поэтому

$$\varphi_{\frac{S_n}{n}}(t) = \left[1 + i\frac{t}{n}m + o\left(\frac{1}{n}\right)\right]^n \to e^{itm}.$$

Функция $\varphi(t) = e^{itm}$ непрерывна в нуле и является характеристической функцией вырожденного распределения вероятностей, сосредоточенного в точке m. Поэтому

$$\frac{S_n}{n} \xrightarrow{d} m,$$

значит (см. задачу 7 в § 10 гл. II),

$$\frac{S_n}{n} \xrightarrow{\mathsf{P}} m.$$

Теорема доказана.

Теорема 3 (центральная предельная теорема для независимых одинаково распределенных случайных величин). *Пусть* $\xi_1, \xi_2, \ldots$ — *последовательность независимых одинаково распределенных (невырожденных) случайных величин с* $\mathsf{M}\xi_1^2 < \infty$ *и* $S_n = \xi_1 + \ldots + \xi_n$. *Тогда при* $n \to \infty$

$$\mathsf{P}\left\{\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} \leqslant x\right\} \to \Phi(x), \qquad x \in R, \tag{5}$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

Доказательство. Пусть $\mathsf{M}\xi_1 = m$, $\mathsf{D}\xi_1 = \sigma^2$ и

$$\varphi(t) = \mathsf{M}e^{it(\xi_1 - m)}.$$

Тогда, если обозначить

$$\varphi_n(t) = \mathsf{M}e^{it\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}},$$

то получим, что

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

Но в силу (II.12.14)

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \qquad t \to 0.$$

Поэтому для любого фиксированного t и $n \to \infty$

$$\varphi_n(t) = \left[1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + o\left(\frac{1}{n}\right)\right]^n \to e^{-t^2/2}.$$

Функция $e^{-t^2/2}$ является характеристической функцией нормально распределенной случайной величины (обозначим ее $\mathscr{N}(0, 1)$) с нулевым средним и единичной дисперсией, что в силу теоремы 1 и доказывает требуемое утверждение (5). В соответствии с замечанием к теореме 1 это утверждение записывают также в следующем виде:

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} \xrightarrow{d} \mathscr{N}(0, 1). \tag{6}$$

Теорема доказана.

Предыдущие две теоремы относились к асимптотическому поведению вероятностей (нормированных и центрированных) сумм $S_n = \xi_1 + \ldots + \xi_n$ независимых одинаково распределенных случайных величин. Однако, чтобы сформулировать теорему Пуассона (§ 6 гл. I) приходится привлекать к рассмотрению более общую модель, называемую *схемой серий* случайных величин.

Именно, будем предполагать, что для каждого $n \geqslant 1$ задана последовательность независимых случайных величин $\xi_{n,1}, \ldots, \xi_{n,n}$. Иначе говоря, пусть задана треугольная таблица

$$\begin{pmatrix} \xi_{1,1} & & \\ \xi_{2,1}, & \xi_{2,2} & \\ \xi_{3,1}, & \xi_{3,2}, & \xi_{3,3} \\ \cdots & \cdots & \cdots \end{pmatrix}.$$

случайных величин, которые в каждой строчке независимы между собой. Положим $S_n = \xi_{n,1} + \ldots + \xi_{n,n}$.

Теорема 4 (теорема Пуассона). *Пусть при каждом $n \geqslant 1$ независимые одинаково распределенные случайные величины $\xi_{n,1}, \ldots \ldots, \xi_{n,n}$ таковы, что*

$$\mathsf{P}(\xi_{n,k} = 1) = p_n, \ \mathsf{P}(\xi_{n,k} = 0) = q_n, \ 1 \leqslant k \leqslant n,$$

где $p_n + q_n = 1$ и $p_n \to 0$ так, что $np_n \to \lambda > 0$, $n \to \infty$. Тогда

$$\mathsf{P}(S_n = m) \to \frac{e^{-\lambda}\lambda^m}{m!}, \quad m = 0, 1, \ldots \tag{7}$$

Доказательство. Поскольку для $1 \leqslant k \leqslant n$

$$\mathsf{M}e^{\xi_{n,k}} = p_n e^{it} + q_n,$$

то

$$\varphi_{S_n}(t) = \mathsf{M}e^{itS_n} = (p_n e^{it} + q_n)^n = \\ = (1 + p_n(e^{it} - 1))^n \to \exp\{\lambda(e^{it} - 1)\}, \quad n \to \infty.$$

Функция $\varphi(t) = \exp\{\lambda(e^{it} - 1)\}$ является характеристической функцией пуассоновского распределения (II.12.11), что и доказывает (7).

Если через $\pi(\lambda)$ обозначить пуассоновскую случайную величину с параметром λ, то по аналогии с (6) утверждение (7) можно записать также в следующем виде:

$$S_n \xrightarrow{d} \pi(\lambda).$$

Теорема доказана.

4. Задачи.

1. Доказать справедливость утверждений теоремы 1 для случая пространств R^n, $n \geqslant 2$.

2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечными средними значениями $\mathsf{M}|\xi_n|$ и дисперсиями $\mathsf{D}\xi_n$ такими, что $\mathsf{D}\xi_n \leqslant K < \infty$, где K — некоторая константа. Используя неравенство Чебышева, доказать справедливость закона больших чисел (1).

3. В следствии к теореме 1 установить, что семейство $\{\varphi_n\}$ *равностепенно непрерывно* и сходимость $\varphi_n \to \varphi$ равномерна на каждом ограниченном интервале.

4. Пусть ξ_n, $n \geqslant 1$, случайные величины с характеристическими функциями $\varphi_{\xi_n}(t)$, $n \geqslant 1$. Показать, что $\xi_n \xrightarrow{d} 0$ тогда и только тогда, когда в некоторой окрестности точки $t = 0$ $\varphi_{\xi_n}(t) \to 1$, $n \to \infty$.

5. Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных векторов (со значениями в R^k), имеющих нулевое среднее

и (конечную) матрицу ковариаций Г. Показать, что

$$\frac{X_1+\ldots+X_n}{\sqrt{n}} \xrightarrow{d} \mathscr{N}(0,\ \Gamma).$$

(Ср. с теоремой 3.)

§ 4. Центральная предельная теорема

1. Теорема 1. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечными вторыми моментами. Пусть $m_k = \mathsf{M}\xi_k$, $\sigma_k^2 = \mathsf{D}\xi_k > 0$, $S_n = \xi_1 + \ldots + \xi_n$, $D_n^2 = \sum_{k=1}^{n} \sigma_k^2$ и $F_k = F_k(x)$ — функция распределения случайной величины ξ_k.*

Предположим, что выполнено «условие Линдеберга»: для всякого $\varepsilon > 0$

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int\limits_{\{x:\ |x-m_k| \geqslant \varepsilon D_n\}} (x-m_k)^2\, dF_k(x) \to 0, \qquad n \to \infty. \tag{1}$$

Тогда

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} \xrightarrow{d} \mathscr{N}(0,\ 1). \tag{2}$$

Доказательство. Без ограничения общности можно считать $m_k = 0$, $k \geqslant 1$. Обозначим $\varphi_k(t) = \mathsf{M}e^{it\xi_k}$, $T_n = \frac{S_n}{\sqrt{\mathsf{D}S_n}} = \frac{S_n}{D_n}$, $\varphi_{S_n}(t) = \mathsf{M}e^{itS_n}$, $\varphi_{T_n}(t) = \mathsf{M}e^{itT_n}$.

Тогда

$$\varphi_{T_n}(t) = \mathsf{M}e^{itT_n} = \mathsf{M}e^{i\frac{t}{D_n}S_n} = \varphi_{S_n}\left(\frac{t}{D_n}\right) = \prod_{k=1}^{n} \varphi_k\left(\frac{t}{D_n}\right) \tag{3}$$

и для доказательства (2) достаточно (в силу теоремы 1 из § 3) установить, что для каждого $t \in R$

$$\varphi_{T_n}(t) \to e^{-t^2/2}, \qquad n \to \infty. \tag{4}$$

Возьмем некоторое $t \in R$ и будем считать его фиксированным на протяжении всего доказательства. В силу разложений

$$e^{iy} = 1 + iy + \frac{\theta_1 y^2}{2},$$

справедливых для каждого действительного y с $\theta_1 = \theta_1(y)$, $\theta_2 = \theta_2(y)$, такими, что $|\theta_1| \leqslant 1$, $|\theta_2| \leqslant 1$, находим, что

$$\varphi_k(t) = \mathsf{M} e^{it\xi_k} =$$
$$= \int_{-\infty}^{\infty} e^{itx}\, dF_k(x) = \int_{|x| \geqslant \varepsilon D_n} \left(1 + itx + \frac{\theta_1 (tx)^2}{2}\right) dF_k(x) +$$
$$+ \int_{|x| < \varepsilon D_n} \left(1 + itx - \frac{t^2x^2}{2} + \frac{\theta_2 |tx|^3}{6}\right) dF_k(x) =$$
$$= 1 + \frac{t^2}{2} \int_{|x| \geqslant \varepsilon D_n} \theta_1 x^2\, dF_k(x) - \frac{t^2}{2} \int_{|x| < \varepsilon D_n} x^2\, dF_k(x) +$$
$$+ \frac{|t|^3}{6} \int_{|x| < \varepsilon D_n} \theta_2 |x|^3\, dF_k(x)$$

(здесь мы воспользовались также тем, что, согласно предположению, $m_k = \int_{-\infty}^{\infty} x\, dF_k(x) = 0$).

Следовательно,

$$\varphi_k\left(\frac{t}{D_n}\right) = 1 - \frac{t^2}{2D_n^2} \int_{|x| < \varepsilon D_n} x^2\, dF_k(x) +$$
$$+ \frac{t^2}{2D_n^2} \int_{|x| \geqslant \varepsilon D_n} \theta_1 x^2\, dF_k(x) + \frac{|t|^3}{6D_n^3} \int_{|x| < \varepsilon D_n} \theta_2 |x|^3\, dF_k(x). \tag{5}$$

Поскольку

$$\left| \frac{1}{2} \int_{|x| \geqslant \varepsilon D_n} \theta_1 x^2\, dF_k(x) \right| \leqslant \frac{1}{2} \int_{|x| \geqslant \varepsilon D_n} x^2\, dF_k(x),$$

то

$$\frac{1}{2} \int_{|x| \geqslant \varepsilon D_n} \theta_1 x^2\, dF_k(x) = \tilde{\theta}_1 \int_{|x| \geqslant \varepsilon D_n} x^2\, dF_k(x), \tag{6}$$

где $\tilde{\theta}_1 = \tilde{\theta}_1(t, k, n)$ и $|\tilde{\theta}_1| \leqslant 1/2$.

Точно так же

$$\left| \frac{1}{6} \int_{|x| < \varepsilon D_n} \theta_2 |x|^3\, dF_k(x) \right| \leqslant \frac{1}{6} \int_{|x| < \varepsilon D_n} \frac{\varepsilon D_n}{|x|} \cdot |x|^3\, dF_k(x) \leqslant$$
$$\leqslant \frac{1}{6} \int_{|x| < \varepsilon D_n} \varepsilon D_n x^2\, dF_k(x),$$

и, значит,

$$\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\theta_2|x|^3\,dF_k(x)=\tilde{\theta}_2\int\limits_{|x|<\varepsilon D_n}\varepsilon D_n x^2\,dF_k(x), \tag{7}$$

где $\tilde{\theta}_2=\tilde{\theta}_2(t,\ k,\ n)$ и $|\tilde{\theta}_2|\leqslant 1/6$.

Положим теперь

$$A_{kn}=\frac{1}{D_n^2}\int\limits_{|x|<\varepsilon D_n}x^2\,dF_k(x),$$

$$B_{kn}=\frac{1}{D_n^2}\int\limits_{|x|\geqslant\varepsilon D_n}x^2\,dF_k(x).$$

Тогда в силу (5) — (7)

$$\varphi_k\left(\frac{t}{D_n}\right)=1-\frac{t^2A_{kn}}{2}+t^2\tilde{\theta}_1B_{kn}+|t|^3\varepsilon\tilde{\theta}_2A_{kn}=1+C_{kn}. \tag{8}$$

Заметим, что

$$\sum_{k=1}^{n}(A_{kn}+B_{kn})=1 \tag{9}$$

и, согласно условию (1),

$$\sum_{k=1}^{n}B_{kn}\to 0,\qquad n\to\infty. \tag{10}$$

Поэтому для достаточно больших n

$$\max_{1\leqslant k\leqslant n}|C_{kn}|\leqslant t^2\varepsilon^2+\varepsilon|t|^3 \tag{11}$$

и

$$\sum_{k=1}^{n}|C_{kn}|\leqslant t^2+\varepsilon|t|^3. \tag{12}$$

Воспользуемся теперь тем, что для любых комплексных чисел z с $|z|\geqslant 1/2$

$$\ln(1+z)=z+\theta|z|^2,$$

где $\theta=\theta(z)$ с $|\theta|\leqslant 1$ и $\ln$ обозначает главное значение логарифма. Тогда для достаточно больших n из (8) и (11) следует, что для достаточно малых $\varepsilon>0$

$$\ln\varphi_k\left(\frac{t}{D_n}\right)=\ln(1+C_{kn})=C_{kn}+\theta_{kn}|C_{kn}|^2,$$

где $|\theta_{kn}|\leqslant 1$. Следовательно, из (3)

$$\frac{t^2}{2}+\ln\varphi_{T_n}(t)=\frac{t^2}{2}+\sum_{k=1}^{n}\ln\varphi_k\left(\frac{t}{D_n}\right)=$$
$$=\frac{t^2}{2}+\sum_{k=1}^{n}C_{kn}+\sum_{k=1}^{n}\theta_{kn}|C_{kn}|^2.$$

Но

$$\frac{t^2}{2}+\sum_{k=1}^{n}C_{kn}=\frac{t^2}{2}\left(1-\sum_{k=1}^{n}A_{kn}\right)+t^2\sum_{k=1}^{n}\tilde\theta_1(t,\ k,\ n)B_{kn}+$$
$$+\varepsilon|t|^3\sum_{k=1}^{n}\tilde\theta_2(t,\ k,\ n)A_{kn},$$

и в силу (9), (10) для любого $\delta>0$ можно найти столь большое n_0 и $\varepsilon>0$, что для всех $n\geqslant n_0$

$$\left|\frac{t^2}{2}+\sum_{k=1}^{n}C_{kn}\right|\leqslant\frac{\delta}{2}.$$

Далее, в силу (11) и (12)

$$\left|\sum_{k=1}^{n}\theta_{kn}|C_{kn}|^2\right|\leqslant\max_{1\leqslant k\leqslant n}|C_{kn}|\sum_{k=1}^{n}|C_{kn}|\leqslant(t^2\varepsilon^2+\varepsilon|t|^3)(t^2+\varepsilon|t|^3).$$

Поэтому для достаточно больших n за счет выбора $\varepsilon>0$ можно добиться того, что

$$\left|\sum_{k=1}^{n}\theta_{kn}|C_{kn}|^2\right|\leqslant\frac{\delta}{2}$$

и, следовательно,

$$\left|\frac{t^2}{2}+\ln\varphi_{T_n}(t)\right|\leqslant\delta.$$

Таким образом, для любого действительного t

$$\varphi_{T_n}(t)e^{t^2/2}\to 1,\qquad n\to\infty$$

и, значит,

$$\varphi_{T_n}(t)\to e^{-t^2/2},\qquad n\to\infty.$$

Теорема доказана.

2. Остановимся на некоторых частных случаях, в которых выполнено условие Линдеберга (1) и, следовательно, справедлива центральная предельная теорема.

a) Пусть выполнено «условие Ляпунова»: для некоторого $\delta > 0$

$$\frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \mathsf{M} |\xi_k - m_k|^{2+\delta} \to 0, \qquad n \to \infty. \tag{13}$$

Пусть $\varepsilon > 0$, тогда

$$\mathsf{M} |\xi_k - m_k|^{2+\delta} = \int_{-\infty}^{\infty} |x - m_k|^{2+\delta} \, dF_k(x) \geqslant$$
$$\geqslant \int_{\{x:\, x\,|x - m_k| \geqslant \varepsilon D_n\}} |x - m_k|^{2+\delta} \, dF_k(x) \geqslant$$
$$\geqslant \varepsilon^\delta D_n^\delta \int_{\{x:\, |x - m_k| \geqslant \varepsilon D_n\}} (x - m_k)^2 \, dF_k(x)$$

и, значит,

$$\frac{1}{D_n^2} \sum_{k=1} \int_{\{x:\, |x - m_k| \geqslant \varepsilon D_n\}} (x - m_k)^2 \, dF_k(x) \leqslant$$
$$\leqslant \frac{1}{\varepsilon^\delta} \cdot \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \mathsf{M} |\xi_k - m_k|^{2+\delta}.$$

Следовательно, «условие Ляпунова» обеспечивает выполнение «условия Линдеберга».

b) Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $m = \mathsf{M}\xi_1$ и дисперсией $0 < \sigma^2 \equiv \mathsf{D}\xi_1 < \infty$. Тогда

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x:\, |x - m| \geqslant \varepsilon D_n\}} |x - m|^2 \, dF_k(x) =$$
$$= \frac{n}{n\sigma^2} \int_{\{x:\, |x - m| \geqslant \varepsilon \sigma^2 \sqrt{n}} |x - m|^2 \, dF_1(x) \to 0,$$

поскольку $\{x: |x - m| \geqslant \varepsilon\sigma^2 \sqrt{n}\} \downarrow \varnothing$, $n \to \infty$, а $\sigma^2 = \mathsf{M} |\xi_1 - m|^2 < \infty$.

Таким образом, «условие Линдеберга» выполнено и, следовательно, теорема 3 из § 3 вытекает из доказанной теоремы 1.

c) Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины такие, что для всех $n \geqslant 1$

$$|\xi_k| \leqslant K < \infty,$$

где K — некоторая постоянная, и $D_n \to \infty$, $n \to \infty$, Тогда из неравенства Чебышева

$$\int\limits_{\{x:\ |x-m_k|\geqslant \varepsilon D_n\}} |x-m_k|^2 dF_k(x) = \mathsf{M}[(\xi_k-m_k)^2 I(|\xi_k-m_k|\geqslant \varepsilon D_n)] \leqslant$$

$$\leqslant (2K)^2 \mathsf{P}\{|\xi_k-m_k|\geqslant \varepsilon D_n\} \leqslant (2K)^2 \frac{\sigma_k^2}{\varepsilon^2 D_n^2}$$

и, значит,

$$\frac{1}{D_n^2}\sum_{k=1}^{n}\int\limits_{\{x:\ |x-m_k|\geqslant \varepsilon D_n\}} |x-m_k|^2\, dF_k(x) \leqslant \frac{(2K)^2}{\varepsilon^2 D_n^2} \to 0, \qquad n\to\infty.$$

Следовательно, снова выполнено условие Линдеберга и, значит, справедлива центральная предельная теорема.

3. Замечание 1. Условие Линдеберга достаточно для справедливости центральной предельной теоремы. Оказывается, что при некотором дополнительном условии (асимптотической малости величин $\frac{\xi_k-\mathsf{M}\xi_k}{D_n}$) условие Линдеберга оказывается и необходимым.

Будем говорить, что величины $\frac{\xi_k-\mathsf{M}\xi_k}{D_n}$, $1\leqslant k\leqslant n$, $n\geqslant 1$, *асимптотически малы*, если для любого $\varepsilon>0$

$$\max_{1\leqslant k\leqslant n} \mathsf{P}\left\{\left|\frac{\xi_k-\mathsf{M}\xi_k}{D_n}\right|\geqslant \varepsilon\right\}\to 0, \qquad n\to\infty. \tag{14}$$

Поскольку

$$\frac{1}{D_n^2}\sum_{k=1}^{n}\int\limits_{\{x:\ |x-m_k|\geqslant \varepsilon D_n\}} (x-m_k)^2\, dF_k(x) \geqslant$$

$$\geqslant \varepsilon^2 \sum_{k=1}^{n} \mathsf{P}\{|\xi_k-m_k|\geqslant \varepsilon D_n\} \geqslant \varepsilon^2 \max_{1\leqslant k\leqslant n} \mathsf{P}\{|\xi_k-m_k|\geqslant \varepsilon D_n\},$$

то из условия Линдеберга вытекает условие (14).

Вместе с теоремой 1 это показывает, что условие Линдеберга достаточно для выполнения центральной предельной теоремы и условия асимптотической малости. Следующая теорема, приводимая без доказательства, показывает, что условие Линдеберга является и необходимым.

Теорема 2. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечным вторым моментом. Условие Линдеберга* (1) *является необходимым и достаточным для* (2) *и* (14).

Замечание 2. Пусть $T_n = \frac{S_n-\mathsf{M}S_n}{D_n}$ и $F_{T_n}(x)=\mathsf{P}(T_n\leqslant x)$. Тогда утверждение (2) означает, что для всякого $x\in R$

$$F_{T_n}(x)\to\Phi(x), \qquad n\to\infty.$$

Поскольку функция $\Phi(x)$ непрерывна, то на самом деле сходимость здесь равномерная (задача 5 в § 1):

$$\sup_{x \in R} |F_{T_n}(x) - \Phi(x)| \to 0, \quad n \to \infty. \tag{15}$$

В частности, отсюда следует, что

$$\mathsf{P}\{S_n \leqslant x\} - \Phi\left(\frac{x - \mathsf{M}S_n}{D_n}\right) \to 0, \quad n \to \infty.$$

Это утверждение часто выражают словами, что при достаточно большом n величина S_n примерно нормально распределена со средним $\mathsf{M}S_n$ и дисперсией $D_n^2 \equiv \mathsf{D}S_n$.

Замечание 3. Поскольку в соответствии с предыдущим замечанием сходимость $F_{T_n}(x) \to \Phi(x)$, $n \to \infty$, равномерна по x, то естественно поставить вопрос о скорости сходимости в (15). В том случае, когда величины $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены и $\mathsf{M}|\xi_1|^3 < \infty$, ответ на этот вопрос дается *неравенством Берри — Эссеена*:

$$\sup_x |F_{T_n}(x) - \Phi(x)| \leqslant C \frac{\mathsf{M}|\xi_1 - \mathsf{M}\xi_1|^3}{\sigma^3 \sqrt{n}}, \tag{16}$$

где абсолютная константа C такова, что

$$\frac{1}{\sqrt{2\pi}} \leqslant C < 0{,}8.$$

Важно подчеркнуть, что без дополнительных предположений о природе суммируемых случайных величин порядок оценки (16) не может быть улучшен (см. задачу 3).

4. Задачи.

1. Показать, что в доказательстве теоремы 1 в самом деле без ограничения общности можно считать $m_k = 0$, $k \geqslant 1$.

2. Пусть $\xi_1, \xi_2, \ldots$ последовательность независимых нормально распределенных случайных величин с $\mathsf{M}\xi_k = 0$, $k \geqslant 1$, и $\mathsf{D}\xi_1 = 1$, $\mathsf{D}\xi_k = 2^{k-2}$, $k \geqslant 2$. Показать, что в этом случае условие Линдеберга не выполнено, но в то же самое время центральная предельная теорема (2) справедлива.

3. Показать, что в схеме Бернулли величина $\sup_x |F_{T_n}(x) - \Phi(x)|$ имеет порядок $\frac{1}{\sqrt{n}}$, $n \to \infty$.

4. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{M}\xi_1 = 0$, $\mathsf{M}\xi_1^2 = 1$. Покажите, что $\max\left(\frac{|\xi_1|}{\sqrt{n}}, \ldots, \frac{|\xi_n|}{\sqrt{n}}\right) \xrightarrow{d} 0$, $n \to \infty$.

§ 5. Безгранично делимые и устойчивые распределения

1. В § 3 отмечалось, что для формулирования теоремы Пуассона приходится прибегать к рассмотрению так называемой схемы серий, считая, что при каждом $n \geqslant 1$ задана последовательность независимых случайных величин $\{\xi_{n,k}\}$, $1 \leqslant k \leqslant n$.

Положим

$$T_n = \xi_{n,1} + \ldots + \xi_{n,n}, \qquad n \geqslant 1. \tag{1}$$

Понятие безгранично делимого распределения возникает в связи со следующим вопросом: как охарактеризовать все те распределения, которые могут выступать в качестве предельных для последовательности распределений случайных величин T_n, $n \geqslant 1$?

Вообще говоря, при такой общей постановке вопроса предельное распределение может быть произвольным. Действительно, если ξ — некоторая случайная величина и $\xi_{n,1} = \xi$, $\xi_{n,k} = 0$, $1 < k \leqslant n$, то $T_n \equiv \xi$ и, следовательно, предельное распределение совпадает с распределением ξ, которое может быть взято произвольным.

Чтобы сделать задачу о предельных распределениях более содержательной, будем всюду в этом параграфе предполагать, что при каждом $n \geqslant 1$ величины $\xi_{n,1}, \ldots, \xi_{n,n}$ не только *независимы*, но и *одинаково распределены*.

Напомним, что именно такая ситуация имела место в теореме Пуассона (теорема 4 из § 3). К этой схеме относится и центральная предельная теорема (теорема 3 из § 3) для сумм $S_n = \xi_1 + \ldots + \xi_n$, $n \geqslant 1$, независимых и одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots$ В самом деле, если положить

$$\xi_{n,k} = \frac{\xi_k - \mathsf{M}\xi_k}{\mathsf{D}_n}, \qquad D_n^2 = \mathsf{D}S_n,$$

то тогда

$$T_n = \sum_{k=1}^{n} \xi_{n,k} = \frac{S_n - \mathsf{M}S_n}{D_n}.$$

Таким образом, нормальное и пуассоновское распределения могут выступать в качестве предельных в схеме серий. Если $T_n \xrightarrow{d} T$, то интуитивно понятно, что, поскольку T_n есть сумма независимых одинаково распределенных случайных величин, то предельная величина T должна быть также суммой независимых одинаково распределенных случайных величин. Имея это в виду, введем такое

Определение 1. Случайная величина T (а также ее функция распределения F_T и ее характеристическая функция φ_T) называется *безгранично делимой*, если для каждого $n \geqslant 1$ можно найти такие независимые одинаково распределенные случайные величины $\eta_1, \ldots, \eta_n$, что *) $T \overset{d}{=} \eta_1 + \ldots + \eta_n$ (или, что то же самое, $F_T = F_{\eta_1} * \ldots * F_{\eta_n}$, или $\varphi_T = (\varphi_{\eta_1})^n$).

Теорема 1. *Случайная величина T может быть пределом по распределению сумм $T_n = \sum_{k=1}^{n} \xi_{n,k}$ в том и только том случае, когда T безгранично делима.*

Доказательство. Если ξ безгранично делима, то для каждого $n \geqslant 1$ существуют независимые одинаково распределенные случайные величины $\xi_{n,1}, \ldots, \xi_{n,k}$ такие, что $\xi \overset{d}{=} \xi_{n,1} + \ldots + \xi_{n,k}$, а это и означает, что $\xi \overset{d}{=} T_n$, $n \geqslant 1$.

Обратно, пусть $T_n \xrightarrow{d} T$. Покажем, что тогда T безгранично делима, т. е. для любого k найдутся независимые одинаково распределенные случайные величины $\eta_1, \ldots, \eta_k$ такие, что $T \overset{d}{=} \eta_1 + \ldots + \eta_k$.

Зафиксируем некоторое $k \geqslant 1$ и представим величину T_{nk} в виде $\zeta_n^{(1)} + \ldots + \zeta_n^{(k)}$, где

$$\zeta_n^{(1)} = \xi_{nk,1} + \ldots + \xi_{nk,n}, \ \ldots, \ \zeta_n^{(k)} = \xi_{nk,n(k-1)+1} + \ldots + \xi_{nk,nk}.$$

Поскольку $T_{nk} \xrightarrow{d} T$, $n \to \infty$, то последовательность функций распределений, соответствующих случайным величинам T_{nk}, $n \geqslant 1$, относительно компактна и, значит, по теореме Прохорова плотна. Далее,

$$[\mathrm{P}(\zeta_n^{(1)} > z)]^k = \mathrm{P}(\zeta_n^{(1)} > z, \ \ldots, \ \zeta_n^{(k)} > z) \leqslant \mathrm{P}(T_{nk} > kz)$$

и

$$[\mathrm{P}(\zeta_n^{(1)} < -z)]^k = \mathrm{P}(\zeta_n^{(1)} < -z, \ \ldots, \ \zeta_n^{(k)} < -z) \leqslant \mathrm{P}(T_{nk} < -kz).$$

Из этих двух неравенств и плотности семейства распределений для T_{nk}, $n \geqslant 1$, вытекает плотность семейства распределений для $\zeta_n^{(1)}$, $n \geqslant 1$. Поэтому найдется подпоследовательность $\{n_i\} \subseteq \{n\}$ и случайная величина η_1 такая, что $\zeta_{n_i}^{(1)} \xrightarrow{d} \eta_1$, $n_i \to \infty$. Поскольку величины $\zeta_n^{(1)}, \ldots, \zeta_n^{(k)}$ одинаково распределены, то $\zeta_{n_i}^{(2)} \xrightarrow{d} \eta_2, \ldots, \zeta_{n_i}^{(k)} \xrightarrow{d} \eta_k$, где $\eta_1 \overset{d}{=} \eta_2 \overset{d}{=} \ldots \overset{d}{=} \eta_k$. В силу независимости величин $\zeta_n^{(1)}, \ldots, \zeta_n^{(k)}$ из следствия к теореме 1 из § 3 вытекает, что величины

*) Запись $\xi \overset{d}{=} \eta$ означает, что случайные величины ξ и η совпадают *по распределению*, т. е. $F_\xi(x) = F_\eta(x)$, $x \in R$.

$\eta_1, \ldots, \eta_k$ независимы и

$$T_{n_i k} = \zeta_{n_i}^{(1)} + \ldots + \zeta_{n_i}^{(k)} \xrightarrow{d} \eta_1 + \ldots + \eta_k.$$

Но $T_{n_i k} \xrightarrow{d} T$, поэтому (задача 1)

$$\xi \stackrel{d}{=} \eta_1 + \ldots + \eta_k.$$

Теорема доказана.

Замечание. Утверждение теоремы остается в силе, если условие, что при каждом $n \geqslant 1$ величины $\xi_{n,1}, \ldots, \xi_{n,n}$ одинаково распределены, заменить на условие равномерной асимптотической малости (4.14).

2. При проверке того, является ли данная случайная величина T безгранично делимой, проще всего исходить из вида ее характеристической функции $\varphi(t)$. Если для любого $n \geqslant 1$ можно найти такие характеристические функции $\varphi_n(t)$, что $\varphi(t) = [\varphi_n(t)]^n$, то T безгранично делима.

В гауссовском случае

$$\varphi(t) = e^{itm} e^{-\frac{t^2\sigma^2}{2}},$$

и, полагая

$$\varphi_n(t) = e^{it\frac{m}{n}} e^{-\frac{t^2 \frac{\sigma^2}{n}}{2}},$$

сразу находим, что $\varphi(t) = [\varphi_n(t)]^n$.

В пуассоновском случае

$$\varphi(t) = e^{\lambda(e^{it}-1)},$$

и если положить $\varphi_n(t) = e^{\frac{\lambda}{n}(e^{it}-1)}$, то $\varphi(t) = [\varphi_n(t)]^n$.

Если случайная величина T имеет Г-распределение с плотностью

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

то, как нетрудно показать, ее характеристическая функция равна

$$\varphi(t) = \frac{1}{(1 - i\beta t)^\alpha}.$$

Следовательно, $\varphi(t) = (\varphi_n(t))^n$, где

$$\varphi_n(t) = \frac{1}{(1 - i\beta t)^{\alpha/n}},$$

и, значит, T безгранично делима.

Приведем без доказательства следующий результат об общем виде характеристической функции безгранично делимых распределений.

Теорема 2 (представление Леви — Хинчина). *Случайная величина T является безгранично делимой тогда и только тогда, когда* $\varphi(t)=\exp\psi(t)$ *с*

$$\psi(t)=it\beta-\frac{t^2\sigma^2}{2}+\int_{-\infty}^{\infty}\left(e^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\,d\lambda(x), \tag{2}$$

где $\beta\in R$, $\sigma^2\geqslant 0$ *и* λ — *некоторая конечная мера на* $(R, \mathscr{B}(R))$ *с* $\lambda\{0\}=0$.

3. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых, одинаково распределенных случайных величин и $S_n=\xi_1+\ldots+\xi_n$. Предположим, что существуют такие константы b_n, $a_n>0$ и случайная величина T, что

$$\frac{S_n-b_n}{a_n}\xrightarrow{d}T. \tag{3}$$

Спрашивается, как охарактеризовать все распределения (случайных величин T), которые могут возникать в виде предельных распределений в (3)?

Если независимые одинаково распределенные случайные величины $\xi_1, \xi_2, \ldots$ таковы, что $0<\sigma^2\equiv \mathsf{D}\xi_1<\infty$, то, полагая $b_n=nM\xi_1$ и $a_n=\sigma\sqrt{n}$, согласно § 4, находим, что T имеет нормальное распределение $\mathscr{N}(0, 1)$.

Если $f(x)=\dfrac{\theta}{\pi(x^2+\theta^2)}$ — плотность распределения Коши (с параметром $\theta>0$) и $\xi_1, \xi_2, \ldots$ — независимые случайные величины с плотностью $f(x)$, то характеристическая функция $\varphi_{\xi_1}(t)$ равна $e^{-\theta|t|}$ и, значит, $\varphi_{S_n/n}(t)=\left(e^{-\frac{\theta}{n}|t|}\right)^n=e^{-\theta|t|}$, т. е. величина S_n/n имеет также распределение Коши (с тем же самым параметром θ).

Таким образом, в качестве предельных распределений, помимо нормального, могут появляться и другие распределения (как, например, распределение Коши).

Если положить $\xi_{n,k}=\dfrac{\xi_k}{a_n}-\dfrac{b_n}{na_n}$, $1\leqslant k\leqslant n$, то найдем, что

$$\frac{S_n-b_n}{a_n}=\sum_{k=1}^{n}\xi_{n,k}\quad (=T_n).$$

Таким образом, все мыслимые распределения для T, которые могут появляться в качестве предельных в (3), обязательно являются (в соответствии с теоремой 1) безгранично делимыми. Однако специфика рассматриваемых величин $T_n=\dfrac{S_n-b_n}{a_n}$ дает возможность получить дополнительную информацию о структуре возникающих здесь предельных распределений.

С этой целью введем такое

Определение 2. Случайная величина T (а также ее функция распределения $F(x)$ и характеристическая функция $\varphi(t)$) называется *устойчивой*, если для любого $n \geqslant 1$ найдутся такие константы $a_n > 0$, b_n и такие независимые случайные величины $\xi_1, \ldots, \xi_n$, распределенные как T, что

$$a_n T + b_n \stackrel{d}{=} \xi_1 + \ldots + \xi_n \tag{4}$$

или, что то же самое, $F\left(\frac{x-b_n}{a_n}\right) = F * \ldots * F(x)$, или

$$[\varphi(t)]^n = [\varphi(a_n t)] e^{itb_n}. \tag{5}$$

Теорема 3. *Случайная величина T может быть пределом по распределению случайных величин $\frac{S_n - b_n}{a_n}$, $a_n > 0$, тогда и только тогда, когда T является устойчивой.*

Доказательство. Если T устойчива, то, согласно (4),

$$T \stackrel{d}{=} \frac{S_n - b_n}{a_n},$$

где $S_n = \xi_1 + \ldots + \xi_n$, и, следовательно, $\frac{S_n - b_n}{a_n} \xrightarrow{d} T$.

Обратно, пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, $S_n = \xi_1 + \ldots + \xi_n$ и $\frac{S_n - b_n}{a_n} \xrightarrow{d} T$, $a_n > 0$. Покажем, что T является устойчивой случайной величиной.

Если T — вырожденная случайная величина, то она, очевидно, устойчива. Будем поэтому предполагать, что T является невырожденной случайной величиной.

Зафиксируем $k \geqslant 1$ и обозначим

$$S_n^{(1)} = \xi_1 + \ldots + \xi_n, \ \ldots, \ S_n^{(k)} = \xi_{(k-1)n+1} + \ldots + \xi_{kn},$$

$$T_n^{(1)} = \frac{S_n^{(1)} - b_n}{a_n}, \ \ldots, \ T_n^{(k)} = \frac{S_n^{(k)} - b_n}{a_n}.$$

Ясно, что по распределению все величины $T_n^{(1)}, \ldots, T_n^{(k)}$ совпадают и

$$T_n^{(i)} \xrightarrow{d} T, \quad n \to \infty, \quad i = 1, \ldots, k.$$

Обозначим

Тогда

$$U_n^{(k)} \xrightarrow{d} T^{(1)} + \ldots + T^{(k)},$$

где $T^{(1)} \overset{d}{=} \ldots \overset{d}{=} T^{(k)} \overset{d}{=} T$.

С другой стороны,

$$U_n^{(k)} = \frac{\xi_1 + \ldots + \xi_{kn} - kb_n}{a_n} =$$

$$= \frac{a_{kn}}{a_n}\left(\frac{\xi_1 + \ldots + \xi_{kn} - b_{kn}}{a_{kn}}\right) + \frac{b_{kn} - kb_n}{a_n} = \alpha_n^{(k)} V_{kn} + \beta_n^{(k)}, \tag{6}$$

где

$$\alpha_n^{(k)} = \frac{a_{kn}}{a_n}, \quad \beta_n^{(k)} = \frac{b_{kn} - kb_n}{a_n}$$

и

$$V_{kn} = \frac{\xi_1 + \ldots + \xi_{kn} - b_{kn}}{a_{kn}}.$$

Из (6) ясно, что

$$V_{kn} = \frac{U_n^{(k)} - \beta_n^{(k)}}{\alpha_n^{(k)}},$$

где $V_{kn} \xrightarrow{d} T$, $U_n^{(k)} \xrightarrow{d} T^{(1)} + \ldots + T^{(k)}$, $n \to \infty$.

Из приводимой ниже леммы следует, что найдутся такие константы $\alpha^{(k)} > 0$ и $\beta^{(k)}$, что $\alpha_n^{(k)} \to \alpha^{(k)}$, $\beta_n^{(k)} \to \beta^{(k)}$, $n \to \infty$. Поэтому

$$T \overset{d}{=} \frac{T^{(1)} + \ldots + T^{(k)} - \beta^{(k)}}{\alpha^{(k)}},$$

что и доказывает, что T является устойчивой случайной величиной.

Теорема доказана.

Сформулируем и докажем упомянутую выше лемму.

Лемма. *Пусть $\xi_n \xrightarrow{d} \xi$ и существуют такие константы $a_n > 0$ и b_n, что*

$$a_n \xi_n + b_n \xrightarrow{d} \tilde{\xi},$$

причем случайные величины ξ и $\tilde{\xi}$ не вырождены. Тогда найдутся такие константы $a > 0$ и b, что $\lim a_n = a$, $\lim b_n = b$ и

$$\tilde{\xi} \overset{d}{=} a\xi + b.$$

Доказательство. Пусть φ_n, φ и $\tilde{\varphi}$ — характеристические функции ξ_n, ξ и $\tilde{\xi}$ соответственно. Тогда $\varphi_{a_n\xi_n + b_n}(t)$, характеристическая функция $a_n\xi_n + b_n$, равна $e^{itb_n}\varphi_n(a_n t)$, и согласно следствию к теореме 1 и задаче 3 из § 3,

$$e^{itb_n}\varphi_n(a_n t) \to \tilde{\varphi}(t), \tag{7}$$

$$\varphi_n(t) \to \varphi(t) \tag{8}$$

равномерно на каждом конечном интервале изменения t.

Пусть $\{n_i\}$ — подпоследовательность $\{n\}$ такая, что $a_{n_i} \to a$. Покажем прежде всего, что $a < \infty$. Пусть $a = \infty$. В силу (7) для любого $c > 0$

$$\sup_{|t| \leqslant c} \big|\, |\varphi_n(a_n t)| - |\tilde{\varphi}(t)| \,\big| \to 0, \qquad n \to \infty.$$

Возьмем вместо t величину $t_{n_i} = \dfrac{t_0}{a_{n_i}}$. Тогда, поскольку $a_{n_i} \to \infty$, то

$$\left|\varphi_{n_i}\left(a_{n_i}\frac{t_0}{a_{n_i}}\right)\right| - \left|\tilde{\varphi}\left(\frac{t_0}{a_{n_i}}\right)\right| \to 0$$

и, значит,

$$|\varphi_{n_i}(t_0)| \to |\tilde{\varphi}(0)| = 1.$$

Но $|\varphi_{n_i}(t_0)| \to |\varphi(t_0)|$. Поэтому $|\varphi(t_0)| = 1$ для любого $t_0 \in R$, и, следовательно, согласно теореме 5 из § 12 гл. II, случайная величина ξ должна быть вырождена, что противоречит предположению леммы.

Итак, $a < \infty$. Предположим теперь, что существуют две подпоследовательности $\{n_i\}$ и $\{n_i'\}$ такие, что $a_{n_i} \to a$, $a_{n_i'} \to a'$, где $a \neq a'$ и для определенности $0 \leqslant a' < a$. Тогда из (7) и (8)

$$|\varphi_{n_i}(a_{n_i}t)| \to |\varphi(at)|, \qquad |\varphi_{n_i}(a_{n_i}t)| \to |\tilde{\varphi}(t)|$$

и

$$|\varphi_{n_i'}(a_{n_i'}t)| \to |\varphi(a't)|, \qquad |\varphi_{n_i'}(a_{n_i'}(t)| \to |\tilde{\varphi}(t)|.$$

Следовательно,

$$|\varphi(at)| = |\varphi(a't)|,$$

и, значит, для любого $t \in R$

$$|\varphi(t)| = \left|\varphi\left(\frac{a'}{a}\,t\right)\right| = \ldots = \left|\varphi\left(\left(\frac{a'}{a}\right)^n t\right)\right| \to 1, \qquad n \to \infty.$$

Поэтому $|\varphi(t)| \equiv 1$ и, согласно теореме 5 из § 12 гл. II, отсюда вытекает, что ξ — вырожденная случайная величина. Полученное противоречие показывает, что $a = a'$ и, значит, существует конечный предел $\lim a_n = a$, причем $a \geqslant 0$.

Покажем теперь, что существует предел $\lim b_n = b$ и $a > 0$. Поскольку (8) выполнено равномерно на каждом конечном интервале, то

$$\varphi_n(a_n t) \to \varphi(at),$$

и, значит, в силу (7) существует $\lim\limits_{n\to\infty} e^{itb_n}$ для всех тех t, для которых $\varphi(at) \neq 0$. Пусть $\delta > 0$ таково, что для всех $|t| < \delta$ $\varphi(at) \neq 0$. Тогда для таких t существует $\lim e^{itb_n}$ и, значит, $\overline{\lim} |b_n| < \infty$.

Пусть существуют две подпоследовательности $\{n_i\}$ и $\{n_i'\}$, такие, что $\lim b_{n_i} = b$ и $\lim b_{n_i'} = b'$. Тогда для $|t| < \delta$

$$e^{itb} = e^{itb'}$$

и, следовательно, $b = b'$. Итак, существует конечный предел $b = \lim b_n$ и, согласно (7),

$$\tilde{\varphi}(t) = e^{itb}\varphi(at),$$

что означает, что $\tilde{\xi} \overset{d}{=} a\xi + b$. Поскольку $\tilde{\xi}$ не вырождена, то $a > 0$. Лемма доказана.

4. Приведем теперь (без доказательства) теорему об общем виде характеристической функции устойчивых распределений.

Теорема 4 (представление Леви — Хинчина). *Случайная величина T является устойчивой тогда и только тогда, когда ее характеристическая функция $\varphi(t)$ имеет вид $\varphi(t) = \exp\psi(t)$,*

$$\psi(t) = it\beta - d|t|^\alpha \left(1 + i\theta \frac{t}{|t|} G(t, \alpha)\right), \tag{9}$$

где $0 < \alpha < 2$, $\beta \in R$, $d \geq 0$, $|\theta| \leq 1$, $\frac{t}{|t|} = 0$ при $t = 0$ и

$$G(t, \alpha) = \begin{cases} \operatorname{tg}\frac{\pi}{2}\alpha, & \text{если } \alpha \neq 1, \\ \frac{2}{\pi}\log|t|, & \text{если } \alpha = 1. \end{cases} \tag{10}$$

Отметим, что особо просто устроены характеристические функции симметричных устойчивых распределений:

$$\varphi(t) = e^{-d|t|^\alpha}, \tag{11}$$

где $0 < \alpha \leq 2$, $d \geq 0$.

5. Задачи.

1. Показать, что если $\xi_n \xrightarrow{d} \xi$ и $\xi_n \xrightarrow{d} \eta$, то $\xi \overset{d}{=} \eta$.

2. Показать, что если φ_1 и φ_2 — две безгранично делимые характеристические функции, то $\varphi_1 \cdot \varphi_2$ — также безгранично делимая характеристическая функция.

ГЛАВА IV

ПОСЛЕДОВАТЕЛЬНОСТИ И СУММЫ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 1. Законы «нуля или единицы»

1. Ряд $\sum_{n=1}^{\infty} \frac{1}{n}$ расходится, а ряд $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n}$ сходится. Поставим следующий вопрос. Что можно сказать о сходимости или расходимости ряда $\sum_{n=1}^{\infty} \frac{\xi_n}{n}$, где $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных бернуллиевских случайных величин с $\mathsf{P}(\xi_1 = +1) = \mathsf{P}(\xi_1 = -1) = 1/2$? Иначе говоря, что можно сказать о сходимости ряда с общим членом $\pm 1/n$, где знаки $+$ и $-$ «разбросаны» в случайном порядке в соответствии с рассматриваемой последовательностью $\xi_1, \xi_2, \ldots$?

Обозначим

$$A_1 = \left\{\omega:\ \sum_{n=1}^{\infty} \frac{\xi_n}{n}\ \textit{сходится}\right\} -$$

множество тех элементарных исходов, где ряд $\sum_{n=1}^{\infty} \frac{\xi_n}{n}$ сходится (к конечным значениям) и рассмотрим вероятность $\mathsf{P}(A_1)$ этого множества. Заранее не ясно, какие значения может принимать эта вероятность. Замечательным оказывается, однако, то обстоятельство, что a priori можно утверждать, что эта вероятность может принимать только два значения 0 или 1. Этот результат является следствием так называемого *закона «нуля или единицы»* («0 или 1») *Колмогорова*, формулировка и доказательство которого составляют основное содержание данного параграфа.

2. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство, $\xi_1, \xi_2, \ldots$ — некоторая последовательность случайных величин. Обозначим

$\mathscr{F}_n^\infty = \sigma(\xi_n, \xi_{n+1}, \ldots)$ — σ-алгебру, порожденную случайными величинами $\xi_n, \xi_{n+1}, \ldots$, и пусть

$$\mathscr{X} = \bigcap_{n=1}^{\infty} \mathscr{F}_n^\infty.$$

Поскольку пересечение σ-алгебр есть снова σ-алгебра, то $\mathscr{X}$ — есть σ-алгебра. Эта σ-алгебра будет называться «хвостовой» или «остаточной», в связи с тем, что всякое событие $A \in \mathscr{X}$ не зависит от значений случайных величин $\xi_1, \ldots, \xi_n$ при любом конечном числе n, а определяется лишь «поведением бесконечно далеких значений последовательности $\xi_1, \xi_2, \ldots$».

Поскольку для любого $k \geqslant 1$

$$A_1 \equiv \left\{ \sum_{n=1}^{\infty} \frac{\xi_n}{n} \ \textit{сходится} \right\} = \left\{ \sum_{n=k}^{\infty} \frac{\xi_n}{n} \ \text{сходится} \right\} \in \mathscr{F}_k^\infty,$$

то $A_1 \in \bigcap_k \mathscr{F}_k^\infty \equiv \mathscr{X}$. Точно так же, если $\xi_1, \xi_2, \ldots$ — произвольная последовательность, то

$$A_2 = \left\{ \sum_n \xi_n \ \text{сходится} \right\} \in \mathscr{X}.$$

Следующие события также являются «хвостовыми»:

$$A_3 = \{\xi_n \in I_n \ \textit{для бесконечно многих} \ n\},$$

где $I_n \in \mathscr{B}(R)$, $n \geqslant 1$;

$$A_4 = \left\{ \overline{\lim_n}\, \xi_n < \infty \right\};$$

$$A_5 = \left\{ \overline{\lim_n}\, \frac{\xi_1 + \ldots + \xi_n}{n} < \infty \right\};$$

$$A_6 = \left\{ \overline{\lim_n}\, \frac{\xi_1 + \ldots + \xi_n}{n} < c \right\},$$

$$A_7 = \left\{ \frac{S_n}{n} \ \text{сходится} \right\};$$

$$A_8 = \left\{ \overline{\lim_n}\, \frac{S_n}{\sqrt{2n \log n}} = 1 \right\}.$$

С другой стороны,

$$B_1 = \{\xi_n = 0 \quad \text{для всех} \quad n \geqslant 1\},$$

$$B_2 = \left\{ \lim_n (\xi_1 + \ldots + \xi_n) \ \text{существует и меньше} \ c \right\}$$

являются примерами событий, не принадлежащих $\mathscr{X}$.

Будем теперь предполагать, что рассматриваемые случайные величины являются *независимыми*. При этом допущении из леммы

Бореля — Кантелли следует, что

$$\mathsf{P}(A_3)=0 \Longleftrightarrow \sum \mathsf{P}(\xi_n \in I_n) < \infty,$$
$$\mathsf{P}(A_3)=1 \Longleftrightarrow \sum \mathsf{P}(\xi_n \in I_n) = \infty.$$

Таким образом, вероятность события A_3 может принимать лишь два значения 0 или 1 в зависимости от сходимости или расходимости ряда $\sum \mathsf{P}(\xi_n \in I_n)$. Это утверждение носит название *закона «0 или 1» Бореля.*

Теорема 1 (закон «0 или 1» Колмогорова). *Пусть* ξ_1, ξ_2, ... — *последовательность независимых случайных величин и* $A \in \mathscr{X}$. *Тогда вероятность* $\mathsf{P}(A)$ *может принимать лишь два значения: нуль или единица.*

Доказательство. Идея доказательства состоит в том, чтобы показать, что каждое «хвостовое» событие A не зависит от самого себя и, значит, $\mathsf{P}(A \cap A) = \mathsf{P}(A)\cdot\mathsf{P}(A)$, т. е. $\mathsf{P}(A)=\mathsf{P}^2(A)$, откуда $\mathsf{P}(A)=0$ или 1.

Если $A \in \mathscr{X}$, то $A \in \mathscr{F}_1^\infty = \sigma\{\xi_1, \xi_2, \ldots\} = \sigma\left(\bigcup_n \mathscr{F}_1^n\right)$, где $\mathscr{F}_1^n = \sigma\{\xi_1, \ldots, \xi_n\}$, и можно найти (задача 8 из § 3 гл. II) такие множества $A_n \in \mathscr{F}_1^n$, $n \geqslant 1$, что $\mathsf{P}(A \Delta A_n) \to 0$, $n \to \infty$. Отсюда следует, что

$$\mathsf{P}(A_n) \to \mathsf{P}(A), \quad \mathsf{P}(A_n \cap A) \to \mathsf{P}(A). \tag{1}$$

Но если $A \in \mathscr{X}$, то для каждого $n \geqslant 1$ события A_n и A независимы:

$$\mathsf{P}(A \cap A_n) = \mathsf{P}(A)\,\mathsf{P}(A_n),$$

откуда в силу (1) следует, что $\mathsf{P}(A)=\mathsf{P}^2(A)$, и, значит, $\mathsf{P}(A)=0$ или 1.

Теорема доказана.

Следствие. Пусть η — случайная величина, измеримая относительно «хвостовой» σ-алгебры $\mathscr{X}$, т. е. $\{\eta \in B\} \in \mathscr{X}$, $B \in \mathscr{B}(R)$. Тогда η является вырожденной случайной величиной, т. е. существует константа c такая, что $P(\eta = c) = 1$.

3. Приводимая ниже теорема 2 служит иллюстрацией нетривиального применения закона «нуля или единицы» Колмогорова.

Пусть ξ_1, ξ_2, ... — последовательность независимых бернуллиевских случайных величин с $\mathsf{P}(\xi_n = 1) = p$, $\mathsf{P}(\xi_n = -1) = q$, $p + q = 1$, $n \geqslant 1$, и $S_n = \xi_1 + \ldots + \xi_n$. Интуитивно понятно, что в симметричном случае, $p = 1/2$, «типичные» траектории случайного блуждания S_n, $n \geqslant 1$, бесконечно много раз проходят через нуль, а в случае $p \neq 1/2$ «уходят» в бесконечность. Сформулируем теперь точный результат.

Теорема 2. a) *Если* $p = 1/2$, *то* $\mathsf{P}(S_n = 0$ *б. ч.*$) = 1$.

b) Если $p \neq 1/2$, *то* $\mathsf{P}(S_n = 0$ *б. ч.*$) = 0$.

Доказательство. Прежде всего отметим, что событие $B = (S_n = 0$ *б. ч.*) не является «хвостовым», т. е. $B \notin \mathscr{X} = \bigcap \mathscr{F}_n^\infty$, $\mathscr{F}_n^\infty = \sigma\{\xi_n, \xi_{n+1}, \ldots\}$. Поэтому в принципе не ясно, что вероятность события B принимает лишь значения 0 или 1.

Утверждение b) легко доказывается применением (первой части) леммы Бореля — Кантелли. Действительно, если $B_{2n} = \{S_{2n} = 0\}$, то по формуле Стирлинга

$$\mathsf{P}(B_{2n}) = C_{2n}^n p^n q^n \sim \frac{(4pq)^n}{\sqrt{\pi n}},$$

и, значит, $\sum \mathsf{P}(B_{2n}) < \infty$. Поэтому $\mathsf{P}(S_n = 0 \text{ б. ч.}) = 0$.

Для доказательства утверждения а) достаточно доказать, что событие

$$A = \left\{\overline{\lim} \frac{S_n}{\sqrt{n}} = \infty,\ \underline{\lim} \frac{S_n}{\sqrt{n}} = -\infty\right\}$$

имеет вероятность 1, поскольку $A \subseteq B$.

Пусть

$$A_c = \left\{\overline{\lim} \left|\frac{S_n}{\sqrt{n}}\right| > c\right\}.$$

Тогда $A_c \downarrow A$, $c \to \infty$, при этом как событие A, так и все события A_c являются «хвостовыми». Покажем, что для каждого $c > 0$ $\mathsf{P}(A_c) = 1$. Поскольку $A_c \in \mathscr{X}$, то достаточно лишь установить, что $\mathsf{P}(A_c) > 0$. Но

$$\mathsf{P}\left\{\overline{\lim} \left|\frac{S_n}{\sqrt{n}}\right| > c\right\} \geqslant \overline{\lim}\, \mathsf{P}\left(\left|\frac{S_n}{\sqrt{n}}\right| > c\right) > 0,$$

где последнее неравенство следует из теоремы Муавра — Лапласа.

Итак, для всех $c > 0$ $\mathsf{P}(A_c) = 1$ и, значит, $\mathsf{P}(A) = \lim\limits_{c \to \infty} \mathsf{P}(A_c) = 1$.

Теорема доказана.

4. Отметим еще раз, что событие $B = \{S_n = 0 \text{ б. ч.}\}$ не является «хвостовым». Тем не менее из теоремы 2 следует, что для схемы Бернулли вероятность этого события, как и в случае «хвостовых» событий, принимает лишь два значения 0 или 1. Оказывается, что это обстоятельство неслучайно и является следствием так называемого закона «0 или 1» Хьювитта и Сэвиджа, который обобщает для случая независимых одинаково распределенных случайных величин результат теоремы 1 на класс так называемых «перестановочных» событий (включающий в себя и класс «хвостовых» событий).

Введем необходимые определения. Взаимно однозначное отображение $\pi = (\pi_1, \pi_2, \ldots)$ множества $(1, 2, \ldots)$ в себя назовем конечной перестановкой, если $\pi_n = n$ для всех n, за исключением, быть может, конечного числа.

Если $\xi=(\xi_1, \xi_2, \ldots)$ — последовательность случайных величин, то через $\pi(\xi)$ будем обозначать последовательность $(\xi_{\pi_1}, \xi_{\pi_2}, \ldots)$. Если событие $A=\{\xi \in B\}$, $B \in \mathscr{B}(R^\infty)$, то через $\pi(A)$ обозначим событие $\{\pi(\xi) \in B\}$, $B \in \mathscr{B}(R^\infty)$.

Назовем событие $A=\{\xi \in B\}$, $B \in \mathscr{B}(R^\infty)$, *перестановочным*, если для любой конечной перестановки π событие $\pi(A)$ совпадает с A.

Примером перестановочного события является событие $A = \{S_n=0 \text{ б. ч.}\}$, где $S_n=\xi_1+\ldots+\xi_n$. Более того, можно показать (задача 4), что каждое событие из «хвостовой» σ-алгебры $\mathscr{X}(S)=\bigcap \mathscr{F}_n^\infty(S)$, $\mathscr{F}_n^\infty(S)=\sigma\{\omega\colon S_n, S_{n+1}, \ldots\}$, порожденной величинами $S_1=\xi_1$, $S_2=\xi_1+\xi_2$, ..., является перестановочным.

Теорема 3 (закон «0 или 1» Хьюитта и Сэвиджа). *Пусть ξ_1, ξ_2, ... — последовательность независимых одинаково распределенных случайных величин и $A=\{\omega\colon (\xi_1, \xi_2, \ldots) \in B\}$ — перестановочное событие. Тогда* $\mathsf{P}(A)=0$ *или* 1.

Доказательство. Пусть $A=\{\xi \in B\}$ — перестановочное событие. Выберем множества $B_n \in \mathscr{B}(R^n)$ такими, что для $A_n = \{\omega\colon (\xi_1, \ldots, \xi_n) \in B_n\}$

$$\mathsf{P}(A \,\Delta A_n) \to 0, \qquad n \in \infty. \tag{2}$$

Поскольку случайные величины $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены, то распределения вероятностей $P_\xi(B) \equiv \mathsf{P}(\xi \in B)$ и $P_{\pi_n(\xi)}(B) \equiv \mathsf{P}(\pi_n(\xi) \in B)$ совпадают. Значит,

$$\mathsf{P}(A \,\Delta A_n) = P_\xi(B \,\Delta B_n) = P_{\pi_n(\xi)}(B \,\Delta B_n). \tag{3}$$

Раз событие A является перестановочным, то

$$A \equiv \{\xi \in B\} = \pi_n(A) \equiv \{\pi_n(\xi) \in B\}.$$

Поэтому

$$\begin{aligned} P_{\pi_n(\xi)}(B \,\Delta B_n) &= \mathsf{P}\{\pi_n(\xi) \in B) \,\Delta\, (\pi_n(\xi) \in B_n)\} = \\ &= \mathsf{P}\{(\xi \in B) \,\Delta\, (\pi_n(\xi) \in B_n)\} = \mathsf{P}\{A \,\Delta \pi_n(A_n)\}. \end{aligned} \tag{4}$$

Итак, из (3) и (4)

$$\mathsf{P}(A \,\Delta A_n) = \mathsf{P}(A \,\Delta \pi_n(A_n)). \tag{5}$$

В силу (2) отсюда следует, что

$$\mathsf{P}(A \Delta (A_n \cap \pi_n(A_n))) \to 0, \qquad n \to \infty. \tag{6}$$

Поэтому из (2), (5) и (6) заключаем, что

$$\begin{gathered} \mathsf{P}(A_n) \to \mathsf{P}(A), \quad \mathsf{P}(\pi_n(A_n)) \to \mathsf{P}(A), \\ \mathsf{P}(A_n \cap \pi_n(A_n)) \to \mathsf{P}(A). \end{gathered} \tag{7}$$

Далее, в силу независимости случайных величин $\xi_1, \xi_2, \ldots$

$$\begin{aligned}\mathsf{P}(A_n \cap \pi_n(A_n)) &= \mathsf{P}\{(\xi_1, \ldots, \xi_n) \in B_n,\ (\xi_{n+1}, \ldots, \xi_{2n}) \in B_n\} = \\ &= \mathsf{P}\{(\xi_1, \ldots, \xi_n) \in B_n\} \cdot \mathsf{P}\{(\xi_{n+1}, \ldots, \xi_{2n}) \in B_n\} = \\ &= \mathsf{P}(A_n)\,\mathsf{P}(\pi_n(A_n)),\end{aligned}$$

откуда в силу (7)

$$\mathsf{P}(A) = \mathsf{P}^2(A)$$

и, значит, $\mathsf{P}(A) = 0$ или 1.

Теорема доказана.

5. Задачи.

1. Доказать следствие к теореме 1.

2. Показать, что если (ξ_n) — последовательность независимых случайных величин, то случайные величины $\overline{\lim}\, \xi_n$ и $\underline{\lim}\, \xi_n$ являются вырожденными.

3. Пусть (ξ_n) — последовательность независимых случайных величин, $S_n = \xi_1 + \ldots + \xi_n$, и константы b_n таковы, что $0 < b_n \uparrow \infty$. Показать, что случайные величины $\overline{\lim}\, \frac{S_n}{b_n}$ и $\underline{\lim}\, \frac{S_n}{b_n}$ являются вырожденными.

4. Пусть $S_n = \xi_1 + \ldots + \xi_n$, $n \geqslant 1$ и $\mathscr{X}(S) = \bigcap \mathscr{F}_n^\infty(S)$, $\mathscr{F}_n^\infty(S) = \sigma\{\omega\colon S_n, S_{n+1}, \ldots\}$. Показать, что каждое событие из $\mathscr{X}(S)$ является перестановочным.

§ 2. Сходимость рядов

1. Будем предполагать, что $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, $S_n = \xi_1 + \ldots + \xi_n$ и A — множество тех элементарных исходов ω, где ряд $\sum \xi_n(\omega)$ сходится к конечному пределу. Из закона «0 или 1» Колмогорова следует, что вероятность $\mathsf{P}(A) = 0$ или 1, т. е. с вероятностью единица ряд $\sum \xi_n$ сходится или расходится. Цель настоящего параграфа — дать критерии, позволяющие определять, сходится или расходится ряд из независимых случайных величин.

Теорема 1 (Колмогоров и Хинчин). a) *Пусть* $\mathsf{M}\xi_n = 0$, $n \geqslant 1$. *Тогда, если*

$$\sum \mathsf{M}\xi_n^2 < \infty, \tag{1}$$

то ряд $\sum \xi_n$ *сходится с вероятностью единица.*

b) *Если к тому же случайные величины* ξ_n, $n \geqslant 1$, *равномерно ограничены* ($\mathsf{P}(|\xi_n| \leqslant c) = 1$, $c < \infty$), *то верно и обратное: из сходимости с вероятностью единица ряда* $\sum \xi_n$ *следует условие* (1).

Доказательство этой теоремы существенно опирается на

Неравенства Колмогорова. a) *Пусть* $\xi_1, \xi_2, \ldots, \xi_n$ — *независимые случайные величины с* $\mathsf{M}\xi_i = 0$, $\mathsf{M}\xi_i^2 < \infty$, $i \leqslant n$. *Тогда* *для всякого* $\varepsilon>0$

$$\mathsf{P}\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon\right\}\leqslant\frac{\mathsf{M}S_n^2}{\varepsilon^2}. \tag{2}$$

b) *Если к тому же* $\mathsf{P}(|\xi_i|\leqslant c)=1$, $i\leqslant n$, *то*

$$\mathsf{P}\left\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon\right\}\geqslant 1-\frac{(c+\varepsilon)^2}{\mathsf{M}S_n^2}. \tag{3}$$

Доказательство. а) Обозначим

$$A=\{\max|S_k|\geqslant\varepsilon\},$$
$$A_k=\{|S_i|<\varepsilon,\ i=1,\ \ldots,\ k-1,\ |S_k|\geqslant\varepsilon\},\qquad 1\leqslant k\leqslant n.$$

Тогда $A=\sum A_k$ и

$$\mathsf{M}S_n^2\geqslant\mathsf{M}S_n^2I_A=\sum\mathsf{M}S_n^2I_{A_k}.$$

Но

$$\mathsf{M}S_n^2I_{A_k}=\mathsf{M}(S_k+(\xi_{k+1}+\ldots+\xi_n))^2I_{A_k}=$$
$$=\mathsf{M}S_k^2I_{A_k}+2\mathsf{M}S_k(\xi_{k+1}+\ldots+\xi_n)I_{A_k}+\mathsf{M}(\xi_{k+1}+\ldots+\xi_n)^2I_{A_k}\geqslant$$
$$\geqslant\mathsf{M}S_k^2I_{A_k},$$

поскольку $\mathsf{M}S_k(\xi_{k+1}+\ldots+\xi_n)I_{A_k}=\mathsf{M}S_kI_{A_k}\cdot\mathsf{M}(\xi_{k+1}+\ldots+\xi_n)=0$ в силу предположенной независимости и условий $\mathsf{M}\xi_i=0$, $i\leqslant n$. Поэтому

$$\mathsf{M}S_n^2\geqslant\sum\mathsf{M}S_k^2I_{A_k}\geqslant\varepsilon^2\sum\mathsf{P}(A_k)=\varepsilon^2\mathsf{P}(A),$$

что и доказывает первое неравенство.

Для доказательства (3) заметим, что

$$\mathsf{M}S_n^2I_A=\mathsf{M}S_n^2-\mathsf{M}S_n^2I_{\bar A}\geqslant\mathsf{M}S_n^2-\varepsilon^2\mathsf{P}(\bar A)=\mathsf{M}S_n^2-\varepsilon^2+\varepsilon^2\mathsf{P}(A). \tag{4}$$

С другой стороны, на множестве A_k

$$|S_{k-1}|\leqslant\varepsilon,\qquad |S_k|\leqslant|S_{k-1}|+|\xi_k|\leqslant\varepsilon+c$$

и, значит,

$$\mathsf{M}S_n^2I_A=\sum_k\mathsf{M}S_k^2I_{A_k}+\sum_k\mathsf{M}\left(I_{A_k}(S_n-S_k)^2\right)\leqslant$$
$$\leqslant(\varepsilon+c)^2\sum_k\mathsf{P}(A_k)+\sum_{k=1}^n\mathsf{P}(A_k)\sum_{j=k+1}^n\mathsf{M}\xi_j^2\leqslant$$
$$\leqslant\mathsf{P}(A)\left[(\varepsilon+c)^2+\sum_{j=1}^n\mathsf{M}\xi_j^2\right]=\mathsf{P}(A)\left[(\varepsilon+c)^2+\mathsf{M}S_n^2\right]. \tag{5}$$

Из (4) и (5) находим, что

$$\mathsf{P}(A)\geqslant\frac{\mathsf{M}S_n^2-\varepsilon^2}{(\varepsilon+c)^2+\mathsf{M}S_n^2-\varepsilon^2}=1-\frac{(\varepsilon+c)^2}{(\varepsilon+c)^2+\mathsf{M}S_n^2-\varepsilon^2}\geqslant1-\frac{(\varepsilon+c)^2}{\mathsf{M}S_n^2}.$$

Неравенство (3) доказано.

Доказательство теоремы 1. а) Согласно теореме 4 из § 10 гл. II последовательность (S_n), $n \geqslant 1$, сходится с вероятностью единица тогда и только тогда, когда эта последовательность фундаментальна с вероятностью единица. По теореме 1 из § 10 гл. II последовательность (S_n), $n \geqslant 1$, фундаментальна (P-п. н.) в том и только том случае, когда

$$\mathsf{P}\left\{\sup_{k \geqslant 1} |S_{n+k} - S_k| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{6}$$

В силу (2)

$$\mathsf{P}\left\{\sup_{k \geqslant 1} |S_{n+k} - S_n| \geqslant \varepsilon\right\} = \lim_{N \to \infty} \mathsf{P}\left\{\max_{1 \leqslant k \leqslant N} |S_{n+k} - S_n| \geqslant \varepsilon\right\} \leqslant$$
$$\leqslant \lim_{N \to \infty} \frac{\sum_{k=n}^{n+N} \mathsf{M}\xi_k^2}{\varepsilon^2} = \frac{\sum_{k=n}^{\infty} \mathsf{M}\xi_k^2}{\varepsilon^2}.$$

Поэтому, если $\sum_{k=1}^{\infty} \mathsf{M}\xi_k^2 < \infty$, то выполнено условие (6) и, следовательно, ряд $\sum \xi_k$ сходится с вероятностью единица.

b) Пусть ряд $\sum \xi_k$ сходится. Тогда в силу (6) для достаточно больших n

$$\mathsf{P}\left\{\sup_{k \geqslant 1} |S_{n+k} - S_n| \geqslant \varepsilon\right\} < \frac{1}{2}. \tag{7}$$

В силу (3)

$$\mathsf{P}\left\{\sup_{k \geqslant 1} |S_{n+k} - S_n| \geqslant \varepsilon\right\} \geqslant 1 - \frac{(c+\varepsilon)^2}{\sum_{k=n}^{\infty} \mathsf{M}\xi_k^2}.$$

Поэтому, если допустить, что $\sum_{k=1}^{\infty} \mathsf{M}\xi_k^2 = \infty$, то получим

$$\mathsf{P}\left\{\sup_{k \geqslant 1} |S_{n+k} - S_n| \geqslant \varepsilon\right\} = 1,$$

что противоречит неравенству (7).

Теорема доказана.

Пример. Если $\xi_1, \xi_2, \ldots$ — последовательность независимых бернуллиевских случайных величин с $\mathsf{P}(\xi_n = +1) = \mathsf{P}(\xi_n = -1) = 1/2$, то ряд $\sum \xi_n a_n$, где $|a_n| \leqslant c$, сходится с вероятностью единица тогда и только тогда, когда $\sum a_n^2 < \infty$.

2. Теорема 2 (теорема о «двух рядах»). *Для сходимости с вероятностью единица ряда $\sum \xi_n$ из независимых случайных величин достаточно, чтобы одновременно сходились два ряда $\sum \mathsf{M}\xi_n$ и $\sum \mathsf{D}\xi_n$. Если к тому же $\mathsf{P}(|\xi_n| \leqslant c) = 1$, $n \geqslant 1$, то это условие является и необходимым.*

Доказательство. Если $\sum \mathsf{D}\xi_n < \infty$, то по теореме 1 ряд $\sum (\xi_n - \mathsf{M}\xi_n)$ сходится (Р-п. н.). Но по предположению ряд $\sum \mathsf{M}\xi_n$ сходится, поэтому сходится (Р-п. н.) и ряд $\sum \xi_n$.

Для доказательства необходимости воспользуемся следующим приемом «симметризации». Наряду с последовательностью $\xi_1, \xi_2, \ldots$ рассмотрим не зависящую от нее последовательность независимых случайных величин $\tilde{\xi}_1, \tilde{\xi}_2, \ldots$ таких, что $\tilde{\xi}_n$ имеет то же распределение, что и ξ_n, $n \geqslant 1$. (Когда исходное пространство элементарных событий предполагается достаточно «богатым», существование такой последовательности следует из теоремы 1 § 9 гл. II. В свою очередь можно показать, что это предположение не ограничивает общности.)

Тогда, если сходится (Р-п. н.) ряд $\sum \xi_n$, то сходится и ряд $\sum \tilde{\xi}_n$, а значит, и ряд $\sum (\xi_n - \tilde{\xi}_n)$. Но $\mathsf{M}(\xi_n - \tilde{\xi}_n) = 0$ и $\mathsf{P}(|\xi_n - \tilde{\xi}_n| \leqslant 2c) = 1$. Поэтому по теореме 1 $\sum \mathsf{D}(\xi_n - \tilde{\xi}_n) < \infty$. Далее

$$\sum \mathsf{D}\xi_n = \frac{1}{2} \sum \mathsf{D}(\xi_n - \tilde{\xi}_n) < \infty.$$

Поэтому по теореме 1 с вероятностью единица сходится ряд $\sum (\xi_n - \mathsf{M}\xi_n)$, а значит, сходится и ряд $\sum \mathsf{M}\xi_n$.

Итак, из сходимости (Р-п. н.) ряда $\sum \xi_n$ (в предположении $\mathsf{P}(|\xi_n| \leqslant c) = 1$, $n \geqslant 1$) вытекает, что оба ряда $\sum \mathsf{M}\xi_n$ и $\sum \mathsf{D}\xi_n$ сходятся.

Теорема доказана.

3. Следующая теорема дает необходимое и достаточное условие сходимости ряда $\sum \xi_n$ без предположений об ограниченности случайных величин.

Пусть c — некоторая константа и

$$\xi^c = \begin{cases} \xi, & |\xi| \leqslant c, \\ 0, & |\xi| > c. \end{cases}$$

Теорема 3 (теорема Колмогорова о «трех рядах»). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин. Для сходимости с вероятностью единица ряда $\sum \xi$ необходимо, чтобы для любого $c > 0$ сходились ряды*

$$\sum \mathsf{M}\xi_n^c, \quad \sum \mathsf{D}\xi_n^c, \quad \sum \mathsf{P}(|\xi_n| \geqslant c)$$

и достаточно, чтобы эти ряды сходились при некотором $c > 0$.

Доказательство. Достаточность. По теореме о «двух рядах» ряд $\sum \xi_n^c$ сходится с вероятностью единица. Но если $\sum \mathsf{P}(|\xi_n| \geqslant c) < \infty$, то по лемме Бореля — Кантелли с вероятностью единица $\sum I(|\xi_n| \geqslant c) < \infty$, а значит, $\xi_n = \xi_n^c$ для всех n, за исключением, быть может, конечного числа. Поэтому ряд $\sum \xi_n$ также сходится (Р-п. н.).

Необходимость. Если ряд $\sum \xi_n$ сходится (Р-п. н.), то $\xi_n \to 0$ (Р-п. н.) и, значит, для всякого $c > 0$ может произойти (Р-п. н.) не более конечного числа событий $\{|\xi_n| \geqslant c\}$. Поэтому $\sum I(|\xi_n| \geqslant c) < \infty$ (Р-п. н.) и по второй части леммы Бореля — Кантелли $\sum \mathsf{P}(|\xi_n| > c) < \infty$. Далее, из сходимости ряда $\sum \xi_n$ следует и сходимость ряда $\sum \xi_n^c$. Поэтому по теореме о «двух рядах» каждый из рядов $\sum \mathsf{M}\xi_n^c$ и $\sum \mathsf{D}\xi_n^c$ сходится.

Теорема доказана.

Следствие. Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины с $\mathsf{M}\xi_n = 0$. Тогда, если

$$\sum \mathsf{M} \frac{\xi_n^2}{1+|\xi_n|} < \infty,$$

то ряд $\sum \xi_n$ сходится с вероятностью единица.

Для доказательства заметим, что

$$\sum \mathsf{M} \frac{\xi_n^2}{1+|\xi_n|} < \infty \Leftrightarrow \sum \mathsf{M}[\xi_n^2 I(|\xi_n| \leqslant 1) + |\xi_n| I(|\xi_n| > 1)] < \infty.$$

Поэтому, если $\xi_n^1 = \xi_n I(|\xi_n| \leqslant 1)$, то

$$\sum \mathsf{M}(\xi_n^1)^2 < \infty.$$

Поскольку $\mathsf{M}\xi_n = 0$, то

$$\sum |\mathsf{M}\xi_n^1| = \sum |\mathsf{M}\xi_n I(|\xi_n| \leqslant 1)| = \sum |\mathsf{M}\xi_n I(|\xi_n| > 1)| \leqslant$$
$$\leqslant \sum \mathsf{M}|\xi_n| I(|\xi_n| > 1) < \infty.$$

Значит, каждый из рядов $\sum \mathsf{M}\xi_n^1$ и $\sum \mathsf{D}\xi_n^1$ сходится. Далее, по неравенству Чебышева

$$\mathsf{P}\{|\xi_n| > 1\} = \mathsf{P}\{|\xi_n| I(|\xi_n| > 1) > 1\} \leqslant \mathsf{M}(|\xi_n| I(|\xi_n| > 1)).$$

Поэтому $\sum \mathsf{P}(|\xi_n| > 1) < \infty$. Тем самым сходимость ряда $\sum \xi_n$ следует из теоремы о «трех рядах».

4. Задачи.

1. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, $S_n = \xi_1 + \ldots + \xi_n$. Используя теорему о «трех рядах», показать, что: a) если $\sum \xi_n^2 < \infty$ (Р-п. н.), то ряд $\sum \xi_n$ сходится с вероятностью единица в том и только том случае, когда сходится ряд $\sum \mathsf{M}\xi_i I(|\xi_i| \leqslant 1)$; b) если ряд $\sum \xi_n$ сходится (Р-п. н.), то ряд $\sum \xi_n^2 < \infty$ (Р-п. н.) в том и только том случае, когда

$$\sum (\mathsf{M}|\xi_n| I(|\xi_n| \leqslant 1))^2 < \infty.$$

2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин. Показать, что $\sum \xi_n^2 < \infty$ (Р-п. н.) тогда и только

тогда, когда

$$\sum \mathsf{M} \frac{\xi_n^2}{1+\xi_n^2} < \infty.$$

3. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин. Показать, что ряд $\sum \xi_n$ сходится (P-п. н.) тогда и только тогда, когда он сходится по вероятности.

§ 3. Усиленный закон больших чисел

1. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечными вторыми моментами, $S_n = \xi_1 + \ldots + \xi_n$. Согласно задаче 2 из § 3 гл. III, если дисперсии $\mathsf{D}\xi_i$ равномерно ограничены, то имеет место закон больших чисел:

$$\frac{S_n - \mathsf{M}S_n}{n} \xrightarrow{\mathsf{P}} 0, \quad n \to \infty. \tag{1}$$

Усиленным законом больших чисел называется утверждение, в котором сходимость по вероятности в (1) заменяются *сходимостью с вероятностью единица.*

Один из первых результатов в этом направлении дается следующей теоремой.

Теорема 1 (Кантелли). *Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины с конечным четвертым моментом и такие, что для некоторой константы C*

$$\mathsf{M}|\xi_n - \mathsf{M}\xi_n|^4 \leq C, \quad n \geq 1.$$

Тогда при $n \to \infty$

$$\frac{S_n - \mathsf{M}S_n}{n} \to 0 \quad (\mathsf{P}\text{-п. н.}). \tag{2}$$

Доказательство. Не ограничивая общности, будем считать $\mathsf{M}\xi_n = 0$, $n \geq 1$. По следствию к теореме 1 из § 10 гл. II для сходимости $\frac{S_n}{n} \to 0$ (P-п. н.) достаточно, чтобы для любого $\varepsilon > 0$

$$\sum \mathsf{P}\left\{\left|\frac{S_n}{n}\right| \geq \varepsilon\right\} < \infty.$$

В свою очередь, в силу неравенства Чебышева, для этого достаточно выполнения условия

$$\sum \mathsf{M}\left|\frac{S_n}{n}\right|^4 < \infty.$$

Покажем, что при сделанных предположениях это условие действительно выполнено.

Имеем

$$S_n^4=(\xi_1+\ldots+\xi_n)^4=\sum_{i=1}^{n}\xi_i^4-\sum_{\substack{i,j\\ i<j}}\frac{4!}{2!\,2!}\xi_i^2\xi_j^2+$$

$$+\sum_{\substack{i\neq j\\ i\neq k\\ j<k}}\frac{4!}{2!\,1!\,1!}\xi_i^2\xi_j\xi_k+\sum_{i<j<k<l}4!\,\xi_i\xi_j\xi_k\xi_l+\sum_{i\neq j}\frac{4!}{3!\,1!}\xi_i^3\xi_j.$$

Тогда, учитывая, что $\mathsf{M}\xi_k=0$, $k\leqslant n$, отсюда находим

$$\mathsf{M}S_n^4=\sum_{i=1}^{n}\mathsf{M}\xi_i^4+6\sum_{i,j=1}^{n}\mathsf{M}\xi_i^2\mathsf{M}\xi_j^2\leqslant nC+6\sum_{\substack{i,j=1\\ i<j}}\sqrt{\mathsf{M}\xi_i^4\cdot\mathsf{M}\xi_j^4}\leqslant$$

$$\leqslant nC+\frac{6n(n-1)}{2}C=(3n^2-2n)C<3n^2C.$$

Следовательно,

$$\sum\mathsf{M}\left(\frac{S_n}{n}\right)^4\leqslant 3C\sum\frac{1}{n^2}<\infty.$$

Теорема доказана.

2. Привлечение более тонких методов позволяет существенно ослабить предположения, сделанные в теореме 1, для справедливости усиленного закона больших чисел.

Теорема 2 (Колмогоров). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечными вторыми моментами, положительные числа b_n таковы, что $b_n\uparrow\infty$ и*

$$\sum\frac{\mathsf{D}\xi_n}{b_n^2}<\infty. \tag{3}$$

Тогда

$$\frac{S_n-\mathsf{M}S_n}{b_n}\to 0\quad(\mathsf{P}\text{-п. н.}). \tag{4}$$

В частности, если

$$\sum\frac{\mathsf{D}\xi_n}{n^2}<\infty, \tag{5}$$

то

$$\frac{S_n-\mathsf{M}S_n}{n}\to 0\quad(\mathsf{P}\text{-п. н.}). \tag{6}$$

Для доказательства этой теоремы, а также нижеследующей теоремы 3 нам понадобятся следующие два вспомогательных утверждения.

Лемма 1 (Теплиц). *Пусть $\{a_n\}$ — последовательность неотрицательных чисел, $b_n=\sum_{i=1}^{n}a_i$, $b_n>0$ для всех $n\geqslant 1$ и $b_n\uparrow\infty$,*

$n \to \infty$. *Пусть также* $\{x_n\}$ — *последовательность чисел, сходящаяся к некоторому числу* x. *Тогда*

$$\frac{1}{b_n} \sum_{j=1}^{n} a_j x_j \to x. \tag{7}$$

В частности, если $a_n = 1$, *то*

$$\frac{x_1 + \ldots + x_n}{n} \to x. \tag{8}$$

Доказательство. Пусть $\varepsilon > 0$ и $n_0 = n_0(\varepsilon)$ таково, что для всех $n \geqslant n_0$ $|x_n - x| \leqslant \varepsilon/2$. Выберем $n_1 > n_0$ так, что

$$\frac{1}{b_{n_1}} \sum_{j=1}^{n_0} |x_j - x| < \varepsilon/2.$$

Тогда для $n > n_1$

$$\left| \frac{1}{b_n} \sum_{j=1}^{n} a_j x_j - x \right| \leqslant \frac{1}{b_n} \sum_{j=1}^{n} a_j |x_j - x| =$$

$$= \frac{1}{b_n} \sum_{j=1}^{n_0} a_j |x_j - x| + \frac{1}{b_n} \sum_{j=n_0+1}^{n} a_j |x_j - x| \leqslant$$

$$\leqslant \frac{1}{b_{n_1}} \sum_{j=1}^{n_0} a_j |x_j - x| + \frac{1}{b_n} \sum_{j=n_0+1}^{n} a_j |x_j - x| \leqslant \frac{\varepsilon}{2} + \frac{n-n_0}{n} \frac{\varepsilon}{2} \leqslant \varepsilon.$$

Лемма доказана.

Лемма 2 (Кронекер). *Пусть* $\{b_n\}$ — *последовательность положительных возрастающих чисел,* $b_n \uparrow \infty$, $n \to \infty$, *и* $\{x_n\}$ — *последовательность чисел таких, что ряд* $\sum x_n$ *сходится. Тогда*

$$\frac{1}{b_n} \sum_{j=1}^{n} b_j x_j \to 0, \qquad n \to \infty. \tag{9}$$

В частности, если $b_n = n$, $x_n = \frac{y_n}{n}$ *и ряд* $\sum \frac{y_n}{n}$ *сходится, то*

$$\frac{y_1 + \ldots + y_n}{n} \to 0, \qquad n \to \infty. \tag{10}$$

Доказательство. Пусть $b_0 = 0$, $S_0 = 0$, $S_n = \sum_{j=1}^{n} x_j$. Тогда («суммирование по частям»)

$$\sum_{j=1}^{n} b_j x_j = \sum_{j=1}^{n} b_j (S_j - S_{j-1}) = b_n S_n - b_0 S_0 - \sum_{j=1}^{n} S_{j-1} (b_j - b_{j-1})$$

и, значит,

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j = S_n - \frac{1}{b_n}\sum_{j=1}^{n} S_{j-1}a_j \to 0,$$

поскольку, если $S_n \to x$, то по лемме Теплица

$$\frac{1}{b_n}\sum_{j=1}^{n} S_{j-1}a_j \to x.$$

Лемма доказана.

Доказательство теоремы 1. Поскольку

$$\frac{S_n - \mathsf{M}S_n}{b_n} = \frac{1}{b_n}\sum_{k=1}^{n} b_k\left(\frac{\xi_k - \mathsf{M}\xi_k}{b_k}\right),$$

то в силу леммы Кронекера для выполнения (4) достаточно, чтобы (P-п. н.) сходился ряд $\sum \frac{\xi_k - \mathsf{M}\xi_k}{b_k}$. Но этот ряд действительно сходится в силу условия (3) и теоремы 1 из § 2.

Теорема доказана.

Пример 1. Пусть $\xi_1, \xi_2, \ldots$ — последовательность бернуллиевских независимых случайных величин с $\mathsf{P}(\xi_n = 1) = \mathsf{P}(\xi_n = -1) = 1/2$. Тогда, поскольку $\sum \frac{1}{n \log^2 n} < \infty$, то

$$\frac{S_n}{\sqrt{n}\log n} \to 0 \quad (\mathsf{P}\text{-п. н.}). \tag{11}$$

3. В том случае, когда величины $\xi_1, \xi_2, \ldots$ не только независимы, но и к тому же одинаково распределены, для справедливости усиленного закона больших чисел нет надобности требовать (как в теореме 2) существования второго момента, а достаточно лишь существования первого абсолютного момента.

Теорема 3 (Колмогоров). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{M}|\xi_1| < \infty$. Тогда*

$$\frac{S_n}{n} \to m \quad (\mathsf{P}\text{-п. н.}), \tag{12}$$

где $m = \mathsf{M}\xi_1$.

Для доказательства нам понадобится следующая

Лемма 3. *Пусть ξ — неотрицательная случайная величина. Тогда*

$$\sum_{n=1}^{\infty} \mathsf{P}(\xi \geqslant n) \leqslant \mathsf{M}\xi \leqslant 1 + \sum_{n=1}^{\infty} \mathsf{P}(\xi \geqslant n). \tag{13}$$

Доказательство следует из следующей цепочки неравенств:

$$\sum_{n=1}^{\infty} \mathsf{P}(\xi \geqslant n) = \sum_{n=1}^{\infty} \sum_{k \geqslant n} \mathsf{P}(k \leqslant \xi < k+1) =$$

$$= \sum_{k=1}^{\infty} k\mathsf{P}(k \leqslant \xi < k+1) = \sum_{k=0}^{\infty} \mathsf{M}[kI(k \leqslant \xi < k+1)] \leqslant$$

$$\leqslant \sum_{k=0}^{\infty} \mathsf{M}[\xi I(k \leqslant \xi < k+1)] =$$

$$= \mathsf{M}\xi \leqslant \sum_{k=0}^{\infty} \mathsf{M}[(k+1)I(k \leqslant \xi < k+1)] =$$

$$= \sum_{k=0}^{\infty} (k+1)\mathsf{P}(k \leqslant \xi < k+1) =$$

$$= \sum_{n=1}^{\infty} \mathsf{P}(\xi \geqslant n) + \sum_{k=0}^{\infty} \mathsf{P}(k \leqslant \xi < k+1) = \sum_{n=1}^{\infty} \mathsf{P}(\xi \geqslant n) + 1.$$

Доказательство теоремы 3. В силу леммы 3 и леммы Бореля — Кантелли

$$\mathsf{M}|\xi_1| < \infty \Leftrightarrow \sum \mathsf{P}\{|\xi_1| \geqslant n\} < \infty \Leftrightarrow$$
$$\Leftrightarrow \sum \mathsf{P}\{|\xi_n| \geqslant n\} < \infty \Leftrightarrow \mathsf{P}\{|\xi_n| \geqslant n \text{ б. ч.}\} = 0.$$

Поэтому с вероятностью единица для всех n, за исключением лишь конечного числа, $|\xi_n| < n$.

Обозначим

$$\tilde{\xi}_n = \begin{cases} \xi_n, & |\xi_n| < n, \\ 0, & |\xi_n| \geqslant n, \end{cases}$$

и будем считать, что $\mathsf{M}\xi_n = 0$, $n \geqslant 1$. Тогда $\frac{\xi_1 + \ldots + \xi_n}{n} \to 0$ (P-п. н.), если и только если $\frac{\tilde{\xi}_1 + \ldots + \tilde{\xi}_n}{n} \to 0$ (P-п. н.). Заметим, что, вообще говоря, $\mathsf{M}\tilde{\xi}_n \neq 0$, но

$$\mathsf{M}\tilde{\xi}_n = \mathsf{M}\xi_n I(|\xi_n| < n) = \mathsf{M}\xi_1 I(|\xi_1| < n) \to \mathsf{M}\xi_1 = 0.$$

Поэтому по лемме Теплица

$$\frac{1}{r} \sum_{k=1}^{n} \mathsf{M}\tilde{\xi}_k \to 0, \qquad n \to \infty,$$

и, следовательно, $\frac{\xi_1+\ldots+\xi_n}{n}\to 0$ (Р-п. н.) в том и только том случае, когда (Р-п. н.)

$$\frac{(\tilde{\xi}_1-\mathsf{M}\tilde{\xi}_1)+\ldots+(\tilde{\xi}_n-\mathsf{M}\tilde{\xi}_n)}{n}\to 0,\qquad n\to\infty. \tag{14}$$

Обозначим $\bar{\xi}_n=\tilde{\xi}_n-\mathsf{M}\tilde{\xi}_n$. В силу леммы Кронекера для выполнения (14) достаточно лишь установить, что ряд $\sum\frac{\bar{\xi}_n}{n}$ сходится (Р-п. н.). В свою очередь, согласно теореме 1 из § 2, для этого достаточно показать, что предположение $\mathsf{M}|\xi_1|<\infty$ обеспечивает сходимость ряда $\sum\frac{\mathsf{D}\bar{\xi}_n}{n^2}$.

Имеем

$$\sum\frac{\mathsf{D}\bar{\xi}_n}{n^2}\leqslant\sum_{n=1}^{\infty}\frac{\mathsf{M}\tilde{\xi}_n^2}{n^2}=\sum_{n=1}^{\infty}\frac{1}{n^2}\mathsf{M}[\xi_n I(|\xi_n|<n)]^2=$$

$$=\sum_{n=1}^{\infty}\frac{1}{n^2}\mathsf{M}[\xi_1^2 I(|\xi_1|<n)]=\sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=1}^{n}\mathsf{M}[\xi_1^2 I(k-1\leqslant|\xi_1|<k)]=$$

$$=\sum_{k=1}^{\infty}\mathsf{M}[\xi_1^2 I(k-1\leqslant|\xi_1|<k)]\cdot\sum_{n=k}^{\infty}\frac{1}{n^2}\leqslant$$

$$\leqslant 2\sum_{k=1}^{\infty}\frac{1}{k}\mathsf{M}[\xi_1^2 I(k-1\leqslant|\xi_1|<k)]\leqslant$$

$$\leqslant 2\sum_{k=1}^{\infty}\mathsf{M}[|\xi_1|\,I(k-1\leqslant|\xi_1|<k)]=2\mathsf{M}|\xi_1|<\infty.$$

Теорема доказана.

З а м е ч а н и е 1. Утверждение теоремы допускает обращение в следующем смысле. Пусть $\xi_1,\xi_2,\ldots$ — последовательность независимых одинаково распределенных случайных величин, для которых с вероятностью единица

$$\frac{\xi_1+\ldots+\xi_n}{n}\to C,$$

где C — некоторая (конечная) константа. Тогда $\mathsf{M}|\xi_1|<\infty$ и $C=\mathsf{M}\xi_1$.

В самом деле, если $\frac{S_n}{n}\to C$ (Р-п. н.), то

$$\frac{\xi_n}{n}=\frac{S_n}{n}-\left(\frac{n-1}{n}\right)\frac{S_{n-1}}{n-1}\to 0\ \text{(Р-п. н)}$$

и, значит, $\mathsf{P}(|\xi_n|>n\ \text{б. ч.})=0$. По лемме Бореля — Кантелли

$$\sum\mathsf{P}(|\xi_1|>n)<\infty$$

и в силу леммы 3 $\mathsf{M}|\xi_1|<\infty$. Тогда из доказанной теоремы следует, что $C=\mathsf{M}\xi_1$.

Таким образом, для независимых одинаково распределенных случайных величин условие $\mathsf{M}|\xi_1|<\infty$ является необходимым и достаточным для сходимости (с вероятностью единица) отношений S_n/n к конечному пределу.

Замечание 2. Если математическое ожидание $m=\mathsf{M}\xi_1$ существует, но не обязательно конечно, то утверждение (11) теоремы также остается в силе.

В самом деле, пусть, например, $\mathsf{M}\xi_1^-<\infty$ и $\mathsf{M}\xi_1^+=\infty$. Положим для $C>0$

$$S_n^C=\sum_{i=1}^n \xi_i I(\xi_i\leqslant C).$$

Тогда (P-п. н.)

$$\varliminf_n \frac{S_n}{n}\geqslant \varliminf_n \frac{S_n^C}{n}=\mathsf{M}\xi_1 I(\xi_1\leqslant C).$$

Но при $C\to\infty$

$$\mathsf{M}\xi_1 I(\xi_1\leqslant C)\to\mathsf{M}\xi_1=\infty,$$

поэтому $\frac{S_n}{n}\to+\infty$ (P-п. н.).

4. Остановимся на некоторых применениях усиленного закона больших чисел.

Пример 1 (*применение к теории чисел*). Пусть $\Omega=[0, 1)$, $\mathscr{B}$ — борелевская система подмножеств Ω и P — мера Лебега на $[0, 1)$. Рассмотрим двоичное разложение $\omega=0, \omega_1\omega_2\ldots$ чисел $\omega\in\Omega$ (с бесконечным количеством нулей) и определим случайные величины $\xi_1(\omega), \xi_2(\omega), \ldots$, полагая $\xi_n(\omega)=\omega_n$. Поскольку для любого $n\geqslant 1$ и любых $x_1, \ldots, x_n$, принимающих значения 0 или 1,

$$\{\omega\colon \xi_1(\omega)=x_1, \ldots, \xi_n(\omega)=x_n\}=$$
$$=\left\{\omega\colon \frac{x_1}{2}+\frac{x_2}{2^2}+\ldots+\frac{x_n}{2^n}\leqslant\omega<\frac{x_1}{2}+\ldots+\frac{x_n}{2^n}+\frac{1}{2^n}\right\},$$

то P-мера этого множества равна $1/2^n$. Отсюда вытекает, что $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с

$$\mathsf{P}(\xi_1=0)=\mathsf{P}(\xi_1=1)=\frac{1}{2}.$$

Отсюда и из усиленного закона больших чисел вытекает следующий результат Бореля: *почти все числа интервала* $[0, 1)$ *нормальны в том смысле, что с вероятностью единица доля нулей и единиц в их двоичном разложении стремится к* $1/2$, т. е.

$$\frac{1}{n}\sum_{k=1}^n I(\xi_k=1)\to\frac{1}{2}\quad(\mathsf{P}\text{-п. н.}).$$

Пример 2 (*применение к «методу Монте Карло»*). Пусть $f(x)$ — непрерывная функция, заданная на интервале $[0, 1]$ и принимающая значения из $[0, 1]$. Следующие рассуждения лежат в основе статистического метода численного вычисления интегралов $\int_0^1 f(x)\,dx$ («метод Монте Карло»).

Пусть $\xi_1, \eta_1, \xi_2, \eta_2, \ldots$ — последовательность независимых случайных величин, равномерно распределенных на $[0, 1]$. Положим

$$\rho_i = \begin{cases} 1, & \text{если } f(\xi_i) > \eta_i, \\ 0, & \text{если } f(\xi_i) \leqslant \eta_i. \end{cases}$$

Ясно, что

$$\mathsf{M}\rho_1 = \mathsf{P}\{f(\xi_1) > \eta_1\} = \int_0^1 f(x)\,dx.$$

В силу усиленного закона больших чисел (теорема 3)

$$\frac{1}{n}\sum_{i=1}^{n} \rho_i \to \int_0^1 f(x)\,dx \quad (\mathsf{P}\text{-п. н.}).$$

Таким образом, численный подсчет интеграла $\int_0^1 f(x)\,dx$ можно осуществлять с помощью моделирования пар случайных чисел (ξ_i, η_i), $i \geqslant 1$, с последующим подсчетом величин ρ_i и $\frac{1}{n}\sum_{i=1}^{n} \rho_i$.

5. Задачи.

1. Показать, что $\mathsf{M}\xi^2 < \infty$ тогда и только тогда, когда $\sum_{n=1}^{\infty} n\mathsf{P}(|\xi| > n) < \infty$.

2. Предполагая, что $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены, показать, что если $\mathsf{M}|\xi_1|^\alpha < \infty$ для некоторого $0 < \alpha < 1$, то $\frac{S_n}{n^{1/\alpha}} \to 0$ (P-п. н.), и если $\mathsf{M}|\xi_1|^\beta < \infty$ для некоторого $1 \leqslant \beta < 2$, то $\frac{S_n - n\mathsf{M}\xi_1}{n^{1/\beta}} \to 0$ (P-п. н.).

3. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{M}|\xi_1| = \infty$. Показать, что для любой последовательности констант $\{a_n\}$

$$\varlimsup_n \left| \frac{S_n}{n} - a_n \right| = \infty \quad (\mathsf{P}\text{-п. н.}).$$

4. Показать, что все рациональные числа из $[0, 1)$ не являются нормальными (в смысле примера 1 в п. 4).

§ 4. Закон повторного логарифма

1. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых бернуллиевских случайных величин с $\mathsf{P}(\xi_n=1)=\mathsf{P}(\xi_n=-1)=-1/2$, $S_n=\xi_1+\ldots+\xi_n$. Из доказательства теоремы 2 в § 1 следует, что с вероятностью единица

$$\overline{\lim}\frac{S_n}{\sqrt{n}}=+\infty,\quad \underline{\lim}\frac{S_n}{\sqrt{n}}=-\infty. \tag{1}$$

С другой стороны, согласно (3.11),

$$\frac{S_n}{\sqrt{n}\log n}\to 0\ (\mathsf{P}\text{-п. н.}). \tag{2}$$

Сравним эти два результата.

Из (1) следует, что с вероятностью единица траектории $(S_n)_{n\geqslant 1}$ бесконечное число раз пересекают «кривые» $\pm\varepsilon\sqrt{n}$, где ε — любое положительное число, но в то же самое время они в силу (2) лишь конечное число раз выходят из внутренности области, ограниченной кривыми $\pm\varepsilon\sqrt{n}\log n$. Эти два результата дают весьма полезную информацию о характере «размаха» колебаний симметричного случайного блуждания $(S_n)_{n\geqslant 1}$. Приводимый ниже закон повторного логарифма существенно уточняет эти представления о «размахе» колебаний $(S_n)_{n\geqslant 1}$.

Введем такое

Определение. Функция $\varphi^*=\varphi^*(n)$, $n\geqslant 1$, называется *верхней* (для $(S_n)_{n\geqslant 1}$), если с вероятностью единица $S_n\leqslant\varphi^*(n)$ для всех n, начиная с некоторого $n=n_0(\omega)$.

Функция $\varphi_*=\varphi_*(n)$, $n\geqslant 1$, называется *нижней* (для $(S_n)_{n\geqslant 1}$), если с вероятностью единица $S_n>\varphi_*(n)$ для бесконечно многих n.

В соответствии с этим определением и в силу (1) и (2) можно сказать, что каждая из функций $\varphi^*=\varepsilon\sqrt{n}\log n$, $\varepsilon>0$, является верхней, а функция $\varphi_*=\varepsilon\sqrt{n}$ — нижней, $\varepsilon>0$.

Пусть $\varphi=\varphi(n)$ — некоторая функция и $\varphi_\varepsilon^*=(1+\varepsilon)\varphi$, $\varphi_{*\varepsilon}=(1-\varepsilon)\varphi$, где $\varepsilon>0$. Тогда нетрудно видеть, что

$$\left\{\overline{\lim}\frac{S_n}{\varphi(n)}\leqslant 1\right\}=\left\{\lim_n\left[\sup_{m\geqslant n}\frac{S_m}{\varphi(n)}\right]\leqslant 1\right\}\Leftrightarrow$$

$$\Leftrightarrow\left\{\sup_{m\geqslant n_1(\varepsilon)}\frac{S_m}{\varphi(m)}\leqslant 1+\varepsilon \ \begin{array}{l}\text{для всякого } \varepsilon>0, \text{ начи-}\\ \text{ная с некоторого } n_1(\varepsilon)\end{array}\right\}\Leftrightarrow$$

$$\Leftrightarrow\left\{S_m\leqslant(1+\varepsilon)\varphi(m) \ \begin{array}{l}\text{для всякого } \varepsilon>0, \text{ начи-}\\ \text{ная с некоторого } n_1(\varepsilon)\end{array}\right\}. \tag{3}$$

Точно так же

$$\left\{\overline{\lim}\frac{S_n}{\varphi(n)}\geqslant 1\right\}=\left\{\lim_n\left[\sup_{m\geqslant n}\frac{S_m}{\varphi(m)}\right]\geqslant 1\right\}\Leftrightarrow$$

$$\Leftrightarrow\left\{\sup_{m\geqslant n_2(\varepsilon)}\frac{S_m}{\varphi(m)}\geqslant 1-\varepsilon \text{ для всякого } \varepsilon>0, \text{ начиная с некоторого } n_2(\varepsilon)\right\}\Leftrightarrow$$

$$\Leftrightarrow\left\{\begin{array}{l} S_m\geqslant(1-\varepsilon)\varphi(m) \text{ для всякого } \varepsilon>0 \text{ и для} \\ \text{бесконечно многих значений } m, \text{ больших} \\ \text{некоторого } n_3(\varepsilon)\geqslant n_2(\varepsilon)\end{array}\right\}. \qquad (4)$$

Из (3) и (4) вытекает, что для того, чтобы проверить, что каждая из функций $\varphi_\varepsilon^*=(1+\varepsilon)\varphi$, $\varepsilon>0$, является верхней, надо доказать, что

$$\mathsf{P}\left\{\overline{\lim}\frac{S_n}{\varphi(n)}\leqslant 1\right\}=1. \qquad (5)$$

А для того чтобы доказать, что функции $\varphi_{*\varepsilon}=(1-\varepsilon)\varphi$, $\varepsilon>0$, являются нижними, надо установить, что

$$\mathsf{P}\left\{\overline{\lim}\frac{S_n}{\varphi(n)}\geqslant 1\right\}=1. \qquad (6)$$

2. Теорема 1 (закон повторного логарифма). *Пусть $\xi_1, \xi_2, \ldots$ последовательность независимых одинаково распределенных случайных величин с* $\mathsf{M}\xi_i=0$ *и* $\mathsf{M}\xi_i^2=\sigma^2>0$. *Тогда*

$$\mathsf{P}\left\{\overline{\lim}\frac{S_n}{\psi(n)}=1\right\}=1, \qquad (7)$$

где

$$\psi(n)=\sqrt{2\sigma^2 n\log\log n}. \qquad (8)$$

Для случая равномерно ограниченных случайных величин закон повторного логарифма был установлен Хинчиным (1924 г.). В 1929 г. Колмогоров обобщил этот результат на широкий класс независимых случайных величин. В условиях, сформулированных в теореме 1, закон повторного логарифма установлен Хартманом и Винтнером (1941 г.).

Поскольку доказательство этой теоремы довольно сложно, ограничимся рассмотрением лишь частного случая, когда случайные величины ξ_n являются нормально распределенными, $\xi_n\sim\mathcal{N}(0, 1)$, $n\geqslant 1$.

Начнем с доказательства двух вспомогательных результатов.

Лемма 1. *Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины с симметричным распределением* ($\mathsf{P}(\xi_k\in B)=\mathsf{P}(-\xi_k\in B)$ *для* *каждого* $B \in \mathscr{B}(R)$, $k \leqslant n$). *Тогда для любого действительного* a

$$\mathsf{P}\left(\max_{1 \leqslant k \leqslant n} S_k > a\right) \leqslant 2\mathsf{P}(S_n > a). \tag{9}$$

Доказательство. Пусть $A = \left\{\max_{1 \leqslant k \leqslant n} S_k > a\right\}$, $A_k = \{S_i \leqslant a,\ i \leqslant k-1;\ S_k > a\}$ и $B = \{S_n > a\}$. Поскольку на множестве A_k $S_n > a$ (так как $S_k \leqslant S_n$), то

$$\mathsf{P}(B \cap A_k) \geqslant \mathsf{P}(A_k \cap \{S_n \geqslant S_k\}) = \mathsf{P}(A_k)\,\mathsf{P}(S_n \geqslant S_k) =$$
$$= \mathsf{P}(A_k)\,\mathsf{P}(\xi_{k+1} + \ldots + \xi_n \geqslant 0).$$

В силу симметричности распределений вероятностей случайных величин $\xi_1, \ldots, \xi_n$

$$\mathsf{P}(\xi_{k+1} + \ldots + \xi_n > 0) = \mathsf{P}(\xi_{k+1} + \ldots + \xi_n < 0).$$

Поэтому $\mathsf{P}(\xi_{k+1} + \ldots + \xi_n \geqslant 0) \geqslant 1/2$ и, значит,

$$\mathsf{P}(B) \geqslant \sum_{k=1}^{n} \mathsf{P}(A_k \cap B) \geqslant \frac{1}{2} \sum_{k=1}^{n} \mathsf{P}(A_k) = \frac{1}{2}\mathsf{P}(A),$$

что и доказывает (9).

Лемма 2. *Пусть* $S_n \sim \mathscr{N}(0, \sigma^2(n))$, $\sigma^2(n) \uparrow \infty$ *и числа* $a(n)$, $n \geqslant 1$, *таковы, что* $\frac{a(n)}{\sigma(n)} \to \infty$, $n \to \infty$. *Тогда*

$$\mathsf{P}(S_n > a(n)) \sim \frac{\sigma(n)}{\sqrt{2\pi}\, a(n)} e^{-\frac{a^2(n)}{2\sigma^2(n)}}. \tag{10}$$

Доказательство следует из того, что при $x \to \infty$

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-x^2/2} dy \sim \frac{1}{\sqrt{2\pi}\, x} e^{-x^2/2},$$

а случайная величина $S_n/\sigma(n) \sim \mathscr{N}(0, 1)$.

Доказательство теоремы 1 (в случае $\xi_i \sim \mathscr{N}(0, 1)$). Установим сначала соотношение (5). Пусть $\varepsilon > 0$, $\lambda = 1 + \varepsilon$, $n_k = \lambda^k$, где $k \geqslant k_0$, а k_0 выбирается так, чтобы $\ln \ln k_0$ был определен. Обозначим также

$$A_k = \{S_n > \lambda \psi(n) \text{ для некоторого } n \in (n_k, n_{k+1}]\}, \tag{11}$$

и пусть

$$A=\{A_k \text{ б. ч.}\}=\{S_n>\lambda\psi(n) \text{ для бесконечно многих } n\}.$$

В соответствии с (3) для доказательства (5) достаточно доказать, что $\mathsf{P}(A)=0$.

Покажем, что $\sum \mathsf{P}(A_k)<\infty$. Тогда по лемме Бореля — Кантелли $\mathsf{P}(A)=0$.

Из (11), (9) и (10) находим, что

$$\mathsf{P}(A_k)\leqslant \mathsf{P}\{S_n>\lambda\psi(n_k) \text{ для некоторого } n\in(n_k,\ n_{k+1})\}\leqslant$$
$$\leqslant \mathsf{P}\{S_n>\lambda\psi(n_k) \text{ для некоторого } n\leqslant n_{k+1}\}\leqslant$$
$$\leqslant 2\mathsf{P}\{S_{n_{k+1}}>\lambda\psi(n_k)\}\sim\frac{2}{\sqrt{2\pi}\frac{\lambda\psi(n_k)}{\sqrt{n_k}}}e^{-\frac{1}{2}\left(\frac{\lambda\psi(n_k)}{\sqrt{n_k}}\right)^2}\leqslant$$
$$\leqslant C_1e^{-\lambda\ln\ln\lambda^k}\leqslant C_2e^{-\lambda\ln k}=C_2k^{-\lambda},$$

где C_1 и C_2 — некоторые константы. Но $\sum\limits_{k=1} k^{-\lambda}<\infty$, поэтому $\sum \mathsf{P}(A_k)<\infty$.

Итак, соотношение (5) доказано.

Перейдем к доказательству (6). В соответствии с (4) надо показать, что для $\lambda=1-\varepsilon$, $\varepsilon>0$, с вероятностью единица $S_n\geqslant\geqslant\lambda\psi(n)$ для бесконечно многих n. Применим доказанное соотношение (5) к последовательности $(-S_n)_{n\geqslant 1}$. Тогда получим, что для всех n, за исключением, быть может, конечного числа (P-п. н.) $-S_n\leqslant 2\psi(n)$. Следовательно, если $n_k=N^k$, $N>1$, то для достаточно больших k

$$S_{n_{k-1}}\geqslant-2\psi(n_{k-1})$$

или

$$S_{n_k}\geqslant Y_k-2\psi(n_{k-1}),\tag{12}$$

где $Y_k=S_{n_k}-S_{n_{k-1}}$.

Поэтому, если доказать, что для бесконечно многих k

$$Y_k>\lambda\psi(n_k)+2\psi(n_{k-1}),\tag{13}$$

то вместе с (12) это даст, что (P-п. н.) для бесконечно многих k $S_{n_k}>\lambda\psi(n_k)$. Возьмем некоторое $\lambda'\in(\lambda,\ 1)$. Тогда можно найти такое $N>1$, что для всех k

$$\lambda'[2(N^k-N^{k-1})\ln\ln N^k]^{1/2}>\lambda(2N^k\ln\ln N^k)^{1/2}+$$
$$+2(2N^{k-1}\ln\ln N^{k-1})^{1/2}\equiv\lambda\psi(N^k)+2\psi(N^{k-1}).$$

Теперь достаточно показать, что для бесконечно многих k

$$Y_k > \lambda' [2(N^k - N^{k-1}) \ln \ln N^k]^{1/2}. \tag{14}$$

Очевидно, $Y_k \sim \mathscr{N}(0, N^k - N^{k-1})$. Поэтому в силу леммы 2

$$\mathsf{P}\{Y_k > \lambda' [2(N^k - N^{k-1}) \ln \ln N^k]^{1/2}\} \sim$$

$$\sim \frac{1}{\sqrt{2\pi}\lambda' (2 \ln \ln N^k)^{1/2}} e^{-(\lambda')^2 \ln \ln N^k} \geqslant \frac{C_1}{(\ln k)^{1/2}} k^{-(\lambda')^2} \geqslant \frac{C_2}{k \ln k}.$$

Так как $\sum \frac{1}{k \ln k} = \infty$, то, следовательно, по второй части леммы Бореля — Кантелли с вероятностью единица для бесконечно многих k выполнено (14), что и доказывает соотношение (6).

Теорема доказана.

Замечание 1. Применяя (7) к случайным величинам $(-S_n)_{n \geqslant 1}$, находим, что

$$\underline{\lim} \frac{S_n}{\varphi(n)} = -1. \tag{15}$$

Из (7) и (15) следует, что закону повторного логарифма можно придать также следующую форму:

$$\mathsf{P}\left\{\overline{\lim} \frac{|S_n|}{\varphi(n)} = 1\right\} = 1. \tag{16}$$

Замечание 2. Закон повторного логарифма говорит о том, что для любого $\varepsilon > 0$ каждая из функций $\psi_\varepsilon^* = (1+\varepsilon)\psi$ является верхней, а функция $\psi_{*\varepsilon} = (1-\varepsilon)\psi$ — нижней.

Утверждение (7) закона повторного логарифма эквивалентно также тому, что для всякого $\varepsilon > 0$

$$\mathsf{P}\{|S_n| \geqslant (1-\varepsilon)\psi(n) \text{ б. ч.}\} = 1,$$
$$\mathsf{P}\{|S_n| \geqslant (1+\varepsilon)\psi(n) \text{ б. ч.}\} = 0.$$

3. Задачи.

1. Пусть $\xi_1, \xi_2, \ldots$ последовательность независимых случайных величин, $\xi_n \sim N(0, 1)$. Показать, что

$$\mathsf{P}\left\{\overline{\lim} \frac{\xi_n}{\sqrt{2 \ln n}} = 1\right\} = 1.$$

2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, распределенных по закону Пуассона с параметром $\lambda > 0$. Показать, что (независимо от λ)

$$\mathsf{P}\left\{\overline{\lim} \frac{\xi_n \ln \ln n}{\ln n} = 1\right\} = 1.$$

3. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с

$$\mathsf{M} e^{it\xi_1} = e^{-|t|^\alpha}, \qquad 0 < \alpha < 2.$$

Показать, что

$$\mathsf{P}\left\{\overline{\lim}\left|\frac{S_n}{n^{1/\alpha}}\right|^{\frac{1}{\ln\ln n}}=e^{1/\alpha}\right\}=1.$$

4. Установить справедливость следующего обобщения неравенства (9). Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины. Тогда для всякого действительного a справедливо *неравенство Леви*

$$\mathsf{P}\left\{\max_{1\leqslant k\leqslant n}[S_k+\mu(S_n-S_k)]>a\right\}\leqslant 2\mathsf{P}(S_n>a),$$

где $\mu(\xi)$ — медиана случайной величины ξ, т. е. такая константа, что

$$\mathsf{P}(\xi\geqslant\mu(\xi))\geqslant\frac{1}{2},\quad \mathsf{P}(\xi\leqslant\mu(\xi))\geqslant\frac{1}{2}.$$

ГЛАВА V

СТАЦИОНАРНЫЕ (В УЗКОМ СМЫСЛЕ) СЛУЧАЙНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ И ЭРГОДИЧЕСКАЯ ТЕОРИЯ

§ 1. Стационарные (в узком смысле) случайные последовательности. Сохраняющие меру преобразования

1. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство и $\xi = (\xi_1, \xi_2, \ldots)$ — некоторая последовательность случайных величин, или *случайная последовательность*. Обозначим через $\theta_k\xi$ последовательность $(\xi_{k+1}, \xi_{k+2}, \ldots)$.

Определение 1. Случайная последовательность ξ называется *стационарной* (*в узком смысле*), если для любого $k \geqslant 1$ распределения вероятностей $\theta_k\xi$ и ξ совпадают:

$$\mathsf{P}((\xi_1, \xi_2, \ldots) \in B) = \mathsf{P}((\xi_{k+1}, \xi_{k+2}, \ldots) \in B), \qquad B \in \mathscr{B}(R^\infty).$$

Простейшим примером такой последовательности ξ является последовательность $\xi = (\xi_1, \xi_2, \ldots)$, состоящая из независимых одинаково распределенных случайных величин. Отправляясь от такой последовательности, можно сконструировать широкий класс стационарных последовательностей $\eta = (\eta_1, \eta_2, \ldots)$, если взять произвольную борелевскую функцию $g(x_1, \ldots, x_n)$ и положить $\eta_k = g(\xi_k, \xi_{k+1}, \ldots, \xi_{k+n})$.

Если $\xi = (\xi_1, \xi_2, \ldots)$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{M}|\xi_1| < \infty$ и $\mathsf{M}\xi_1 = m$, то, согласно усиленному закону больших чисел с вероятностью единица,

$$\frac{\xi_1 + \ldots + \xi_n}{n} \to m, \qquad n \to \infty.$$

В 1931 г. Биркгоф получил замечательное обобщение этого результата на случай стационарных последовательностей. Именно доказательство теоремы Биркгофа и составляет основное содержание настоящей главы.

Последующее изложение будет вестись с привлечением понятия «сохраняющего меру преобразования», что даст возможность как познакомиться с одной из интересных ветвей анализа — эргодической теорией, так и установить ее связь с теорией стационарных случайных последовательностей.

Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — некоторое вероятностное пространство.

Определение 2. Отображение T пространства Ω в себя называется *измеримым*, если для всякого $A \in \mathscr{F}$

$$T^{-1}A = \{\omega\colon T\omega \in A\} \in \mathscr{F}.$$

Определение 3. Измеримое отображение T называется *сохраняющим меру преобразованием* (морфизмом), если для всякого $A \in \mathscr{F}$

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

Пусть T — сохраняющее меру преобразование, T^n — его n-я степень и $\xi_1 = \xi_1(\omega)$ — некоторая случайная величина. Положим $\xi_n(\omega) = \xi_1(T^{n-1}\omega)$, $n \geqslant 2$, и рассмотрим последовательность $\xi = (\xi_1, \xi_2, \ldots)$. Мы утверждаем, что эта последовательность является стационарной.

В самом деле, пусть $A = \{\omega\colon \xi \in B\}$, $A_1 = \{\omega\colon \theta_1\xi \in B\}$, где $B \in \mathscr{B}(R^\infty)$. Поскольку $A = \{\omega\colon (\xi_1(\omega), \xi_1(T\omega), \ldots) \in B\}$, а $A_1 = \{\omega\colon (\xi_1(T\omega), \xi_1(T^2\omega), \ldots) \in B\}$, то $\omega \in A_1$ в том и только том случае, когда $T\omega \in A$, или $A_1 = T^{-1}A$. Но $\mathsf{P}(T^{-1}A) = \mathsf{P}A$, и поэтому $\mathsf{P}(A_1) = \mathsf{P}(A)$. Аналогичным образом $\mathsf{P}(A_k) = \mathsf{P}(A)$ для любого $A_k = \{\omega\colon \theta_k\xi \in B\}$, $k \geqslant 2$.

Итак, введение сохраняющего меру преобразования дает возможность построения стационарных (в узком смысле) случайных последовательностей.

В определенном смысле верен и обратный результат: для каждой стационарной последовательности ξ, рассматриваемой на $(\Omega, \mathscr{F}, \mathsf{P})$, можно указать новое вероятностное пространство $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathsf{P}})$, случайную величину $\tilde{\xi}_1(\tilde{\omega})$ и сохраняющее меру преобразование $\tilde{T}$ такие, что распределение случайной последовательности $\tilde{\xi} = \{\tilde{\xi}_1(\tilde{\omega}), \tilde{\xi}_1(\tilde{T}\tilde{\omega}), \ldots\}$ совпадает с распределением последовательности ξ.

Действительно, возьмем в качестве $\tilde{\Omega}$ «координатное» пространство R^∞ и положим $\tilde{\mathscr{F}} = \mathscr{B}(R^\infty)$, $\tilde{\mathsf{P}} = P_\xi$, где $P_\xi(B) = \mathsf{P}\{\omega\colon \xi \in B\}$, $B \in \mathscr{B}(R^\infty)$. Преобразование $\tilde{T}$, действующее в $\tilde{\Omega}$, определим по формуле

$$\tilde{T}(x_1, x_2, \ldots) = (x_2, x_3, \ldots).$$

Положим также для $\tilde{\omega} = (x_1, x_2, \ldots)$

$$\tilde{\xi}_1(\tilde{\omega}) = x_1, \quad \tilde{\xi}_n(\tilde{\omega}) = \tilde{\xi}_1(\tilde{T}^{n-1}\tilde{\omega}), \qquad n \geqslant 2.$$

Пусть теперь $A=\{\tilde{\omega}\colon (x_1, \ldots, x_k)\in B\}$, $B\in\mathscr{B}(R^k)$ и $\tilde{T}^{-1}A = \{\tilde{\omega}\colon (x_2, \ldots, x_{k+1})\in B\}$. Тогда в силу стационарности

$$\tilde{\mathsf{P}}(A)=\mathsf{P}\{\omega\colon (\xi_1, \ldots, \xi_k)\in B\}=\mathsf{P}\{\omega\colon (\xi_2, \ldots, \xi_{k+1})\in B\}=\tilde{\mathsf{P}}(\tilde{T}^{-1}A),$$

т. е. $\tilde{T}$ — сохраняющее меру преобразование. Поскольку $\tilde{\mathsf{P}}\{\tilde{\omega}\colon (\tilde{\xi}_1, \ldots, \tilde{\xi}_k)\in B\}=\mathsf{P}\{\omega\colon (\xi_1, \ldots, \xi_k)\in B\}$ для любого k, то отсюда следует, что распределения ξ и $\tilde{\xi}$ совпадают.

Приведем примеры сохраняющих меру преобразований.

Пример 1. Пусть $\Omega=\{\omega_1, \ldots, \omega_n\}$ — множество, состоящее из конечного числа точек, $n\geqslant 2$, $\mathscr{F}$ — все его подмножества, $T\omega_i=\omega_{i+1}$, $1\leqslant i\leqslant n-1$ и $T\omega_n=\omega_1$. Если $\mathsf{P}(\omega_i)=1/n$, то T — сохраняющее меру преобразование.

Пример 2. Если $\Omega=[0, 1)$, $\mathscr{F}=\mathscr{B}([0, 1))$, P — мера Лебега, $\lambda\in[0, 1)$, то $Tx=(x+\lambda) \bmod 1$ и $Tx=2x \bmod 1$ являются сохраняющими меру преобразованиями.

2. Остановимся на физических предпосылках, приводящих к изучению преобразований, сохраняющих меру.

Будем представлять себе Ω как фазовое пространство состояний ω некоторой системы, эволюционирующей (в дискретном времени) в соответствии с заданным законом движения. Тогда, если ω есть состояние в момент $n=1$, то $T^n\omega$, где T — оператор сдвига (индуцируемый данным законом движения), есть то состояние, в которое перейдет система через n шагов. Далее, если A — какое-то множество состояний ω, то $T^{-1}A=\{\omega\colon T\omega\in A\}$ есть по своему определению множество тех «начальных» состояний ω, которые через один шаг окажутся в множестве A. Поэтому, если интерпретировать Ω как «несжимаемую жидкость», то условие $\mathsf{P}(T^{-1}A)=\mathsf{P}(A)$ можно рассматривать как вполне естественное условие сохранения «объема». (Для классических консервативных гамильтоновых систем известная теорема Лиувилля утверждает, что соответствующее преобразование T является преобразованием, сохраняющим меру Лебега.)

3. Одним из первых результатов относительно преобразований, сохраняющих меру, была следующая теорема Пуанкаре (1912) о «возвратности».

Теорема 1. *Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — некоторое вероятностное пространство, T — преобразование, сохраняющее меру, и $A\in\mathscr{F}$. Тогда для почти каждой точки $\omega\in A$ $T^n\omega\in A$ для бесконечно многих $n\geqslant 1$.*

Доказательство. Обозначим $C=\{\omega\in A\colon T^n\omega\notin A$, *для всех* $n\geqslant 1\}$. Поскольку для любого $n\geqslant 1$ $C\cap T^{-n}C=\varnothing$, то $T^{-m}C\cap T^{-(m+n)}C=T^{-m}(C\cap T^{-n}C)=\varnothing$. Таким образом, последовательность $\{T^{-n}C\}$ состоит из непересекающихся множеств, P-мера которых одна и та же. Поэтому $\sum_{n=0}^{\infty} \mathsf{P}(C) = \sum_{n=0}^{\infty} \mathsf{P}(T^{-n}C) \leqslant \mathsf{P}(\Omega) = 1$ и, следовательно, $\mathsf{P}(C) = 0$. Таким образом, для почти каждой точки $\omega \in A$, по крайней мере для одного $n \geqslant 1$, $T^n\omega \in A$. Выведем отсюда, что тогда и для бесконечно многих n $T^n\omega \in A$.

Применим предшествующий результат к преобразованиям T^k, $k \geqslant 1$. Тогда для каждой точки $\omega \in A \setminus N$, где N — множество нулевой вероятности, являющееся объединением соответствующих множеств, отвечающих разным k, найдется такое n_k, что $(T^k)^{n_k}\omega \in A$. Отсюда, разумеется, следует, что $T^n\omega \in A$ для бесконечно многих n.

Теорема доказана.

Следствие. Пусть $\xi(\omega) \geqslant 0$. Тогда на множестве $\{\omega: \xi(\omega) > 0\}$

$$\sum_{k=0}^{\infty} \xi(T^k\omega) = \infty \quad (\mathsf{P}\text{-п. н.}).$$

В самом деле, пусть $A_n = \left\{\omega: \xi(\omega) \geqslant \frac{1}{n}\right\}$. Тогда, согласно теореме на множестве A_n (P-п. н.) $\sum_{k=0}^{\infty} \xi(T^k\omega) = \infty$, и требуемый результат следует, если положить $n \to \infty$.

Замечание. Теорема сохраняет свою силу, если вместо вероятностной меры P рассмотреть любую конечную меру μ, $\mu(\Omega) < \infty$.

4. Задачи.

1. Пусть T — сохраняющее меру преобразование и $\xi = \xi(\omega)$ — случайная величина с существующим математическим ожиданием $\mathsf{M}\xi(\omega)$. Показать, что $\mathsf{M}\xi(\omega) = \mathsf{M}\xi(T\omega)$.

2. Показать, что в примерах 1 и 2 преобразования T являются преобразованиями, сохраняющими меру.

3. Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$ и P — некоторая мера с непрерывной функцией распределения. Показать, что преобразования $Tx = \lambda x$, $0 < \lambda < 1$, и $Tx = x^2$ не являются преобразованиями, сохраняющими меру.

§ 2. Эргодичность и перемешивание

1. На протяжении всего данного параграфа будем через T обозначать сохраняющее меру преобразование, действующее на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$.

Определение 1. Множество $A \in \mathscr{F}$ называется *инвариантным*, если $T^{-1}A = A$. Множество $A \in \mathscr{F}$ называется *почти инвариантным*, если A и $T^{-1}A$ отличаются на множество меры нуль, т. е. $\mathsf{P}(A \Delta T^{-1}A) = 0$.

Нетрудно проверить, что класс инвариантных (почти инвариантных) множеств $\mathscr{I}$ (соответственно $\mathscr{I}^*$) образует σ-алгебру.

Определение 2. Сохраняющее меру преобразование T называется *эргодическим* (или *метрически транзитивным*), если каждое инвариантное множество A имеет меру нуль или единица.

Определение 3. Случайная величина $\xi=\xi(\omega)$ называется *инвариантной* (*почти инвариантной*), если $\xi(\omega)=\xi(T\omega)$ для всех $\omega\in\Omega$ (для почти всех $\omega\in\Omega$).

Следующая лемма устанавливает связь между инвариантными и почти инвариантными множествами.

Лемма 1. *Если A является почти инвариантным множеством, то найдется такое инвариантное множество B, что* $\mathsf{P}(A\Delta B)=0$.

Доказательство. Положим $B=\overline{\lim}\,T^{-n}A$. Тогда $T^{-1}B=\overline{\lim}\,T^{-(n+1)}A=B$, т. е. $B\in\mathscr{I}$. Нетрудно убедиться в том, что $A\Delta B\subseteq\bigcup\limits_{k=0}^{\infty}(T^{-k}A\Delta T^{-(k+1)}A)$. Но $\mathsf{P}(T^{-k}A\Delta T^{-(k+1)}A)=\mathsf{P})A\Delta T^{-1}A)=0$. Поэтому $\mathsf{P}(A\Delta B)=0$.

Лемма 2. *Преобразование T эргодично тогда и только тогда, когда каждое почти инвариантное множество имеет меру нуль или единица.*

Доказательство. Пусть $A\in\mathscr{I}^*$; тогда по лемме 1 найдется инвариантное множество B такое, что $\mathsf{P}(A\Delta B)=0$. Но T эргодично и, значит, $\mathsf{P}(B)=0$ или 1. Поэтому $\mathsf{P}(A)=0$ или 1. Обратное очевидно, поскольку $\mathscr{I}\subseteq\mathscr{I}^*$.

Лемма доказана.

Теорема 1. *Пусть T — сохраняющее меру преобразование. Следующие условия эквивалентны:*

(1) *T эргодично;*

(2) *каждая почти инвариантная случайная величина есть (P-п. н.) константа;*

(3) *каждая инвариантная случайная величина есть (P-п. н.) константа.*

Доказательство. (1) $\Leftrightarrow$ (2). Пусть T эргодично и ξ почти инвариантна, т. е. (P-п. н.) $\xi(\omega)=\xi(T\omega)$. Тогда для любого $c\in R$ множество $A_c=\{\omega\colon\ \xi(\omega)\le c\}\in\mathscr{I}^*$ и по лемме 2 $\mathsf{P}(A_c)=0$ или 1. Пусть $C=\sup\{c\colon\ \mathsf{P}(A_c)=0\}$. Поскольку $A_c\uparrow\Omega$ при $c\uparrow\infty$ и $A_c\downarrow\varnothing$ при $c\downarrow-\infty$, то $|C|<\infty$. Тогда

$$\mathsf{P}\{\omega\colon\ \xi(\omega)<C\}=\mathsf{P}\left\{\bigcup_{n=1}^{\infty}\left\{\xi(\omega)\le C-\frac{1}{n}\right\}\right\}=0$$

и аналогично $\mathsf{P}\{\omega\colon\ \xi(\omega)>c\}=0$. Тем самым $\mathsf{P}\{\omega\colon\ \xi(\omega)=C\}=1$.

(2) $\Rightarrow$ (3). Очевидно.

(3) $\Rightarrow$ (1). Пусть $A \in \mathscr{J}$, тогда I_A — инвариантная случайная величина и, значит, (Р-п. н.) $I_A = 0$ или $I_A = 1$, откуда $\mathsf{P}(A) = 0$ или 1.

Замечание. Утверждение теоремы остается в силе и в том случае, когда рассматриваемые случайные величины ограничены.

В качестве иллюстрации применения этой теоремы рассмотрим следующий

Пример. Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$, P — мера Лебега и $T\omega = (\omega + \lambda) \bmod 1$. Покажем, что T эргодично в том и только том случае, когда λ иррационально.

Пусть $\xi = \xi(\omega)$ — случайная величина с $\mathsf{M}\xi^2(\omega) < \infty$. Тогда известно, что ряд Фурье $\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n \omega}$ функции $\xi(\omega)$ сходится в среднеквадратическом смысле, $\sum |c_n|^2 < \infty$ и

$$\xi(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n \omega} \quad (\mathsf{P}\text{-п. н.}).$$

Отсюда

$$\xi(T\omega) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n \omega} e^{2\pi i n \lambda},$$

и, если ξ инвариантна, то $c_n(1 - e^{2\pi i n \lambda}) = 0$. По предположению λ иррационально и, значит, для всех $n \neq 1$ $e^{2\pi i n \lambda} \neq 1$. Поэтому $c_n = 0$, $n \neq 1$, $\xi(\omega) = c_0$ (Р-п. н.) и по теореме 1 преобразование T эргодично.

С другой стороны, пусть λ рационально, т. е., $\lambda = k/m$, где k и m — целые. Рассмотрим множество $A = \left\{\omega\colon 0 \leqslant \omega < 1/m,\ 2/m \leqslant \omega < 3/m, \ldots \frac{2m-2}{m} \leqslant \omega < \frac{2m-1}{m}\right\}$. Ясно, что это множество является инвариантным, но $P(A) = 1/2$. Следовательно, T не эргодично.

2. Определение 4. Сохраняющее меру преобразование T называется *перемешиванием* (обладающим свойством перемешивания), если для любых $A, B \in \mathscr{F}$

$$\lim_{n \to \infty} \mathsf{P}(A \cap T^{-n}B) = \mathsf{P}(A)\,\mathsf{P}(B). \tag{1}$$

Следующая теорема устанавливает связь между эргодичностью и перемешиванием.

Теорема 2. *Всякое преобразование T, обладающее свойством перемешивания, является эргодическим.*

Доказательство. Пусть $A \in \mathscr{F}$, $B \in \mathscr{J}$. Тогда $B = T^{-n}B$, $n \geqslant 1$, и, значит, $\mathsf{P}(A \cap T^{-n}B) = \mathsf{P}(A \cap B)$ для всех $n \geqslant 1$. В силу

(1) $\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)$. Поэтому при $A = B$ находим, что $\mathsf{P}(B) = \mathsf{P}^2(B)$ и, следовательно, $\mathsf{P}(B) = 0$ или 1. Теорема доказана.

3. Задачи.

1. Показать, что случайная величина ξ является инвариантной тогда и только тогда, когда она $\mathscr{J}$-измерима.

2. Показать, что множество A является почти инвариантным тогда и только тогда, когда или $\mathsf{P}(T^{-1}A \setminus A) = 0$ или $\mathsf{P}(A \setminus T^{-1}A) = 0$.

3. Показать, что преобразование, рассмотренное в примере п. 1, не обладает свойством перемешивания.

4. Показать, что преобразование T есть перемешивание в том и только том случае, когда для любых двух случайных величин ξ и η с $\mathsf{M}\xi^2 < \infty$, $\mathsf{M}\eta^2 < \infty$

$$\mathsf{M}\xi(T^n\omega)\,\eta(\omega) \to \mathsf{M}\xi(\omega)\,\mathsf{M}\eta(\omega), \qquad n \to \infty.$$

§ 3. Эргодические теоремы

1. Теорема 1 (Биркгоф и Хинчин). *Пусть T — сохраняющее меру преобразование и $\xi = \xi(\omega)$ — случайная величина с $\mathsf{M}|\xi| < \infty$. Тогда* (P-*п. н.*)

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k\omega) = \mathsf{M}(\xi \mid \mathscr{J}). \tag{1}$$

Если к тому же T эргодично, то (P-*п. н.*)

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k\omega) = \mathsf{M}\xi. \tag{2}$$

Приводимое ниже доказательство существенно опирается на следующее предложение, простое доказательство которого было найдено А. Гарсиа (1965).

Лемма (максимальная эргодическая теорема). *Пусть T — сохраняющее меру преобразование, ξ — случайная величина с $\mathsf{M}|\xi| < \infty$ и*

$$S_k(\omega) = \xi(\omega) + \xi(T\omega) + \ldots + \xi(T^{k-1}\omega),$$
$$M_k(\omega) = \max\{0,\ S_1(\omega), \ldots,\ S_k(\omega)\}.$$

Тогда для любого $n \geq 1$

$$\mathsf{M}\left[\xi(\omega)\, I_{\{M_n > 0\}}(\omega)\right] \geq 0.$$

Доказательство. Если $n \geq k$, то $M_n(T\omega) \geq S_k(T\omega)$ и, значит, $\xi(\omega) + M_n(T\omega) \geq \xi(\omega) + S_k(T\omega) = S_{k+1}(\omega)$. Так как, очевидно, что $\xi(\omega) \geqslant S_1(\omega) - M_n(T\omega)$, то

$$\xi(\omega) \geqslant \max\{S_1(\omega), \ldots, S_n(\omega)\} - M_n(T\omega).$$

Поэтому

$$\mathsf{M}\left[\xi(\omega) I_{\{M_n>0\}}(\omega)\right] \geqslant \mathsf{M}(\max(S_1(\omega), \ldots, S_n(\omega)) - M_n(T\omega)).$$

Но на множестве $\{M_n > 0\}$ $\max(S_1, \ldots, S_n) = M_n$. Следовательно,

$$\mathsf{M}\left[\xi(\omega) I_{\{M_n>0\}}(\omega)\right] \geqslant \mathsf{M}\left\{(M_n(\omega) - M_n(T\omega)) I_{\{M_n(\omega)>0\}}\right\} \geqslant$$
$$\geqslant \mathsf{M}\{M_n(\omega) - M_n(T\omega)\} = 0,$$

так как, если T — сохраняющее меру преобразование, то $\mathsf{M}M_n(\omega) = \mathsf{M}M_n(T\omega)$ (задача 1 из § 1).

Лемма доказана.

Доказательство теоремы. Будем предполагать $\mathsf{M}(\xi \mid \mathscr{J}) = 0$ (в противном случае от ξ надо перейти к $\xi - \mathsf{M}(\xi \mid \mathscr{J})$).

Пусть $\bar{\eta} = \overline{\lim} \frac{S_n}{n}$ и $\underline{\eta} = \underline{\lim} \frac{S_n}{n}$. Для доказательства достаточно установить, что (P-п. н.)

$$0 \leqslant \underline{\eta} \leqslant \bar{\eta} \leqslant 0.$$

Рассмотрим случайную величину $\bar{\eta} = \bar{\eta}(\omega)$. Поскольку $\bar{\eta}(\omega) = \bar{\eta}(T\omega)$, то $\bar{\eta}$ инвариантна и, следовательно, для каждого $\varepsilon > 0$ множество $A_\varepsilon = \{\bar{\eta}(\omega) > \varepsilon\}$ также является инвариантным. Введем новую случайную величину

$$\xi^*(\omega) = (\xi(\omega) - \varepsilon) I_{A_\varepsilon}(\omega),$$

и пусть

$$S_k^*(\omega) = \xi^*(\omega) + \ldots + \xi^*(T^{k-1}\omega), \quad M_k^*(\omega) = \max(0, S_1^*, \ldots, S_k^*).$$

Тогда, согласно лемме для любого $n \geqslant 1$,

$$\mathsf{M}\left[\xi^* I_{\{M_n^*>0\}}\right] \geqslant 0.$$

Но при $n \to \infty$

$$\{M_n^* > 0\} = \left\{\max_{1\leqslant k\leqslant n} S_k^* > 0\right\} \uparrow \left\{\sup_{k\geqslant 1} S_k^* > 0\right\} = \left\{\sup_{k\geqslant 1} \frac{S_k^*}{k} > 0\right\} =$$
$$= \left\{\sup_{k\geqslant 1} \frac{S_k}{k} > \varepsilon\right\} \cap A_\varepsilon = A_\varepsilon,$$

где последнее равенство следует из того, что $\sup\limits_{k\geqslant 1} \frac{S_k}{k} \geqslant \bar{\eta}$, а $A_\varepsilon = \{\omega\colon \bar{\eta} > \varepsilon\}$.

Далее, $\mathsf{M}|\xi^*| \leqslant \mathsf{M}|\xi| + \varepsilon$. Поэтому по теореме о мажорируемой сходимости

$$0 \leqslant \mathsf{M}[\xi^* I_{\{M_n^*>0\}}] \to \mathsf{M}[\xi^* I_A].$$

Итак,

$$0 \leqslant \mathsf{M}\left[\xi^* I_{A_\varepsilon}\right] = \mathsf{M}\left[(\xi - \varepsilon) I_{A_\varepsilon}\right] = \mathsf{M}\left[\xi I_{A_\varepsilon}\right] - \varepsilon \mathsf{P}(A_\varepsilon) = \\ = \mathsf{M}\left[\mathsf{M}(\xi \mid \mathscr{J}) I_{A_\varepsilon}\right] - \varepsilon \mathsf{P}(A_\varepsilon) = -\varepsilon \mathsf{P}(A_\varepsilon),$$

откуда $\mathsf{P}(A_\varepsilon) = 0$ и, значит, $\mathsf{P}(\overline{\eta} \leqslant 0) = 1$.

Аналогично, рассматривая вместо $\xi(\omega)$ величины $-\xi(\omega)$, найдем, что

$$\overline{\lim}\left(-\frac{S_n}{n}\right) = -\underline{\lim}\,\frac{S_n}{n} = -\underline{\eta}$$

и $\mathsf{P}(-\underline{\eta} \leqslant 0) = 1$, т. е. $\mathsf{P}(\underline{\eta} \geqslant 0) = 1$. Тем самым $0 \leqslant \underline{\eta} \leqslant \overline{\eta} \leqslant 0$ (P-п. н.), что и доказывает первое утверждение теоремы.

Для доказательства второго утверждения достаточно заметить, что поскольку $\mathsf{M}(\xi \mid \mathscr{J})$ — инвариантная случайная величина, то в эргодическом случае $\mathsf{M}(\xi \mid \mathscr{J}) = \mathsf{M}\xi$ (P-п. н.).

Теорема доказана.

С л е д с т в и е. Сохраняющее меру преобразование T эргодично в том и только том случае, когда для любых $A, B \in \mathscr{F}$

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \mathsf{P}(A \cap T^{-k}B) = \mathsf{P}(A)\,\mathsf{P}(B). \tag{3}$$

Для доказательства эргодичности T положим в (3) $A = B \in \mathscr{J}$. Тогда $A \cap T^{-k}B = B$ и, значит, $\mathsf{P}(B) = \mathsf{P}^2(B)$, т. е. $\mathsf{P}(B) = 0$ или 1. Обратно, пусть T эргодично. Тогда, применяя (2) к случайной величине $\xi = I_B(\omega)$, где $B \in \mathscr{F}$, найдем, что (P-п. н.)

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} I_{T^{-k}B}(\omega) = \mathsf{P}(B),$$

откуда, интегрируя обе части по множеству $A \in \mathscr{F}$ и используя теорему о мажорируемой сходимости, получаем требуемое соотношение (3).

2. Покажем теперь, что в условиях теоремы 1 в (1) и (2) имеет место не только сходимость почти наверное, но и в среднем. (Этот результат будет использован далее в доказательстве теоремы 3.)

Т е о р е м а 2. *Пусть T — сохраняющее меру преобразование и $\xi = \xi(\omega)$ — случайная величина с $\mathsf{M}|\xi| < \infty$. Тогда*

Если к тому же T эргодично, то

$$\mathsf{M}\left|\frac{1}{n}\sum_{k=0}^{n-1}\xi(T^k\omega)-\mathsf{M}\xi\right|\to 0, \qquad n\to\infty. \tag{5}$$

Доказательство. Для всякого $\varepsilon>0$ можно найти такую ограниченную случайную величину η ($|\eta(\omega)|\leqslant M$), что $\mathsf{M}|\xi-\eta|\leqslant\varepsilon$. Тогда

$$\mathsf{M}\left|\frac{1}{n}\sum_{k=0}^{n-1}\xi(T^k\omega)-\mathsf{M}(\xi\mid\mathscr{J})\right|\leqslant\mathsf{M}\left|\frac{1}{n}\sum_{k=0}^{n-1}(\xi(T^k\omega)-\eta(T^k\omega))\right|+$$
$$+\mathsf{M}\left|\frac{1}{n}\sum_{k=0}^{n-1}(\eta(T^k\omega)-\mathsf{M}(\eta\mid\mathscr{J})\right|+\mathsf{M}\,|\,\mathsf{M}(\xi\mid\mathscr{J})-\mathsf{M}(\eta\mid\mathscr{J})\,|. \tag{6}$$

Поскольку $|\eta|\leqslant M$, то по теореме о мажорируемой сходимости и в силу (1) находим, что второй член в правой части (6) стремится к нулю при $n\to\infty$. Что же касается первого и третьего членов, то каждый из них меньше или равен ε. Поэтому для достаточно больших n левая часть в (6) меньше 2ε, что и доказывает (4). Наконец, если T эргодично, то (5) следует из (4) и того замечания, что $\mathsf{M}(\xi\mid\mathscr{J})=\mathsf{M}\xi$ (P-п. н.).

Теорема доказана.

3. Перейдем теперь к вопросу о справедливости эргодической теоремы для стационарных (в узком смысле) случайных последовательностей $\xi=(\xi_1, \xi_2, \ldots)$, заданных на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$. Вообще говоря, на $(\Omega, \mathscr{F}, \mathsf{P})$ может и не существовать сохраняющее меру преобразование, так что непосредственное применение теоремы 1 невозможно. Однако в § 1 уже отмечалось, что можно построить (координатное) вероятностное пространство $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde{\mathsf{P}})$, случайную последовательность $\tilde\xi=(\tilde\xi_1, \tilde\xi_2, \ldots)$ и сохраняющее меру преобразование $\tilde T$ такие, что $\tilde\xi_n(\tilde\omega)=\tilde\xi_1(\tilde T^{n-1}\tilde\omega)$ и по распределению ξ и $\tilde\xi$ совпадают. Поскольку такие свойства, как сходимость почти наверное и в среднем, определяются лишь распределениями вероятностей, то из сходимости $\frac{1}{n}\sum_{k=1}^{n}\tilde\xi_1(T^{k-1}\tilde\omega)$ (P-п. н. и в среднем) к некоторой случайной величине $\tilde\eta$ следует, что $\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega)$ также сходятся (P-п. н. и в среднем) к некоторой случайной величине η такой, что $\eta\overset{d}{=}\tilde\eta$. Из теоремы 1 следует, что если $\tilde{\mathsf{M}}|\tilde\xi_1|<\infty$, то $\tilde\eta=\tilde{\mathsf{M}}(\tilde\xi_1\mid\mathscr{J})$,

где $\mathscr{J}$ — совокупность инвариантных множеств ($\tilde{\mathsf{M}}$ — усреднение по мере $\tilde{\mathsf{P}}$). Опишем теперь структуру величины η.

Определение 1. Множество $A \in \mathscr{F}$ будем называть *инвариантным* по отношению к последовательности ξ, если найдется такое множество $B \in \mathscr{B}(R^\infty)$, что для любого $n \geqslant 1$

$$A = \{\omega: (\xi_n, \xi_{n+1}, \ldots) \in B\}.$$

Совокупность таких инвариантных множеств образует σ-алгебру, которую обозначим $\mathscr{J}_\xi$.

Определение 2. Стационарная последовательность ξ называется *эргодической*, если мера любого инвариантного множества принимает лишь два значения 0 или 1.

Покажем теперь, что исследуемая случайная величина η может быть взята равной $\mathsf{M}(\xi_1 \mid \mathscr{J}_\xi)$. В самом деле, пусть $A \in \mathscr{J}_\xi$. Тогда поскольку $\mathsf{M}\left|\frac{1}{n}\sum_{k=1}^{n-1}\xi_k - \eta\right| \to 0$, то

$$\frac{1}{n}\sum_{k=1}^{n}\int_A \xi_k \, d\mathsf{P} \to \int_A \eta \, d\mathsf{P}. \tag{7}$$

Пусть $B \in \mathscr{B}(R^\infty)$ таково, что $A = \{\omega: (\xi_k, \xi_{k+1}, \ldots) \in B\}$ для любого $k \geqslant 1$. Тогда в силу стационарности ξ

$$\int_A \xi_k \, d\mathsf{P} = \int_{\{\omega: (\xi_k, \xi_{k+1}, \ldots) \in B\}} \xi_k \, d\mathsf{P} = \int_{\{\omega: (\xi_1, \xi_2, \ldots) \in B\}} \xi_1 \, d\mathsf{P} = \int_A \xi_1 \, d\mathsf{P}.$$

Поэтому из (7) следует, что для любого $A \in \mathscr{J}_\xi$

$$\int_A \xi_1 \, d\mathsf{P} = \int_A \eta \, d\mathsf{P},$$

что означает (см. § 7 гл. II), что $\eta = \mathsf{M}(\xi_1 \mid \mathscr{J}_\xi)$. При этом $\mathsf{M}(\xi_1 \mid \mathscr{J}_\xi) = \mathsf{M}\xi_1$, если последовательность ξ является эргодической.

Итак, доказана следующая

Теорема 3 (эргодическая теорема). *Пусть $\xi = (\xi_1, \xi_2, \ldots)$ — стационарная (в узком смысле) случайная последовательность с $\mathsf{M}|\xi_1| < \infty$. Тогда ($\mathsf{P}$-п. н. и в среднем)*

$$\lim \frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) = \mathsf{M}(\xi_1 \mid \mathscr{J}_\xi).$$

Если к тому же ξ — эргодическая последовательность, то (P-п. н. и в среднем)

$$\lim \frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) = \mathsf{M}\xi_1.$$

ГЛАВА VI

СТАЦИОНАРНЫЕ (В ШИРОКОМ СМЫСЛЕ) СЛУЧАЙНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ. L^2-ТЕОРИЯ

§ 1. Спектральное представление ковариационной функции

1. Согласно определению, данному в предшествующей главе, случайная последовательность $\xi=(\xi_1, \xi_2, \ldots)$ называется стационарной в узком смысле, если для любого множества $B \in \mathscr{B}(R^\infty)$ и любого $n \geqslant 1$

$$\mathsf{P}\{(\xi_1, \xi_2, \ldots) \in B\} = \mathsf{P}\{(\xi_{n+1}, \xi_{n+2}, \ldots) \in B\}. \qquad (1)$$

Отсюда, в частности, вытекает, что если $\mathsf{M}\xi_1^2 < \infty$, то $\mathsf{M}\xi_n$ не зависит от n:

$$\mathsf{M}\xi_n = \mathsf{M}\xi_1, \qquad (2)$$

а ковариация $\operatorname{cov}(\xi_{n+m}\xi_n) = \mathsf{M}(\xi_{n+m} - \mathsf{M}\xi_{n+m})(\xi_n - \mathsf{M}\xi_n)$ зависит лишь от m:

$$\operatorname{cov}(\xi_{n+m}, \xi_n) = \operatorname{cov}(\xi_{1+m}, \xi_1). \qquad (3)$$

В настоящей главе будут исследоваться так называемые стационарные в широком смысле последовательности (с конечным вторым моментом), для которых условие (1) заменяется (более слабыми) условиями (2) и (3).

Рассматриваемые случайные величины ξ_n будут предполагаться определенными для $n \in \mathbb{Z} = \{0, \pm 1, \ldots\}$ и к тому же комплекснозначными. Последнее предположение не только не усложняет теорию, но и наоборот — делает ее более изящной. При этом, разумеется, результаты для действительных случайных величин легко могут быть получены в качестве частного случая из соответствующих результатов для комплексных величин.

Пусть $H^2 = H^2(\Omega, \mathscr{F}, \mathsf{P})$ — пространство (комплекснозначных) случайных величин $\xi = \alpha + i\beta$, α, $\beta \in R$, с $\mathsf{M}|\xi|^2 < \infty$, где $|\xi|^2 = \alpha^2 + \beta^2$. Если ξ, $\eta \in H^2$, то положим

$$(\xi, \eta) = \mathsf{M}\xi\bar{\eta}, \qquad (4)$$

где $\bar\eta = \alpha - i\beta$ — комплексно-сопряженная величина к $\eta = \alpha + i\beta$ и

$$\|\xi\| = (\xi, \xi)^{1/2}. \tag{5}$$

Как и для действительных случайных величин, пространство H^2 (точнее, пространство классов эквивалентных случайных величин; ср. с §§ 10 и 11 из гл. II) со скалярным произведением (ξ, η) и нормой $\|\xi\|$ является полным. В соответствии с терминологией функционального анализа пространство H^2 называется унитарным (иначе — комплексным) гильбертовым пространством (случайных величин, рассматриваемых на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$).

Если $\xi, \eta \in H^2$, то их *ковариацией* назовем величину

$$\operatorname{cov}(\xi, \eta) = \mathsf{M}(\xi - \mathsf{M}\xi)\overline{(\eta - \mathsf{M}\eta)}. \tag{6}$$

Из (4) и (6) следует, что если $\mathsf{M}\xi = \mathsf{M}\eta = 0$, то

$$\operatorname{cov}(\xi, \eta) = (\xi, \eta). \tag{7}$$

Определение 1. Последовательность комплексных случайных величин $\xi = (\xi_n)_{n \in \mathbb{Z}}$ с $\mathsf{M}|\xi_n|^2 < \infty$, $n \in \mathbb{Z}$, называется *стационарной* (*в широком смысле*), если для всех $n \in \mathbb{Z}$

$$\begin{aligned} \mathsf{M}\xi_n &= \mathsf{M}\xi_0, \\ \operatorname{cov}(\xi_{k+n}, \xi_k) &= \operatorname{cov}(\xi_n, \xi_0), \quad k \in \mathbb{Z}. \end{aligned} \tag{8}$$

Для простоты изложения в дальнейшем будем предполагать $\mathsf{M}\xi_0 = 0$. Это предположение не умаляет общности, но в то же самое время дает возможность (согласно (7)), отождествляя кова риацию со скалярным произведением, применять методы и результаты теории гильбертовых пространств.

Обозначим

$$R(n) = \operatorname{cov}(\xi_n, \xi_0), \quad n \in \mathbb{Z}, \tag{9}$$

и (в предположении $R(0) = \mathsf{M}|\xi_0|^2 \neq 0$)

$$\rho(n) = \frac{R(n)}{R(0)}, \quad n \in \mathbb{Z}. \tag{10}$$

Функцию $R(n)$ будем называть *ковариационной функцией*, а $\rho(n)$ — *корреляционной функцией* (стационарной в широком смысле) последовательности ξ.

Непосредственно из определения (9) следует, что ковариационная функция $R(n)$ является неотрицательно-определенной, т. е. для любых комплексных чисел $a_1, \ldots, a_m$ и $t_1, \ldots, t_m \in \mathbb{Z}$, $m \geq 1$

$$\sum_{i,j=1}^{m} a_i \bar a_j R(t_i - t_j) \geq 0. \tag{11}$$

В свою очередь отсюда (или непосредственно из (9)) нетрудно вывести (задача 1) следующие свойства ковариационной функции:

$$R(0) \geqslant 0, \quad R(-n) = \overline{R(n)}, \quad |R(n)| \leqslant R(0), \tag{12}$$
$$|R(n) - R(m)|^2 \leqslant 2R(0)[R(0) - \operatorname{Re} R(n-m)].$$

2. Приведем некоторые примеры стационарных последовательностей $\xi = (\xi_n)_{n \in \mathbb{Z}}$. (В дальнейшем слова «в широком смысле», а также указание на то, что $n \in \mathbb{Z}$, часто будут опускаться.)

Пример 1. Пусть $\xi_n = \xi_0 \cdot g(n)$, где $\mathsf{M}\xi_0 = 0$, $\mathsf{M}\xi_0^2 = 1$ и $g = g(n)$ — некоторая функция. Последовательность $\xi = (\xi_n)$ будет стационарной в том и только том случае, когда функция $g(k+n)\overline{g(k)}$ зависит лишь от n. Отсюда нетрудно вывести, что найдется такое λ, что

$$g(n) = g(0)\, e^{i\lambda n}.$$

Таким образом, последовательность случайных величин

$$\xi_n = \xi_0 \cdot g(0)\, e^{i\lambda n}$$

является стационарной с

$$R(n) = |g(0)|^2 e^{i\lambda n}.$$

В частности, случайная «константа» $\xi_n \equiv \xi_0$ образует стационарную последовательность.

Пример 2. *Почти периодическая последовательность*. Пусть

$$\xi_n = \sum_{k=1}^{N} z_k e^{i\lambda_k n}, \tag{13}$$

где $z_1, \ldots, z_N$ — ортогональные ($\mathsf{M}z_i\bar{z}_j = 0$, $i \neq j$) случайные величины с нулевыми средними и $\mathsf{M}|z_k|^2 = \sigma_k^2 > 0$; $-\pi \leqslant \lambda_k < \pi$, $k = 1, \ldots, N$; $\lambda_i \neq \lambda_j$, $i \neq j$. Последовательность $\xi = (\xi_n)$ является стационарной с

$$R(n) = \sum_{k=1}^{N} \sigma_k^2 e^{i\lambda_k n}. \tag{14}$$

В обобщение (13) предположим теперь, что

$$\xi_n = \sum_{k=-\infty}^{\infty} z_k e^{i\lambda_k n}, \tag{15}$$

где величины z_k, $k \in \mathbb{Z}$, обладают теми же свойствами, что и в (13). Если предположить, что $\sum_{k=-\infty}^{\infty} \sigma_k^2 < \infty$, то ряд в правой части формулы (15) сходится в среднеквадратическом смысле и

$$R(n) = \sum_{k=-\infty}^{\infty} \sigma_k^2 e^{i\lambda_k n}. \tag{16}$$

Введем функцию

$$F(\lambda) = \sum_{\{k:\ \lambda_k \leqslant \lambda\}} \sigma_k^2. \tag{17}$$

Тогда ковариационная функция (16) может быть записана в виде интеграла Лебега — Стилтьеса

$$R(n) = \int_{-\pi}^{\pi} e^{i\lambda n}\, dF(\lambda). \tag{18}$$

Стационарные последовательности (15) образованы как суммы «гармоник» $e^{i\lambda_k n}$ с «частотой» λ_k и случайными «амплитудами» z_k «интенсивности» $\sigma_k^2 = \mathsf{M}\,|z_k|^2$. Таким образом, знание функции $F(\lambda)$ дает исчерпывающую информацию о структуре «спектра» последовательности ξ, т. е. о величине интенсивностей, с которыми те или иные частоты входят в представление (15). Согласно (18) знание функции $F(\lambda)$ полностью определяет также и структуру ковариационной функции $R(n)$.

С точностью до постоянного множителя (невырожденная) функция $F(\lambda)$ является, очевидно, функцией распределения, причем в рассматриваемом примере эта функция кусочно-постоянна. Весьма примечательно, что ковариационная функция любой стационарной в широком смысле случайной последовательности может быть представлена (см. теорему в п. 3) в виде (18), где $F(\lambda)$ — некоторая (с точностью до нормировки) функция распределения, носитель которой сосредоточен на множестве $[-\pi, \pi)$, т. е. $F(\lambda) = 0$ для $\lambda < -\pi$ и $F(\lambda) = F(\pi)$ для $\lambda > \pi$.

Результат об интегральном представлении ковариационной функции, сопоставленный с (15) и (16), наводит на мысль, что произвольная стационарная последовательность также допускает «интегральное» представление. Так оно на самом деле и есть, что будет показано в § 3 с помощью так называемых стохастических интегралов по ортогональным стохастическим мерам (§ 2).

Пример 3. *Белый шум.* Пусть $\varepsilon = (\varepsilon_n)$ — последовательность ортонормированных случайных величин, $\mathsf{M}\varepsilon_n = 0$, $\mathsf{M}\varepsilon_i\varepsilon_j = \delta_{ij}$, где δ_{ij} — символ Кронекера. Понятно, что такая последовательность является стационарной и

Отметим, что функция $R(n)$ может быть представлена в виде

$$R(n) = \int_{-\pi}^{\pi} e^{i\lambda n}\, dF(\lambda), \tag{19}$$

где

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)\, d\nu, \quad f(\lambda) = \frac{1}{2\pi}, \quad -\pi \leqslant \lambda < \pi. \tag{20}$$

Сравнение «спектральных» функций (17) и (20) показывает, что если в примере 2 «спектр» был дискретным, то в настоящем примере он оказался абсолютно непрерывным с постоянной «спектральной» плотностью $f(\lambda) \equiv 1/2\pi$. В этом смысле можно сказать, что последовательность $\varepsilon = (\varepsilon_n)$ «составлена из гармоник, интенсивность которых одна и та же». Именно это обстоятельство и послужило поводом называть последовательность $\varepsilon = (\varepsilon_n)$ «белым шумом» по аналогии с белым цветом, составленным из различных цветов одной и той же интенсивности.

Пример 4. *Последовательности скользящего среднего.* Отправляясь от белого шума $\varepsilon = (\varepsilon_n)$, введенного в примере 3, образуем новую последовательность

$$\xi_n = \sum_{k=-\infty}^{\infty} a_k \varepsilon_{n-k}, \tag{21}$$

где a_k — комплексные числа такие, что $\sum_{k=-\infty}^{\infty} |a_k|^2 < \infty$. В силу равенства Парсеваля

$$\operatorname{cov}(\xi_{n+m}, \xi_m) = \operatorname{cov}(\xi_n, \xi_0) = \sum_{k=-\infty}^{\infty} a_{n+k} \bar{a}_k,$$

так что $\xi = (\xi_k)$ является стационарной последовательностью, которую принято называть последовательностью, образованной с помощью *(двустороннего) скользящего среднего* из последовательности $\varepsilon = (\varepsilon_k)$.

В том частном случае, когда все a_k с отрицательными индексами равны нулю, т. е.

$$\xi_n = \sum_{k=-\infty}^{\infty} a_k \varepsilon_{n-k},$$

последовательность $\xi = (\xi_n)$ называют последовательностью *одностороннего скользящего среднего*. Если к тому же все $a_k = 0$ при $k > p$, т. е. если

$$\xi_n = a_0 \varepsilon_n + a_1 \varepsilon_{n-1} + \ldots + a_p \varepsilon_{n-p}, \tag{22}$$

то $\xi=(\xi_n)$ называется последовательностью *скользящего среднего порядка* p.

Можно показать (задача 5), что для последовательности (22) ковариационная функция $R(n)$ имеет вид $R(n)=\int_{-\pi}^{\pi} e^{i\lambda n} f(\lambda)\,d\lambda$, где спектральная плотность равна

$$f(\lambda)=\frac{1}{2\pi}\,|P(e^{-i\lambda})|^2 \tag{23}$$

с

$$P(z)=a_0+a_1z+\ldots+a_pz^p.$$

Пример 5. *Авторегрессионная схема*. Пусть снова $\varepsilon=(\varepsilon_n)$ — белый шум. Будем говорить, что случайная последовательность $\xi=(\xi_n)$ подчиняется *авторегрессивной схеме* порядка q, если

$$\xi_n+b_1\xi_{n-1}+\ldots+b_q\xi_{n-q}=\varepsilon_n. \tag{24}$$

При каких условиях на коэффициенты $b_1, \ldots, b_q$ можно утверждать, что уравнение (24) имеет стационарное решение? Чтобы ответить на этот вопрос, рассмотрим сначала случай $q=1$:

$$\xi_n=\alpha\xi_{n-1}+\varepsilon_n, \tag{25}$$

где $\alpha=-b_1$. Если $|\alpha|<1$, то нетрудно проверить, что стационарная последовательность $\tilde{\xi}=(\tilde{\xi}_n)$ с

$$\tilde{\xi}_n=\sum_{j=0}^{\infty}\alpha^j\varepsilon_{n-j} \tag{26}$$

является решением уравнения (25). (Ряд в правой части (26) сходится в среднеквадратическом смысле.) Покажем теперь, что в классе стационарных последовательностей $\xi=(\xi_n)$ (с конечным вторым моментом) это решение является единственным. В самом деле, из (25) последовательными итерациями находим, что

$$\xi_n=\alpha\xi_{n-1}+\varepsilon_n=\alpha[\alpha\xi_{n-2}+\varepsilon_{n-1}]+\varepsilon_n=\ldots=\alpha^k\xi_{n-k}+\sum_{j=0}^{k-1}\alpha^j\varepsilon_{n-j}.$$

Отсюда следует, что

$$\mathsf{M}\left[\xi_n-\sum_{j=0}^{k-1}\alpha^j\varepsilon_{n-j}\right]^2=\mathsf{M}[\alpha_k\xi_{n-k}]^2=\alpha^{2k}\mathsf{M}\xi_{n-k}^2=$$
$$=\alpha^{2k}\mathsf{M}\xi_0^2\to 0,$$
$$k\to\infty.$$

Таким образом, при $|\alpha|<1$ стационарное решение уравнения (25) существует и представляется в виде одностороннего скользящего среднего (26).

Аналогичный результат имеет место и в случае произвольного $q>1$: если все нули полинома

$$Q(z)=1+b_1z+\ldots+b_qz^q \tag{27}$$

лежат вне единичного круга, то уравнение авторегрессии (24) имеет, и притом единственное, стационарное решение, представимое в виде одностороннего скользящего среднего (задача 2). При этом ковариационная функция $R(n)$ представима (задача 5) в виде

$$R(n)=\int_{-\pi}^{\pi} e^{i\lambda n}\,dF(\lambda),\quad F(\lambda)=\int_{-\pi}^{\lambda} f(\nu)\,d\nu, \tag{28}$$

где

$$f(\lambda)=\frac{1}{2\pi}\cdot\frac{1}{|Q(e^{-i\lambda})|^2}. \tag{29}$$

В частном случае $q=1$ из (25) легко находим, что $\mathsf{M}\xi_0=0$,

$$\mathsf{M}\xi_0^2=\frac{1}{1-|\alpha|^2}$$

и

$$R(n)=\frac{\alpha^n}{1-|\alpha|^2},\qquad n\geqslant 0$$

(для $n<0$ $R(n)=\overline{R(-n)}$). При этом

$$f(\lambda)=\frac{1}{2\pi}\cdot\frac{1}{|1-\alpha e^{-i\lambda}|^2}.$$

Пример 6. Этот пример иллюстрирует возникновение авторегрессионных схем при построении вероятностных моделей в гидрологии. Рассмотрим некоторый водный бассейн (например, Каспийское море) и постараемся построить вероятностную модель, описывающую отклонения уровня в этом бассейне от среднего значения, вызванные колебаниями в стоке и испарением с водной поверхности.

Если за единицу измерения взять один год и через H_n обозначить уровень в бассейне в n-й год, то получим следующее *уравнение баланса*:

$$H_{n+1}=H_n-KS(H_n)+\Sigma_{n+1}, \tag{30}$$

где через Σ_{n+1} обозначена величина стока в $(n+1)$-й год, $S(H)$ — величина поверхности водного бассейна на уровне H, а K — коэффициент испарения.

Обозначим через $\xi_n=H_n-\overline{H}$ отклонение от среднего уровня $\overline{H}$ (который находится по результатам многолетних наблюдений) и предположим, что $S(H)=S(\overline{H})+c(H-\overline{H})$. Тогда из уравнения баланса следует, что величины ξ_n подчиняются уравнениям

$$\xi_{n+1} = \alpha\xi_n + \varepsilon_{n+1} \tag{31}$$

с $\alpha = 1 - cK$, $\varepsilon_n = \sum_n - KS(\bar{H})$. Случайные величины ε_n естественно считать имеющими нулевые средние и в первом приближении некоррелированными и одинаково распределенными. Тогда, как это было показано в примере 5, уравнение (31) (при $|\alpha| < 1$) имеет единственное стационарное решение, которое следует считать решением, описывающим установившийся (с годами) режим колебаний уровня в рассматриваемом бассейне.

В качестве тех практических выводов, которые можно сделать из (теоретической) модели (31), укажем на возможность построения прогноза отклонений уровня на следующий год по результатам наблюдений за настоящий и предшествующий годы. А именно, оказывается (см. далее пример 2 в § 6), что оптимальной (в среднеквадратическом смысле) линейной оценкой величины ξ_{n+1} по значениям $\ldots, \xi_{n-1}, \xi_n$ служит просто величина $\alpha\xi_n$.

Пример 7. *Смешанная модель авторегрессии и скользящего среднего.* Если предположить, что в правой части уравнения (24) вместо ε_n стоит величина $a_0\varepsilon_n + a_1\varepsilon_{n-1} + \ldots + a_{-p}\varepsilon_{n-p}$, то получим так называемую смешанную модель авторегрессии и скользящего среднего порядка (p, q):

$$\xi_n + b_1\xi_{n-1} + \ldots + b_q\xi_{n-q} = a_0\varepsilon_n + a_1\varepsilon_{n-1} + \ldots + a_p\varepsilon_{n-p}. \tag{32}$$

При тех же предположениях относительно нулей полинома $Q(z)$, что и в примере 5, далее показывается (следствие 2 к теореме 3 § 3), что уравнение (32) имеет стационарное решение $\xi = (\xi_n)$, для которого ковариационная функция равна $R(n) = \int_{-\pi}^{\pi} e^{i\lambda n}\, dF(\lambda)$ с $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)\, d\nu$, где

$$f(\lambda) = \frac{1}{2\pi} \cdot \left| \frac{P(e^{-i\lambda})}{Q(e^{-i\lambda})} \right|^2.$$

3. Теорема (Герглотц). *Пусть $R(n)$ — ковариационная функция стационарной (в широком смысле) случайной последовательности с нулевым средним. Тогда на $([-\pi, \pi), \mathscr{B}([-\pi, \pi)))$ найдется такая конечная мера $F = F(B)$, $B \in \mathscr{B}([-\pi, \pi))$, что для любого $n \in \mathbb{Z}$*

$$R(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda). \tag{33}$$

Доказательство. Положим для $N \geqslant 1$ и $\lambda \in [-\pi, \pi]$

$$f_N(\lambda) = \frac{1}{2\pi N} \sum_{k=1}^{N} \sum_{l=1}^{N} R(k-l)\, e^{-ik\lambda} e^{il\lambda}. \tag{34}$$

В силу неотрицательной определенности $R(n)$ функция $f_N(\lambda)$ неотрицательна. Поскольку число тех пар (k, l), для которых $k - l = m$, есть $N - |m|$, то

$$f_N(\lambda) = \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) R(m)\, e^{-im\lambda}. \tag{35}$$

Пусть

$$F_N(B) = \int_B f_N(\lambda)\, d\lambda, \qquad B \in \mathscr{B}([-\pi, \pi)).$$

Тогда

$$\int_{-\pi}^{\pi} e^{i\lambda n} F_N(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} f_N(\lambda)\, d\lambda = \begin{cases} \left(1 - \frac{|n|}{N}\right) R(n), & |n| < N, \\ 0, & |n| \geqslant N. \end{cases} \tag{36}$$

Меры F_N, $N \geqslant 1$, сосредоточены на интервале $[-\pi, \pi]$ и $F_N([-\pi, \pi]) = R(0) < \infty$ для любого $N \geqslant 1$. Следовательно, семейство мер $\{F_N\}$, $N \geqslant 1$, плотно, и по теореме Прохорова (теорема 1 § 2 гл. III) существуют подпоследовательность $\{N_k\} \subseteq \{N\}$ и мера F такие, что $F_{N_k} \xrightarrow{w} F$. (Понятия плотности, относительной компактности, слабой сходимости и теорема Прохорова очевидным образом с вероятностных мер переносятся на любые конечные меры.)

Тогда из (36) следует, что

$$\int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda) = \lim_{N_k \to \infty} \int_{-\pi}^{\pi} e^{i\lambda n} F_{N_k}(d\lambda) = R(n).$$

Построенная мера F сосредоточена на интервале $[-\pi, \pi]$. Не изменяя интеграла $\int_{-\infty}^{\infty} e^{i\lambda n} F(d\lambda)$, можно переопределить меру F, перенеся «массу» $F(\{\pi\})$, сосредоточенную в точке π, в точку $-\pi$. Так полученная новая мера (обозначим ее снова через F) будет уже сосредоточенной на интервале $[-\pi, \pi)$.

Теорема доказана.

Замечание 1. Меру $F = F(B)$, участвующую в представлении (33), называют спектральной мерой, а функцию $F(\lambda) = F([-\pi, \lambda])$ — спектральной функцией стационарной последовательности с ковариационной функцией $R(n)$.

В рассмотренном выше примере 2 спектральная мера оказалась дискретной (сосредоточенной в точках λ_k, $k=0, \pm 1, \ldots$). В примерах 3—6 спектральная мера абсолютно непрерывна.

Замечание 2. Спектральная мера F однозначно определяется по ковариационной функции. В самом деле, пусть F_1 и F_2 — две спектральные меры и

$$\int_{-\pi}^{\pi} e^{i\lambda n} F_1(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} F_2(d\lambda), \qquad n \in \mathbb{Z}.$$

Поскольку любая ограниченная непрерывная функция $g(\lambda)$ может быть равномерно приближена на $[-\pi, \pi)$ тригонометрическими полиномами, то

$$\int_{-\pi}^{\pi} g(\lambda) F_1(d\lambda) = \int_{-\pi}^{\pi} g(\lambda) F_2(d\lambda),$$

откуда (ср. с доказательством в теореме 2 § 12 гл. II) следует, что $F_1(B) = F_2(B)$ для любых $B \in \mathscr{B}([-\pi, \pi))$.

Замечание 3. Если $\xi = (\xi_n)$ — стационарная последовательность, состоящая из действительных случайных величин ξ_n, то

$$R(n) = \int_{-\pi}^{\pi} \cos \lambda n F(d\lambda).$$

4. Задачи.

1. Вывести свойства (12) из (11).

2. Доказать, что уравнение авторегрессии (24) имеет стационарное решение, если все нули полинома $Q(z)$, определенного в (27), лежат вне единичного круга.

3. Доказать, что ковариационная функция (28) допускает представление (29) со спектральной плотностью, задаваемой формулой (30).

4. Показать, что последовательность $\xi = (\xi_n)$ случайных величин

$$\xi_n = \sum_{k=1}^{\infty} (\alpha_k \sin \lambda_k n + \beta_k \cos \lambda_k n)$$

с действительными случайными величинами α_k и β_k может быть представлена в виде

$$\xi_n = \sum_{k=-\infty}^{\infty} z_k e^{i\lambda_k n}$$

с $z_k = \frac{1}{2}(\beta_k - i\alpha_k)$ для $k \geqslant 0$ и $z_k = \bar{z}_{-k}$, $\lambda_k = -\lambda_{-k}$ для $k < 0$.

5. Показать, что для последовательностей (22) и (24) их спектральные функции имеют плотности, задаваемые соответственно формулами (23) и (29).

6. Показать, что если $\sum |R(n)| < \infty$, то спектральная функция $F(\lambda)$ имеет плотность $f(\lambda)$, определяемую формулой

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-i\lambda n} R(n).$$

§ 2. Ортогональные стохастические меры и стохастические интегралы

1. Как уже отмечалось в § 1, интегральное представление ковариационной функции и пример стационарной последовательности

$$\xi_n = \sum_{k=-\infty}^{\infty} z_k e^{i\lambda_k n} \tag{1}$$

с попарно ортогональными случайными величинами z_k, $k \in \mathbb{Z}$, наводит на мысль о возможности представления произвольной стационарной последовательности в виде соответствующего интегрального обобщения суммы (1).

Если положить

$$Z(\lambda) = \sum_{\{k:\ \lambda_k \leq \lambda\}} z_k, \tag{2}$$

то (1) запишется в виде

$$\xi_n = \sum_{k=-\infty}^{\infty} e^{i\lambda_k n} \Delta Z(\lambda_k), \tag{3}$$

где $\Delta Z(\lambda_k) \equiv Z(\lambda_k) - Z(\lambda_k -) = z_k$.

Правая часть (3) напоминает интегральную сумму для «интеграла типа Римана — Стилтьеса» $\int_{-\pi}^{\pi} e^{i\lambda n}\, dZ(\lambda)$. Однако в рассматриваемом нами случае функция $Z(\lambda)$ является случайной (зависящей также и от ω). При этом выясняется, что для интегрального представления произвольной стационарной последовательности приходится привлекать к рассмотрению и такие функции $Z(\lambda)$, которые при каждом ω имеют неограниченную вариацию. Поэтому простое понимание интеграла $\int_{-\pi}^{\pi} e^{i\lambda n}\, dZ(\lambda)$ как интеграла Римана — Стилтьеса для каждого ω становится неприемлемым.

2. По аналогии с общей концепцией интегралов Лебега, Лебега — Стилтьеса и Римана — Стилтьеса (§ 6 гл. II) рассмотрение стохастического случая начнем с определения стохастической меры.

Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство, E — некоторое множество с алгеброй $\mathscr{E}_0$ его подмножеств и σ-алгеброй $\mathscr{E}$.

Определение 1. Комплекснозначная функция $Z(\Delta) = Z(\omega; \Delta)$, определенная для $\omega \in \Omega$ и $\Delta \in \mathscr{E}_0$, называется *конечно-аддитивной стохастической мерой*, если:

1) для любого $\Delta \in \mathscr{E}_0$ $\mathsf{M}|Z(\Delta)|^2 < \infty$;

2) для любых двух непересекающихся множеств Δ_1 и Δ_2 из $\mathscr{E}_0$

$$Z(\Delta_1 + \Delta_2) = Z(\Delta_1) + Z(\Delta_2) \quad (\mathsf{P}\text{-п. н.}). \tag{4}$$

Определение 2. Конечно-аддитивная стохастическая мера $Z(\Delta)$ называется *элементарной стохастической мерой*, если для любых непересекающихся множеств $\Delta_1, \Delta_2, \ldots$ из $\mathscr{E}_0$ таких, что $\Delta = \sum_{k=1}^{\infty} \Delta_k \in \mathscr{E}_0$,

$$\mathsf{M}\left| Z(\Delta) - \sum_{k=1}^{n} Z(\Delta_k) \right|^2 \to 0, \quad n \to \infty. \tag{5}$$

Замечание 1. В данном определении элементарной стохастической меры, заданной на множествах из $\mathscr{E}_0$, предполагается, что ее значения принадлежат гильбертову пространству $H^2 = H^2(\Omega, \mathscr{F}, \mathsf{P})$, а счетная аддитивность выполнена в среднеквадратическом смысле (5). Существуют и другие определения стохастических мер, в которых отсутствует требование существования второго момента, а счетная аддитивность понимается, например, в смысле сходимости по вероятности или с вероятностью единица.

Замечание 2. По аналогии с неслучайными мерами можно показать, что для конечно-аддитивных стохастических мер условие (5) счетной аддитивности (в среднеквадратическом смысле) эквивалентно непрерывности (в среднеквадратическом смысле) в «нуле»:

$$\mathsf{M}|Z(\Delta)|^2 \to 0, \quad \Delta_n \downarrow \varnothing, \quad \Delta_n \in \mathscr{E}_0. \tag{6}$$

В классе элементарных стохастических мер особо важны меры, являющиеся ортогональными в смысле следующего определения.

Определение 3. Элементарная стохастическая мера $Z(\Delta)$, $\Delta \in \mathscr{E}_0$, называется *ортогональной* (или *мерой с ортогональными значениями*), если для любых двух непересекающихся множеств Δ_1 и Δ_2 из $\mathscr{E}_0$

$$\mathsf{M}Z(\Delta_1)\overline{Z(\Delta_2)} = 0, \tag{7}$$

или, что эквивалентно, если для любых Δ_1 и Δ_2 из $\mathscr{E}_0$

$$\mathsf{M}Z(\Delta_1)\bar{Z}(\Delta_2) = \mathsf{M}|Z(\Delta_1 \cap \Delta_2)|^2. \tag{8}$$

Обозначим

$$m(\Delta) = \mathsf{M}|Z(\Delta)|^2, \quad \Delta \in \mathscr{E}_0. \tag{9}$$

Для элементарных ортогональных стохастических мер функция множеств $m=m(\Delta)$, $\Delta\in\mathscr{E}_0$, является, как легко видеть, конечной мерой и, следовательно, по теореме Каратеодори (§ 3 гл. II) она может быть продолжена на $(E, \mathscr{E})$. Так полученную меру будем снова обозначать через $m=m(\Delta)$ и называть *структурной функцией* (элементарной ортогональной стохастической меры $Z=Z(\Delta)$, $\Delta\in\mathscr{E}_0$).

Теперь естественным образом возникает следующий вопрос: раз функция множеств $m=m(\Delta)$, определенная на $(E, \mathscr{E}_0)$, допускает продолжение на $(E, \mathscr{E})$, где $\mathscr{E}=\sigma(\mathscr{E}_0)$, то нельзя ли элементарную ортогональную стохастическую меру $Z=Z(\Delta)$, $\Delta\in\mathscr{E}_0$, продолжить на множества Δ из $\mathscr{E}$, причем так, чтобы $\mathsf{M}|Z(\Delta)|^2=m(\Delta)$, $\Delta\in\mathscr{E}$.

Ответ на этот вопрос утвердительный, что вытекает из нижеследующих конструкций, приводящих в то же самое время и к построению стохастического интеграла, необходимого для интегрального представления стационарных последовательностей.

3. Итак, пусть $Z=Z(\Delta)$ — элементарная ортогональная стохастическая мера, $\Delta\in\mathscr{E}_0$, со структурной функцией $m=m(\Delta)$, $\Delta\in\mathscr{E}$. Для каждой функции

$$f(\lambda)=\sum f_k I_{\Delta_k},\quad \Delta_k\in\mathscr{E}_0, \tag{10}$$

принимающей лишь конечное число различных (комплексных) значений, определим случайную величину

$$\mathscr{I}(f)=\sum f_k Z(\Delta_k).$$

Пусть $L^2=L^2(E, \mathscr{E}, m)$ — гильбертово пространство комплекснозначных функций со скалярным произведением

$$\langle f, g\rangle=\int_E f(\lambda)\bar{g}(\lambda)\,m(d\lambda)$$

и нормой $\|f\|=\langle f, f\rangle^{1/2}$, а $H^2=H^2(\Omega, \mathscr{F}, \mathsf{P})$ — гильбертово пространство комплекснозначных случайных величин со скалярным произведением

$$(\xi, \eta)=\mathsf{M}\xi\bar{\eta}$$

и нормой $\|\xi\|=(\xi, \xi)^{1/2}$.

Тогда очевидным образом для любых двух функций f и g вида (10)

$$(\mathscr{I}(f), \mathscr{I}(g))=\langle f, g\rangle$$

и

$$\|\mathscr{I}(f)\|^2=\|f\|^2=\int_E |f(\lambda)|^2\,m(d\lambda).$$

Пусть теперь $f \in L^2$ и $\{f_n\}$ — функции типа (10) такие, что $\|f-f_n\| \to 0$, $n \to \infty$ (существование таких функций следует из задачи 1). Поэтому

$$\|\mathcal{J}(f_n)-\mathcal{J}(f_m)\| = \|f_n - f_m\| \to 0, \qquad n, m \to \infty.$$

Следовательно, последовательность $\{\mathcal{J}(f_n)\}$ фундаментальна в среднеквадратическом смысле, и в силу теоремы 7 из § 10 гл. II найдется случайная величина (обозначим ее $\mathcal{J}(f)$) такая, что $\mathcal{J}(f) \in H^2$ и $\|\mathcal{J}(f_n)-\mathcal{J}(f)\| \to 0$, $n \to \infty$.

Так построенная случайная величина $\mathcal{J}(f)$ определяется однозначно (с точностью до стохастической эквивалентности) и не зависит от выбора аппроксимирующей последовательности $\{f_n\}$. Назовем ее *стохастическим интегралом* от функции $f \in L^2$ по элементарной ортогональной стохастической мере Z и будем записывать

$$\mathcal{J}(f) = \int_E f(\lambda)\, Z(d\lambda).$$

Отметим следующие основные свойства стохастического интеграла $\mathcal{J}(f)$, непосредственно вытекающие из его конструкции (задача 1). Пусть функции $g, f, f_n \in L^2$. Тогда

$$(\mathcal{J}(f), \mathcal{J}(g)) = \langle f, g \rangle; \tag{11}$$

$$\|\mathcal{J}(f)\| = \|f\|; \tag{12}$$

$$\mathcal{J}(af+bg) = a\mathcal{J}(f) + b\mathcal{J}(g) \qquad (\mathsf{P}\text{-п. н.}), \tag{13}$$

где a и b — константы;

$$\|\mathcal{J}(f_n)-\mathcal{J}(f)\| \to 0, \tag{14}$$

если $\|f_n - f\| \to 0$, $n \to \infty$.

4. Используем определенный выше стохастический интеграл для *продолжения* элементарной ортогональной стохастической меры $Z(\Delta)$, $\Delta \in \mathscr{E}_0$, на множества из $\mathscr{E} = \sigma(\mathscr{E}_0)$.

Поскольку мера m предполагается конечной, то функция $I_\Delta = I_\Delta(\lambda) \in L^2$ для всякого $\Delta \in \mathscr{E}$. Обозначим $\tilde{Z}(\Delta) = \mathcal{J}(I_\Delta)$. Ясно, что для $\Delta \in \mathscr{E}_0$ $\tilde{Z}(\Delta) = Z(\Delta)$. Из (13) следует, что если $\Delta_1 \cap \Delta_2 = \varnothing$, $\Delta_1, \Delta_2 \in \mathscr{E}$, то

$$\tilde{Z}(\Delta_1 + \Delta_2) = \tilde{Z}(\Delta_1) + \tilde{Z}(\Delta_2) \quad (\mathsf{P}\text{-п. н.}),$$

а из (12) вытекает, что

$$\mathsf{M}\,|\tilde{Z}(\Delta)|^2 = m(\Delta), \qquad \Delta \in \mathscr{E}.$$

Покажем, что случайная функция множеств $\tilde{Z}(\Delta)$, $\Delta \in \mathscr{E}$, является счетно-аддитивной в среднеквадратическом смысле.

В самом деле, пусть $\Delta_k \in \mathscr{E}$ и $\Delta = \sum_{k=1}^{\infty} \Delta_k$. Тогда

$$\tilde{Z}(\Delta) - \sum_{k=1}^{n} \tilde{Z}(\Delta_k) = \mathscr{I}(g_n),$$

где

$$g_n(\lambda) = I_\Delta(\lambda) - \sum_{k=1}^{n} I_{\Delta_k}(\lambda) = I_{\sum\limits_{k=n+1}^{\infty} \Delta_k}(\lambda).$$

Но

$$\mathsf{M}\,|\mathscr{I}(g_n)|^2 = \|g_n\|^2 = m\left(\sum_{k=n+1}^{\infty} \Delta_k\right) \downarrow 0, \qquad n \to \infty,$$

т. е.

$$\mathsf{M}\,\left|\tilde{Z}(\Delta) - \sum_{k=1}^{n} \tilde{Z}(\Delta_k)\right|^2 \to 0, \quad n \to \infty.$$

Из (11) следует также, что для $\Delta_1 \cap \Delta_2 = \varnothing$, Δ_1, $\Delta_2 \in \mathscr{E}$,

$$\mathsf{M}\tilde{Z}(\Delta_1)\bar{\tilde{Z}}(\Delta_2) = 0.$$

Итак, построенная случайная функция $\tilde{Z}(\Delta)$, определенная на множествах $\Delta \in \mathscr{E}$, является счетно-аддитивной в среднеквадратическом смысле и на множествах $\Delta \in \mathscr{E}_0$ совпадает с $Z(\Delta)$. Будем называть $\tilde{Z}(\Delta)$, $\Delta \in \mathscr{E}$, ортогональной стохастической мерой (являющейся продолжением элементарной ортогональной стохастической меры $Z(\Delta)$) со структурной функцией $m(\Delta)$, $\Delta \in \mathscr{E}$, а определенный выше интеграл $\mathscr{I}(f) = \int_E f(\lambda)\,\tilde{Z}(d\lambda)$ — стохастическим интегралом по этой мере.

5. Обратимся теперь к наиболее важному для наших целей случаю $(E, \mathscr{E}) = (R, \mathscr{B}(R))$. Как известно (§ 3 гл. II), всякая конечная мера $m = m(\Delta)$ на $(R, \mathscr{B}(R))$ находится во взаимно-однозначном соответствии с некоторой (обобщенной) функцией распределения $G = G(x)$, причем $m(a, b] = G(b) - G(a)$.

Оказывается, нечто подобное справедливо и для ортогональных стохастических мер. Введем

Определение 4. Совокупность (комплекснозначных) случайных величин $\{Z_\lambda\}$, $\lambda \in R$, заданных на $(\Omega, \mathscr{F}, \mathsf{P})$, назовем *случайным процессом с ортогональными приращениями*, если

1) $\mathsf{M}\,|Z_\lambda|^2 < \infty$, $\lambda \in R$;

2) для каждого $\lambda \in R$

$$\mathsf{M}\,|Z_\lambda - Z_{\lambda_n}|^2 \to 0, \quad \lambda_n \downarrow \lambda, \quad \lambda_n \in R;$$

3) для любых $\lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$

$$\mathsf{M}(Z_{\lambda_4} - Z_{\lambda_3})\overline{(Z_{\lambda_2} - Z_{\lambda_1})} = 0.$$

Условие (3) является условием ортогональности приращений. Условие 1) означает, что $Z_\lambda \in H^2$. Наконец, условие 2) носит технический характер и является требованием непрерывности справа (в среднеквадратическом смысле) в каждой точке $\lambda \in R$.

Пусть $Z = Z(\Delta)$ — ортогональная стохастическая мера со структурной функцией $m = m(\Delta)$, являющейся конечной мерой с (обобщенной) функцией распределения $G(\lambda)$. Положим

$$Z_\lambda = Z(-\infty, \lambda].$$

Тогда $\mathsf{M}|Z_\lambda|^2 = m(-\infty, \lambda] = G(\lambda) < \infty$, $\mathsf{M}|Z_\lambda - Z_{\lambda_n}|^2 = m(\lambda_n, \lambda] \downarrow 0$, $\lambda_n \downarrow \lambda$, и, очевидно, выполнено также условие 3). Таким образом, построенный процесс $\{Z_\lambda\}$ является процессом *с ортогональными приращениями.*

С другой стороны, если $\{Z_\lambda\}$ — такой процесс с $\mathsf{M}|Z_\lambda|^2 = G(\lambda)$, $G(-\infty) = 0$, $G(+\infty) < \infty$, то положим для $\Delta = (a, b]$

$$Z(\Delta) = Z(b) - Z(a).$$

Пусть $\mathscr{E}_0$ — алгебра множеств $\Delta = \sum_{k=1}^{n} (a_k, b_k]$ и $Z(\Delta) = \sum_{k=1}^{n} Z(a_k, b_k]$. Ясно, что

$$\mathsf{M}|Z(\Delta)|^2 = m(\Delta),$$

где $m(\Delta) = \sum_{k=1}^{n} [G(b_k) - G(a_k)]$, и для непересекающихся интервалов $\Delta_1 = (a_1, b_1]$ и $\Delta_2 = (a_2, b_2]$

$$\mathsf{M}Z(\Delta_1)\bar{Z}(\Delta_2) = 0.$$

Таким образом, $Z = Z(\Delta)$, $\Delta \in \mathscr{E}_0$, является элементарной стохастической мерой с ортогональными значениями. Функция множеств $m = m(\Delta)$, $\Delta \in \mathscr{E}_0$, однозначно продолжается до меры на $\mathscr{E} = \mathscr{B}(R)$, и из предшествующих конструкций следует, что тогда $Z = Z(\Delta)$, $\Delta \in \mathscr{E}_0$, также можно продолжить на множестве $\Delta \in \mathscr{E}$, где $\mathscr{E} = \mathscr{B}(R)$, при этом $\mathsf{M}|Z(\Delta)|^2 = m(\Delta)$, $\Delta \in \mathscr{B}(R)$.

Тем самым между процессами $\{Z_\lambda\}$, $\lambda \in R$, с ортогональными приращениями и $\mathsf{M}|Z_\lambda|^2 = G(\lambda)$, $G(-\infty) = 0$, $G(+\infty) < \infty$, и ортогональными стохастическими мерами $Z = Z(\Delta)$, $\Delta \in \mathscr{B}(R)$, со структурной функцией $m = m(\Delta)$ существует взаимно однозначное соответствие, при котором

$$Z_\lambda = Z(-\infty, \lambda], \quad G(\lambda) = m(-\infty, \lambda]$$

и

$$Z(a, b] = Z_b - Z_a, \quad m(a, b] = G_b - G_a.$$

По аналогии с обозначениями, принятыми в теории интеграла Римана — Стилтьеса, под стохастическим интегралом $\int_R f(\lambda)\, dZ_\lambda$, где $\{Z_\lambda\}$ — некоторый процесс с ортогональными приращениями, понимается стохастический интеграл $\int\limits_R f(\lambda)\,Z(d\lambda)$ по соответствующей этому процессу ортогональной стохастической мере.

6. Задачи.

1. Доказать эквивалентность условий (5) и (6).

2. Пусть функция $f\in L^2$. Используя результаты гл. II (теорема 1 в § 4, следствие к теореме 3 § 6 и задачу 9 в § 3), доказать, что найдется последовательность функций f_n вида (10) таких, что $\|f-f_n\|\to 0$, $n\to\infty$.

3. Установить справедливость следующих свойств ортогональной стохастической меры $Z(\Delta)$ со структурной функцией $m(\Delta)$:

$$\mathsf{M}\,|Z(\Delta_1)-Z(\Delta_2)|^2=m(\Delta_1\triangle\Delta_2),$$

$$Z(\Delta_1\setminus\Delta_2)=Z(\Delta_1)-Z(\Delta_1\cap\Delta_2)\quad(\mathsf{P}\text{-п. н.}),$$

$$Z(\Delta_1\triangle\Delta_2)=Z(\Delta_1)+Z(\Delta_2)-2Z(\Delta_1\cap\Delta_2)\quad(\mathsf{P}\text{-п. н.}).$$

§ 3. Спектральное представление стационарных (в широком смысле) последовательностей

1. Если $\xi=(\xi_n)$ — стационарная последовательность с $\mathsf{M}\xi_n=0$, $n\in\mathbb{Z}$, то, согласно теореме из § 1, найдется такая конечная мера $F=F(\Delta)$ на $([-\pi,\pi),\ \mathscr{B}([-\pi,\pi)))$, что ковариационная функция $R(n)=\operatorname{cov}(\xi_{k+n},\xi_k)$ допускает спектральное представление

$$R(n)=\int\limits_{-\pi}^{\pi}e^{i\lambda n}F(d\lambda). \tag{1}$$

Следующий результат дает соответствующее спектральное представление самой последовательности $\xi=(\xi_n)$, $n\in\mathbb{Z}$.

Теорема 1. *Существует такая ортогональная стохастическая мера $Z=Z(\Delta)$, $\Delta\in\mathscr{B}([-\pi,\pi))$, что для каждого $n\in\mathbb{Z}$* (P-п. н.)

$$\xi_n=\int\limits_{-\pi}^{\pi}e^{i\lambda n}Z(d\lambda). \tag{2}$$

При этом $\mathsf{M}\,|Z(\Delta)|^2=F(\Delta)$.

Доказательство проще всего провести, опираясь на некоторые факты теории гильбертовых пространств.

Пусть $L^2(F)=L^2(E,\mathscr{E},F)$ — гильбертово пространство комплекснозначных функций, $E=[-\pi,\pi)$, $\mathscr{E}=\mathscr{B}([-\pi,\pi))$, со скалярным произведением

$$\langle f,g\rangle=\int\limits_{-\pi}^{\pi}f(\lambda)\,\overline{g}(\lambda)\,F(d\lambda), \tag{3}$$

и $L_0^2(F)$ — линейное многообразие ($L_0^2(F) \subseteq L^2(F)$), порожденное функциями $e_n = e_n(\lambda)$, $n \in \mathbb{Z}$, где $e_n(\lambda) = e^{i\lambda n}$.

Заметим, что поскольку $E = [-\pi, \pi)$ и мера F конечна, то замыкание $\overline{L}_0^2(F)$ совпадает (задача 1) с $L^2(F)$:

$$\overline{L}_0^2(F) = L^2(F).$$

Пусть, далее, $L_0^2(\xi)$ — линейное многообразие, порожденное случайными величинами ξ_n, $n \in \mathbb{Z}$, и $L^2(\xi)$ — его замыкание в среднеквадратическом смысле (по мере P).

Установим между элементами $L_0^2(F)$ и $L_0^2(\xi)$ взаимнооднозначное соответствие «$\leftrightarrow$», полагая

$$e_n \leftrightarrow \xi_n, \qquad n \in \mathbb{Z}, \tag{4}$$

и доопределяя для произвольных элементов (точнее — классов эквивалентных элементов) по линейности:

$$\sum \alpha_n e_n \leftrightarrow \sum \alpha_n \xi_n \tag{5}$$

(здесь предполагается, что только конечное число комплексных чисел α_n отлично от нуля).

Отметим, что соответствие (5) корректно определено в том смысле, что $\sum \alpha_n e_n = 0$ почти всюду по мере F тогда и только тогда, когда $\sum \alpha_n \xi_n = 0$ (P-п. н.).

Так определенное соответствие «$\leftrightarrow$» является *изометрическим*, т. е. сохраняющим скалярные произведения. В самом деле, в силу (3)

$$\langle e_n, e_m \rangle = \int_{-\pi}^{\pi} e_n(\lambda) \overline{e_m}(\lambda) F(d\lambda) =$$

$$= \int_{-\pi}^{\pi} e^{i\lambda(n-m)} F(d\lambda) = R(n-m) = \mathsf{M}\xi_n \overline{\xi}_m = (\xi_n, \xi_m)$$

и аналогично

$$\langle \sum \alpha_n e_n, \sum \beta_n e_n \rangle = (\sum \alpha_n \xi_n, \sum \beta_n \xi_n). \tag{6}$$

Пусть теперь $\eta \in L^2(\xi)$. Поскольку $L^2(\xi) = \overline{L}_0^2(\xi)$, то найдется такая последовательность $\{\eta_n\}$, что $\eta_n \in L_0^2(\xi)$ и $\|\eta_n - \eta\| \to 0$, $n \to \infty$. Следовательно, последовательность $\{\eta_n\}$ фундаментальна и, значит, таковой же является и последовательность функций $\{f_n\}$, где $f_n \in L_0^2(F)$ и $f_n \leftrightarrow \eta_n$. Пространство $L^2(F)$ полно и, следовательно, найдется такая функция $f \in L^2(F)$, что $\|f_n - f\| \to 0$.

Очевидным образом верно и обратное: если $f \in L^2(F)$ и $\|f - f_n\| \to 0$, $f_n \in L_0^2(F)$, то найдется такой элемент $\eta \in L^2(\xi)$, что $\|\eta - \eta_n\| \to 0$, $\eta_n \in L_0^2(\xi)$ и $\eta_n \leftrightarrow f_n$.

До сих пор (изометрическое) соответствие «$\leftrightarrow$» было определено лишь между элементами из $L_0^2(\xi)$ и $L_0^2(F)$. Доопределим его по непрерывности, полагая $f \leftrightarrow \eta$, где f и η — рассмотренные выше элементы Нетрудно проверить, что так установленное соответствие является взаимнооднозначным (между классами эквивалентных случайных величин и функций), линейным и сохраняющим скалярное произведение

Рассмотрим функцию $f(\lambda) = I_\Delta(\lambda)$, где $\Delta \in \mathscr{B}([-\pi, \pi))$, и пусть $Z(\Delta)$ — элемент из $L^2(\xi)$ такой, что $I_\Delta(\lambda) \leftrightarrow Z(\Delta)$. Ясно, что $\| I_\Delta(\lambda) \|^2 = F(\Delta)$ и, значит, $\mathsf{M}\,|Z(\Delta)|^2 = F(\Delta)$. Далее, если $\Delta_1 \cap \Delta_2 = \varnothing$, то $\mathsf{M}Z(\Delta_1)\bar{Z}(\Delta_2) = 0$ и $\mathsf{M}\left| Z(\Delta) - \sum_{k=1}^{n} Z(\Delta_k)\right|^2 \to 0$, $n \to \infty$, где $\Delta = \sum_{k=1}^{\infty} \Delta_k$

Тем самым совокупность элементов $Z(\Delta)$, $\Delta \in \mathscr{B}([-\pi, \pi))$, образует ортогональную стохастическую меру, по которой (согласно § 2) можно определить стохастический интеграл

$$\mathscr{I}(f) = \int_{-\pi}^{\pi} f(\lambda)\, Z(d\lambda), \qquad f \in L^2(F).$$

Пусть $f \in L^2(F)$ и $\eta \leftrightarrow f$. Обозначим элемент η через $\Phi(f)$ (точнее говоря, выберем по одному представителю из соответствующих классов эквивалентных случайных величин и функций). Покажем, что (P-п. н.)

$$\mathscr{I}(f) = \Phi(f). \tag{7}$$

Действительно, если

$$f(\lambda) = \sum \alpha_k I_{\Delta_k}(\lambda) \tag{8}$$

есть конечная линейная комбинация функций $I_{\Delta_k}(\lambda)$, $\Delta_k = (a_k, b_k]$, то по самому определению стохастического интеграла $\mathscr{I}(f) = \sum \alpha_k Z(\lambda_k)$, что, очевидно, равно $\Phi(f)$. Таким образом, (7) справедливо для функций вида (8). Но если $f \in L^2(F)$ и $\|f_n - f\| \to 0$, где f_n — функции вида (8), то $\|\Phi(f_n) - \Phi(f)\| \to 0$ и $\|\mathscr{I}(f_n) - \mathscr{I}(f)\| \to 0$ (согласно (2.14)). Значит, $\Phi(f) = \mathscr{I}(f)$ (P-п. н.).

Возьмем функцию $f(\lambda) = e^{i\lambda n}$. Тогда в силу (4) $\Phi(e^{i\lambda n}) = \xi_n$ и, с другой стороны, $\mathscr{I}(e^{i\lambda n}) = \int_{-\pi}^{\pi} e^{i\lambda n} Z(d\lambda)$. Поэтому в силу (7) ($\mathsf{P}$-п. н.)

$$\xi_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z(d\lambda), \qquad n \in \mathbb{Z}.$$

Теорема доказана.

Следствие 1. Пусть $\xi = (\xi_n)$ — стационарная последовательность, состоящая из действительных случайных величин ξ_n, $n \in \mathbb{Z}$.

Тогда стохастическая мера $Z=Z(\Delta)$, участвующая в спектральном представлении (2), такова, что для любого $\Delta \in \mathscr{B}([-\pi, \pi))$

$$Z(\Delta)=\overline{Z(-\Delta)}, \tag{9}$$

где множество $-\Delta=\{\lambda: -\lambda \in \Delta\}$.

В самом деле, пусть $f(\lambda)=\sum \alpha_k e^{i\lambda k}$ и $\eta=\sum \alpha_k \xi_k$ (суммы конечные). Тогда $f \leftrightarrow \eta$ и, значит,

$$\bar{\eta}=\sum \bar{\alpha}_k \xi_k \leftrightarrow \sum \bar{\alpha}_k e^{i\lambda k}=\overline{f(-\lambda)}. \tag{10}$$

Поскольку $I_\Delta(\lambda) \leftrightarrow Z(\Delta)$, то из (10) вытекает, что $I_\Delta(-\lambda) \leftrightarrow \bar{Z}(\Delta)$ или $I_{-\Delta}(\lambda) \leftrightarrow \bar{Z}(\Delta)$. Но, с другой стороны, $I_{-\Delta}(\lambda) \leftrightarrow Z(-\Delta)$. Поэтому $\bar{Z}(\Delta)=Z(-\Delta)$ (P-п. н.).

Следствие 2. Пусть снова $\xi=(\xi_n)$ — стационарная последовательность, где ξ_n — действительные случайные величины, и $Z(\Delta)=Z_1(\Delta)+iZ_2(\Delta)$. Тогда для любых Δ_1 и Δ_2

$$\mathsf{M} Z_1(\Delta_1) Z_2(\Delta_2)=0, \tag{11}$$

и если $\Delta_1 \cap \Delta_2=\varnothing$, то

$$\mathsf{M} Z_1(\Delta_1) Z_1(\Delta_2)=0, \quad \mathsf{M} Z_2(\Delta_1) Z_2(\Delta_2)=0. \tag{12}$$

Действительно, поскольку $Z(\Delta)=\bar{Z}(-\Delta)$, то

$$Z_1(-\Delta)=Z_1(\Delta), \quad Z_2(-\Delta)=-Z_2(\Delta). \tag{13}$$

Далее, так как $\mathsf{M} Z(\Delta_1)\bar{Z}(\Delta_2)=\mathsf{M}\,|Z(\Delta_1 \cap \Delta_2)|^2$, то $\operatorname{Im} \mathsf{M} Z(\Delta_1) \times \bar{Z}(\Delta_2)=0$, т. е.

$$\mathsf{M} Z_1(\Delta_1) Z_2(\Delta_2)+\mathsf{M} Z_2(\Delta_1) Z_1(\Delta_2)=0. \tag{14}$$

Взяв вместо Δ_1 интервал $-\Delta_1$, находим отсюда

$$\mathsf{M} Z_1(-\Delta_1) Z_2(\Delta_2)+\mathsf{M} Z_2(-\Delta_1) Z_1(\Delta_2)=0,$$

что в силу (13) преобразуется к виду

$$\mathsf{M} Z_1(\Delta_1) Z_2(\Delta_2)-\mathsf{M} Z_2(\Delta_1) Z_1(\Delta_2)=0. \tag{15}$$

Из (14) и (15) получаем равенство (11).

Если же $\Delta_1 \cap \Delta_2=\varnothing$, то $\mathsf{M} Z(\Delta_1)\bar{Z}(\Delta_2)=0$, откуда $\operatorname{Re} \mathsf{M} Z(\Delta_1) \times \bar{Z}(\Delta_2)=0$ и $\operatorname{Re} \mathsf{M} Z(-\Delta_1)\bar{Z}(\Delta_2)=0$, что вместе с (13) очевидным образом доказывает равенства (12).

Следствие 3. Пусть $\xi=(\xi_n)$ — гауссовская последовательность. Тогда для любого набора $\Delta_1, \ldots, \Delta_k$ вектор $(Z_1(\Delta_1), \ldots$ $\ldots, Z_1(\Delta_k), Z_2(\Delta_1), \ldots, Z_2(\Delta_k))$ имеет нормальное распределение.

В самом деле, линейное многообразие $L_0^2(\xi)$ состоит из (комплекснозначных) гауссовских случайных величин η, т. е. вектор $(\operatorname{Re}\eta, \operatorname{Im}\eta)$ имеет гауссовское распределение. Тогда в соответствии с п. 5 § 13 гл. II замыкание $\bar{L}_0^2(\xi)$ также состоит из гауссовских величин. Отсюда и из следствия 2 вытекает, что в случае гауссовской последовательности $\xi=(\xi_n)$ действительные и мнимые части Z_1 и Z_2 независимы в том смысле, что любые наборы случайных величин $(Z_1(\Delta_1), \ldots, Z_1(\Delta_k))$ и $(Z_2(\Delta_1), \ldots, Z_2(\Delta_k))$ независимы между собой. Из (12) следует также, что для непересекающихся множеств $\Delta_1, \ldots, \Delta_k$ случайные величины $Z_i(\Delta_1), \ldots, Z_i(\Delta_k)$ независимы в совокупности, $i=1, 2$.

Следствие 4. Если $\xi=(\xi_n)$ — стационарная последовательность действительных случайных величин, то (Р-п. н.)

$$\xi_n=\int_{-\pi}^{\pi}\cos\lambda n Z_1(d\lambda)+\int_{-\pi}^{\pi}\sin\lambda n Z_2(d\lambda). \tag{16}$$

Замечание. Если $\{Z_\lambda\}$, $\lambda\in[-\pi, \pi)$, — процесс с ортогональными приращениями, соответствующий ортогональной стохастической мере $Z=Z(\Delta)$, то спектральное представление (2) можно (в соответствии с § 2) записать также в следующем виде:

$$\xi_n=\int_{-\pi}^{\pi} e^{i\lambda n}\,dZ_\lambda, \qquad n\in\mathbb{Z}. \tag{17}$$

Пусть $\xi=(\xi_n)$ — стационарная последовательность со спектральным разложением (2), и пусть $\eta\in L^2(\xi)$. Следующая теорема описывает структуру таких случайных величин.

Теорема 2. *Если $\eta\in L^2(\xi)$, то найдется такая функция $\varphi\in L^2(F)$, что* (Р-п. н.)

$$\eta=\int_{-\pi}^{\pi}\varphi(\lambda)\,Z(d\lambda). \tag{18}$$

Доказательство. Если

$$\eta_n=\sum_{|k|\leqslant n}\alpha_k\xi_k, \tag{19}$$

то в силу (2)

$$\eta_n=\int_{-\pi}^{\pi}\Bigl(\sum_{|k|\leqslant n}\alpha_k e^{i\lambda k}\Bigr)Z(d\lambda), \tag{20}$$

т. е. (18) выполнено с функцией

$$\varphi_n(\lambda)=\sum_{|k|\leqslant n}\alpha_n e^{i\lambda k}. \tag{21}$$

В общем случае, если $\eta\in L^2(\xi)$, то найдутся такие величины η_n типа (19), что $\|\eta-\eta_n\|\to 0$, $n\to\infty$. Но тогда $\|\varphi_n-\varphi_m\|=\|\eta_n-\eta_m\|\to 0$, $n, m\to\infty$ и, следовательно, последовательность $\{\varphi_n\}$ фундаментальна в $L^2(F)$ и, значит, найдется такая функция $\varphi\in L^2(F)$, что $\|\varphi-\varphi_n\|\to 0$, $n\to\infty$.

В соответствии со свойством (2.14) $\|\mathscr{I}(\varphi_n)-\mathscr{I}(\varphi)\|\to 0$, и так как $\eta_n=\mathscr{I}(\varphi_n)$, то $\eta=\mathscr{I}(\varphi)$ (Р-п. н.).

Теорема доказана.

Замечание. Пусть $H_0(\xi)$ и $H_0(F)$ — замкнутые линейные многообразия, порожденные величинами ξ_n и функциями e_n соответственно для $n\leqslant 0$. Тогда, если $\eta\in H_0(\xi)$, то найдется такая функция $\varphi\in H_0(F)$, что (Р-п. н.) $\eta=\int_{-\pi}^{\pi}\varphi(\lambda)\,Z(d\lambda)$.

3. Формула (18) описывает структуру тех случайных величин, которые получаются из ξ_n, $n\in\mathbb{Z}$, с помощью линейных преобразований, т. е. в виде конечных сумм (19) и их пределов в среднеквадратическом смысле.

Частный, но важный класс таких линейных преобразований задается с помощью так называемых (линейных) *фильтров*. Предположим, что в момент времени m на вход некоторой системы (фильтр) подается сигнал x_m, при этом реакция системы на этот сигнал такова, что на ее выходе в момент времени n получается сигнал $h(n-m)\,x_m$, где $h=h(s)$, $s\in\mathbb{Z}$, — некоторая комплексно-значная функция, называемая *импульсной переходной функцией* (фильтра).

Таким образом, суммарный сигнал y_n на выходе системы представляется в виде

$$y_n=\sum_{m=-\infty}^{\infty}h(n-m)\,x_m. \tag{22}$$

Для физически осуществимых систем значение выходного сигнала в момент времени n определяется лишь «прошлыми» значениями входного сигнала, т. е. значениями x_m при $m\leqslant n$. Естественно поэтому фильтр с импульсной переходной функцией $h=h(s)$ назвать *физически осуществимым*, если $h(s)=0$ для всех $s<0$, иначе говоря, если

$$y_n=\sum_{m=-\infty}^{n}h(n-m)\,x_m=\sum_{m=0}^{\infty}h(m)\,x_{n-m}. \tag{23}$$

Важной *спектральной характеристикой* фильтра с импульсной переходной функцией h является ее преобразование Фурье

$$\varphi(\lambda)=\sum_{m=-\infty}^{\infty}e^{-i\lambda m}h(m), \tag{24}$$

называемое *частотной характеристикой* фильтра.

Остановимся теперь на условиях сходимости рядов в (22) и (24), о которых до сих пор ничего не говорилось. Предположим, что на вход фильтра подается стационарная случайная последовательность $\xi=(\xi_n)$, $n\in\mathbb{Z}$, с ковариационной функцией $R(n)$ и спектральным разложением (2). Тогда, если

$$\sum_{k,\,l=-\infty}^{\infty} h(k)\,R(k-l)\,\bar{h}(l)<\infty, \tag{25}$$

то ряд $\sum\limits_{k=-\infty}^{\infty} h(n-m)\,\xi_m$ сходится в среднеквадратическом смысле и, следовательно, определена стационарная последовательность $\eta=(\eta_n)$ с

$$\eta_n=\sum_{m=-\infty}^{\infty} h(n-m)\,\xi_m=\sum_{m=-\infty}^{\infty} h(m)\,\xi_{n-m}. \tag{26}$$

В спектральных терминах условие (25), очевидно, эквивалентно тому, что $\varphi(\lambda)\in L^2(F)$, т. е.

$$\int_{-\pi}^{\pi} |\varphi(\lambda)|^2\,F(d\lambda)<\infty. \tag{27}$$

При условии (25) или (27) из (26) и (2) находим спектральное разложение последовательности η:

$$\eta_n=\int_{-\pi}^{\pi} e^{i\lambda n}\varphi(\lambda)\,Z(d\lambda). \tag{28}$$

Следовательно, ковариационная функция $R_\eta(n)$ последовательности η определяется формулой

$$R_\eta(n)=\int_{-\pi}^{\pi} e^{i\lambda n}\,|\varphi(\lambda)|^2\,F(d\lambda). \tag{29}$$

В частности, если на вход фильтра с частотной характеристикой $\varphi=\varphi(\lambda)$ подается белый шум $\varepsilon=(\varepsilon_n)$, то на его выходе будет получаться стационарная последовательность (скользящего среднего)

$$\eta_n=\sum_{m=-\infty}^{\infty} h(m)\,\varepsilon_{n-m} \tag{30}$$

со спектральной плотностью

$$f_\eta(\lambda)=\frac{1}{2\pi}\,|\varphi(\lambda)|^2.$$

Следующая теорема показывает, что в определенном смысле всякая стационарная последовательность со спектральной плотностью есть последовательность, полученная с помощью скользящего среднего.

Теорема 3. *Пусть $\eta=(\eta_n)$ — стационарная последовательность со спектральной плотностью $f_\eta(\lambda)$. Тогда (быть может, за счет расширения исходного вероятностного пространства) можно найти такую последовательность $\varepsilon=(\varepsilon_n)$, являющуюся белым шумом, и такой фильтр, что справедливо представление* (30).

Доказательство. По заданной (неотрицательной) функции $f_\eta(\lambda)$ найдем такую функцию $\varphi(\lambda)$, что $f_\eta(\lambda)=\frac{1}{2\pi}|\varphi(\lambda)|^2$. Поскольку $\int_{-\pi}^{\pi} f_\eta(\lambda)\,d\lambda<\infty$, то $\varphi(\lambda)\in L^2(\mu)$, где μ — мера Лебега на $[-\pi, \pi)$. Поэтому ее можно представить в виде ряда Фурье (24) с $h(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{im\lambda}\varphi(\lambda)\,d\lambda$, причем сходимость понимается в том смысле, что

$$\int_{-\pi}^{\pi}\left|\varphi(\lambda)-\sum_{|m|\leqslant n} e^{-i\lambda m}h(m)\right|^2 d\lambda\to 0,\qquad n\to\infty.$$

Пусть

$$\eta_n=\int_{-\pi}^{\pi} e^{i\lambda n}Z(d\lambda),\qquad n\in\mathbb{Z}.$$

Наряду с мерой $Z=Z(\Delta)$ введем в рассмотрение не зависящую от нее новую ортогональную стохастическую меру $\tilde Z=\tilde Z(\Delta)$ с $\mathsf{M}|\tilde Z(a, b]|^2=\frac{b-a}{2\pi}$. (Возможность построения такой меры предполагает, вообще говоря, что исходное вероятностное пространство является достаточно «богатым».) Положим

$$\bar Z(\Delta)=\int_\Delta \varphi^\oplus(\lambda)Z(d\lambda)+\int_\Delta[1-\varphi^\oplus(\lambda)\varphi(\lambda)]\tilde Z(d\lambda),$$

где

$$a^\oplus=\begin{cases} a^{-1}, & \text{если } a\neq 0,\\ 0, & \text{если } a=0.\end{cases}$$

Стохастическая мера $\bar Z=\bar Z(\Delta)$ является мерой с ортогональными значениями, при этом для всякого $\Delta=(a, b]$

$$\mathsf{M}|\bar Z(\Delta)|^2=$$
$$=\frac{1}{2\pi}\int_\Delta|\varphi^\oplus(\lambda)|^2|\varphi(\lambda)|^2\,d\lambda+\frac{1}{2\pi}\int_\Delta|1-\varphi^\oplus(\lambda)\varphi(\lambda)|^2\,d\lambda=\frac{|\Delta|}{2\pi},$$

где $|\Delta|=b-a$. Поэтому стационарная последовательность $\varepsilon=(\varepsilon_n)$, $n\in\mathbb{Z}$, с

$$\varepsilon_n=\int_{-\pi}^{\pi} e^{i\lambda n}\bar Z(d\lambda)$$

является белым шумом.

Заметим теперь, что

$$\int_{-\pi}^{\pi} e^{i\lambda n}\varphi(\lambda)\,\bar Z(d\lambda)=\int_{-\pi}^{\pi} e^{in\lambda}Z(d\lambda)=\eta_n \tag{31}$$

и, с другой стороны, по свойству (2.14) (Р-п. н.)

$$\int_{-\pi}^{\pi} e^{in\lambda}\varphi(\lambda)\,\bar Z(d\lambda)=\int_{-\pi}^{\pi} e^{i\lambda n}\left(\sum_{m=-\infty}^{\infty} e^{-i\lambda m}h(m)\right)\bar Z(d\lambda)=$$

$$=\sum_{m=-\infty}^{\infty} h(m)\int_{-\pi}^{\pi} e^{i\lambda(n-m)}\bar Z(d\lambda)=\sum_{m=-\infty}^{\infty} h(m)\,\varepsilon_{n-m},$$

что вместе с (31) доказывает представление (30).

Теорема доказана.

Замечание. Если $f_\eta(\lambda)>0$ (почти всюду по мере Лебега), то введение вспомогательной меры $\tilde Z=\tilde Z(\Delta)$ становится излишним (поскольку тогда $1-\varphi^{\oplus}(\lambda)\,\varphi(\lambda)=0$ почти всюду по мере Лебега) и оговорка относительно необходимости расширения исходного вероятностного пространства может быть опущена.

Следствие 1. Пусть спектральная плотность $f_\eta(\lambda)>0$ (почти всюду по мере Лебега) и

$$f_\eta(\lambda)=\frac{1}{2\pi}|\varphi(\lambda)|^2,$$

где

$$\varphi(\lambda)=\sum_{k=0}^{\infty} e^{-i\lambda k}h(k),\qquad \sum_{k=0}^{\infty}|h(k)|^2<\infty.$$

Тогда последовательность η допускает представление в виде одностороннего скользящего среднего

$$\eta_n=\sum_{m=0}^{\infty} h(m)\,\varepsilon_{n-m}.$$

В частности, пусть $P(z)=a_0+a_1z+\ldots+a_pz^p$ — полином, не имеющий нулей на множестве $\{z\colon |z|=1\}$. Тогда последовательность $\eta=(\eta_n)$ со спектральной плотностью

$$f_\eta(\lambda)=\frac{1}{2\pi}|P(e^{-i\lambda})|^2$$

представима в виде

$$\eta_n=a_0\varepsilon_n+a_1\varepsilon_{n-1}+\ldots+a_p\varepsilon_{n-p}.$$

Следствие 2. Пусть $\xi=(\xi_n)$ — стационарная последовательность с рациональной спектральной плотностью

$$f_\xi(\lambda)=\frac{1}{2\pi}\left|\frac{P(e^{-i\lambda})}{Q(e^{-i\lambda})}\right|^2, \tag{32}$$

где $P(z)=a_0+a_1z+\ldots+a_pz^p$, $Q(z)=1+b_1z+\ldots+b_qz^q$.

Покажем, что если полиномы $P(z)$ и $Q(z)$ не имеют нулей на множестве $\{z\colon |z|=1\}$, то найдется такой белый шум $\varepsilon=(\varepsilon_n)$, что (P-п. н.)

$$\xi_n+b_1\xi_{n-1}+\ldots+b_q\xi_{n-q}=a_0\varepsilon_n+a_1\varepsilon_{n-1}+\ldots+a_p\varepsilon_{n-p}. \tag{33}$$

Обратно, всякая стационарная последовательность $\xi=(\xi_n)$, удовлетворяющая такому уравнению с некоторым белым шумом $\varepsilon=(\varepsilon_n)$ и полиномом $Q(z)$, не имеющим нулей на множестве $\{z\colon |z|=1\}$, имеет спектральную плотность (32).

Действительно, пусть $\eta_n=\xi_n+b_1\xi_{n-1}+\ldots+b_q\xi_{n-q}$. Тогда $f_\eta(\lambda)=\frac{1}{2\pi}|P(e^{-i\lambda})|^2$ и требуемое представление вытекает из следствия 1.

С другой стороны, если имеет место представление (33) и $F_\xi(\lambda)$ и $F_\eta(\lambda)$ — спектральные функции последовательностей ξ и η, то

$$F_\eta(\lambda)=\int_{-\pi}^{\lambda}|Q(e^{-i\nu})|^2\,dF_\xi(\nu)=\frac{1}{2\pi}\int_{-\pi}^{\lambda}|P(e^{-i\nu})|^2\,d\nu.$$

Поскольку $|Q(e^{-i\nu})|^2>0$, то отсюда следует, что $F_\xi(\lambda)$ имеет плотность, определяемую формулой (32).

4. Следующая эргодическая теорема (в среднеквадратическом смысле) может рассматриваться как аналог закона больших чисел для стационарных (в широком смысле) случайных последовательностей.

Теорема 4. *Пусть* $\xi=(\xi_n)$, $n\in\mathbb{Z}$, — *стационарная последовательность с* $\mathsf{M}\xi_n=0$, *ковариационной функцией* (1) *и спектральным разложением* (2). *Тогда*

$$\frac{1}{n}\sum_{k=0}^{n-1}\xi_k\xrightarrow{L^2}Z(\{0\}) \tag{34}$$

и

$$\frac{1}{n}\sum_{k=0}^{n-1}R(k)\to F(\{0\}). \tag{35}$$

Доказательство. В силу (2)

$$\frac{1}{n}\sum_{k=0}^{n-1}\xi_k=\int_{-\pi}^{\pi}\frac{1}{n}\sum_{k=0}^{n-1}e^{ik\lambda}Z(d\lambda)=\int_{-\pi}^{\pi}\varphi_n(\lambda)\,Z(\lambda),$$

где

$$\varphi_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} e^{ik\lambda} = \begin{cases} 1, & \lambda = 0, \\ \frac{1}{n} \cdot \frac{e^{in\lambda} - 1}{e^{i\lambda} - 1}, & \lambda \neq 0. \end{cases} \tag{36}$$

Поскольку $|\sin\lambda| \geqslant \frac{2}{\pi}|\lambda|$ для $|\lambda| \leqslant \frac{\pi}{2}$, то

$$|\varphi_n(\lambda)| = \left| \frac{\sin \frac{n\lambda}{2}}{n \sin \frac{\lambda}{2}} \right| \leqslant \frac{\pi}{2} \left| \frac{\sin \frac{n\lambda}{2}}{\frac{n\lambda}{2}} \right| \leqslant \frac{\pi}{2}.$$

Далее, $\varphi_n(\lambda) \xrightarrow{L^2(F)} I_{\{0\}}(\lambda)$, поэтому по свойству (2.14)

$$\int_{-\pi}^{\pi} \varphi_n(\lambda)\, Z(d\lambda) \xrightarrow{L^2} \int_{-\pi}^{\pi} I_{\{0\}}(\lambda)\, Z(d\lambda) = Z(\{0\}),$$

что и доказывает (34).

Аналогичным образом доказывается и утверждение (35).

Теорема доказана.

Следствие. Если спектральная функция непрерывна в нуле, т. е. $F(\{0\}) = 0$, то $Z(\{0\}) = 0$ (P-п. н.) и в силу (34), (35)

$$\frac{1}{n} \sum_{k=0}^{n-1} R(k) \to 0 \Rightarrow \frac{1}{n} \sum_{k=0}^{n-1} \xi_k \xrightarrow{L^2} 0.$$

Поскольку

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} R(k) \right|^2 = \left| \mathsf{M} \left(\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \right) \bar{\xi}_0 \right|^2 \leqslant \mathsf{M}\,|\xi_0|^2\, \mathsf{M} \left(\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \right)^2,$$

то верна и обратная импликация:

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \xrightarrow{L^2} 0 \Rightarrow \frac{1}{n} \sum_{k=0}^{n-1} R(k) \to 0.$$

Таким образом, условие $\frac{1}{n} \sum_{k=0}^{n-1} R(k) \to 0$ является необходимым и достаточным для сходимости (в среднеквадратическом смысле) средних арифметических $\frac{1}{n} \sum_{k=0}^{n-1} \xi_k$ к нулю.

Отсюда следует, что если исходная последовательность $\xi=(\xi_n)$ такова, что ее математическое ожидание есть m ($\mathsf{M}\xi_0=m$), то

$$\frac{1}{n}\sum_{k=0}^{n-1} R(k)\to 0 \Longleftrightarrow \frac{1}{n}\sum_{k=0}^{n-1}\xi_k \xrightarrow{L^2} m, \tag{37}$$

где $R(n)=\mathsf{M}(\xi_n-\mathsf{M}\xi_n)(\xi_0-\mathsf{M}\xi_0)$.

Отметим также, что если $Z(\{0\})\neq 0$ (P-п. н.), а $m=0$, то это означает, что последовательность ξ_n содержит «случайную константу α»:

$$\xi_n=\alpha+\eta_n,$$

где $\alpha=Z(\{0\})$, а в спектральном представлении $\eta_n=\int_{-\pi}^{\pi} e^{i\lambda n} Z_\eta(d\lambda)$ мера $Z_\eta=Z_\eta(\Delta)$ уже такова, что $Z_\eta(\{0\})=0$ (P-п. н.). Утверждение (34) означает, что средние арифметические сходятся в среднеквадратическом смысле именно к этой случайной константе α.

5. Задачи.

1. Показать, что $L_0^2(F)=L^2(F)$ (обозначения см. в доказательстве теоремы 1).

2. Пусть $\xi=(\xi_n)$ — стационарная последовательность, обладающая тем свойством, что для некоторого N и всех n $\xi_{n+N}=\xi_n$. Показать, что спектральное представление такой последовательности сводится к представлению (1.13).

3. Пусть $\xi=(\xi_n)$ — стационарная последовательность такая, что $\mathsf{M}\xi_n=0$ и

$$\frac{1}{N^2}\sum_{k=0}^{N}\sum_{l=0}^{N} R(k-l)=\frac{1}{N}\sum_{|k|\leqslant N-1} R(k)\left[1-\frac{|k|}{N}\right]\leqslant CN^{-\alpha}$$

при некоторых $C>0$, $\alpha>0$. Используя лемму Бореля — Кантелли показать, что тогда

$$\frac{1}{N}\sum_{k=0}^{N}\xi_k\to 0 \quad (\mathsf{P}\text{-п. н.}).$$

4. Пусть спектральная плотность $f_\xi(\lambda)$ последовательности $\xi=(\xi_n)$ является рациональной

$$f_\xi(\lambda)=\frac{1}{2\pi}\frac{|P_{n-1}(e^{-i\lambda})|}{|Q_n(e^{-i\lambda})|}, \tag{38}$$

где $P_{n-1}(z)=a_0+a_1z+\ldots+a_{n-1}z^{n-1}$ и $Q_n(z)=1+b_1z+\ldots+b_nz^n$, причем все корни полинома лежат вне единичного круга.

Показать, что найдется такой белый шум $\varepsilon=(\varepsilon_m)$, $m\in\mathbb{Z}$, что последовательность (ξ_m) будет компонентой n-мерной последовательности $(\xi_m^1, \xi_m^2, \ldots, \xi_m^n)$, $\xi_m^1 = \xi_m$, удовлетворяющей системе уравнений

$$\xi_{m+1}^i = \xi_m^{i+1} + \beta_i \varepsilon_{m+1}, \qquad i = 1, \ldots, n-1,$$
$$\xi_{m+1}^n = -\sum_{j=0}^{n-1} b_{n-j} \xi_m^{j+1} + \beta_n \varepsilon_{m+1}, \tag{39}$$

где $\beta_1 = a_0$, $\beta_i = a_{i-1} - \sum_{k=1}^{i-1} \beta_k b_{i-k}$.

§ 4. Статистическое оценивание ковариационной функции и спектральной плотности

1. Задачи статистического оценивания тех или иных характеристик распределений вероятностей стационарных случайных последовательностей возникают в самых разнообразных областях науки (геофизика, медицина, экономика и др.). Материал, излагаемый в настоящем параграфе, дает представление о понятиях и методах оценивания и о тех трудностях, которые здесь возникают.

Итак, пусть $\xi = (\xi_n)$, $n \in \mathbb{Z}$, стационарная в широком смысле (действительная — для простоты) случайная последовательность с математическим ожиданием $\mathsf{M}\xi_n = m$ и ковариацией $R(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda)$.

Пусть $x_0, x_1, \ldots, x_{N-1}$ — полученные в ходе наблюдений значения случайных величин $\xi_0, \xi_1, \ldots, \xi_{N-1}$. Как по ним построить «хорошую» оценку (неизвестного) среднего значения m?

Положим

$$m_N(x) = \frac{1}{N} \sum_{k=0}^{N-1} x_k. \tag{1}$$

Тогда из элементарных свойств математического ожидания следует, что эта оценка является «хорошей» оценкой величины m в том смысле, что «в среднем по всем реализациям $x_0, \ldots, x_{N-1}$» она является *несмещенной*, т. е

$$\mathsf{M} m_N(\xi) = \mathsf{M}\left(\frac{1}{N} \sum_{k=0}^{N-1} \xi_k\right) = m. \tag{2}$$

Более того, из теоремы 4 § 3 вытекает, что при условии $\frac{1}{N} \sum_{k=0}^{N} R(k) \to 0$, $N \to \infty$, рассматриваемая оценка является также и *состоятельной* (в среднеквадратическом смысле), т. е.

$$\mathsf{M}\,|m_N(\xi)-m|^2\to 0,\qquad N\to\infty. \tag{3}$$

Займемся теперь вопросом оценивания ковариационной функции $R(n)$, спектральной функции $F(\lambda)=F([-\pi,\ \lambda])$ и спектральной плотности $f(\lambda)$, предполагая, что $m=0$.

Поскольку $R(n)=\mathsf{M}\xi_{n+k}\xi_n$, то в качестве оценки этой величины по результатам N наблюдений $x_0,\ x_1,\ \ldots,\ x_{N-1}$ естественно взять (для $0\leqslant n<N$) величину

$$\hat{R}_n(n;\ x)=\frac{1}{N-n}\sum_{k=0}^{N-n-1}x_{n+k}x_k.$$

Ясно, что эта оценка является несмещенной в том смысле, что

$$\mathsf{M}\hat{R}_N(n;\ \xi)=R(n),\qquad 0\leqslant n<N.$$

Рассмотрим теперь вопрос о ее состоятельности. Подставляя в (3.37) вместо ξ_k величины $\xi_{n+k}\xi_k$ и предполагая у рассматриваемой последовательности $\xi=(\xi_n)$ существование четвертого момента ($\mathsf{M}\xi_0^4<\infty$), находим, что условие

$$\frac{1}{N}\sum_{k=0}^{N-1}\mathsf{M}\,[\xi_{n+k}\xi_k-R(n)][\xi_n\xi_0-R(n)]\to 0,\qquad N\to\infty, \tag{4}$$

является необходимым и достаточным для того, чтобы

$$\mathsf{M}\,|\hat{R}_N(n;\ \xi)-R(n)|^2\to 0,\qquad N\to\infty. \tag{5}$$

Предположим, что исходная последовательность $\xi=(\xi_n)$ является гауссовской (с нулевым средним и ковариацией $R(n)$). Тогда в силу (II.12.51)

$$\begin{aligned}\mathsf{M}\,[\xi_{n+k}\xi_k-R(n)][\xi_n\xi_0-R(n)]&=\mathsf{M}\xi_{n+k}\xi_k\xi_n\xi_0-R^2(n)=\\&=\mathsf{M}\xi_{n+k}\xi_k\cdot\mathsf{M}\xi_n\xi_0+\mathsf{M}\xi_{n+k}\xi_n\cdot\mathsf{M}\xi_k\xi_0+\mathsf{M}\xi_{n+k}\xi_0\cdot\mathsf{M}\xi_k\xi_n-R^2(n)=\\&=R^2(k)+R(n+k)\,R(n-k).\end{aligned}$$

Поэтому в гауссовском случае условие (4) эквивалентно условию

$$\frac{1}{N}\sum_{k=0}^{N-1}[R^2(k)+R(n+k)\,R(n-k)]\to 0,\qquad N\to\infty. \tag{6}$$

Поскольку $|R(n+k)\,R(n-k)|\leqslant|R(n+k)|^2+|R(n-k)|^2$, то из условия

$$\frac{1}{N}\sum_{k=0}^{N-1}R^2(k)\to 0,\qquad N\to\infty, \tag{7}$$

вытекает и условие (6). В свою очередь, если (6) верно для $n=0$, то выполняется условие (7).

Таким образом доказана следующая

Теорема. *Пусть $\xi=(\xi_n)$ — гауссовская стационарная последовательность с $\mathsf{M}\xi_n=0$ и ковариационной функцией $R(n)$. Тогда выполнение условия* (7) *является необходимым и достаточным для того, чтобы при любом $n\geqslant 0$ оценка $\widehat{R}_N(n;\,x)$ была состоятельной в среднеквадратическом смысле (т. е. чтобы было выполнено условие* (5)).

Замечание. Если воспользоваться спектральным представлением ковариационной функции, то получим

$$\frac{1}{N}\sum_{k=0}^{N-1} R^2(k)=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{N}\sum_{k=0}^{N-1} e^{i(\lambda-\nu)k}F(d\lambda)\,F(d\nu)=$$
$$=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_N(\lambda,\ \nu)\,F(d\lambda)\,F(d\nu),$$

где (ср. с (3.35))

$$f_N(\lambda,\ \nu)=\begin{cases} 1, & \lambda=0,\\ \dfrac{1-e^{i(\lambda-\nu)N}}{N\,[1-e^{i(\lambda-\nu)}]}, & \lambda\neq\nu.\end{cases}$$

Но при $n\to\infty$

$$f_n(\lambda,\ \nu)\to f(\lambda,\ \nu)=\begin{cases} 1, & \lambda=\nu,\\ 0, & \lambda\neq\nu.\end{cases}$$

Поэтому

$$\frac{1}{N}\sum_{k=0}^{N-1} R^2(k)\to\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\lambda,\ \nu)\,F(d\lambda)\,F(d\nu)=$$
$$=\int_{-\pi}^{\pi} F(\{\lambda\})\,F(d\lambda)=\sum_{\lambda} F^2(\{\lambda\}),$$

где сумма по λ не более чем счетна, поскольку мера F конечна.

Тем самым условие (7) эквивалентно условию

$$\sum_{\lambda} F^2(\{\lambda\})=0, \tag{8}$$

означающему, что спектральная функция $F(\lambda)=F([-\pi,\ \lambda])$ является *непрерывной*.

2. Перейдем теперь к вопросу построения оценок для спектральной функции $F(\lambda)$ и спектральной плотности $f(\lambda)$ (в предположении, что она существует).

Естественно напрашивающийся путь построения оценок спектральной плотности следует из проведенного выше доказательства теоремы Герглотца. Напомним, что введенная в § 1 функция

$$f_N(\lambda)=\frac{1}{2\pi}\sum_{|n|<N}\left(1-\frac{|n|}{N}\right)R(n) \tag{9}$$

обладала тем свойством, что построенная по ней функция

$$F_N(\lambda)=\int_{-\pi}^{\lambda} f_N(\nu)\,d\nu$$

сходилась в основном к спектральной функции $F(\lambda)$ Поэтому, если $F(\lambda)$ имеет плотность $f(\lambda)$, то для каждого $\lambda\in[-\pi,\pi)$

$$\int_{-\pi}^{\lambda} f_N(\nu)\,d\nu\to\int_{-\pi}^{\lambda} f(\nu)\,d\nu. \tag{10}$$

Исходя из этих фактов и вспоминая, что в качестве оценки $R(n)$ (по наблюдениям $x_0, x_1, \ldots, x_{N-1}$) брались величины $\hat{R}_N(n, x)$, возьмем в качестве оценки $f(\lambda)$ функцию

$$\hat{f}_N(\lambda, x)=\frac{1}{2\pi}\sum_{|n|<N}\left(1-\frac{|n|}{N}\right)\hat{R}_N(n;x), \tag{11}$$

полагая $\hat{R}_N(n;x)=\hat{R}_N(|n|;x)$ для всех $|n|<N$

Функцию $\hat{f}_N(\lambda;x)$ принято называть *периодограммой*, и нетрудно проверить, что ее можно представить также в следующем несколько более удобном виде:

$$\hat{f}_N(\lambda, x)=\frac{1}{2\pi N}\left|\sum_{n=0}^{N-1} x_n e^{-i\lambda n}\right|^2. \tag{12}$$

Поскольку $\mathsf{M}\hat{R}_N(n;\xi)=R(n)$, $|n|<N$, то

$$\mathsf{M}\hat{f}_N(\lambda;\xi)=f_N(\lambda).$$

Если спектральная функция $F(\lambda)$ имеет плотность $f(\lambda)$, то, учитывая, что $f_N(\lambda)$ может быть записана также в виде (1 34), найдем, что

$$f_N(\lambda)=\frac{1}{2\pi N}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\int_{-\pi}^{\pi} e^{i\nu(k-l)}e^{i\nu(l-k)}f(\nu)\,d\nu=$$

$$=\int_{-\pi}^{\pi}\frac{1}{2\pi N}\left|\sum_{k=0}^{N-1} e^{i(\nu-\lambda)k}\right|^2 f(\nu)\,d\nu.$$

Функция

$$\Phi_N(\lambda)=\frac{1}{2\pi N}\left|\sum_{k=0}^{N-1} e^{i\lambda k}\right|^2=\frac{1}{2\pi N}\left|\frac{\sin\frac{\pi}{2}N}{\sin\lambda/2}\right|^2$$

называется ядром Фейера. Из свойств этой функции известно, что для почти всех λ (по мере Лебега)

$$\int_{-\pi}^{\pi}\Phi_N(\lambda-\nu)f(\nu)\,d\nu\to f(\lambda). \tag{13}$$

Поэтому для почти всех $\lambda\in[-\pi,\ \pi)$

$$\mathsf{M}\hat{f}_N(\lambda;\ \xi)\to f(\lambda), \tag{14}$$

иначе говоря, оценка $\hat{f}_N(\lambda;\ x)$ спектральной плотности $f(\lambda)$ по наблюдениям $x_0, x_1, \ldots, x_{N-1}$ является *асимптотически несмещенной*.

В этом смысле оценку $\hat{f}_N(\lambda;\ x)$ можно было бы считать достаточно «хорошей». Однако на индивидуальных наблюдениях $x_0, \ldots$ $\ldots, x_{N-1}$ значения периодограммы $\hat{f}_N(\lambda;\ x)$ оказываются, как правило, далекими от истинных значений $f(\lambda)$. В самом деле, пусть $\xi=(\xi_n)$ — стационарная последовательность независимых гауссовских случайных величин, $\xi_n\sim\mathscr{N}(0,\ 1)$. Тогда $f(\lambda)\equiv 1/2\pi$, а

$$\hat{f}_N(\lambda;\ \xi)=\frac{1}{2\pi}\left|\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}\xi_k e^{-i\lambda k}\right|^2.$$

Поэтому при $\lambda=0$ $\hat{f}_N(0;\ \xi)$ по распределению совпадает с квадратом гауссовской случайной величины $\eta\sim\mathscr{N}(0,\ 1)$. Отсюда при любом N

$$\mathsf{M}\,|\hat{f}_N(0;\ \xi)-f(0)|^2=\frac{1}{4\pi^2}\mathsf{M}\,|\eta^2-1|^2>0.$$

Более того, несложный подсчет показывает, что если $f(\lambda)$ — спектральная плотность стационарной последовательности $\xi=(\xi_n)$, образованной по схеме скользящего среднего:

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k} \tag{15}$$

с $\sum\limits_{k=0}^{\infty}|a_k|<\infty$, $\sum\limits_{k=0}^{\infty}|a_k|^2<\infty$, где $\varepsilon=(\varepsilon_n)$ — белый шум с $\mathsf{M}\varepsilon_0^4<$ $<\infty$, то

Отсюда становится понятным, что периодограмма не может служить удовлетворительной оценкой спектральной плотности. Чтобы исправить это положение, в качестве оценок для $f(\lambda)$ часто используют оценки вида

$$\hat{f}_N^W(\lambda; x) = \int_{-\pi}^{\pi} W_N(\lambda-\nu)\hat{f}_N(\nu; x)\,d\nu, \tag{17}$$

которые строятся по периодограмме $\hat{f}_N(\lambda; x)$ и некоторым «сглаживающим» функциям $W_N(\lambda)$, называемым *спектральными окнами*. Естественные требования, предъявляемые к функциям $W_N(\lambda)$, состоят в том, чтобы:

a) $W_N(\lambda)$ имели резко выраженный максимум в окрестности точки $\lambda=0$;

b) $\int_{-\pi}^{\pi} W_N(\lambda)\,d\lambda = 1$;

c) $\mathsf{M}\,|\hat{f}_N^W(\lambda; \xi) - f(\lambda)|^2 \to 0, \quad N\to\infty, \quad \lambda\in[-\pi, \pi)$.

В силу (14) и требования b) оценки $\hat{f}_N^W(\lambda; \xi)$ являются асимптотически несмещенными. Требование c) является условием асимптотической состоятельности в среднеквадратическом смысле, что, как было показано выше, нарушается для периодограммы. Наконец, требование a) обеспечивает «вырезание» из периодограммы требуемой частоты λ.

Приведем некоторые примеры оценок вида (17).

Оценка Бартлета основана на выборе спектрального окна

$$W_N(\lambda) = a_N B(a_N\lambda),$$

где $a_N \uparrow \infty$, $a_N/N \to 0$, $N\to\infty$ и

$$B(\lambda) = \frac{1}{2\pi}\left|\frac{\sin\frac{\lambda}{2}}{\lambda/2}\right|^2.$$

Оценка Парзена использует в качестве спектрального окна функцию

$$W_N(\lambda) = a_N P(a_N\lambda),$$

где a_N такие же, что и выше, а

$$P(\lambda) = \frac{3}{8\pi}\left|\frac{\sin\frac{\lambda}{4}}{\lambda/4}\right|^4.$$

Оценки Журбенко строятся с помощью спектральных окон вида

$$W_N(\lambda) = a_N Z(a_N\lambda)$$

с

$$Z(\lambda)=\begin{cases} -\frac{\alpha+1}{2\alpha}|\lambda|^{\alpha}+\frac{\alpha+1}{2\alpha}, & |\lambda|\leqslant 1,\\ 0, & |\lambda|>1,\end{cases}$$

где $0<\alpha\leqslant 2$, а величины a_N подбираются специальным образом.

Не останавливаясь подробнее на вопросах оценивания спектральных плотностей, укажем лишь, что имеется обширная статистическая литература, посвященная построению спектральных окон и сравнению свойств соответствующих им оценок $\hat{f}_N^W(\lambda;\ x)$.

3. Рассмотрим теперь вопрос оценивания спектральной функции $F(\lambda)=F([-\pi,\ \lambda])_\pi$. С этой целью положим

$$F_N(\lambda)=\int_{-\pi}^{\pi} f_N(\nu)\,d\nu,\quad \hat{F}_N(\lambda;\ x)=\int_{-\pi}^{\lambda}\hat{f}_N(\nu;\ x)\,d\nu,$$

где $\hat{f}_N(\nu;\ x)$ — периодограмма, построенная по $(x_0,\ x_1,\ \ldots,\ x_{N-1})$.

Из доказательства теоремы Герглотца (§ 1) следует, что для любого $n\in\mathbb{Z}$

$$\int_{-\pi}^{\pi} e^{i\lambda n}\,dF_N(\lambda)\to\int_{-\pi}^{\pi} e^{i\lambda n}\,dF(\lambda).$$

Отсюда (ср. со следствием к теореме 1 § 3 гл. III) следует, что $F_N\Rightarrow F$, т. е. $F_N(\lambda)$ сходятся к $F(\lambda)$ в каждой точке непрерывности функции $F(\lambda)$.

Заметим, что для всех $|n|<N$

$$\int_{-\pi}^{\pi} e^{i\lambda n}\,d\hat{F}_N(\lambda;\ \xi)=\hat{R}_N(n;\ \xi)\left(1-\frac{|n|}{N}\right).$$

Поэтому, если предположить, что $\hat{R}_N(n;\ \xi)$ сходятся с вероятностью единица к $R(n)$ при $N\to\infty$, то тогда

$$\int_{-\pi}^{\pi} e^{i\lambda n}\,d\hat{F}_N(\lambda;\ \xi)\to\int_{-\pi}^{\pi} e^{i\lambda n}\,dF(\lambda)\quad (\mathsf{P}\text{-п. н.})$$

и, значит, (P-п. н.) $\hat{F}_N(\lambda;\ \xi)\Rightarrow F(\lambda)$.

Отсюда нетрудно вывести (переходя в случае необходимости от последовательностей к подпоследовательностям), что если $\hat{R}_N(n;\ \xi)\to R(n)$ по вероятности, то тогда и $\hat{F}_N(\lambda;\ \xi)\Rightarrow F(\lambda)$ по вероятности.

4. Задачи.

1. Пусть в схеме (1.5) величины $\varepsilon_n\sim\mathcal{N}(0,\ 1)$. Показать, что для любого n и $N\to\infty$

$$(N-n)\,\mathsf{D}\hat{R}_N(n;\ \xi)\to 2\pi\int_{-\pi}^{\pi}(1+e^{2in\lambda})\,f^2(\lambda)\,d\lambda.$$

2. Установить справедливость формулы (16) и следующего ее обобщения

$$\lim_{N\to\infty} \operatorname{cov}(\hat{f}_N(\lambda;\ \xi),\ \hat{f}_N(\nu;\ \xi)) = \begin{cases} 2f^2(0), & \lambda=\nu=0,\ \pm\pi, \\ f^2(\lambda), & \lambda=\nu\neq 0,\ \pm\pi, \\ 0, & \lambda\neq\pm\nu. \end{cases}$$

§ 5. Разложение Вольда

1. В отличие от представления (3 2) дающего разложение стационарной последовательности в *частотной* области, рассматриваемое ниже разложение Вольда действует во *временной* области. Суть этого разложения сводится к тому, что стационарная последовательность $\xi=(\xi_n)$, $n\in\mathbb{Z}$, представляется в виде суммы двух стационарных последовательностей, одна из которых полностью предсказуема (в том смысле, что ее значения полностью восстанавливаются по «прошлому»), а вторая этим свойством не обладает.

Введем прежде всего некоторые обозначения. Пусть $H_n(\xi) = \overline{L}^2(\xi^n)$ и $H(\xi)=\overline{L}^2(\xi)$ — замкнутые линейные многообразия, порожденные величинами $\xi^n=(\ldots,\ \xi_{n-1},\ \xi_n)$ и $\xi=(\ldots\xi_{n-1},\ \xi_n,\ \ldots)$ соответственно. Пусть также

$$S(\xi)=\bigcap_n H_n(\xi).$$

Очевидно, что

$$H_n(\xi)\uparrow H(\xi),\qquad n\to\infty$$

и

$$H_n(\xi)\downarrow S(\xi),\qquad n\to-\infty.$$

Для любого элемента $\eta\in H(\xi)$ обозначим через

$$\hat{\pi}_n(\eta)=\hat{\mathsf{M}}(\eta\mid H_n(\xi))$$

проекцию элемента η на подпространство $H_n(\xi)$ (см. § 11 гл. II). Будем обозначать также

$$\hat{\pi}_{-\infty}(\eta)=\hat{\mathsf{M}}(\eta\mid S(\xi)).$$

Каждый элемент $\eta\in H(\xi)$ можно представить следующим образом:

$$\eta=\hat{\pi}_{-\infty}(\eta)+(\eta-\hat{\pi}_{-\infty}(\eta)),$$

где $\eta-\hat{\pi}_{-\infty}(\eta)\perp\hat{\pi}_{-\infty}(\eta)$. Поэтому пространство $H(\xi)$ представляется в виде ортогональной суммы

$$H(\xi)=S(\xi)\oplus R(\xi),$$

где $S(\xi)$ состоит из элементов $\hat{\pi}_{-\infty}(\eta)$ с $\eta \in H(\xi)$, а $R(\xi)$ — из элементов вида $\eta - \hat{\pi}_{-\infty}(\eta)$.

Всюду в дальнейшем будем предполагать, что $\mathsf{M}\xi_n = 0$ и $\mathsf{D}\xi_n > 0$. Тем самым пространство $H(\xi)$ заведомо является нетривиальным (содержит элементы, отличные от нулевого).

Определение 1. Стационарная последовательность $\xi = (\xi_n)$ называется *регулярной*, если

$$H(\xi) = R(\xi),$$

и *сингулярной*, если

$$H(\xi) = S(\xi).$$

Замечание. Сингулярные последовательности называют также *детерминированными*, регулярные — *чисто* или *вполне недетерминированными*. Если $S(\xi)$ есть собственное подпространство пространства $H(\xi)$, то последовательность ξ называют *недетерминированной*.

Теорема 1. *Всякая стационарная в широком смысле случайная последовательность ξ допускает и притом единственное разложение*

$$\xi_n = \xi_n^r + \xi_n^s, \tag{1}$$

где $\xi^r = (\xi_n^r)$ — регулярная, а $\xi^s = (\xi_n^s)$ — сингулярная последовательности. При этом ξ^r и ξ^s ортогональны ($\xi_n^r \perp \xi_m^s$ для всех n и m).

Доказательство. По определению положим

$$\xi_n^s = \hat{\mathsf{M}}(\xi_n / S(\xi)), \quad \xi_n^r = \xi_n - \xi_n^s.$$

Поскольку $\xi_n^r \perp S(\xi)$ для любого n, то $S(\xi^r) \perp S(\xi)$. С другой стороны, $S(\xi^r) \subseteq S(\xi)$, и, значит, $S(\xi^r)$ тривиально (содержит лишь случайные величины, совпадающие почти наверное с нулем). Следовательно, процесс ξ^r является регулярным.

Далее, $H_n(\xi) \subseteq H_n(\xi^s) \oplus H_n(\xi^r)$ и $H_n(\xi^s) \subseteq H_n(\xi)$, $H_n(\xi^r) \subseteq H_n(\xi)$. Поэтому $H_n(\xi) = H_n(\xi^s) \oplus H_n(\xi^r)$, и, значит, для любого n

$$S(\xi) \subseteq H_n(\xi^s) \oplus H_n(\xi^r). \tag{2}$$

Поскольку $\xi_n^r \perp S(\xi)$, то из (2) следует, что

$$S(\xi) \subseteq H_n(\xi^s),$$

и, значит, $S(\xi) \subseteq S(\xi^s) \subseteq H(\xi^s)$. Но $\xi_n^s \in S(\xi)$, поэтому $H(\xi^s) \subseteq S(\xi)$ и, следовательно,

$$S(\xi) = S(\xi^s) = H(\xi^s),$$

что означает сингулярность последовательности ξ^s.

Ортогональность последовательностей ξ^s и ξ^r следует очевидным образом из того, что $\xi_n^s \in S(\xi)$, а $\xi_n^r \perp S(\xi)$.

Докажем теперь единственность разложения (1). Пусть $\xi_n = \eta_n^r + \eta_n^s$, где η^r и η^s — регулярные и сингулярные ортогональные последовательности. Тогда, поскольку $H_n(\eta^s) = H(\eta^s)$, то

$$H_n(\xi) = H_n(\eta^r) \oplus H_n(\eta^s) = H_n(\eta^r) \oplus H(\eta^s),$$

и поэтому $S(\xi) = S(\eta^r) \oplus H(\eta^s)$. Но $S(\eta^r)$ тривиально и, значит, $S(\xi) = H(\eta^s)$.

Поскольку $\eta_n^s \in H(\eta^s) = S(\xi)$, а $\eta_n^r \perp H(\eta^s) = S(\xi)$, то $\hat{\mathsf{M}}(\xi_n \mid S(\xi)) = \hat{\mathsf{M}}(\eta_n^r + \eta_n^s \mid S(\xi)) = \eta_n^s$, т. е. совпадает с ξ_n^s, что и доказывает единственность разложения (1).

Теорема доказана.

2. Определение 2. Пусть $\xi = (\xi_n)$ — невырожденная стационарная последовательность. Случайную последовательность $\varepsilon = (\varepsilon_n)$ назовем *обновляющей* последовательностью (для ξ), если:

a) $\varepsilon = (\varepsilon_n)$ состоит из попарно ортогональных случайных величин с $\mathsf{M}\varepsilon_n = 0$, $\mathsf{M}|\varepsilon_n|^2 = 1$;

b) $H_n(\xi) = H_n(\varepsilon)$ для любого $n \in \mathbb{Z}$.

Замечание. Смысл термина «обновление» обусловлен ассоциацией с тем, что ε_{n+1} как бы привносит новую «информацию», не содержащуюся в $H_n(\xi)$ (иначе — «обновляет информацию» в $H_n(\xi)$, которая необходима для образования $H_{n+1}(\xi)$.

Следующая важная теорема устанавливает связь между введенными выше (пример 4 в § 1) последовательностями одностороннего скользящего среднего и регулярными последовательностями.

Теорема 2. *Для того чтобы невырожденная последовательность ξ была регулярной, необходимо и достаточно, чтобы нашлась такая обновляющая последовательность $\varepsilon = (\varepsilon_n)$ и последовательность комплексных чисел (a_n), $n \geq 0$, с $\sum_{n=0}^{\infty} |a_n|^2 < \infty$, что* ($\mathsf{P}$-п. н.)

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}. \tag{3}$$

Доказательство. Необходимость. Представим $H_n(\xi)$ в виде

$$H_n(\xi) = H_{n-1}(\xi) \oplus B_n(\xi_n),$$

где $B_n(\xi_n)$ есть пространство случайных величин вида $\beta \cdot \xi_n$, где β — комплексные числа. Пространство $H_n(\xi)$ не может совпадать с $H_{n-1}(\xi)$ ни при одном n. В самом деле, если при каком-то n $B_n(\xi_n)$ тривиально, то в силу стационарности тривиальными будут пространства $B_k(\xi_k)$ при всех k, а, значит, тогда $H(\xi) = S(\xi)$, что противоречит предположению о регулярности последовательности ξ. Итак, пространство $B_n(\xi_n)$ содержит заведомо хотя бы один ненулевой элемент, скажем, η_n. Положим

$$\varepsilon_n = \frac{\eta_n}{\|\eta_n\|},$$

где $\|\eta_n\|^2 = \mathsf{M}|\eta_n|^2 > 0$.

Для фиксированных n и $k \geqslant 0$ рассмотрим разложения

$$H_n(\xi) = H_{n-k}(\xi) \oplus B_{n-k+1}(\xi_{n-k+1}) \oplus \ldots \oplus B_n(\xi_n).$$

Тогда $\varepsilon_{n-k}, \ldots, \varepsilon_n$ образуют ортонормированный базис в $B_{n-k+1}(\xi_{n-k+1}) \oplus \ldots \oplus B_n(\varepsilon_n)$ и

$$\xi_n = \sum_{j=0}^{k-1} a_j \varepsilon_{n-j} + \hat{\pi}_{n-k}(\xi_n), \tag{4}$$

где $a_j = \mathsf{M}\xi_n \bar{\varepsilon}_{n-j}$.

В силу неравенства Бесселя (II.11.6)

$$\sum_{j=0}^{\infty} |a_j|^2 \leqslant \|\xi_n\|^2 < \infty.$$

Отсюда следует, что ряд $\sum_{j=0}^{\infty} a_j \varepsilon_{n-j}$ сходится в среднеквадратическом смысле, и в силу (4) для доказательства (3) осталось лишь доказать, что $\hat{\pi}_{n-k}(\xi_n) \xrightarrow{L^2} 0$, $k \to \infty$.

Достаточно рассмотреть случай $n = 0$. Поскольку

$$\hat{\pi}_{-k} = \hat{\pi}_0 + \sum_{i=0}^{k} [\hat{\pi}_{-i} - \hat{\pi}_{-i+1}],$$

а слагаемые, участвующие в сумме, ортогональны, то для любого $k \geqslant 0$

$$\sum_{i=0}^{k} \|\hat{\pi}_{-i} - \hat{\pi}_{-i+1}\|^2 = \left\| \sum_{i=0}^{k} (\hat{\pi}_{-i} - \hat{\pi}_{-i+1}) \right\|^2 =$$
$$= \|\hat{\pi}_{-k} - \hat{\pi}_0\|^2 \leqslant 4\|\xi_0\|^2 < \infty.$$

Поэтому существует (в среднеквадратическом смысле) предел $\lim_{k\to\infty} \hat{\pi}_{-k}$. Для каждого k $\hat{\pi}_{-k} \in H_{-k}(\xi)$, и, значит, рассматриваемый предел должен принадлежать подпространству $\bigcap_{k\geqslant 0} H_{-k}(\xi) = S(\xi)$. Но по предположению $S(\xi)$ тривиально, и поэтому $\hat{\pi}_{-k} \xrightarrow{L^2} 0$, $k \to \infty$.

Достаточность. Пусть невырожденная последовательность ξ допускает представление в виде (3), где $\varepsilon = (\varepsilon_n)$ — некоторая ортонормированная система (не обязательно удовлетворяющая условию $H_n(\xi) = H_n(\varepsilon)$, $n \in \mathbb{Z}$). Тогда $H_n(\xi) \subseteq H_n(\varepsilon)$ и, значит, $S(\xi) =$

$=\bigcap_k H_k(\xi) \subseteq H_n(\varepsilon)$ для любого n. Но $\varepsilon_{n+1} \perp H_n(\varepsilon)$, поэтому $\varepsilon_{n+1} \perp S(\xi)$ и в то же самое время $\varepsilon = (\varepsilon_n)$ является базисом в $H(\xi)$. Отсюда следует, что подпространство $S(\xi)$ является тривиальным, и, следовательно, последовательность ξ регулярна.

Теорема доказана.

Замечание. Из проведенного доказательства следует, что невырожденная последовательность ξ является регулярной тогда и только тогда, когда она допускает представление в виде одностороннего скользящего среднего

$$\xi_n = \sum_{k=0}^{\infty} \tilde{a}_k \tilde{\varepsilon}_{n-k}, \tag{5}$$

где $\tilde{\varepsilon} = (\tilde{\varepsilon}_n)$ — некоторая ортонормированная система, которая (это важно подчеркнуть!) не обязательно удовлетворяет условию $H_n(\xi) = H_n(\tilde{\varepsilon})$, $n \in \mathbb{Z}$. В этом смысле утверждение теоремы 2 говорит о большем, а именно о том, что для регулярной последовательности ξ найдутся такие $a = (a_n)$ и ортонормированная система $\varepsilon = (\varepsilon_n)$, что наряду с (5) будет справедливо представление (3), для которого $H_n(\xi) = H_n(\varepsilon)$, $n \in \mathbb{Z}$.

Из теорем 1 и 2 непосредственно вытекает

Теорема 3 (разложение Вольда). *Если $\xi = (\xi_n)$ — невырожденная стационарная последовательность, то*

$$\xi_n = \xi_n^s + \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}, \tag{6}$$

где $\sum_{k=0}^{\infty} |a_k|^2 < \infty$ и $\varepsilon = (\varepsilon_n)$ — некоторая обновляющая последовательность (для ξ^r).

3. Смысл введенных выше понятий регулярной и сингулярной последовательностей становится особенно ясным при рассмотрении следующей задачи (линейной) экстраполяции, для общего решения которой оказывается весьма полезным использование разложения Вольда (5).

Пусть $H_0(\xi) = \overline{L}^2(\xi^0)$ — замкнутое линейное многообразие, порожденное величинами $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$. Рассмотрим задачу построения *оптимальной* (в среднеквадратическом смысле) *линейной оценки* $\hat{\xi}_n$ величины ξ_n по «прошлым» наблюдениям $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$.

Из § 11 гл. II следует, что

$$\hat{\xi}_n = \hat{\mathsf{M}}(\xi_n \mid H_0(\xi)). \tag{7}$$

(В обозначениях п. 1 $\hat{\xi}_n = \hat{\pi}_0(\xi_n)$.) Поскольку ξ^r и ξ^s ортогональны и $H_0(\xi) = H_0(\xi^r) \oplus H_0(\xi^s)$, то с учетом (6) находим

$$\begin{aligned}\hat{\xi}_n &= \hat{\mathsf{M}}(\xi_n^s + \xi_n^r \mid H_0(\xi)) = \hat{\mathsf{M}}(\xi_n^s \mid H_0(\xi)) + \hat{\mathsf{M}}(\xi_n^r \mid H_0(\xi)) = \\ &= \hat{\mathsf{M}}(\xi_n^s \mid H_0(\xi^r) \oplus H_0(\xi^s)) + \hat{\mathsf{M}}(\xi_n^r \mid H_0(\xi^r) \oplus H_0(\xi^s)) = \\ &= \hat{\mathsf{M}}(\xi_n^s \mid H_0(\xi^s)) + \hat{\mathsf{M}}(\xi_n^r \mid H_0(\xi^r)) = \\ &= \xi_n^s + \hat{\mathsf{M}}\left(\sum_{k=0}^{\infty} a_k \varepsilon_{n-k} \mid H_0(\xi^r)\right).\end{aligned}$$

В (6) последовательность $\varepsilon = (\varepsilon_n)$ является обновляющей для $\xi^r = (\xi_n^r)$ и, значит, $H_0(\xi^r) = H_0(\varepsilon)$. Поэтому

$$\hat{\xi}_n = \xi_n^s + \hat{\mathsf{M}}\left(\sum_{k=0}^{\infty} a_k \varepsilon_{n-k} \mid H_0(\varepsilon)\right) = \xi_n^s + \sum_{k=n}^{\infty} a_k \varepsilon_{n-k} \tag{8}$$

и среднеквадратическая ошибка предсказания ξ_n по $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$ равна

$$\sigma_n^2 = \mathsf{M} \left| \xi_n - \hat{\xi}_n \right|^2 = \sum_{k=0}^{n-1} |a_k|^2. \tag{9}$$

Отсюда вытекают следующие два важных вывода.

a) Если последовательность ξ *сингулярна*, то для любого $n \geqslant 1$ ошибка (экстраполяции) σ_n^2 равна нулю, иначе говоря, возможно безошибочное предсказание ξ_n по «прошлому» $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$.

b) Если последовательность ξ *регулярна*, то $\sigma_n^2 \leqslant \sigma_{n+1}^2$ и

$$\lim_{n\to\infty} \sigma_n^2 = \sum_{k=0}^{\infty} |a_k|^2. \tag{10}$$

Поскольку

$$\sum_{k=0}^{\infty} |a_k|^2 = \mathsf{M} |\xi_n|^2,$$

то из (10) и (9) следует, что

$$\hat{\xi}_n \xrightarrow{L^2} 0 \qquad n \to \infty,$$

т. е. с ростом n прогноз величины ξ_n по $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$ становится тривиальным (совпадающим просто с $\mathsf{M}\xi_n = 0$).

4. Будем предполагать, что ξ — невырожденная регулярная стационарная последовательность. Согласно теореме 2 всякая такая последовательность допускает представление в виде одностороннего скользящего среднего

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}, \tag{11}$$

где $\sum_{k=0}^{\infty} |a_k|^2 < \infty$ и ортонормированная последовательность $\varepsilon = (\varepsilon_n)$ обладает тем важным свойством, что

$$H_n(\xi) = H_n(\varepsilon), \qquad n \in \mathbb{Z}. \tag{12}$$

Представление (11) означает (см. п. 3 § 3), что ξ_n можно рассматривать как сигнал на выходе физически осуществимого фильтра с импульсной переходной функцией $a = (a_k)$, $k \geqslant 0$, когда на вход подается последовательность $\varepsilon = (\varepsilon_n)$.

Как и всякая последовательность двустороннего скользящего среднего, регулярная последовательность имеет спектральную плотность $f(\lambda)$. Но то обстоятельство, что регулярная последовательность допускает представление в виде одностороннего скользящего среднего, позволяет получить дополнительную информацию о свойствах спектральной плотности.

Прежде всего ясно, что

$$f(\lambda) = \frac{1}{2\pi} |\varphi(\lambda)|^2,$$

где

$$\varphi(\lambda) = \sum_{k=0}^{\infty} e^{-i\lambda k} a_k, \qquad \sum_{k=0}^{\infty} |a_k|^2 < \infty. \tag{13}$$

Положим

$$\Phi(z) = \sum_{k=0}^{\infty} a_k z^k. \tag{14}$$

Эта функция является аналитической в открытой области $|z| < 1$ и в силу условия $\sum_{k=0}^{\infty} |a_k|^2 < \infty$ принадлежит так называемому *классу Харди* H^2, т. е. классу аналитических в области $|z| < 1$ функций $g = g(z)$, для которых

$$\sup_{0 \leqslant r < 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(re^{i\theta})|^2 \, d\theta < \infty. \tag{15}$$

Действительно,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\Phi(re^{i\theta})|^2 \, d\theta = \sum_{k=0}^{\infty} |a_k|^2 r^{2k}$$

и

$$\sup_{0 \leqslant r < 1} \sum |a_k|^2 r^{2k} \leqslant \sum |a_k|^2 < \infty.$$

В теории функций комплексного переменного доказывается, что граничное значение $\Phi(e^{i\lambda})$, $-\pi \leqslant \lambda < \pi$, тождественно не равной нулю функции $\Phi \in H^2$ обладает тем свойством, что

$$\int_{-\pi}^{\pi} \ln |\Phi(e^{-i\lambda})| \, d\lambda > -\infty. \tag{16}$$

В рассматриваемом нами случае

$$f(\lambda) = \frac{1}{2\pi} |\Phi(e^{-i\lambda})|^2,$$

где $\Phi \in H^2$. Поэтому

$$\ln f(\lambda) = -\ln 2\pi + 2 \ln |\Phi(e^{-i\lambda})|,$$

и, следовательно, спектральная плотность $f(\lambda)$ регулярного процесса удовлетворяет условию

$$\int_{-\pi}^{\pi} \ln f(\lambda) \, d\lambda > -\infty. \tag{17}$$

С другой стороны, пусть спектральная плотность $f(\lambda)$ такова, что выполнено условие (17). Опять-таки из теории функций комплексного переменного следует, что тогда найдется такая функция $\Phi(z) = \sum_{k=0}^{\infty} a_k z^k$, принадлежащая классу Харди H^2, что (почти всюду по мере Лебега)

$$f(\lambda) = \frac{1}{2\pi} |\Phi(e^{-i\lambda})|^2.$$

Поэтому, полагая $\varphi(\lambda) = \Phi(e^{-i\lambda})$, получаем

$$f(\lambda) = \frac{1}{2\pi} |\varphi(\lambda)|^2,$$

где $\varphi(\lambda)$ задается формулой (13). Тогда из следствия к теореме 3 § 3 вытекает, что последовательность ξ допускает представление в виде одностороннего скользящего среднего (11); где $\varepsilon = (\varepsilon_n)$ — некоторая ортонормированная последовательность. Отсюда и из замечания к теореме 2 следует, что последовательность ξ регулярна.

Итак, имеет место

Теорема 4 (Колмогоров). *Пусть ξ — невырожденная регулярная стационарная последовательность. Тогда существует спектральная плотность $f(\lambda)$ такая, что*

$$\int_{-\pi}^{\pi} \ln f(\lambda) \, d\lambda > -\infty. \tag{18}$$

В частности, $f(\lambda) > 0$ (почти всюду по мере Лебега).

Обратно, если ξ — некоторая стационарная последовательность, имеющая спектральную плотность, удовлетворяющую условию (18), *то эта последовательность является регулярной.*

5. Задачи.

1. Показать, что стационарная последовательность с дискретным спектром (спектральная функция $F(\lambda)$ — кусочно-постоянна) является сингулярной.

2. Пусть $\sigma_n^2 = \mathsf{M}\,|\xi_n - \hat{\xi}_n|^2$, $\hat{\xi}_n = \hat{\mathsf{M}}(\xi_n \,|\, H_0(\xi))$. Показать, что если для некоторого $n \geqslant 1$ $\sigma_n^2 = 0$, то последовательность ξ является сингулярной; если же при $n \to \infty$ $\sigma_n^2 \to R(0)$, то — регулярной.

3. Показать, что стационарная последовательность $\xi = (\xi_n)$, $\xi_n = e^{in\varphi}$, где φ — равномерная случайная величина на $[0, 2\pi]$, является регулярной. Найти оценку $\hat{\xi}_n$, величину σ_n^2 и показать, что *нелинейная* оценка

$$\tilde{\xi}_n = \left(\frac{\xi_0}{\xi_{-1}}\right)^n$$

дает безошибочный прогноз ξ_n по «прошлому» $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$, т. е.

$$\mathsf{M}\,|\tilde{\xi}_n - \xi_n|^2 = 0, \qquad n \geqslant 1.$$

§ 6. Экстраполяция, интерполяция и фильтрация

1. Экстраполяция. В соответствии с результатами предыдущего параграфа сингулярные последовательности допускают безошибочный прогноз (экстраполяцию) величин ξ_n, $n \geqslant 1$, по «прошлому» $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$. Естественно поэтому при рассмотрении задач экстраполяции для произвольных стационарных последовательностей изучить сначала случай регулярных последовательностей.

Согласно теореме 2 из § 5 всякая регулярная последовательность $\xi = (\xi_n)$ допускает представление в виде одностороннего скользящего среднего,

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k} \tag{1}$$

с $\sum_{k=0}^{\infty} |a_k|^2 < \infty$ и некоторой обновляющей последовательностью $\varepsilon = (\varepsilon_n)$. Представление (1), как следует из § 5, решает задачу нахождения оптимальной (линейной) оценки $\hat{\xi} = \hat{\mathsf{M}}(\xi_n \,|\, H_0(\xi))$, поскольку, согласно (5.8),

$$\hat{\xi}_n = \sum_{k=n}^{\infty} a_k \varepsilon_{n-k} \tag{2}$$

и

$$\sigma_n^2 = \mathsf{M}\,|\xi_n - \hat{\xi}_n|^2 = \sum_{k=0}^{n-1} |a_k|^2. \tag{3}$$

Однако это решение можно считать лишь принципиальным решением в силу следующего обстоятельства.

Обычно рассматриваемые последовательности задаются не представлением (1), а с помощью задания их ковариационной функции $R(n)$ или спектральной плотности $f(\lambda)$ (которая существует для регулярных последовательностей). Поэтому решение (2) можно признать удовлетворительным, если коэффициенты a_k будут выражены через значения $R(n)$ или $f(\lambda)$, а величины ε_k — через значения $\ldots \xi_{k-1}, \xi_k$.

Не затрагивая эту проблему в ее общем виде, ограничимся рассмотрением одного частного (но интересного для приложений) случая, когда спектральная плотность представляется в виде

$$f(\lambda) = \frac{1}{2\pi} |\Phi(e^{-i\lambda})|^2, \tag{4}$$

где функция $\Phi(z) = \sum_{k=0}^{\infty} b_k z^k$ имеет радиус сходимости $r > 1$ и не имеет нулей в области $|z| \leqslant 1$.

Пусть

$$\xi_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z(d\lambda) \tag{5}$$

— спектральное представление последовательности $\xi = (\xi_n)$, $n \in \mathbb{Z}$.

Теорема 1. *Если спектральная плотность последовательности ξ представима в виде* (4), *то оптимальная* (*линейная*) *оценка $\hat{\xi}_n$ величины ξ_n по $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$ задается формулой*

$$\hat{\xi}_n = \int_{-\pi}^{\pi} \hat{\varphi}_n(\lambda) Z(d\lambda), \tag{6}$$

где

$$\hat{\varphi}_n(\lambda) = e^{i\lambda n} \frac{\Phi_n(e^{-i\lambda})}{\Phi(e^{-i\lambda})} \tag{7}$$

и

$$\Phi_n(z) = \sum_{k=n}^{\infty} b_k z^k.$$

Доказательство. Согласно замечанию к теореме 2 § 3 всякая величина $\tilde{\xi}_n \in H_0(\xi)$ допускает представление в виде

$$\tilde{\xi}_n = \int_{-\pi}^{\pi} \tilde{\varphi}_n(\lambda) Z(d\lambda), \qquad \tilde{\varphi}_n \in H_0(F), \tag{8}$$

где $H_0(F)$ — замкнутое линейное многообразие, порожденное функциями $e_n = e^{i\lambda n}$ с $n \leqslant 0$ $\left(F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)\, d\nu\right)$.

Поскольку

$$\mathsf{M}\,|\xi_n - \tilde{\xi}_n|^2 = \mathsf{M}\,\left|\int_{-\pi}^{\pi} (e^{i\lambda n} - \tilde{\varphi}_n(\lambda))\, Z(d\lambda)\right|^2 = \int_{-\pi}^{\pi} |e^{i\lambda n} - \tilde{\varphi}_n(\lambda)|^2 f(\lambda)\, d\lambda,$$

то доказательство оптимальности оценки (6) сводится к доказательству того, что

$$\inf_{\tilde{\varphi}_n \in H_0(F)} \int_{-\pi}^{\pi} |e^{i\lambda n} - \tilde{\varphi}_n(\lambda)|^2 f(\lambda)\, d\lambda = \int_{-\pi}^{\pi} |e^{i\lambda n} - \hat{\varphi}_n(\lambda)|^2 f(\lambda)\, d\lambda. \tag{9}$$

Из теории гильбертовых пространств (§ 11 гл. II) следует, что оптимальная (в смысле (9)) функция $\hat{\varphi}_n(\lambda)$ определяется двумя условиями:

1) $\hat{\varphi}_n(\lambda) \in H_0(F)$, (10)

2) $e^{i\lambda n} - \hat{\varphi}_n(\lambda) \perp H_0(F)$.

Поскольку

$$e^{i\lambda n}\Phi_n(e^{-i\lambda}) = e^{i\lambda n}[b_n e^{-i\lambda n} + b_{n+1} e^{-i\lambda(n+1)} + \ldots] \in H_0(F)$$

и аналогичным образом $\frac{1}{\Phi(e^{-i\lambda})} \in H_0(F)$, то функция $\hat{\varphi}_n(\lambda)$, определенная в (7), принадлежит классу $H_0(F)$. Поэтому для доказательства оптимальности функции $\hat{\varphi}_n(\lambda)$ достаточно лишь проверить, что для любого $m \geqslant 0$

$$e^{i\lambda n} - \hat{\varphi}_n(\lambda) \perp e^{i\lambda m},$$

т. е.

$$I_{n,m} \equiv \int_{-\pi}^{\pi} [e^{i\lambda n} - \hat{\varphi}_n(\lambda)]\, e^{-i\lambda m} f(\lambda)\, d\lambda = 0, \qquad m \geqslant 0.$$

Следующая цепочка равенств показывает, что это действительно так:

$$\begin{aligned} I_{n,m} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda(n-m)}\left[1 - \frac{\Phi_n(e^{-i\lambda})}{\Phi(e^{-i\lambda})}\right] |\Phi(e^{-i\lambda})|^2\, d\lambda = \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda(n-m)}\,[\Phi(e^{-i\lambda}) - \Phi_n(e^{-i\lambda})]\,\overline{\Phi(e^{-i\lambda})}\, d\lambda = \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda(n-m)}\left(\sum_{k=0}^{n-1} b_k e^{-i\lambda k}\right)\left(\sum_{l=0}^{\infty} \bar{b}_l e^{i\lambda l}\right) d\lambda = \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\lambda m}\left(\sum_{k=0}^{n-1} b_k e^{i\lambda(n-k)}\right)\left(\sum_{l=0}^{\infty} \bar{b}_l e^{i\lambda l}\right) d\lambda = 0, \end{aligned}$$

где последнее равенство следует из того, что для $m \geqslant 0$ и $r > 1$

$$\int_{-\pi}^{\pi} e^{-i\lambda m} e^{i\lambda r}\, d\lambda = 0.$$

Теорема доказана.

Замечание 1. Разлагая функцию $\hat{\varphi}_n(\lambda)$ в ряд Фурье

$$\hat{\varphi}_n(\lambda) = C_0 + C_{-1}e^{-i\lambda} + C_{-2}e^{-2i\lambda} + \ldots,$$

находим, что прогноз $\hat{\xi}_n$ величины ξ_n, $n \geqslant 1$, по прошлому $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$ определяется формулой

$$\hat{\xi}_n = C_0\xi_0 + C_{-1}\xi_{-1} + C_{-2}\xi_{-2} + \ldots$$

Замечание 2. Типичным примером спектральной плотности, представимой в виде (4), является *рациональная* функция

$$f(\lambda) = \frac{1}{2\pi}\left|\frac{P(e^{-i\lambda})}{Q(e^{-i\lambda})}\right|^2,$$

где полиномы $P(z) = a_0 + a_1 z + \ldots + a_p z^p$ и $Q(z) = 1 + b_1 z + \ldots + b_q z^q$ не имеют нулей в области $\{z\colon |z| \leqslant 1\}$.

Действительно, в этом случае достаточно положить $\Phi(z) = P(z)/Q(z)$. Тогда $\Phi(z) = \sum_{k=0}^{\infty} C_k z^k$, причем радиус сходимости этого ряда больше единицы.

Приведем два примера, иллюстрирующих теорему 1.

Пример 1. Пусть спектральная плотность

$$f(\lambda) = \frac{1}{2\pi}(5 + 4\cos\lambda).$$

Соответствующая ковариационная функция $R(n)$ имеет «треугольный» вид:

$$R(0) = 5, \qquad R(\pm 1) = 2, \qquad R(n) = 0 \text{ при } |n| \geqslant 2. \tag{11}$$

Поскольку рассматриваемая спектральная плотность может быть представлена в виде

$$f(\lambda) = \frac{1}{2\pi}|2 + e^{-i\lambda}|^2,$$

то возможно применение теоремы 1. Легко находим, что

$$\hat{\varphi}_1(\lambda) = e^{i\lambda}\frac{e^{-i\lambda}}{2 + e^{-i\lambda}}, \quad \hat{\varphi}_n(\lambda) = 0 \text{ при } n \geqslant 2. \tag{12}$$

Поэтому для всех $n \geqslant 2$ $\hat{\xi}_n = 0$, т. е. (линейный) прогноз значения ξ_n по $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$ является тривиальным, что совсем неудивительно, если заметить, что, согласно (11), корреляция между ξ_n и любой из величин $\xi_0, \xi_{-1}, \ldots$ равна нулю для $n \geqslant 2$.

Для $n=1$ из (6) и (12) находим, что

$$\hat{\xi}_1=\int_{-\pi}^{\pi} e^{i\lambda}\frac{e^{-i\lambda}}{2+e^{-i\lambda}}Z(d\lambda)=$$
$$=\frac{1}{2}\int_{-\pi}^{\pi}\frac{1}{\left(1+\frac{e^{-i\lambda}}{2}\right)}Z(d\lambda)=\sum_{k=0}^{\infty}\frac{(-1)^k}{2^{k+1}}\int_{-\pi}^{\pi}e^{-ik\lambda}Z(d\lambda)=$$
$$=\sum_{k=0}^{\infty}\frac{(-1)^k\xi_k}{2^{k+1}}=\frac{1}{2}\xi_0-\frac{1}{4}\xi_{-1}+\ldots$$

Пример 2. Пусть ковариационная функция

$$R(n)=a^n,\quad |a|<1.$$

Тогда (см. пример 5 в § 1)

$$f(\lambda)=\frac{1}{2\pi}\frac{1-|a|^2}{|1-ae^{-i\lambda}|^2},$$

т. е.

$$f(\lambda)=\frac{1}{2\pi}|\Phi(e^{-i\lambda})|^2,$$

где

$$\Phi(z)=\frac{(1-|a|^2)^{1/2}}{1-az}=(1-|a|^2)^{1/2}\sum_{k=0}^{\infty}(az)^k,$$

откуда $\hat{\varphi}_n(\lambda)=a^n$ и, значит,

$$\hat{\xi}_n=\int_{-\pi}^{\pi}a^nZ(d\lambda)=a^n\xi_0.$$

Иначе говоря, для прогнозирования величины ξ_n по наблюдениям $\xi^0=(\ldots,\xi_{-1},\xi_0)$ достаточно знания лишь последнего наблюдения ξ_0.

Замечание 3. Из разложения Вольда регулярной последовательности $\xi=(\xi_n)$ с

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k} \tag{13}$$

следует, что спектральная плотность $f(\lambda)$ допускает представление

$$f(\lambda)=\frac{1}{2\pi}|\Phi(e^{-i\lambda})|^2, \tag{14}$$

где

$$\Phi(z)=\sum_{k=0}^{\infty}a_kz^k. \tag{15}$$

Очевидно, что и обратно, если $f(\lambda)$ допускает представление (14) с функцией $\Phi(z)$ вида (15), то разложение Вольда для ξ_n имеет вид (13). Таким образом, задача представления спектральной плотности $f(\lambda)$ в виде (14) и задача отыскания коэффициентов a_k в разложении Вольда эквивалентны.

Сделанные в теореме I предположения относительно функции $\Phi(z)$ (отсутствие нулей в области $|z| \leqslant 1$ и $r > 1$) на самом деле не нужны для ее справедливости. Иначе говоря, если спектральная плотность регулярной последовательности представлена в виде (14), то оптимальная в среднеквадратическом смысле оценка $\hat{\xi}_n$ величины ξ_n по $\xi^0 = (\ldots, \xi_{-1}, \xi_0)$ определяется формулами (6) и (7).

Замечание 4. Теорема 1 (вместе с предшествующим замечанием) дает решение задачи прогноза для регулярной последовательности. Покажем, что на самом деле тот же ответ остается в силе и для произвольной стационарной последовательности. Точнее, пусть $\xi_n = \xi_n^s + \xi_n^r$, $\xi_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z(d\lambda)$, $F(\Delta) = \mathsf{M}|Z(\Delta)|^2$ и $f^r(\lambda) = \frac{1}{2\pi}|\Phi(e^{-i\lambda})|^2$ — спектральная плотность регулярной последовательности $\xi^r = (\xi_n^r)$. Тогда оценка $\hat{\xi}_n$ определяется формулами (6) и (7).

В самом деле (см. п. 3 § 5), пусть

$$\hat{\xi}_n = \int_{-\pi}^{\pi} \hat{\varphi}_n(\lambda)\, Z(d\lambda), \qquad \hat{\xi}_n^r = \int_{-\pi}^{\pi} \hat{\varphi}_n^r(\lambda)\, Z^r(d\lambda),$$

где $Z^r(\Delta)$ — ортогональная стохастическая мера в представлении регулярной последовательности ξ^r. Тогда

$$\begin{aligned} \mathsf{M}|\xi_n - \hat{\xi}_n|^2 &= \int_{-\pi}^{\pi} |e^{i\lambda n} - \hat{\varphi}_n(\lambda)|^2 F(d\lambda) \geqslant \\ &\geqslant \int_{-\pi}^{\pi} |e^{i\lambda n} - \hat{\varphi}_n(\lambda)|^2 f^r(\lambda)\, d\lambda \geqslant \int_{-\pi}^{\pi} |e^{i\lambda n} - \hat{\varphi}_n^r(\lambda)|^2 f^r(\lambda)\, d\lambda = \\ &= \mathsf{M}|\xi_n^r - \hat{\xi}_n^r|^2. \end{aligned} \tag{16}$$

Но $\xi_n - \hat{\xi}_n = \xi_n^r - \hat{\xi}^r$, поэтому $\mathsf{M}|\xi_n - \hat{\xi}_n|^2 = \mathsf{M}|\xi_n^r - \hat{\xi}_n^r|^2$, и из (16) следует, что в качестве $\hat{\varphi}_n(\lambda)$ можно взять функцию $\hat{\varphi}_n^r(\lambda)$.

2. Интерполяция. Будем предполагать, что $\xi = (\xi_n)$ — регулярная последовательность со спектральной плотностью $f(\lambda)$. Простейшей задачей интерполяции является задача построения оптимальной (в среднеквадратическом смысле) линейной оценки по результатам наблюдений $\{\xi_n, n = \pm 1, \pm 2, \ldots\}$ «пропущенного» значения ξ_0.

Обозначим через $H^0(\xi)$ — замкнутое линейное многообразие, порожденное величинами ξ_n, $n \neq 0$. Тогда в соответствии с теоремой 2 § 3 всякая случайная величина $\eta \in H^0(\xi)$ представима в виде

$$\eta = \int_{-\pi}^{\pi} \varphi(\lambda) Z(d\lambda),$$

где φ принадлежит $H^0(F)$ — замкнутому линейному многообразию, порожденному функциями $e^{i\lambda n}$, $n \neq 0$, и оценка

$$\check{\xi}_0 = \int_{-\pi}^{\pi} \check{\varphi}(\lambda) Z(d\lambda) \tag{17}$$

будет оптимальной тогда и только тогда, когда

$$\inf_{\eta \in H^0(\xi)} \mathsf{M} |\xi_0 - \eta|^2 = \inf_{\varphi \in H^0(F)} \int_{-\pi}^{\pi} |1 - \varphi(\lambda)|^2 F(d\lambda) = \\ = \int_{-\pi}^{\pi} |1 - \check{\varphi}(\lambda)|^2 F(d\lambda) = \mathsf{M} |\xi_0 - \check{\xi}_0|^2.$$

Из свойств «перпендикуляров» в гильбертовом пространстве $H^0(F)$ вытекает, что функция $\check{\varphi}(\lambda)$ полностью определяется (ср. с (16)) двумя условиями

$$\begin{aligned} &1)\ \check{\varphi}(\lambda) \in H^0(F), \\ &2)\ 1 - \check{\varphi}(\lambda) \perp H^0(F). \end{aligned} \tag{18}$$

Теорема 2 (Колмогоров). *Пусть $\xi = (\xi_n)$ — регулярная последовательность с*

$$\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)} < \infty. \tag{19}$$

Тогда

$$\check{\varphi}(\lambda) = 1 - \frac{\alpha}{f(\lambda)}, \tag{20}$$

где

$$\alpha = \frac{2\pi}{\displaystyle\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}}, \tag{21}$$

и ошибка интерполяции $\delta^2 = \mathsf{M} |\xi_0 - \check{\xi}_0|^2$ задается формулой $\delta^2 = 2\pi \cdot \alpha$.

Доказательство проведем лишь при весьма строгих предположениях относительно спектральной плотности, считая, что

$$0 < c \leqslant f(\lambda) \leqslant C < \infty. \tag{22}$$

Из условия 2) в (18) следует, что для любого $n \neq 0$

$$\int_{-\pi}^{\pi} [1 - \check{\varphi}(\lambda)] e^{in\lambda} f(\lambda)\, d\lambda = 0. \tag{23}$$

В силу предположения (22) функция $[1-\check{\varphi}(\lambda)] f(\lambda)$ принадлежит гильбертову пространству $L^2([-\pi, \pi], \mathscr{B}[-\pi, \pi], \mu)$ с мерой Лебега μ. В этом пространстве система функций $\left\{\frac{e^{i\lambda n}}{\sqrt{2\pi}},\ n=0, \pm 1, \ldots\right\}$ образует ортонормированный базис (задача 7 § 11 гл. II). Поэтому из (23) следует, что функция $[1-\check{\varphi}(\lambda)] f(\lambda)$ есть константа, которую обозначим α.

Итак, второе условие в (18) приводит к тому, что

$$\check{\varphi}(\lambda) = 1 - \frac{\alpha}{f(\lambda)}. \tag{24}$$

Исходя из первого условия в (18), определим теперь константу α.

В силу (22) $\check{\varphi} \in L^2$ и условие $\check{\varphi} \in H^0(F)$ равносильно условию, что $\check{\varphi}$ принадлежит замкнутому (в смысле нормы в L^2) линейному многообразию, порожденному функциями $e^{i\lambda n}$, $n \neq 0$. Отсюда ясно, что нулевой коэффициент в разложении функции $\check{\varphi}(\lambda)$ должен быть равен нулю. Поэтому

$$0 = \int_{-\pi}^{\pi} \check{\varphi}(\lambda)\, d\lambda = 2\pi - \alpha \int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}$$

и, значит, константа α определяется формулой (21).

Наконец,

$$\delta^2 = \mathsf{M}\,|\xi_0 - \check{\xi}_0|^2 = \int_{-\pi}^{\pi} |1 - \check{\varphi}(\lambda)|^2 f(\lambda)\, d\lambda =$$

$$= |\alpha|^2 \int_{-\pi}^{\pi} \frac{f(\lambda)}{f^2(\lambda)}\, d\lambda = \frac{4\pi^2}{\displaystyle\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}}.$$

Теорема (при дополнительном предположении (22)) доказана.

Следствие. Если

$$\check{\varphi}(\lambda) = \sum_{0 < |k| \leqslant N} c_k e^{i\lambda k},$$

то

Пример 3. Пусть $f(\lambda)$ — спектральная плотность из рассмотренного выше примера 2. Тогда нетрудно подсчитать, что

$$\check{\xi}_0 = \int_{-\pi}^{\pi} \frac{a}{1+|a|^2}[e^{i\lambda}+e^{-i\lambda}]Z(d\lambda) = \frac{a}{1+|a|^2}[\xi_1+\xi_{-1}],$$

а ошибка интерполяции равна

$$\delta^2 = \frac{1-|\alpha|^2}{1+|\alpha|^2}.$$

3. **Фильтрация.** Пусть $(\theta, \xi) = ((\theta_n), (\xi_n))$, $n \in \mathbb{Z}$, — *частично наблюдаемая последовательность,* где $\theta = (\theta_n)$ — ненаблюдаемая, а $\xi = (\xi_n)$ — наблюдаемая компонента. Каждая из последовательностей θ и ξ будет предполагаться стационарной (в широком смысле) с нулевыми средними и спектральными представлениями

$$\theta_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z_\theta(d\lambda), \qquad \xi_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z_\xi(d\lambda)$$

соответственно. Обозначим

$$F_\theta(\Delta) = \mathsf{M}|Z_\theta(\Delta)|^2, \qquad F_\xi(\Delta) = \mathsf{M}|Z_\xi(\Delta)|^2$$

и

$$F_{\theta\xi}(\Delta) = \mathsf{M}Z_\theta(\Delta)\overline{Z}_\xi(\Delta).$$

Кроме того, будем считать также, что θ и ξ *стационарно связаны,* т. е. их функция ковариации $\operatorname{cov}(\theta_n, \xi_m) = \mathsf{M}\theta_n\bar{\xi}_m$ зависит лишь от разности $n-m$. Обозначим $R_{\theta\xi}(n) = \mathsf{M}\theta_n\bar{\xi}_0$; тогда

$$R_{\xi\theta}(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F_{\theta\xi}(d\lambda).$$

Рассматриваемая задача фильтрации состоит в построении оптимальной (в среднеквадратическом смысле) линейной оценки $\hat{\theta}_n$ величины θ_n по тем или иным наблюдениям последовательности ξ.

Совсем просто эта задача решается в предположении, что оценка $\hat{\theta}_n$ строится *по всем* значениям ξ_m, $m \in \mathbb{Z}$. Действительно, поскольку $\hat{\theta}_n = \hat{\mathsf{M}}(\theta_n | H(\xi))$, то найдется такая функция $\hat{\varphi}_n(\lambda)$, что

$$\hat{\theta}_n = \int_{-\pi}^{\pi} \hat{\varphi}_n(\lambda) Z_\xi(d\lambda). \tag{25}$$

Как и в пп. 1 и 2, условия, которым должна удовлетворять «оптимальная» функция $\hat{\varphi}_n(\lambda)$, состоят в том, что:

1) $\hat{\varphi}_n(\lambda) \in H(F_\xi)$,

2) $\theta_n - \hat{\theta}_n \perp H(\xi)$.

Из последнего условия находим, что для любого $m \in \mathbb{Z}$

$$\int_{-\pi}^{\pi} e^{i\lambda(n-m)} F_{\theta\xi}(d\lambda) - \int_{-\pi}^{\pi} e^{-i\lambda m} \hat{\varphi}_n(\lambda) F_\xi(d\lambda) = 0. \tag{26}$$

Поэтому, если предположить, что функции $F_{\theta\xi}(\lambda)$ и $F_\xi(\lambda)$ имеют плотности $f_{\theta\xi}(\lambda)$ и $f_\xi(\lambda)$, то из (26) получим

$$\int_{-\pi}^{\pi} e^{i\lambda(n-m)} [f_{\theta\xi}(\lambda) - e^{-i\lambda n} \hat{\varphi}_n(\lambda) f_\xi(\lambda)]\, d\lambda = 0.$$

Если $f_\xi(\lambda) > 0$ (почти всюду по мере Лебега), то отсюда сразу находим, что

$$\hat{\varphi}_n(\lambda) = e^{i\lambda n} \hat{\varphi}(\lambda), \tag{27}$$

где

$$\hat{\varphi}(\lambda) = f_{\theta\xi}(\lambda) \cdot f_\xi^{\oplus}(\lambda)$$

и $f_\xi^{\oplus}(\lambda)$ — «псевдообращение» $f_\xi(\lambda)$, т. е.

$$f_\xi^{\oplus}(\lambda) = \begin{cases} f_\xi^{-1}(\lambda), & f_\xi(\lambda) > 0, \\ 0, & f_\xi(\lambda) = 0. \end{cases}$$

При этом ошибка фильтрации

$$\mathsf{M}\,|\theta_n - \hat{\theta}_n|^2 = \int_{-\pi}^{\pi} [f_\theta(\lambda) - f_{\theta\xi}^2(\lambda) f_\xi^{\oplus}(\lambda)]\, d\lambda. \tag{28}$$

Как нетрудно проверить, $\hat{\varphi}_n \in H(F_\xi)$ и, следовательно, оценка (25) с функцией (27) является оптимальной.

Пример 4. *Выделение сигнала из смеси с шумом.* Пусть $\xi_n = \theta_n + \eta_n$, где сигнал $\theta = (\theta_n)$ и шум $\eta = (\eta_n)$ являются некоррелированными последовательностями со спектральными плотностями $f_\theta(\lambda)$ и $f_\eta(\lambda)$. Тогда

$$\hat{\theta}_n = \int_{-\pi}^{\pi} e^{i\lambda n} \hat{\varphi}(\lambda)\, Z_\xi(d\lambda),$$

где

$$\hat{\varphi}(\lambda) = f_\theta(\lambda)\, [f_\theta(\lambda) + f_\eta(\lambda)]^{\oplus},$$

а ошибка фильтрации

$$\mathsf{M}\,|\theta_n - \hat{\theta}_n|^2 = \int_{-\pi}^{\pi} [f_\theta(\lambda) f_\eta(\lambda)]\, [f_\theta(\lambda) + f_\eta(\lambda)]^{\oplus}\, d\lambda.$$

Полученное решение (25) можно теперь использовать для построения оптимальной оценки $\tilde{\theta}_{n+m}$ величины θ_{n+m} по результатам наблюдений ξ_k, $k \leqslant n$, где m — некоторое заданное число из $\mathbb{Z}$. Предположим, что последовательность $\xi = (\xi_n)$ регулярна со спектральной плотностью

$$f(\lambda) = \frac{1}{2\pi} |\Phi(e^{-i\lambda})|^2,$$

где $\Phi(z) = \sum_{k=0}^{\infty} a_k z^k$. Согласно разложению Вольда

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k},$$

где $\varepsilon = (\varepsilon_k)$ — белый шум со спектральным разложением

$$\varepsilon_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z_\varepsilon(d\lambda).$$

Поскольку

$$\tilde{\theta}_{n+m} = \hat{\mathsf{M}}[\theta_{n+m} \mid H_n(\xi)] = \hat{\mathsf{M}}[\hat{\mathsf{M}}[\theta_{n+m} \mid H(\xi)] \mid H_n(\xi)] = \hat{\mathsf{M}}[\hat{\theta}_{n+m} \mid H_n(\xi)]$$

и

$$\hat{\theta}_{n+m} = \int_{-\pi}^{\pi} e^{i\lambda(n+m)} \hat{\varphi}(\lambda) \Phi(e^{-i\lambda}) Z_\varepsilon(d\lambda) = \sum_{k \leqslant n+m} \hat{a}_{n+m-k} \varepsilon_k,$$

где

$$\hat{a}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda k} \hat{\varphi}(\lambda) \Phi(e^{-i\lambda}) \, d\lambda, \tag{29}$$

то

$$\tilde{\theta}_{n+m} = \hat{\mathsf{M}}\left[\sum_{k \leqslant n+m} \hat{a}_{n+m-k} \varepsilon_k \mid H_n(\xi)\right].$$

Но $H_n(\xi) = H_n(\varepsilon)$, и, значит,

$$\tilde{\theta}_{n+m} = \sum_{k \leqslant n} \hat{a}_{n+m-k} \varepsilon_k = \int_{-\pi}^{\pi} \left[\sum_{k \leqslant n} \hat{a}_{n+m-k} e^{i\lambda k}\right] Z_\varepsilon(d\lambda) =$$
$$= \int_{-\pi}^{\pi} e^{i\lambda n} \left[\sum_{l=0}^{\infty} \hat{a}_{l+m} e^{-i\lambda l}\right] \Phi^{\oplus}(e^{-i\lambda}) Z_\xi(d\lambda),$$

где $\Phi^{\oplus}$ — псевдообращение Φ.

Итак, доказана следующая

Теорема 3. *Если наблюдаемая последовательность $\xi = (\xi_n)$ является регулярной, то оптимальная (в среднеквадратическом смысле) линейная оценка $\tilde{\theta}_{n+m}$ величины θ_{n+m} по ξ_k, $k \leqslant n$, задается* *формулой*

$$\tilde{\theta}_{n+m} = \int_{-\pi}^{\pi} e^{i\lambda n} H_m(e^{-i\lambda}) Z_\xi(d\lambda), \tag{30}$$

где

$$H_m(e^{-i\lambda}) = \sum_{l=0}^{\infty} \hat{a}_{l+m} e^{-i\lambda l} \Phi^{\oplus}(e^{-i\lambda}) \tag{31}$$

и коэффициенты $\hat{a}_k$ *определяются в* (29).

4. Задачи.

1. Пусть ξ — невырожденная регулярная последовательность со спектральной плотностью (4). Показать, что $\Phi(z)$ не имеет нулей при $|z| < 1$.

2. Доказать, что утверждение теоремы 1 сохраняет свою силу и без предположений, что $\Phi(z)$ имеет радиус сходимости $r > 1$, а нули $\Phi(z)$ лежат только в области $|z| > 1$.

3. Показать, что для регулярного процесса функция $\Phi(z)$, входящая в (4), может быть представлена в виде

$$\Phi(z) = \sqrt{2\pi} \exp\left\{\frac{1}{2} c_0 + \sum_{k=1}^{\infty} c_k z^k\right\}, \qquad |z| < 1,$$

где

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \ln f(\lambda)\, d\lambda.$$

Вывести отсюда и из (5.9), что ошибка прогноза на один шаг $\sigma_1^2 = \mathsf{M}|\hat{\xi}_1 - \xi_1|^2$ задается *формулой Сеге — Колмогорова*

$$\sigma_1^2 = 2\pi \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda\right\}.$$

4. Дать доказательство теоремы 2 без предположения (22).

5. Пусть некоррелированные сигнал θ и шум η имеют спектральные плотности

$$f_\theta(\lambda) = \frac{1}{2\pi} \cdot \frac{1}{|1 + b_1 e^{-i\lambda}|^2}, \quad f_\eta(\lambda) = \frac{1}{2\pi} \cdot \frac{1}{|1 + b_2 e^{-i\lambda}|^2}.$$

Опираясь на теорему 3, найти оценку $\tilde{\theta}_{n+m}$ величины θ_{n+m} по значениям ξ_k, $k \leqslant n$, где $\xi_k = \theta_k + \eta_k$. Рассмотреть ту же задачу для спектральных плотностей

$$f_\theta(\lambda) = \frac{1}{2\pi} |2 + e^{-i\lambda}|^2, \quad f_\eta(\lambda) = \frac{1}{2\pi}.$$

§ 7. Фильтр Калмана — Бьюси и его обобщения

1. С вычислительной точки зрения данное выше решение задачи фильтрации ненаблюдаемой компоненты θ по наблюдениям ξ не является удобным, поскольку, будучи выраженным в спектральных терминах, оно для своей реализации требует обращения к аналоговым устройствам. В схеме, предложенной Калманом и Бьюси, синтезирование оптимального фильтра осуществляется рекуррентным способом, что дает возможность реализации с помощью цифровых вычислительных устройств. Есть и другие причины, обусловившие широкое применение фильтра Калмана — Бьюси. Одна из них состоит в том, что он «работает» и без предположения *стационарности* последовательностей (θ, ξ).

Ниже будет рассматриваться не только традиционная схема Калмана — Бьюси, но также и ее обобщения, состоящие в том, что в рекуррентных уравнениях, определяющих (θ, ξ), коэффициенты могут зависеть от всех прошлых наблюдаемых данных.

Итак, будем предполагать, что $(\theta, \xi) = ((\theta_n), (\xi_n))$ есть частично наблюдаемая последовательность, причем

$$\theta_n = (\theta_1(n), \ldots, \theta_k(n)), \quad \xi_n = (\xi_1(n), \ldots, \xi_l(n))$$

управляются рекуррентными уравнениями

$$\begin{aligned} \theta_{n+1} &= a_0(n, \xi) + a_1(n, \xi)\theta_n + b_1(n, \xi)\varepsilon_1(n+1) + \\ &\qquad\qquad + b_2(n, \xi)\varepsilon_2(n+1), \\ \xi_{n+1} &= A_0(n, \xi) + A_1(n, \xi)\theta_n + B_1(n, \xi)\varepsilon_1(n+1) + \\ &\qquad\qquad + B_2(n, \xi)\varepsilon_2(n+1). \end{aligned} \tag{1}$$

Здесь $\varepsilon_1(n) = (\varepsilon_{11}(n), \ldots, \varepsilon_{1k}(n))$, $\varepsilon_2(n) = (\varepsilon_{21}(n), \ldots, \varepsilon_{2l}(n))$ — независимые гауссовские векторы с независимыми компонентами, каждая из которых имеет нормальное распределение с параметрами 0 и 1; $a_0(n, \xi) = (a_{01}(n, \xi), \ldots, a_{0k}(n, \xi))$ и $A_0(n, \xi) = (A_{01}(n, \xi), \ldots$ $\ldots, A_{0l}(n, \xi))$ — вектор-функции, где зависимость от $\xi = (\xi_0, \xi_1, \ldots)$ входит неупреждающим образом, т. е. для фиксированного n $a_{01}(n, \xi), \ldots, A_{0l}(n, \xi)$ зависят лишь от $\xi_0, \ldots, \xi_n$; матричные функции

$$\begin{aligned} b_1(n, \xi) &= \| b_{ij}^{(1)}(n, \xi) \|, & b_2(n, \xi) &= \| b_{ij}^{(2)}(n, \xi) \|, \\ B_1(n, \xi) &= \| B_{ij}^{(1)}(n, \xi) \|, & B_2(n, \xi) &= \| B_{ij}^{(2)}(n, \xi) \|, \\ a_1(n, \xi) &= \| a_{ij}^{(1)}(n, \xi) \|, & A_1(n, \xi) &= \| A_{ij}^{(1)}(n, \xi) \| \end{aligned}$$

имеют порядок $k \times k$, $k \times l$, $l \times k$, $l \times l$, $k \times k$, $l \times k$, соответственно, и также неупреждающим образом зависят от ξ. Предполагается также, что вектор начальных данных (θ_0, ξ_0) не зависит от последовательностей $\varepsilon_1 = (\varepsilon_1(n))$ и $\varepsilon_2 = (\varepsilon_2(n))$.

Для простоты изложения указание на зависимость коэффициентов от ξ в дальнейшем часто будет опускаться.

Чтобы система (1) имела решение с конечным вторым моментом, будем предполагать, что $\mathsf{M}(\|\theta_0\|^2+\|\xi_0\|^2)<\infty$ $\left(\|x\|^2=\sum_{i=1}^{k} x_i^2,\ x=(x_1, \ldots, x_k)\right)$, $|a_{ij}^{(1)}(n, \xi)|\leqslant C$, $|A_{ij}^{(1)}(n, \xi)|\leqslant C$, и если $g(n, \xi)$ — любая из функций a_{0i}, A_{0j}, $b_{ij}^{(1)}$, $b_{ij}^{(2)}$, $B_{ij}^{(1)}$, $B_{ij}^{(2)}$, то $\mathsf{M}|g(n, \xi)|^2<\infty$, $n=0, 1, \ldots$ В этих допущениях последовательность (θ, ξ) такова, что и $\mathsf{M}(\|\theta_n\|^2+\|\xi_n\|^2)<\infty$, $n\geqslant 0$.

Пусть, далее, $\mathscr{F}_n^\xi=\sigma\{\omega: \xi_0, \ldots, \xi_n\}$ — наименьшая σ-алгебра, порожденная величинами $\xi_0, \ldots, \xi_n$, и

$$m_n=\mathsf{M}(\theta_n|\mathscr{F}_n^\xi), \quad \gamma_n=\mathsf{M}\left[(\theta_n-m_n)(\theta_n-m_n)^*|\mathscr{F}_n^\xi\right].$$

Согласно теореме 1 § 8 гл. II $m_n=(m_1(n), \ldots, m_k(n))$ является оптимальной в среднеквадратическом смысле оценкой вектора $\theta_n=(\theta_1(n), \ldots, \theta_k(n))$, а $\mathsf{M}\gamma_n=\mathsf{M}[(\theta_n-m_n)(\theta_n-m_n)^*]$ есть матрица ошибок оценивания. Отыскание этих величин для произвольных последовательностей (θ, ξ), управляемых уравнениями (1), является весьма трудной задачей. Однако при одном дополнительном предположении относительно (θ_0, ξ_0), состоящем в том, что условное распределение $\mathsf{P}(\theta_0\leqslant a|\xi_0)$ является гауссовским,

$$\mathsf{P}(\theta_0\leqslant a|\xi_0)=\frac{1}{\sqrt{2\pi\gamma_0}}\int_{-\infty}^{a} e^{-\frac{(x-m_0)^2}{2\gamma_0^2}}\,dx \tag{2}$$

с параметрами $m_0=m_0(\xi_0)$, $\gamma_0=\gamma_0(\xi_0)$, для m_n и γ_n можно вывести систему рекуррентных уравнений, включающих в себя и так называемые уравнения фильтра Калмана — Бьюси.

Прежде всего установим один важный вспомогательный результат.

Лемма 1. *При сделанных выше предположениях относительно коэффициентов системы* (1) *и условии* (2) *последовательность* (θ, ξ) *является условно-гауссовской, т. е. условная функция распределения*

$$\mathsf{P}\left\{\theta_0\leqslant a_0, \ldots, \theta_n\leqslant a_n|\mathscr{F}_n^\xi\right\}$$

есть (P-*п. н.*) *функция распределения n-мерного гауссовского вектора, среднее значение и матрица ковариаций которого зависят от* $(\xi_0, \ldots, \xi_n)$.

Доказательство. Ограничимся доказательством гауссовости лишь распределения $\mathsf{P}(\theta_n\leqslant a|\mathscr{F}_n^\xi)$, что достаточно для вывода уравнений для m_n и γ_n.

Прежде всего заметим, что из (1) следует, что условное распределение

$$\mathsf{P}(\theta_{n+1}\leqslant a_1, \xi_{n+1}\leqslant x|\mathscr{F}_n^\xi, \theta_n=b)$$

является гауссовским с вектором средних значений

$$\mathbb{A}_0 + \mathbb{A}_1 b = \begin{pmatrix} a_0 + a_1 b \\ A_0 + A_1 b \end{pmatrix}$$

и матрицей ковариаций

$$\mathbb{B} = \begin{pmatrix} b \circ b & b \circ B \\ (b \circ B)^* & B \circ B \end{pmatrix},$$

где $b \circ b = b_1 b_1^* + b_2 b_2^*$, $b \circ B = b_1 B_1^* + b_2 B_2^*$, $B \circ B = B_1 B_1^* + B_2 B_2^*$.

Обозначим $\zeta_n = (\theta_n, \xi_n)$ и $t = (t_1, \ldots, t_{k+l})$. Тогда

$$\mathsf{M}\left[\exp(it^* \zeta_{n+1}) \mid \mathscr{F}_n^\xi, \theta_n\right] =$$
$$= \exp\left\{it^* (\mathbb{A}_0(n, \xi) + \mathbb{A}_1(n, \xi)\theta_n) - \frac{1}{2} t^* \mathbb{B}(n, \xi) t\right\}. \quad (3)$$

Допустим теперь, что утверждение леммы справедливо для некоторого $n \geqslant 0$. Тогда

$$\mathsf{M}\left[\exp(it^* \mathbb{A}_1(n, \xi)\theta_n) \mid \mathscr{F}_n^\xi\right] =$$
$$= \exp\left(it^* \mathbb{A}_1(n, \xi) m_n - \frac{1}{2} t^* (\mathbb{A}_1(n, \xi)\gamma_n \mathbb{A}_1^*(n, \xi)) t\right). \quad (4)$$

Докажем, что формула (4) останется верной и при замене n на $n+1$.

Из (3) и (4) имеем

$$\mathsf{M}\left[\exp(it^* \zeta_{n+1}) \mid \mathscr{F}_n^\xi\right] = \exp\{it^* (\mathbb{A}_0(n, \xi) + \mathbb{A}_1(n, \xi) m_n) -$$
$$- \frac{1}{2} t^* \mathbb{B}(n, \xi) t - \frac{1}{2} t^* (\mathbb{A}_1(n, \xi)\gamma_n \mathbb{A}_1^*(n, \xi)) t\right\}.$$

Поэтому условные распределения

$$\mathsf{P}\left(\theta_{n+1} \leqslant a, \ \xi_{n+1} \leqslant x \mid \mathscr{F}_n^\xi\right) \quad (5)$$

являются гауссовскими.

Как и при доказательстве теоремы о нормальной корреляции (теорема 2 в § 13 гл. II), проверяется, что существует такая матрица C, что вектор

$$\eta = \left[\theta_{n+1} - \mathsf{M}\left(\theta_{n+1} \mid \mathscr{F}_n^\xi\right)\right] - C\left[\xi_{n+1} - \mathsf{M}\left(\xi_{n+1} \mid \mathscr{F}_n^\xi\right)\right]$$

обладает тем свойством, что (P-п. н.)

$$\mathsf{M}\left[\eta\left(\xi_{n+1} - \mathsf{M}\left(\xi_{n+1} \mid \mathscr{F}_n^\xi\right)\right)^* \mid \mathscr{F}_n^\xi\right] = 0.$$

Отсюда следует, что условно-гауссовские векторы η и ξ_{n+1}, рассматриваемые при условии $\mathscr{F}_n^\xi$, являются независимыми, т. е.

$$\mathsf{P}\left(\eta \in A, \ \xi_{n+1} \in B \mid \mathscr{F}_n^\xi\right) = \mathsf{P}\left(\eta \in A \mid \mathscr{F}_n^\xi\right) \cdot \mathsf{P}\left(\xi_{n+1} \in B \mid \mathscr{F}_n^\xi\right)$$

для любых $A \in \mathscr{B}(R^k)$, $B \in \mathscr{B}(R^l)$.

Поэтому, если $s=(s_1, \ldots, s_k)$, то

$$\begin{aligned}
&\mathsf{M}\left[\exp(is^*\theta_{n+1}) \mid \mathscr{F}_n^\xi, \xi_{n+1}\right]= \\
&=\mathsf{M}\left\{\exp\left(is^*\left[\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)+\eta+C[\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)]\right]\right) \mid \mathscr{F}_n^\xi, \xi_{n+1}\right\}= \\
&\qquad=\exp\left\{is^*\left[\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)+C[\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)]\right]\right\}\times \\
&\qquad\times\mathsf{M}\left[\exp(is^*\eta) \mid \mathscr{F}_n^\xi, \xi_{n+1}\right]=\exp\left\{is^*\left[\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)\right]+\right. \\
&\qquad\qquad\left.+C\left[\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)\right]\right\}\mathsf{M}(\exp(is^*\eta) \mid \mathscr{F}_n^\xi]. \qquad (6)
\end{aligned}$$

В силу (5) условное распределение $\mathsf{P}(\eta \leqslant y \mid \mathscr{F}_n^\xi)$ является гауссовским. Вместе с (6) это доказывает, что условное распределение $\mathsf{P}(\theta_{n+1} \leqslant a \mid \mathscr{F}_{n+1}^\xi)$ также является гауссовским.

Лемма доказана.

Теорема 1. *Пусть* (θ, ξ) — *частично наблюдаемая последовательность, удовлетворяющая системе* (1). *Тогда* (m_n, γ_n) *подчиняются следующим рекуррентным уравнениям:*

$$\begin{aligned}
m_{n+1}=[a_0+a_1m_n]+[b\circ B+a_1\gamma_nA_1^*][B\circ B+A_1\gamma_nA_1^*]^{\oplus}\times& \\
\times[\xi_{n+1}-A_0-A_1m_n],& \qquad (7)\\
\gamma_{n+1}=[a_1\gamma_na_1^*+b\circ b]-[b\circ B+a_1\gamma_nA_1^*]\times& \\
\times[B\circ B+A_1\gamma_nA_1^*]^{\oplus}\cdot[b\circ B+a_1\gamma_nA_1^*]^*.& \qquad (8)
\end{aligned}$$

Доказательство. Из (1)

$$\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)=a_0+a_1m_n, \quad \mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)=A_0+A_1m_n \qquad (9)$$

и

$$\left.\begin{aligned}
\theta_{n+1}-\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)&=a_1[\theta_n-m_n]+b_1\varepsilon_1(n+1)+b_2\varepsilon_2(n+1),\\
\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)&=A_1[\theta_n-m_n]+B_1\varepsilon_1(n+1)+B_2\varepsilon_2(n+1).
\end{aligned}\right\} \qquad (10)$$

Обозначим

$$\begin{aligned}
d_{11}&=\operatorname{cov}(\theta_{n+1}, \theta_{n+1} \mid \mathscr{F}_n^\xi)= \\
&\qquad=\mathsf{M}\left\{[\theta_{n+1}-\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)][\theta_{n+1}-\mathsf{M}(\theta_{n+1} \mid \mathscr{F}_n^\xi)]^*/\mathscr{F}_n^\xi\right\},\\
d_{12}&=\operatorname{cov}(\theta_{n+1}, \xi_{n+1} \mid \mathscr{F}_n^\xi)= \\
&\qquad=\mathsf{M}\left\{[\theta_{n+1}-\mathsf{M}[\theta_{n+1} \mid \mathscr{F}_n^\xi)][\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)]^*/\mathscr{F}_n^\xi\right\},\\
d_{22}&=\operatorname{cov}(\xi_{n+1}, \xi_{n+1} \mid \mathscr{F}_n^\xi)= \\
&\qquad=\mathsf{M}\left\{[\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)][\xi_{n+1}-\mathsf{M}(\xi_{n+1} \mid \mathscr{F}_n^\xi)]^*/\mathscr{F}_n^\xi\right\}.
\end{aligned}$$

Тогда из (10)

$$d_{11}=a_1\gamma_na_1^*+b\circ b, \quad d_{12}=a_1\gamma_nA_1^*+b\circ B, \quad d_{22}=A_1\gamma_nA_1^*+B\circ B. \qquad (11)$$

В силу теоремы о нормальной корреляции (см. теорему 2 задачу 4 в § 13 гл. II)

$$m_{n+1}=\mathsf{M}\left(\theta_{n+1}\mid \mathscr{F}_n^{\xi},\ \xi_{n+1}\right)=\mathsf{M}\left(\theta_{n+1}\mid \mathscr{F}_n^{\xi}\right)+d_{12}d_{22}^{\oplus}\left(\xi_{n+1}-\mathsf{M}\left(\xi_{n+1}\mid \mathscr{F}_n^{\xi}\right)\right)$$

и

$$\gamma_{n+1}=\operatorname{cov}\left(\theta_{n+1},\ \theta_{n+1}\mid \mathscr{F}_n^{\xi},\ \xi_{n+1}\right)=d_{11}-d_{12}d_{22}^{\oplus}d_{12}^{*}.$$

Подставляя сюда выражения для $\mathsf{M}\left(\theta_{n+1}\mid \mathscr{F}_n^{\xi}\right)$, $\mathsf{M}\left(\xi_{n+1}\mid \mathscr{F}_n^{\xi}\right)$ из (9) и для d_{11}, d_{12}, d_{22} из (11) получаем искомые рекуррентные уравнения (7) и (8).

Теорема доказана.

Следствие 1. Если все коэффициенты $a_0(n, \xi), \ldots, B_2(n, \xi)$ в системе (1) не зависят от ξ, то соответствующая схема называется схемой Калмана — Бьюси, а уравнения (7) и (8) для m_n и γ_n — *фильтром Калмана — Бьюси*. Важно подчеркнуть, что в этом случае условная матрица ошибок γ_n совпадает с безусловной, т. е.

$$\gamma_n\equiv \mathsf{M}\gamma_n=\mathsf{M}\left[(\theta_n-m_n)(\theta_n-m_n)^*\right].$$

Следствие 2. Предположим, что частично наблюдаемая последовательность (θ_n, ξ_n) такова, что для θ_n справедливо первое из уравнений в (1), а для ξ_n — уравнение

$$\xi_n=\tilde{A}_0(n-1,\ \xi)+\tilde{A}_1(n-1,\ \xi)\,\theta_n+ \\ +\tilde{B}_1(n-1,\ \xi)\,\varepsilon_1(n)+\tilde{B}_2(n-1,\ \xi)\,\varepsilon_2(n). \quad (12)$$

Тогда, очевидно,

$$\xi_{n+1}=\tilde{A}_0(n,\ \xi)+\tilde{A}_1(n,\ \xi)\left[a_0(n,\ \xi)+a_1(n,\ \xi)\,\theta_n+\right. \\ \left.+b_1(n,\ \xi)\,\varepsilon_1(n+1)+b_2(n,\ \xi)\,\varepsilon_2(n+1)\right]+\tilde{B}_1(n,\ \xi)\,\varepsilon_1(n+1)+ \\ +\tilde{B}_2(n,\ \xi)\,\varepsilon_2(n+1),$$

и, обозначая

$$A_0=A_0+\tilde{A}_1a_0, \quad A_1=\tilde{A}_1a_1,$$
$$B_1=A_1b_1+\tilde{B}_1, \quad B_2=\tilde{A}_1b_2+\tilde{B}_2,$$

получаем, что рассматриваемый случай также укладывается в схему (1), а m_n и γ_n удовлетворяют уравнениям (7) и (8).

2. Обратимся к *линейной* схеме (ср. с (1))

$$\begin{aligned}\theta_{n+1}&=a_0+a_1\theta_n+a_2\xi_n+b_1\varepsilon_1(n+1)+b_2\varepsilon_2(n+1),\\ \xi_{n+1}&=A_0+A_1\theta_n+A_2\xi_n+B_1\varepsilon_1(n+1)+B_2\varepsilon_2(n+1),\end{aligned} \quad (13)$$

где все коэффициенты $a_0, \ldots, B_2$ могут зависеть от n (но не от ξ), а $\varepsilon_{ij}(n)$ — независимые гауссовские случайные величины с $\mathsf{M}\varepsilon_{ij}(n)=0$ и $\mathsf{M}\varepsilon_{ij}^2(n)=1$.

Пусть система (13) решается при начальных значениях (θ_0, ξ_0) таких, что условное распределение $\mathsf{P}(\theta_0\leqslant a\mid \xi_0)$ является гауссовским с параметрами $m_0=\mathsf{M}(\theta_0\mid \xi_0)$ и $\gamma_0=\operatorname{cov}(\theta_0,\ \theta_0\mid \xi_0)=\mathsf{M}\gamma_0$.

Тогда в силу теоремы о нормальной корреляции и (7), (8) оптимальная оценка $m_n = \mathsf{M}(\theta_n \mid \mathscr{F}_n^{\xi})$ является линейной функцией от $\xi_0, \xi_1, \ldots, \xi_n$.

Это замечание позволяет доказать следующее важное утверждение о структуре оптимального линейного фильтра при отказе от предположений гауссовости.

Теорема 2. *Пусть* $(\theta, \xi) = (\theta_n, \xi_n)_{n \geqslant 0}$ — *частично наблюдаемая последовательность, удовлетворяющая системе* (13), *где* $\varepsilon_{ij}(n)$ — *некоррелированные случайные величины с* $\mathsf{M}\varepsilon_{ij}(n) = 0$, $\mathsf{M}\varepsilon_{ij}^2(n) = 1$, *а компоненты вектора начальных значений* (θ_0, ξ_0) *имеют конечный второй момент. Тогда оптимальная линейная оценка* $\hat{m}_n = \hat{\mathsf{M}}(\theta_n \mid \xi_0, \ldots, \xi_n)$ *удовлетворяет уравнениям* (7) *с* $a_0(n, \xi) = a_0(n) + a_2(n)\xi_n$, $A_0(n, \xi) = A_0(n) + A_2(n)\xi_n$, *а матрица ошибок* $\hat{\gamma}_n = \hat{\mathsf{M}}[(\theta_n - \hat{m}_n)(\theta_n - \hat{m}_n)^*]$ — *уравнениям* (8) *с начальными данными*

$$
\begin{aligned}
\hat{m}_0 &= \operatorname{cov}(\theta_0, \xi_0)\operatorname{cov}^{\oplus}(\xi_0, \xi_0)\cdot\xi_0,\\
\hat{\gamma}_0 &= \operatorname{cov}(\theta_0, \theta_0) - \operatorname{cov}(\theta_0, \xi_0)\operatorname{cov}^{\oplus}(\xi_0, \xi_0)\operatorname{cov}^*(\theta_0, \xi_0).
\end{aligned}
\tag{14}
$$

Для доказательства этой теоремы понадобится следующая лемма, раскрывающая роль гауссовского случая при отыскании оптимальных линейных оценок.

Лемма 2. *Пусть* (α, β) — *двумерный случайный вектор* $\mathsf{M}(\alpha^2 + \beta^2) < \infty$, *а* $(\tilde{\alpha}, \tilde{\beta})$ — *двумерный гауссовский вектор с теми же первыми и вторыми моментами, что и у* (α, β), *т. е.*

$$\mathsf{M}\tilde{\alpha}^i = \mathsf{M}\alpha^i, \quad \mathsf{M}\tilde{\beta}^i = \mathsf{M}\beta^i, \quad i = 1, 2, \quad \mathsf{M}\tilde{\alpha}\tilde{\beta} = \mathsf{M}\alpha\beta.$$

Пусть $\lambda(b)$ — *линейная функция от* b *такая, что*

$$\lambda(b) = \mathsf{M}(\tilde{\alpha} \mid \tilde{\beta} = b).$$

Тогда $\lambda(\beta)$ *является оптимальной (в среднеквадратическом смысле) линейной оценкой* α *по* β, *т. е.*

$$\hat{\mathsf{M}}(\alpha \mid \beta) = \lambda(\beta).$$

При этом $\mathsf{M}\lambda(\beta) = \mathsf{M}\alpha$.

Доказательство. Прежде всего отметим, что существование линейной функции $\lambda(b)$, совпадающей с $\mathsf{M}\tilde{\alpha}(\mid \tilde{\beta} = b)$, вытекает из теоремы о нормальной корреляции. Далее, пусть $\bar{\lambda}(b)$ — какая-то другая линейная оценка. Тогда

$$\mathsf{M}[\tilde{\alpha} - \bar{\lambda}(\tilde{\beta})]^2 \geqslant \mathsf{M}[\tilde{\alpha} - \lambda(\tilde{\beta})]^2$$

и в силу линейности оценок $\bar{\lambda}(b)$ и $\lambda(b)$ и условий леммы

$$\mathsf{M}[\alpha - \bar{\lambda}(\beta)]^2 = \mathsf{M}[\tilde{\alpha} - \bar{\lambda}(\tilde{\beta})]^2 \geqslant \mathsf{M}[\tilde{\alpha} - \lambda(\tilde{\beta})]^2 = \mathsf{M}[\alpha - \lambda(\beta)]^2,$$

что и доказывает оптимальность $\lambda(\beta)$ в классе линейных оценок. Наконец,

$$\mathsf{M}\lambda(\beta)=\mathsf{M}\lambda(\tilde\beta)=\mathsf{M}[\mathsf{M}(\tilde\alpha\,|\,\tilde\beta)]=\mathsf{M}\tilde\alpha=\mathsf{M}\alpha.$$

Лемма доказана.

Доказательство теоремы 2. Наряду с (13) рассмотрим систему

$$\begin{aligned}\tilde\theta_{n+1}&=a_0+a_1\tilde\theta_n+a_2\tilde\xi_n+b_1\tilde\varepsilon_1(n+1)+b_2\tilde\varepsilon_2(n+1),\\ \tilde\xi_{n+1}&=A_0+A_1\tilde\theta_n+A_2\tilde\xi_n+B_1\tilde\varepsilon_1(n+1)+B_2\tilde\varepsilon_2(n+1),\end{aligned}\tag{15}$$

где $\tilde\varepsilon_{ij}(n)$ — независимые гауссовские случайные величины с $\mathsf{M}\tilde\varepsilon_{ij}(n)=0$ и $\mathsf{M}\tilde\varepsilon_{ij}^2(n)=1$. Пусть также $(\tilde\theta_0,\ \tilde\xi_0)$ — гауссовский вектор, имеющий те же первые моменты и ковариации, что и у $(\theta_0,\ \xi_0)$, и не зависящий от $\tilde\varepsilon_{ij}(n)$. Тогда в силу линейности системы (13) вектор $(\tilde\theta_0,\ldots,\tilde\theta_n,\tilde\xi_0,\ldots,\tilde\xi_n)$ является гауссовским, и, значит, утверждение теоремы следует из леммы 2 (точнее, из ее очевидного многомерного аналога) и теоремы о нормальной корреляции.

Теорема доказана.

3. Рассмотрим несколько примеров, иллюстрирующих теоремы 1 и 2.

Пример 1. Пусть $\theta=(\theta_n)$ и $\eta=(\eta_n)$ — две стационарные (в широком смысле) некоррелированные случайные последовательности с $\mathsf{M}\theta_n=\mathsf{M}\eta_n=0$ и спектральными плотностями

$$f_\theta(\lambda)=\frac{1}{2\pi\,|1+b_1e^{-i\lambda}|^2},\quad f_\eta(\lambda)=\frac{1}{2\pi}\cdot\frac{1}{|1+b_2e^{-i\lambda}|^2},$$

где $|b_1|<1$, $|b_2|<1$.

В дальнейшем будем интерпретировать θ как полезный сигнал, а η — как шум и предполагать, что наблюдению подлежит последовательность $\xi=(\xi_n)$ с

$$\xi_n=\theta_n+\eta_n.$$

Согласно следствию 2 к теореме 3 из § 3 найдутся (некоррелированные между собой) белые шумы $\varepsilon_1=(\varepsilon_1(n))$ и $\varepsilon_2=(\varepsilon_2(n))$ такие, что

$$\theta_{n+1}+b_1\theta_n=\varepsilon_1(n),\quad \eta_{n+1}+b_2\eta_n=\varepsilon_2(n).$$

Тогда

$$\begin{aligned}\xi_{n+1}&=\theta_{n+1}+\eta_{n+1}=-b_1\theta_n-b_2\eta_n+\varepsilon_1(n)+\varepsilon_2(n)=\\ &=-b_2(\theta_n+\eta_n)-\theta_n(b_1-b_2)+\varepsilon_1(n)+\varepsilon_2(n)=\\ &=-b_2\xi_n-(b_1-b_2)\theta_n+\varepsilon_1(n)+\varepsilon_2(n).\end{aligned}$$

Тем самым для θ и ξ справедливы рекуррентные уравнения

$$\begin{aligned} \theta_{n+1} &= -b_1\theta_n + \varepsilon_1(n), \\ \xi_{n+1} &= -(b_1-b_2)\theta_n - b_2\xi_n + \varepsilon_1(n) + \varepsilon_2(n) \end{aligned} \tag{16}$$

и, согласно теореме 2, $m_n = \hat{\mathsf{M}}(\theta_n \mid \xi_0, \ldots, \xi_n)$ и $\gamma_n = \mathsf{M}(\theta_n - m_n)^2$ удовлетворяют следующей системе рекуррентных уравнений оптимальной линейной фильтрации:

$$\begin{aligned} m_{n+1} &= -b_1 m_n + \frac{b_1(b_1-b_2)\gamma_n}{2+(b_1-b_2)^2\gamma_n}[\xi_{n+1} + (b_1-b_2)m_n + b_2\xi_n], \\ \gamma_{n+1} &= b_1^2\gamma_n + 1 - \frac{[1+b_1(b_1-b_2)\gamma_n]^2}{2+(b_1-b_2)^2\gamma_n}. \end{aligned} \tag{17}$$

Найдем начальные значения m_0 и γ_0, при которых должна решаться эта система. Обозначим $d_{11} = \mathsf{M}\theta_n^2$, $d_{12} = \mathsf{M}\theta_n\xi_n$, $d_{22} = \mathsf{M}\xi_n^2$. Тогда из (16) находим, что

$$\begin{aligned} d_{11} &= b_1^2 d_{11} + 1, \\ d_{12} &= b_1(b_1-b_2)d_{11} + b_1 b_2 d_{12} + 1, \\ d_{22} &= (b_1-b_2)^2 d_{11} + b_2^2 d_{22} + 2b_2(b_1-b_2)d_{12} + 2, \end{aligned}$$

откуда

$$d_{11} = \frac{1}{1-b_1^2}, \quad d_{12} = \frac{1}{1-b_1^2}, \quad d_{22} = \frac{2-b_1^2-b_2^2}{(1-b_1^2)(1-b_2^2)},$$

что в силу (14) приводит к следующим значениям начальных данных:

$$\begin{aligned} m_0 &= \frac{d_{12}}{d_{22}}\xi_0 = \frac{1-b_2^2}{2-b_1^2-b_2^2}\xi_0, \\ \gamma_0 &= d_{11} - \frac{d_{12}^2}{d_{22}} = \frac{1}{1-b_1^2} - \frac{1-b_2^2}{(1-b_1^2)(2-b_1^2-b_2^2)} = \frac{1}{2-b_1^2-b_2^2}. \end{aligned} \tag{18}$$

Итак, оптимальная (в среднеквадратическом смысле) линейная оценка m_n сигнала θ_n по $\xi_0, \ldots, \xi_n$ и среднеквадратическая ошибка γ_n определяются из системы рекуррентных уравнений (17), решаемых при начальных условиях (18). Отметим, что уравнение для γ_n не содержит случайных составляющих, и, следовательно, величины γ_n, необходимые для отыскания значений m_n, могут быть рассчитаны заранее — до решения самой задачи фильтрации.

Пример 2. Этот пример поучителен с той точки зрения, что показывает, как результат теоремы 2 может быть применен для отыскания оптимального линейного фильтра в задаче, где последовательности (θ, ξ) подчиняются (нелинейной) системе, не совпадающей с системой (13).

Пусть $\varepsilon_1 = (\varepsilon_1(n))$ и $\varepsilon_2 = (\varepsilon_2(n))$ — две независимые гауссовские последовательности, состоящие из независимых случайных величин с $\mathsf{M}\varepsilon_i(n) = 0$, $\mathsf{M}\varepsilon_i^2(n) = 1$, $n \geqslant 1$. Рассмотрим пару последовательностей $(\theta, \xi)=(\theta_n, \xi_n)$, $n\geqslant 0$, с

$$\begin{aligned}\theta_{n+1}&=a\theta_n+(1+\theta_n)\varepsilon_1(n+1),\\ \xi_{n+1}&=A\theta_n+\varepsilon_2(n+1).\end{aligned}\tag{19}$$

Будем считать, что θ_0 не зависит от $(\varepsilon_1, \varepsilon_2)$ и $\theta_0\sim\mathscr{N}(m_0, \gamma_0)$.

Система (19) является *нелинейной*, и непосредственное применение теоремы 2 невозможно. Однако если положить

$$\tilde{\varepsilon}_1(n+1)=\frac{1+\theta_n}{\sqrt{\mathsf{M}(1+\theta_n)^2}}\varepsilon_1(n+1),$$

то замечаем, что $\mathsf{M}\tilde{\varepsilon}_1(n)=0$, $\mathsf{M}\tilde{\varepsilon}_1(n)\tilde{\varepsilon}_1(m)=0$, $n\neq m$, $\mathsf{M}\tilde{\varepsilon}_1^2(n)=1$. Поэтому наряду с (19) исходная последовательность (θ, ξ) подчиняется также линейной системе

$$\begin{aligned}\theta_{n+1}&=a_1\theta_n+b_1\tilde{\varepsilon}_1(n+1),\\ \xi_{n+1}&=A_1\theta_n+\varepsilon_2(n+1),\end{aligned}\tag{20}$$

где $b_1=\sqrt{\mathsf{M}(1+\theta_n)^2}$, а $\{\tilde{\varepsilon}_1(n)\}$ — некоторая последовательность некоррелированных случайных величин.

Система (20) является линейной системой типа (13) и, следовательно, оптимальная линейная оценка $\hat{m}_n=\hat{\mathsf{M}}(\theta_n\,|\,\xi_0, \ldots, \xi_n)$ и ее ошибка $\hat{\gamma}_n$ могут быть определены в соответствии с теоремой 2 из системы (7), (8), принимающей в рассматриваемом случае следующий вид:

$$\begin{aligned}m_{n+1}&=a_1m_n+\frac{a_1A_1\gamma_n}{1+A_1^2\gamma_n}[\xi_{n+1}-A_1m_n],\\ \gamma_{n+1}&=(a_1^2\gamma_n+b_1^2)-\frac{(a_1A_1\gamma_n)^2}{1+A_1^2\gamma_n},\end{aligned}$$

где $b_1=\sqrt{\mathsf{M}(1+\theta_n)^2}$ должно быть найдено из первого уравнения системы (19).

Пример 3. *Оценка параметров*. Пусть $\theta=(\theta_1, \ldots, \theta_k)$ — гауссовский вектор с $\mathsf{M}\theta=m$ и $\operatorname{cov}(\theta, \theta)=\gamma$. Предположим, что (при известных m и γ) ищется оптимальная оценка θ по результатам наблюдений за l-мерной последовательностью $\xi=(\xi_n)$, $n\geqslant 0$, с

$$\xi_{n+1}=A_0(n, \xi)+A_1(n, \xi)\theta+B_1(n, \xi)\varepsilon_1(n+1), \quad \xi_0=0, \tag{21}$$

где ε_1 — те же, что и в системе (1).

Тогда из (7), (8) для $m_n=\mathsf{M}(\theta\,|\,\mathscr{F}_n^\xi)$ и γ_n находим, что

$$\begin{aligned}m_{n+1}&=m_n+\gamma_nA_1^*(n, \xi)[(B_1B_1^*)(n, \xi)+A_1(n, \xi)\gamma_nA_1^*(n, \xi)]^{\oplus}\times\\ &\qquad\times[\xi_{n+1}-A_0(n, \xi)-A_1(n, \xi)m_n],\\ \gamma_{n+1}&=\gamma_n-\gamma_nA_1^*(n, \xi)[(B_1B_1^*)(n, \xi)+A_1(n, \xi)\gamma_nA_1^*(n, \xi)]^{\oplus}\times\\ &\qquad\times A_1(n, \xi)\gamma_n.\end{aligned}\tag{22}$$

Если матрицы $B_1B_1^*$ являются невырожденными, то решения системы (22) задаются формулами

$$m_{n+1} = \left[E + \gamma \sum_{m=0}^{n} A_1^*(m, \xi)(B_1B_1^*)^{-1}(m, \xi) A_1^*(m, \xi)\right]^{-1} \times$$
$$\times \left[m + \gamma \sum_{m=0}^{n} A_1^*(m, \xi)(B_1B_1^*)^{-1}(m, \xi)(\xi_{m+1} - A_0(m, \xi))\right], \tag{23}$$
$$\gamma_{n+1} = \left[E + \gamma \sum_{m=0}^{n} A_1^*(m, \xi)(B_1B_1^*)^{-1}(m, \xi) A_1(m, \xi)\right]^{-1} \gamma,$$

где E — единичная матрица.

4. Задачи.

1. Показать, что для схемы (1) векторы m_n и $\theta_n - m_n$ не коррелированы:

$$\mathsf{M}\left[m_n^*(\theta - m_n)\right] = 0.$$

2. Пусть в схеме (1) γ_0 и все коэффициенты, за исключением, быть может, коэффициентов $a_0(n, \xi)$, $A_0(n, \xi)$, не зависят от «случая» (т. е. от ξ). Показать, что тогда условная ковариация γ_n также не зависит от «случая»: $\gamma_n = \mathsf{M}\gamma_n$.

3. Показать, что решения системы (22) задаются формулами (23).

4. Пусть $(\theta, \xi) = (\theta_n, \xi_n)$ — гауссовская последовательность, удовлетворяющая следующему частному виду схемы (1):

$$\theta_{n+1} = a\theta_n + b\varepsilon_1(n+1), \qquad \xi_{n+1} = A\theta_n + B\varepsilon_2(n+1).$$

Показать, что если $A \neq 0$, $b \neq 0$, $B \neq 0$, то предельная ошибка фильтрации $\gamma = \lim\limits_{n\to\infty} \gamma_n$ существует и определяется как положительный корень уравнения

$$\gamma^2 + \left[\frac{B^2(1-a^2)}{A^2} - b^2\right]\gamma - \frac{b^2B^2}{A^2} = 0.$$

ГЛАВА VII

ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН, ОБРАЗУЮЩИЕ МАРТИНГАЛ

§ 1. Определения мартингалов и родственных понятий

1. Исследование зависимости между случайными величинами осуществляется в теории вероятностей разными способами. В теории стационарных (в широком смысле) случайных последовательностей основным показателем зависимости является ковариационная функция и все выводы этой теории полностью определяются свойствами этой функции. В теории марковских цепей (§ 12 гл. I и гл. VIII) основной характеристикой зависимости служит переходная функция, которая полностью определяет эволюцию случайных величин, связанных марковской зависимостью.

В настоящей главе (см. также § 11 гл. I) выделяется достаточно обширный класс последовательностей случайных величин (мартингалы и их обобщения), для которых изучение зависимости проводится методами, основанными на исследовании свойств условных математических ожиданий.

2. Будем предполагать заданным вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$ с выделенным на нем семейством $(\mathscr{F}_n)$ σ-алгебр $\mathscr{F}_n$, $n \geqslant 0$, таких, что $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \ldots \subseteq \mathscr{F}$.

Пусть $X_0, X_1, \ldots$ — последовательность случайных величин, заданных на $(\Omega, \mathscr{F}, \mathsf{P})$. Если для каждого $n \geqslant 0$ величины X_n являются *$\mathscr{F}_n$-измеримыми*, то будем говорить, что набор $X = (X_n, \mathscr{F}_n)$, $n \geqslant 0$, или просто $X = (X_n, \mathscr{F}_n)$ образует *стохастическую последовательность*.

Если стохастическая последовательность $X = (X_n, \mathscr{F}_n)$ к тому же такова, что для каждого $n \geqslant 1$ величины X_n являются *$\mathscr{F}_{n-1}$-измеримыми*, то будем это записывать в виде $X = (X_n, \mathscr{F}_{n-1})$, считая $\mathscr{F}_{-1} = \mathscr{F}_0$, и называть X *предсказуемой последовательностью*. Такая последовательность будет называться *возрастающей*, если $X_0 = 0$ и $X_n \leqslant X_{n+1}$ (P-п. н.).

Определение 1. Стохастическая последовательность $X = (X_n, \mathscr{F}_n)$ называется *мартингалом* (*субмартингалом*), если для всех $n \geqslant 0$

$$\mathsf{M}|X_n| < \infty, \tag{1}$$

$$\mathsf{M}(X_{n+1}|\mathscr{F}_n) \underset{(\geqslant)}{=} X_n \ (\mathsf{P}\text{-п. н.}). \tag{2}$$

Стохастическая последовательность $X=(X_n, \mathscr{F}_n)$ называется *супермартингалом*, если последовательность $-X=(-X_n, \mathscr{F}_n)$ есть субмартингал.

В том частном случае, когда $\mathscr{F}_n = \mathscr{F}_n^X$, где $\mathscr{F}_n^X = \sigma\{\omega: X_0, \ldots, X_n\}$, и стохастическая последовательность $X=(X_n, \mathscr{F}_n^X)$ образует мартингал (субмартингал), будем говорить, что сама последовательность $(X_n)_{n\geqslant 0}$ образует мартингал (субмартингал).

Из свойств условных математических ожиданий легко выводится, что условие (2) эквивалентно тому, что для любого $n \geqslant 0$ и $A \in \mathscr{F}_n$

$$\int_A X_{n+1}\, d\mathsf{P} \underset{(\geqslant)}{=} \int_A X_n\, d\mathsf{P}. \tag{3}$$

Пример 1. Если $(\xi_n)_{n\geqslant 0}$ — последовательность независимых случайных величин с $\mathsf{M}\xi_n=0$ и $X_n=\xi_0+\ldots+\xi_n$, $\mathscr{F}_n = \sigma\{\omega: \xi_0, \ldots, \xi_n\}$, то стохастическая последовательность $X=(X_n, \mathscr{F}_n)$ образует мартингал.

Пример 2. Если $(\xi_n)_{n\geqslant 0}$ — последовательность независимых случайных величин с $\mathsf{M}\xi_n=1$, то стохастическая последовательность $X=(X_n, \mathscr{F}_n)$ с $X_n=\prod_{k=0}^{n}\xi_k$, $\mathscr{F}_n=\sigma\{\omega: \xi_0, \ldots, \xi_n\}$ также образует мартингал.

Пример 3. Пусть ξ — случайная величина с $\mathsf{M}|\xi|<\infty$ и $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \ldots \subseteq \mathscr{F}$. Тогда последовательность $X=(X_n, \mathscr{F}_n)$ с $X_n = \mathsf{M}(\xi|\mathscr{F}_n)$ является мартингалом.

Пример 4. Если $(\xi_n)_{n\geqslant 0}$ — последовательность неотрицательных интегрируемых случайных величин, то последовательность (X_n) с $X_n=\xi_0+\ldots+\xi_n$ образует субмартингал.

Пример 5. Если $X=(X_n, \mathscr{F}_n)$ — мартингал и $g(x)$ — выпуклая книзу функция с $\mathsf{M}|g(X_n)|<\infty$, $n\geqslant 0$, то стохастическая последовательность $(g(X_n), \mathscr{F}_n)$ является субмартингалом (что следует из неравенства Иенсена).

Если $X=(X_n, \mathscr{F}_n)$ — субмартингал, а $g(x)$ — выпуклая книзу неубывающая функция с $\mathsf{M}|g(X_n)|<\infty$ для всех $n\geqslant 0$, то $(g(X_n), \mathscr{F}_n)$ также является субмартингалом.

Сделанное в определении 1 предположение (1) гарантирует существование условных математических ожиданий $\mathsf{M}(X_{n+1}|\mathscr{F}_n)$, $n\geqslant 0$. Однако эти условные математические ожидания могут существовать и без предположения $\mathsf{M}|X_{n+1}|<\infty$. Напомним, что, согласно § 7 гл. II, $\mathsf{M}(X_{n+1}^+ | \mathcal{F}_n)$ и $\mathsf{M}(X_{n+1}^- | \mathcal{F}_n)$ определены всегда, и если (мы пишем $A=B$ (Р-п.н.), когда $\mathsf{P}(A \Delta B)=0$)

$$\{\omega: \mathsf{M}(X_{n+1}^+ | \mathcal{F}_n) < \infty\} \cup \{\omega: \mathsf{M}(X_n^- | \mathcal{F}_n) < \infty\} = \Omega \quad (\text{Р-п. н.}),$$

то говорят, что $\mathsf{M}(X_{n+1} | \mathcal{F}_n)$ также определено и по определению полагают

$$\mathsf{M}(X_{n+1} | \mathcal{F}_n) = \mathsf{M}(X_{n+1}^+ | \mathcal{F}_n) - \mathsf{M}(X_{n+1}^- | \mathcal{F}_n).$$

Исходя из этого, становится естественным следующее

Определение 2. Стохастическая последовательность $X = (X_n, \mathcal{F}_n)$ называется *обобщенным мартингалом* (*субмартингалом*), если для каждого $n \geqslant 0$ определены условные математические ожидания $\mathsf{M}(X_{n+1} | \mathcal{F}_n)$ и выполнено условие (2).

Заметим, что из этого определения вытекает, что для обобщенного субмартингала $\mathsf{M}(X_{n+1}^- | \mathcal{F}_n) < \infty$, а для обобщенного мартингала $\mathsf{M}(|X_{n+1}| | \mathcal{F}_n) < \infty$ (Р-п. н.).

3. Вводимое в нижеследующем определении понятие марковского момента играет исключительно важную роль во всей рассматриваемой далее теории.

Определение 3. Случайная величина $\tau = \tau(\omega)$, принимающая значения во множестве $\{0, 1, \ldots, +\infty\}$, называется *марковским моментом* (относительно системы $(\mathcal{F}_n)$) или *случайной величиной, не зависящей от будущего*, если для каждого $n \geqslant 0$

$$\{\tau = n\} \in \mathcal{F}_n. \tag{4}$$

В случае $\mathsf{P}(\tau < \infty) = 1$ марковский момент τ будем называть *моментом остановки*.

Пусть $X = (X_n, \mathcal{F}_n)$ — некоторая стохастическая последовательность и τ — марковский момент (относительно системы $(\mathcal{F}_n)$). Обозначим

$$X_\tau = \sum_{n=0}^{\infty} X_n I_{\{\tau = n\}}(\omega)$$

(тем самым $X_\tau = 0$ на множестве $\{\omega: \tau = \infty\}$).

Тогда для каждого $B \in \mathcal{B}(R)$

$$\{\omega: X_\tau \in B\} = \sum_{n=0}^{\infty} \{X_n \in B, \ \tau = n\} \in \mathcal{F},$$

и, следовательно, X_τ является случайной величиной.

Пример 6. Пусть $X = (X_n, \mathcal{F}_n)$ — некоторая стохастическая последовательность и $B \in \mathcal{B}(R)$. Тогда момент (первого попадания в множество B)

$$\tau_B = \inf\{n \geqslant 0: X_n \in B\}$$

(с $\tau_B=+\infty$, если $\{\cdot\}=\varnothing$) является марковским, поскольку для любого $n\geqslant 0$

$$\{\tau_B=n\}=\{X_0\notin B,\ \ldots,\ X_{n-1}\notin B,\ X_n\in B\}\in\mathscr{F}_n.$$

Пример 7. Пусть $X=(X_n, \mathscr{F}_n)$ — мартингал (субмартингал) и τ — марковский момент (относительно системы $(\mathscr{F}_n)$). Тогда «остановленная» последовательность $X^\tau=(X_{n\wedge\tau}, \mathscr{F}_n)$ также образует мартингал (субмартингал).

В самом деле, из соотношения

$$X_{n\wedge\tau}=\sum_{m=0}^{n-1}X_m I_{\{\tau=m\}}+X_n I_{\{\tau\geqslant n\}}$$

следует, что величины $X_{n\wedge\tau}$ $\mathscr{F}_n$-измеримы, интегрируемы и

$$X_{(n+1)\wedge\tau}-X_{n\wedge\tau}=I_{\{\tau>n\}}\,(X_{n+1}-X_n),$$

откуда

$$\mathsf{M}\,[X_{(n+1)\wedge\tau}-X_{n\wedge\tau}\,|\,\mathscr{F}_n]=I_{\{\tau>n\}}\mathsf{M}\,[X_{n+1}-X_n\,|\,\mathscr{F}_n]\underset{(\geqslant)}{=}0.$$

С каждой системой $(\mathscr{F}_n)$ и марковским моментом τ относительно ее можно связать совокупность множеств

$$\mathscr{F}_\tau=\{A\in\mathscr{F}\colon\ A\cap\{\tau=n\}\in\mathscr{F}_n\ \text{для всех}\ n\geqslant 0\}.$$

Ясно, что $\Omega\in\mathscr{F}_\tau$ и $\mathscr{F}_\tau$ замкнуто относительно взятия счетных объединений. Кроме того, если $A\in\mathscr{F}_\tau$, то $\bar{A}\cap\{\tau=n\}=\{\tau=n\}\setminus(A\cap\{\tau=n\})\in\mathscr{F}_n$ и, значит, $\bar{A}\in\mathscr{F}_\tau$. Отсюда следует, что $\mathscr{F}_\tau$ является σ-алгеброй.

Если трактовать $\mathscr{F}_n$ как совокупность событий, наблюдаемых до момента времени n (включительно), то тогда $\mathscr{F}_\tau$ можно представлять как совокупность событий, наблюдаемых за «случайное» время τ.

Нетрудно показать (задача 3), что случайные величины τ и X_τ являются $\mathscr{F}_\tau$-измеримыми.

4. Определение 4. Стохастическая последовательность $X=(X_n, \mathscr{F}_n)$ называется *локальным мартингалом* (*субмартингалом*), если найдется такая (локализующая) последовательность $(\tau_k)_{k\geqslant 1}$ марковских моментов, что $\tau_k\leqslant\tau_{k+1}$ (Р-п. н.), $\tau_k\uparrow\infty$ (Р-п. н.), $k\to\infty$, и каждая «остановленная» последовательность $X^{\tau_k}=(X_{\tau_k\wedge n}\cdot I_{\{\tau_k>0\}}, \mathscr{F}_n)$ является мартингалом (субмартингалом).

Ниже в теореме 1 показывается, что на самом деле класс локальных мартингалов совпадает с классом обобщенных мартингалов. Более того, каждый локальный мартингал может быть получен с помощью так называемого мартингального преобразования из некоторого мартингала и некоторой предсказуемой последовательности.

Определение 5. Пусть $Y=(Y_n, \mathscr{F}_n)_{n\geqslant 0}$ — стохастическая последовательность и $V=(V_n, \mathscr{F}_{n-1})$ — предсказуемая последовательность ($\mathscr{F}_{-1}=\mathscr{F}_0$). Стохастическая последовательность $V\cdot Y = ((V\cdot Y)_n, \mathscr{F}_n)$ с

$$(V\cdot Y)_n = V_0Y_0+\sum_{i=1}^{n} V_i\,\Delta Y_i, \tag{5}$$

где $\Delta Y_i = Y_i - Y_{i-1}$, называется *преобразованием* Y *с помощью* V. Если к тому же Y — мартингал, то говорят, что $V\cdot Y$ есть *мартингальное преобразование*.

Теорема 1. *Пусть $X=(X_n, \mathscr{F}_n)_{n\geqslant 0}$ — стохастическая последовательность с $X_0=0$ (Р-п. н.). Следующие условия являются эквивалентными:*

a) *X — локальный мартингал;*

b) *X — обобщенный мартингал;*

c) *X — есть мартингальное преобразование, т. е. существуют предсказуемая последовательность $V=(V_n, \mathscr{F}_{n-1})$ с $V_0=0$ и мартингал $Y=(Y_n, \mathscr{F}_n)$ с $Y_0=0$ такие, что $X=V\cdot Y$.*

Доказательство. a) $\Rightarrow$ b). Пусть X — локальный мартингал и (τ_k) — его локализующая последовательность марковских моментов. Тогда для любого $m\geqslant 0$

$$\mathsf{M}[|X_{m\wedge\tau_k}|I_{\{\tau_k>0\}}]<\infty, \tag{6}$$

и тем самым

$$\mathsf{M}[|X_{(n+1)\wedge\tau_k}|I_{\{\tau_k>n\}}]=\mathsf{M}[|X_{n+1}|I_{\{\tau_k>n\}}]<\infty. \tag{7}$$

Случайная величина $I_{\{\tau_k>n\}}$ является $\mathscr{F}_n$-измеримой. Поэтому из (7) следует, что

$$\mathsf{M}[|X_{n+1}|I_{\{\tau_k>n\}}\,|\mathscr{F}_n]=I_{\{\tau_k>n\}}\mathsf{M}[|X_{n+1}|\,|\mathscr{F}_n]<\infty \text{ (Р-п. н.)}.$$

Здесь $I_{\{\tau_k>n\}}\to 1$ (Р-п. н.), $k\to\infty$, и значит,

$$\mathsf{M}[|X_{n+1}|\,|\mathscr{F}_n]<\infty \text{ (Р-п. н.)}. \tag{8}$$

В силу этого условия $\mathsf{M}[X_{n+1}|\mathscr{F}_n]$ определено и осталось лишь показать, что $\mathsf{M}[X_{n+1}|\mathscr{F}_n]=X_n$ (Р-п. н.).

Поскольку X^{τ_k} — мартингалы, то для любого множества $A\in\mathscr{F}_n$ из (6) находим, что величины $X_{n+1}I_{\{\tau_k>n\}}$ и $X_nI_{\{\tau_k>n\}}$ интегрируемы и

Но $\{\tau_k > n\} \uparrow \Omega$, $k \to \infty$, поэтому

$$\int\limits_A X_{n+1}\, d\mathsf{P} = \int\limits_A X_n\, d\mathsf{P}, \qquad A \in \mathscr{F}_n,$$

что и доказывает требуемое равенство $\mathsf{M}[X_{n+1} | \mathscr{F}_n] = X_n$ (P-п. н.).

b) $\Rightarrow$ c). Пусть $\Delta X_n = X_n - X_{n-1}$, $X_0 = 0$ и $V_0 = 0$, $V_n = M[|\Delta X_n| | \mathscr{F}_{n-1}]$, $n \geqslant 1$. Положим $W_n = V_n^{\oplus}$, $Y_0 = 0$ и

$$Y_n = \sum_{i=1}^{n} W_i \Delta X_i, \quad n \geqslant 1.$$

Ясно, что

$$\mathsf{M}[|\Delta Y_n| | \mathscr{F}_{n-1}] \leqslant 1, \quad \mathsf{M}[\Delta Y_n | \mathscr{F}_{n-1}] = 0,$$

и, следовательно, $Y = (Y_n, \mathscr{F}_n)$ есть мартингал. Далее, $X_0 = V_0 \cdot Y_0 = 0$ и $\Delta(V \cdot Y)_n = \Delta X_n$. Поэтому

$$X = V \cdot Y.$$

c) $\Rightarrow$ a). Пусть $X = V \cdot Y$, где V — предсказуемая последовательность, Y — мартингал и $V_0 = Y_0 = 0$. Положим

$$\tau_k = \inf\{n \geqslant 0 \colon |V_{n+1}| > k\},$$

считая $\tau_k = \infty$, если множество $\{\cdot\} = \varnothing$. Поскольку V_{n+1} являются $\mathscr{F}_n$-измеримыми, то для каждого $k \geqslant 1$ величины τ_k являются марковскими моментами.

Рассмотрим «остановленные» последовательности $X^{\tau_k} = ((V \cdot Y)_{n \wedge \tau_k} I_{\{\tau_k > 0\}}, \mathscr{F}_n)$. На множестве $\{\tau_k > 0\}$ действует неравенство: $|V_{n \wedge \tau_k}| \leqslant k$. Отсюда следует, что для любого $n \geqslant 1$ $\mathsf{M}|(V \cdot Y)_{n \wedge \tau_k} I_{\{\tau_k > 0\}}| < \infty$. Далее, для $n \geqslant 1$

$$\mathsf{M}\left\{\left[(V \cdot Y)_{(n+1) \wedge \tau_k} - (V \cdot Y)_{n \wedge \tau_k}\right] I_{\{\tau_k > 0\}} \,|\, \mathscr{F}_n\right\} = \\ = I_{\{\tau_k > 0\}} \cdot V_{(n+1) \wedge \tau_k} \cdot \mathsf{M}\left\{Y_{(n+1) \wedge \tau_k} - Y_{n \wedge \tau_k} \,|\, \mathscr{F}_n\right\} = 0.$$

поскольку (см. пример 7) $\mathsf{M}\left\{Y_{(n+1) \wedge \tau_k} - Y_{n \wedge \tau_k} | \mathscr{F}_n\right\} = 0$.

Итак, для каждого $k \geqslant 1$ стохастические последовательности X^{τ_k} являются мартингалами, $\tau_k \uparrow \infty$ (P-п. н.), и, следовательно, X — локальный мартингал.

Теорема доказана.

5. Пример 8. Пусть $(\eta_n)_{n \geqslant 1}$ — последовательность независимых одинаково распределенных бернуллиевских случайных величин с $\mathsf{P}(\eta_n = 1) = p$, $\mathsf{P}(\eta_n = -1) = q$, $p + q = 0$. Будем интерпретировать событие $\{\eta_n = 1\}$ как успех (выигрыш), а событие $\{\eta_n = -1\}$ как неуспех (проигрыш) некоего игрока в n-й партии.

Предположим, что его ставка в n-й партии есть V_n. Тогда суммарный выигрыш игрока за n партий равен

$$X_n = \sum_{i=1}^{n} V_i \eta_i = X_{n-1} + V_n \eta_n, \quad X_0 = 0.$$

Вполне естественно, что величина ставки V_n в n-й партии может зависеть от результатов предшествующих партий, т. е. от $V_1, \ldots, V_{n-1}$ и $\eta_1, \ldots, \eta_{n-1}$. Иначе говоря, если положить $\mathscr{F}_0 = \{\varnothing, \Omega\}$ и $\mathscr{F}_n = \sigma\{\omega: \eta_1, \ldots, \eta_n\}$, то V_n будет $\mathscr{F}_{n-1}$-измеримой случайной величиной, т. е. последовательность $V = (V_n, \mathscr{F}_{n-1})$, определяющая «стратегию» игрока, является предсказуемой. Полагая $Y_n = \eta_1 + \ldots + \eta_n$, находим, что

$$X_n = \sum_{i=1}^{n} V_i \Delta Y_i,$$

т. е. последовательность $X = (X_n, \mathscr{F}_n)$ с $X_0 = 0$ есть преобразование Y с помощью V.

С точки зрения игрока, рассматриваемая игра является *справедливой* (*благоприятной* или *неблагоприятной*), если на каждом шаге величина ожидаемого выигрыша $\mathsf{M}(X_{n+1} - X_n \mid \mathscr{F}_n) = 0$ ($\geqslant 0$ или $\leqslant 0$). Поэтому ясно, что игра

справедлива, если $p = q = 1/2$,
благоприятна, если $p > q$,
неблагоприятна, если $p < q$.

Поскольку последовательность $X = (X_n, \mathscr{F}_n)$ образует

мартингал, если $p = q = 1/2$,
субмартингал, если $p > q$,
супермартингал, если $p < q$,

то можно сказать, что предположение о справедливости (благоприятности или неблагоприятности) игры соответствует предположению о мартингальности (субмартингальности или супермартингальности) последовательности X.

Рассмотрим сейчас специальный класс «стратегий» $V = (V_n, \mathscr{F}_{n-1})_{n \geqslant 1}$ с $V_1 = 1$ и для $n > 1$ с

$$V_n = \begin{cases} 2^{n-1}, & \text{если } \eta_1 = -1, \ldots, \eta_{n-1} = -1, \\ 0, & \text{в остальных случаях,} \end{cases} \tag{9}$$

смысл которых сводится к тому, что игрок, начиная со ставки $V_1 = 1$, каждый раз увеличивает ставку вдвое при проигрыше и прекращает игру вовсе после первого выигрыша.

Если $\eta_1 = -1, \ldots, \eta_n = -1$, то суммарные потери игрока за n партий будут равны

$$\sum_{i=1}^{n} 2^{i-1} = 2^n - 1.$$

Поэтому, если к тому же $\eta_{n+1}=1$, то

$$X_{n+1}=X_n+V_{n+1}=-(2^n-1)+2^n=1.$$

Обозначим $\tau=\inf\{n\geqslant 1:\ X_n=1\}$. Если $p=q=1/2$, т. е. рассматриваемая игра является справедливой, то $\mathsf{P}(\tau=n)=(1/2)^n$, $\mathsf{P}(\tau<\infty)=1$, $\mathsf{P}(X_\tau=1)=1$ и $\mathsf{M}X_\tau=1$. Таким образом, даже в справедливой игре, придерживаясь «стратегии» (9), игрок за конечное (с вероятностью единица) время может вполне успешно закончить игру, добавив к своему капиталу еще одну единицу ($\mathsf{M}X_\tau=1>X_0=0$).

В игровой практике описанная система игры, заключающаяся в удвоении ставки при проигрыше и прекращении игры при первом выигрыше, называется мартингалом. Именно отсюда ведет свое происхождение математическое понятие «мартингал».

Замечание. В случае $p=q=1/2$ последовательность $X=(X_n,\ \mathscr{F}_n)_{n\geqslant 0}$ с $X_0=0$ является мартингалом и, значит, для любого $n\geqslant 1$

$$\mathsf{M}X_n=\mathsf{M}X_0=0.$$

Можно поэтому ожидать, что это соотношение сохранится, если вместо моментов n рассматривать случайные моменты τ. Как станет ясно из дальнейшего (теорема 1 в § 2), в «типичных» ситуациях $\mathsf{M}X_\tau=\mathsf{M}X_0$. Нарушение же этого равенства (как в рассмотренной выше игре) происходит в тех, так сказать, физически нереализуемых ситуациях, когда или τ, или $|X_n|$ принимают слишком большие значения. (Заметим, что рассмотренная выше игра физически нереализуема, поскольку она предполагает неограниченность времени игры и неограниченность начального капитала игрока).

6. Определение 6. Стохастическая последовательность $\xi=(\xi_n,\ \mathscr{F}_n)_{n\geqslant 0}$ называется *мартингал-разностью*, если $\mathsf{M}|\xi_n|<\infty$ для всех $n\geqslant 0$ и

$$\mathsf{M}(\xi_{n+1}\,|\,\mathscr{F}_n)=0\ \ (\mathsf{P}\text{-п. н.}). \tag{10}$$

Из определений 1 и 6 ясна связь между мартингалами и мартингал-разностями. А именно, если $X=(X_n,\ \mathscr{F}_n)$ — мартингал, то $\xi=(\xi_n,\ \mathscr{F}_n)$ с $\xi_0=X_0$ и $\xi_n=\Delta X_n$, $n\geqslant 1$, является мартингал-разностью. В свою очередь, если $\xi=(\xi_n,\ \mathscr{F}_n)$ есть мартингал-разность, то $X=(X_n,\ \mathscr{F}_n)$ с $X_n=\xi_0+\ldots+\xi_n$ является мартингалом.

В соответствии с этой терминологией всякая последовательность $\xi=(\xi_n)_{n\geqslant 0}$ независимых интегрируемых случайных величин образует мартингал-разность (с $\mathscr{F}_n=\sigma\{\omega:\ \xi_0,\ \xi_1,\ \ldots,\ \xi_n\}$).

7. Следующая теорема проясняет структуру субмартингалов (супермартингалов).

Теорема 2 (Дуб). *Пусть $X=(X_n,\ \mathscr{F}_n)$ — субмартингал. Тогда найдутся мартингал $m=(m_n,\ \mathscr{F}_n)$ и предсказуемая возрастающая последовательность $A=(A_n,\ \mathscr{F}_{n-1})$ такие, что для каждого $n\geqslant 0$*

имеет место разложение Дуба:

$$X_n = m_n + A_n \text{ (P-п. н.).} \tag{11}$$

Разложение подобного типа является единственным.

Доказательство. Положим $m_0 = X_0$, $A_0 = 0$ и

$$m_n = m_0 + \sum_{j=0}^{n-1} [X_{j+1} - \mathsf{M}(X_{j+1} | \mathscr{F}_j)], \tag{12}$$

$$A_n = \sum_{j=0}^{n-1} [\mathsf{M}(X_{j+1} | \mathscr{F}_j) - X_j]. \tag{13}$$

Очевидно, что так определенные m и A обладают требуемыми свойствами. Далее, пусть также $X_n = m'_n + A'_n$, где $m' = (m'_n, \mathscr{F}_n)$ — мартингал, а $A' = (A'_n, \mathscr{F}_n)$ — предсказуемая возрастающая последовательность. Тогда

$$A'_{n+1} - A'_n = (A_{n+1} - A_n) + (m_{n+1} - m_n) - (m'_{n+1} - m'_n),$$

и, беря от обеих частей условные математические ожидания, получаем, что (P-п. н.) $A'_{n+1} - A'_n = A_{n+1} - A_n$. Но $A_0 = A'_0 = 0$, и, значит, $A_n = A'_n$ и $m_n = m'_n$ (P-п. н.) для всех $n \geqslant 0$.

Теорема доказана.

Из разложения (11) вытекает, что последовательность $A = (A_n, \mathscr{F}_{n-1})$ компенсирует $X = (X_n, \mathscr{F}_n)$ до мартингала. Это замечание оправдывает такое

Определение 7. Предсказуемая возрастающая последовательность $A = (A_n, \mathscr{F}_{n-1})$, входящая в разложение Дуба (11), называется *компенсатором* (субмартингала X).

Разложение Дуба играет ключевую роль при исследовании квадратично интегрируемых мартингалов $M = (M_n, \mathscr{F}_n)_{n \geqslant 0}$, т. е. мартингалов, для которых $\mathsf{M} M_n^2 < \infty$, $n \geqslant 0$, что основано на том замечании, что стохастическая последовательность $M^2 = (M_n^2, \mathscr{F}_n)$ является субмартингалом. Согласно теореме 2 найдутся такой мартингал $m = (m_n, \mathscr{F}_n)$ и предсказуемая возрастающая последовательность $\langle M \rangle = (\langle M \rangle_n, \mathscr{F}_{n-1})$, что

$$M_n^2 = m_n + \langle M \rangle_n. \tag{14}$$

Последовательность $\langle M \rangle$ называется *квадратической характеристикой* мартингала M и во многом определяет его структуру и свойства.

Из (12) следует, что

$$\langle M \rangle_n = \sum_{j=1}^{n} \mathsf{M}[(\Delta M_j)^2 | \mathscr{F}_{j-1}] \tag{15}$$

и для всех $l \leqslant k$

$$\mathsf{M}[(M_k - M_l)^2 | \mathscr{F}_l] = \mathsf{M}[M_k^2 - M_l^2 | \mathscr{F}_l] = \mathsf{M}[\langle M \rangle_k - \langle M \rangle_l | \mathscr{F}_l]. \tag{16}$$

В частности, если $M_0=0$ (P-п. н.), то

$$\mathsf{M}M_k^2=\mathsf{M}\langle M\rangle_k. \tag{17}$$

Полезно заметить, что если $M_0=0$ и $M_n=\xi_1+\ldots+\xi_n$, где (ξ_n) — последовательность независимых случайных величин с $\mathsf{M}\xi_i=0$ и $\mathsf{M}\xi_i^2<\infty$, то квадратическая характеристика

$$\langle M\rangle_n=\mathsf{M}M_n^2=\mathsf{D}\xi_1+\ldots+\mathsf{D}\xi_n, \tag{18}$$

является неслучайной и совпадает с дисперсией.

Если $X=(X_n, \mathscr{F}_n)$ и $Y=(Y_n, \mathscr{F}_n)$ — квадратично интегрируемые мартингалы, то положим

$$\langle X, Y\rangle_n=\frac{1}{4}[\langle X+Y\rangle_n-\langle X-Y\rangle_n]. \tag{19}$$

Нетрудно проверить, что $(X_nY_n-\langle X, Y\rangle_n, \mathscr{F}_n)$ есть мартингал и, значит, для $l\leqslant k$

$$\mathsf{M}[(X_k-X_l)(Y_k-Y_l)\,|\,\mathscr{F}_l]=\mathsf{M}[\langle X, Y\rangle_k-\langle X, Y\rangle_l\,|\,\mathscr{F}_l]. \tag{20}$$

В случае, когда $X_n=\xi_1+\ldots+\xi_n$, $Y_n=\eta_1+\ldots+\eta_n$, где (ξ_n) и (η_n) — последовательности независимых случайных величин с $\mathsf{M}\xi_i=\mathsf{M}\eta_i=0$ и $\mathsf{M}\xi_i^2<\infty$, $\mathsf{M}\eta_i^2<\infty$, величина $\langle X, Y\rangle_n$ равна

$$\langle X, Y\rangle_n=\sum_{i=1}^{n}\operatorname{cov}(\xi_i, \eta_i).$$

Последовательность $\langle X, Y\rangle=(\langle X, Y\rangle_n, \mathscr{F}_{n-1})$ часто называют *взаимной характеристикой* (квадратично интегрируемых) мартингалов X и Y.

8. Задачи.

1. Показать эквивалентность условий (2) и (3).

2. Пусть σ и τ — марковские моменты. Показать, что $\tau+\sigma$, $\tau\wedge\sigma$, $\tau\vee\sigma$ также являются марковскими моментами, и если $\mathsf{P}(\sigma\leqslant\tau)=1$, то $\mathscr{F}_\sigma\subseteq\mathscr{F}_\tau$.

3. Показать, что τ и X_τ являются $\mathscr{F}_\tau$-измеримыми.

4. Пусть $Y=(Y_n, \mathscr{F}_n)$ — мартингал (субмартингал), $V=(V_n, \mathscr{F}_{n-1})$ — предсказуемая последовательность и $(V\cdot Y)_n$ — интегрируемые случайные величины, $n\geqslant 0$. Показать, что тогда $V\cdot Y$ есть мартингал (субмартингал).

5. Пусть $\mathscr{F}_1\subseteq\mathscr{F}_2\subseteq\ldots$ — неубывающее семейство σ-алгебр и ξ — интегрируемая случайная величина. Показать, что последовательность $(X_n)_{n\geqslant 1}$ с $X_n=\mathsf{M}(\xi\,|\,\mathscr{F}_n)$ образует мартингал.

6. Пусть $\mathscr{G}_1\supseteq\mathscr{G}_2\supseteq\ldots$ — невозрастающее семейство σ-алгебр и ξ — интегрируемая случайная величина. Показать, что последовательность $(X_n)_{n\geqslant 1}$ с $X_n=\mathsf{M}(\xi\,|\,\mathscr{G}_n)$ образует *обращенный* мартин-

гал, т. е.

$$\mathsf{M}(X_n \mid X_{n+1}, X_{n+2}, \ldots) = X_{n+1} \quad (\mathsf{P}\text{-п. н.})$$

для любого $n \geqslant 1$.

7. Пусть $\xi_1, \xi_2, \xi_3, \ldots$ — независимые случайные величины, $\mathsf{P}(\xi_i = 0) = \mathsf{P}(\xi_i = 2) = \frac{1}{2}$ и $X_n = \prod_{i=1}^{n} \xi_i$. Показать, что не существует такой интегрируемой случайной величины ξ и неубывающего семейства σ-алгебр $(\mathscr{F}_n)$, что $X_n = \mathsf{M}(\xi \mid \mathscr{F}_n)$. (Этот пример показывает, что не каждый мартингал $(X_n)_{n \geqslant 1}$ представим в виде $(\mathsf{M}(\xi \mid \mathscr{F}_n))_{n \geqslant 1}$; ср. с примером 3 § 11 гл. I.)

§ 2. О сохранении свойства мартингальности при замене времени на случайный момент

1. Если $X = (X_n, \mathscr{F}_n)_{n \geqslant 0}$ — мартингал, то для всякого $n \geqslant 1$

$$\mathsf{M}X_n = \mathsf{M}X_0. \tag{1}$$

Сохранится ли это свойство, если вместо момента n взять марковский момент τ? Приведенный в предыдущем параграфе пример 8 показывает, что, вообще говоря, это не так: существует такой мартингал X и марковский момент τ (конечный с вероятностью единица), что

$$\mathsf{M}X_\tau \neq \mathsf{M}X_0. \tag{2}$$

Следующая важная теорема описывает те «типичные» ситуации, для которых, в частности, $\mathsf{M}X_\tau = \mathsf{M}X_0$.

Теорема 1 (Дуб). *Пусть $X = (X_n, \mathscr{F}_n)$ — мартингал (субмартингал), τ_1 и τ_2 — моменты остановки, для которых*

$$\mathsf{M}|X_{\tau_i}| < \infty, \qquad i = 1, 2, \tag{3}$$

$$\varliminf_{n \to \infty} \int_{\{\tau_i > n\}} |X_n| \, d\mathsf{P} = 0, \qquad i = 1, 2. \tag{4}$$

Тогда

$$\mathsf{M}(X_{\tau_2} \mid \mathscr{F}_{\tau_1}) \underset{(\geqslant)}{=} X_{\tau_1} \quad (\{\tau_2 \geqslant \tau_1\};\ \mathsf{P}\text{-п. н.}). \tag{5}$$

Если к тому же $\mathsf{P}(\tau_1 \leqslant \tau_2) = 1$, то

$$\mathsf{M}X_{\tau_2} \underset{(\geqslant)}{=} \mathsf{M}X_{\tau_1}. \tag{6}$$

Доказательство. Достаточно показать, что для всякого $A \in \mathscr{F}_{\tau_1}$

В свою очередь для этого достаточно установить, что для любого $n \geqslant 0$

$$\int_{A \cap \{\tau_2 \geqslant \tau_1\} \cap \{\tau_1 = n\}} X_{\tau_2}\, dP \underset{(\geqslant)}{=} \int_{A \cap \{\tau_2 \geqslant \tau_1\} \cap \{\tau_1 = n\}} X_{\tau_1}\, dP,$$

или, что то же,

$$\int_{B \cap \{\tau_2 \geqslant n\}} X_{\tau_2}\, dP \underset{(\geqslant)}{=} \int_{B \cap \{\tau_2 \geqslant n\}} X_n\, dP, \tag{8}$$

где $B = A \cap \{\tau_1 = n\} \in \mathscr{F}_n$.

Имеем

$$\int_{B \cap \{\tau_2 \geqslant n\}} X_n\, dP = \int_{B \cap \{\tau_2 = n\}} X_n\, dP + \int_{B \cap \{\tau_2 > n\}} X_n\, dP \underset{(\leqslant)}{=} \int_{B \cap \{\tau_2 = n\}} X_n dP +$$
$$+ \int_{B \cap \{\tau_2 > n\}} \mathsf{M}(X_{n+1} \mid \mathscr{F}_n)\, dP = \int_{B \cap \{\tau_2 = n\}} X_{\tau_2} dP + \int_{B \cap \{\tau_2 \geqslant n+1\}} X_{n+1} dP \underset{(\leqslant)}{=}$$
$$\underset{(\leqslant)}{=} \int_{B \cap \{n \leqslant \tau_2 \leqslant n+1\}} X_{\tau_2}\, dP + \int_{B \cap \{\tau_2 \geqslant n+2\}} X_{n+2}\, dP \underset{(\leqslant)}{=} \ldots \underset{(\leqslant)}{=}$$
$$\underset{(\leqslant)}{=} \int_{B \cap \{n \leqslant \tau_2 \leqslant m\}} X_{\tau_2}\, dP + \int_{B \cap \{\tau_2 > m\}} X_m\, dP,$$

откуда

$$\int_{B \cap \{n \leqslant \tau_2 \leqslant m\}} X_{\tau_2}\, dP \underset{(\geqslant)}{=} \int_{B \cap \{n \leqslant \tau_2\}} X_n\, dP - \int_{B \cap \{m < \tau_2\}} X_m\, dP$$

и в силу (4)

$$\int_{B \cap \{\tau_2 \geqslant n\}} X_{\tau_2}\, dP \underset{(\geqslant)}{=} \varlimsup_{m \to \infty} \left[\int_{B \cap \{n \leqslant \tau_2\}} X_n\, dP - \int_{B \cap \{m < \tau_2\}} X_m\, dP \right] =$$
$$= \int_{B \cap \{n \leqslant \tau_2\}} X_n\, dP - \varliminf_{m \to \infty} \int_{B \cap \{m < \tau_2\}} X_m\, dP = \int_{B \cap \{\tau_2 \geqslant n\}} X_n\, dP,$$

что и доказывает (8), а значит, и (5). Наконец, соотношение (6) следует из (5).

Теорема доказана.

Следствие 1. Если существует константа N такая, что $P(\tau_1 \leqslant N) = 1$, $P(\tau_2 \leqslant N) = 1$, то выполнены условия (3), (4). Поэтому, если к тому же $P(\tau_1 \leqslant \tau_2) = 1$ и X — мартингал, то

$$\mathsf{M}X_0 = \mathsf{M}X_{\tau_1} = \mathsf{M}X_{\tau_2} = \mathsf{M}X_N. \tag{9}$$

Следствие 2. Если семейство случайных величин $\{X_n\}$ равномерно интегрируемо (в частности, если с вероятностью единица $|X_n| \leqslant C < \infty$, $n \geqslant 0$), то выполнены условия (3) и (4).

Действительно, $P(\tau_i > n) \to 0$, $n \to \infty$, поэтому условие (4) следует из леммы 2 § 6 гл. II. Далее, поскольку семейство $\{X_n\}$ равномерно интегрируемо, то (см. II.6.16))

$$\sup_N \mathsf{M}|X_N| < \infty. \tag{10}$$

Если τ — некоторый момент остановки и X — субмартингал, то, согласно следствию 1, примененному к ограниченному моменту $\tau_N=\tau\wedge N$,

$$\mathsf{M}X_0\leqslant\mathsf{M}X_{\tau_N}.$$

Поэтому

$$\mathsf{M}|X_{\tau_N}|=2\mathsf{M}X_{\tau_N}^+-\mathsf{M}X_{\tau_N}\leqslant 2\mathsf{M}X_{\tau_N}^+-\mathsf{M}X_0. \tag{11}$$

Последовательность $X^+=(X_n^+, \mathscr{F}_n)$ является субмартингалом (пример 5 из § 1) и, значит,

$$\mathsf{M}X_{\tau_N}^+=\sum_{j=0}^{N}\int_{\{\tau_N=j\}}X_j^+\,d\mathsf{P}+\int_{\{\tau>N\}}X_N^+\,d\mathsf{P}\leqslant\sum_{j=0}^{N}\int_{\{\tau_N=j\}}X_N^+\,d\mathsf{P}+$$
$$+\int_{\{\tau>N\}}X_N^+\,d\mathsf{P}=\mathsf{M}X_N^+\leqslant\mathsf{M}|X_N|\leqslant\sup_N\mathsf{M}|X_N|,$$

что вместе с (11) дает неравенство

$$\mathsf{M}|X_{\tau_N}|\leqslant 3\sup_N\mathsf{M}|X_N|,$$

откуда по лемме Фату

$$\mathsf{M}|X_\tau|\leqslant 3\sup_N\mathsf{M}|X_N|.$$

Поэтому, выбирая $\tau=\tau_i$, $i=1, 2$, и учитывая (10), получаем, что $\mathsf{M}|X_{\tau_i}|<\infty$, $i=1, 2$.

Замечание. В примере 8, рассмотренном в предыдущем параграфе,

$$\int_{\{\tau>n\}}|X_n|\,d\mathsf{P}=(2^n-1)\,\mathsf{P}\{\tau>n\}=(2^n-1)\cdot 2^{-n}\to 1,\quad n\to\infty,$$

и, следовательно, нарушается условие (4) (для $\tau_2=\tau$).

2. Для приложений часто оказывается полезным следующее предложение, выводимое из теоремы 1.

Теорема 2. *Пусть $X=(X_n)$ — мартингал (субмартингал) и τ — момент остановки (относительно $(\mathscr{F}_n^X)$, $\mathscr{F}_n^X=\sigma\{\omega\colon X_0,\ldots,X_n\}$). Предположим, что*

$$\mathsf{M}\tau<\infty,$$

и для любого $n\geqslant 0$ и некоторой константы C

$$\mathsf{M}\{|X_{n+1}-X_n|\,|\,\mathscr{F}_n^X\}\leqslant C\ (\mathsf{P}\text{-п. н.}).$$

Тогда

$$\mathsf{M}|X_\tau|<\infty$$

и

$$\mathsf{M}X_\tau\underset{(\geqslant)}{=}\mathsf{M}X_0. \tag{12}$$

Доказательство. Проверим для $\tau_2=\tau$ выполнение условий (3) и (4) в теореме 1.

Пусть $Y_0=|X_0|$, $Y_j=|X_j-X_{j-1}|$, $j\geqslant 1$. Тогда $|X_\tau|\leqslant\sum_{j=0}^{\tau}Y_j$ и

$$\mathsf{M}|X_\tau|\leqslant\mathsf{M}\left(\sum_{j=0}^{\tau}Y_j\right)=\int_\Omega\left(\sum_{j=0}^{\tau}Y_j\right)d\mathsf{P}=\sum_{n=0}^{\infty}\int_{\{\tau=n\}}\sum_{j=0}^{n}Y_j\,d\mathsf{P}=$$

$$=\sum_{n=0}^{\infty}\sum_{j=0}^{n}\int_{\{\tau=n\}}Y_j\,d\mathsf{P}=\sum_{j=0}^{\infty}\sum_{n=j}^{\infty}\int_{\{\tau=n\}}Y_j\,d\mathsf{P}=\sum_{j=0}^{\infty}\int_{\{\tau\geqslant j\}}Y_j\,d\mathsf{P}.$$

Множество $\{\tau\geqslant j\}=\Omega\diagdown\{\tau<j\}\in\mathscr{F}_{j-1}^X$. Поэтому

$$\int_{\{\tau\geqslant j\}}Y_j\,d\mathsf{P}=\int_{\{\tau\geqslant j\}}\mathsf{M}[Y_j\,|\,X_0,\ \ldots,\ X_{j-1}]\,d\mathsf{P}\leqslant C\mathsf{P}\{\tau\geqslant j\}$$

и в силу (8)

$$\mathsf{M}|X_\tau|\leqslant\mathsf{M}\left(\sum_{j=0}^{\tau}Y_j\right)\leqslant C\sum_{j=0}^{\infty}\mathsf{P}\{\tau\geqslant j\}=C\mathsf{M}\tau<\infty. \tag{13}$$

Далее, если $\tau>n$, то

$$\sum_{j=0}^{n}Y_j\leqslant\sum_{j=0}^{\tau}Y_j,$$

и поэтому

$$\int_{\{\tau>n\}}|X_n|\,d\mathsf{P}\leqslant\int_{\{\tau>n\}}\sum_{j=0}^{\tau}Y_j\,d\mathsf{P}.$$

Отсюда, учитывая, что (согласно (13)) $\mathsf{M}\sum_{j=0}^{\tau}Y_j<\infty$ и что $\{\tau>n\}\downarrow\varnothing$, $n\to\infty$, по теореме о мажорируемой сходимости получаем

$$\varlimsup_{n\to\infty}\int_{\{\tau>n\}}|X_n|\,d\mathsf{P}\leqslant\varlimsup_{n\to\infty}\int_{\{\tau>n\}}\left(\sum_{j=0}^{\tau}Y_j\right)d\mathsf{P}=0.$$

Тем самым выполнены условия теоремы 1, из которой следует требуемое соотношение (12).

Теорема доказана.

3. Остановимся на некоторых применениях доказанных теорем.

Теорема 3 (тождества Вальда). *Пусть $\xi_1,\ \xi_2,\ \ldots$ — независимые одинаково распределенные случайные величины с $\mathsf{M}|\xi_i|<\infty$ и τ — момент остановки (относительно $(\mathscr{F}_n^\xi)$, $\mathscr{F}_n^\xi=\sigma\{\omega\colon\ \xi_1,\ \ldots,\ \xi_n\}$, $\tau\geqslant 1$) с $\mathsf{M}\tau<\infty$. Тогда*

$$\mathsf{M}(\xi_1+\ldots+\xi_\tau)=\mathsf{M}\xi_1\cdot\mathsf{M}\tau. \tag{14}$$

Если к тому же $M\xi_i^2 < \infty$, *то*

$$M\{(\xi_1 + \ldots + \xi_\tau) - \tau M\xi_1\}^2 = D\xi_1 \cdot M\tau. \tag{15}$$

Доказательство. Ясно, что $X = (X_n, \mathscr{F}_n^\xi)_{n \geqslant 1}$ с $X_n = (\xi_1 + \ldots + \xi_n) - nM\xi_1$ есть мартингал с

$$M[|X_{n+1} - X_n| \,|\, X_1, \ldots, X_n] = M[|\xi_{n+1} - M\xi_1| \,|\, \xi_1, \ldots, \xi_n] = M|\xi_{n+1} - M\xi_1| \leqslant 2M|\xi_1| < \infty.$$

Поэтому по теореме 2 $MX_\tau = MX_0 = 0$, что и доказывает (14).

Аналогичные рассмотрения, примененные к мартингалу $Y = (Y_n, \mathscr{F}_n^\xi)$ с $Y_n = X_n^2 - nD\xi_1$, приводят к доказательству соотношения (15).

Следствие. Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$, $S_n = \xi_1 + \ldots + \xi_n$ и $\tau = \inf\{n \geqslant 1\colon S_n = 1\}$. Тогда $P\{\tau < \infty\} = 1$ (см., например, (I.9.20)) и, значит, $P(S_\tau = 1) = 1$, $MS_\tau = 1$. Отсюда и из (14) вытекает, что $M\tau = \infty$.

Теорема 4 (фундаментальное тождество Вальда). *Пусть* $\xi_1, \xi_2, \ldots$ — *последовательность независимых одинаково распределенных случайных величин,* $S_n = \xi_1 + \ldots + \xi_n$, $n \geqslant 1$. *Пусть* $\varphi(t) = Me^{t\xi_1}$, $t \in R$, *причем для некоторого* $t_0 \neq 0$ $\varphi(t_0)$ *существует и* $\varphi(t_0) \geqslant 1$.

Если τ — *момент остановки (относительно* $(\mathscr{F}_n^\xi)$, $\mathscr{F}_n^\xi = \sigma\{\omega\colon \xi_1, \ldots, \xi_n\}$, $\tau \geqslant 1$) *такой, что* $|S_n| \leqslant C$ ($\{\tau \geqslant n\}$; P-*п. н.*) *и* $M\tau < \infty$, *то*

$$M\left[\frac{e^{t_0 S_\tau}}{(\varphi(t_0))^\tau}\right] = 1. \tag{16}$$

Доказательство. Положим

$$Y_n = e^{t_0 S_n}(\varphi(t_0))^{-n}.$$

Тогда $Y = (Y_n, \mathscr{F}_n^\xi)_{n \geqslant 1}$ есть мартингал с $MY_n = 1$ и на множестве $\{\tau \geqslant n\}$

$$M\{|Y_{n+1} - Y_n| \,|\, Y_1, \ldots, Y_n\} = Y_n M\left\{\left|\frac{e^{t_0\xi_{n+1}}}{\varphi(t_0)} - 1\right| \,\middle|\, \xi_1, \ldots, \xi_n\right\} = Y_n \cdot M\{|e^{t_0\xi_1}\varphi^{-1}(t_0) - 1|\} \leqslant B < \infty,$$

где B — некоторая константа. Поэтому применима теорема 2, из которой следует (16), поскольку $MY_1 = 1$.

Теорема доказана.

Пример 1. Этот пример служит иллюстрацией применения вышеизложенных результатов к задачам нахождения вероятностей разорения и средней продолжительности игры (см. § 9 в гл. I).

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых бернуллиевских случайных величин с $P(\xi_i = 1) = p$, $P(\xi_i = -1) = q$, $p + q = 1$,

$S_n = \xi_1 + \ldots + \xi_n$ и

$$\tau = \inf\{n \geqslant 1:\ S_n = B \text{ или } A\}, \tag{17}$$

где $(-A)$ и B — положительные целые числа.

Из (I.9.20) следует, что $\mathsf{P}(\tau < \infty) = 1$ и $\mathsf{M}\tau < \infty$. Тогда, если $\alpha = \mathsf{P}(S_\tau = A)$, $\beta = \mathsf{P}(S_\tau = B)$, то $\alpha + \beta = 1$, и при $p = q = 1/2$ из (14) находим

$$0 = \mathsf{M}S_\tau = \alpha A + \beta B,$$

откуда

$$\alpha = \frac{B}{B + |A|}, \quad \beta = \frac{|A|}{B + |A|}.$$

Применяя (15), получаем

$$\mathsf{M}\tau = \mathsf{M}S_\tau^2 = \alpha A^2 + \beta B^2 = |AB|.$$

Если же $p \neq q$, то, рассматривая мартингал $\left(\left(\frac{q}{p}\right)^{S_n}\right)_{n \geqslant 1}$, находим, что

$$\mathsf{M}\left(\frac{q}{p}\right)^{S_\tau} = \mathsf{M}\left(\frac{q}{p}\right)^{S_1} = 1,$$

и, значит,

$$\alpha\left(\frac{q}{p}\right)^{A} + \beta\left(\frac{q}{p}\right)^{B} = 1.$$

Вместе с равенством $\alpha + \beta = 1$ это дает

$$\alpha = \frac{\left(\frac{q}{p}\right)^{B} - 1}{\left(\frac{q}{p}\right)^{B} - \left(\frac{q}{p}\right)^{|A|}}, \quad \beta = \frac{1 - \left(\frac{q}{p}\right)^{|A|}}{\left(\frac{q}{p}\right)^{B} - \left(\frac{q}{p}\right)^{|A|}}. \tag{18}$$

Наконец, учитывая, что $\mathsf{M}S_\tau = (p - q)\,\mathsf{M}\tau$, находим

$$\mathsf{M}\tau = \frac{\mathsf{M}S_\tau}{p - q} = \frac{\alpha A + \beta B}{p - q},$$

где α и β определяются из (18).

Пример 2. Пусть в рассмотренном выше примере $p = q = 1/2$. Покажем, что для всякого $0 < \lambda < \frac{\pi}{B + |A|}$ и момента τ, определенного в (17),

$$\mathsf{M}(\cos\lambda)^{-\tau} = \frac{\cos\lambda \cdot \frac{B + A}{2}}{\cos\lambda \cdot \frac{B + |A|}{2}}. \tag{19}$$

С этой целью рассмотрим мартингал $X = (X_n, \mathscr{F}_n^\xi)_{n \geqslant 0}$ с

$$X_n = (\cos\lambda)^{-n}\cos\lambda\left(S_n - \frac{B + A}{2}\right)$$

и $S_0=0$. Ясно, что

$$\mathsf{M}X_n=\mathsf{M}X_0=\cos\lambda\frac{B+A}{2}. \tag{20}$$

Покажем, что семейство $\{X_{n\wedge\tau}\}$ является равномерно интегрируемым. Для этого заметим, что в силу следствия 1 к теореме 1 при $0<\lambda<\frac{\pi}{B+|A|}$

$$\mathsf{M}X_0=\mathsf{M}X_{n\wedge\tau}=\mathsf{M}(\cos\lambda)^{-(n\wedge\tau)}\cos\lambda\left(S_{n\wedge\tau}-\frac{B+A}{2}\right)\geqslant$$
$$\geqslant\mathsf{M}(\cos\lambda)^{-(n\wedge\tau)}\cos\lambda\frac{B-A}{2}.$$

Поэтому из (20)

$$\mathsf{M}(\cos\lambda)^{-(n\wedge\tau)}\leqslant\frac{\cos\lambda\frac{B+A}{2}}{\cos\lambda\frac{B+|A|}{2}},$$

и, значит, по лемме Фату

$$\mathsf{M}(\cos\lambda)^{-\tau}\leqslant\frac{\cos\lambda\frac{B+A}{2}}{\cos\lambda\frac{B+|A|}{2}}. \tag{21}$$

Следовательно, согласно (19),

$$|X_{n\wedge\tau}|\leqslant(\cos\lambda)^{-\tau},$$

что вместе с (21) доказывает равномерную интегрируемость семейства $\{X_{n\wedge\tau}\}$. Тогда в силу следствия 2 к теореме 1

$$\cos\lambda\frac{B+A}{2}=\mathsf{M}X_0=\mathsf{M}X_\tau=\mathsf{M}(\cos\lambda)^{-\tau}\cos\lambda\frac{B-A}{2},$$

откуда следует требуемое равенство (18).

4. Задачи.

1. Показать, что в случае субмартингалов теорема 1 остается справедливой, если условие (4) заменить условием

$$\varliminf_{n\to\infty}\int_{\{\tau_i>n\}}X_n^+\,d\mathsf{P}=0,\qquad i=1,\ 2.$$

2. Пусть $X=(X_n,\mathscr{F}_n)_{n\geqslant 0}$ — квадратично интегрируемый мартингал, τ — момент остановки и

$$\varliminf_{n\to\infty}\int_{\{\tau>n\}}X_n^2\,d\mathsf{P}=0,$$
$$\varliminf_{n\to\infty}\int_{\{\tau>n\}}|X_n|\,d\mathsf{P}=0.$$

Показать, что тогда

$$\mathsf{M}X_\tau^2=\mathsf{M}\langle X\rangle_\tau\left(=\mathsf{M}\sum_{j=0}^{\tau}(\Delta X_j)^2\right),$$

где $\Delta X_0=X_0$, $\Delta X_j=X_j-X_{j-1}$, $j\geqslant 1$.

3. Показать, что для каждого мартингала или неотрицательного субмартингала $X=(X_n, \mathscr{F}_n)_{n\geqslant 0}$ и момента остановки τ

$$\mathsf{M}|X_\tau|\leqslant\lim_{n\to\infty}\mathsf{M}|X_n|.$$

4. Пусть $X=(X_n, \mathscr{F}_n)_{n\geqslant 0}$ — супермартингал такой, что $X_n\geqslant\mathsf{M}(\xi|\mathscr{F}_n)$ (Р-п. н.), $n\geqslant 0$, где $\mathsf{M}|\xi|<\infty$. Показать, что если τ_1 и τ_2 — моменты остановки с $\mathsf{P}(\tau_1\leqslant\tau_2)=1$, то

$$X_{\tau_1}\geqslant\mathsf{M}(X_{\tau_2}|\mathscr{F}_{\tau_1})\quad(\text{Р-п. н.}).$$

5. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с $\mathsf{P}(\xi_i=1)=\mathsf{P}(\xi_i=-1)=1/2$, a и b — положительные числа, $b>a$,

$$X_n=a\sum_{k=1}^{n}I(\xi_k=+1)-b\sum_{k=1}^{n}I(\xi_k=-1)$$

и

$$\tau=\inf\{n\geqslant 1:\ X_n\leqslant -r\},\qquad r>0.$$

Показать, что $\mathsf{M}e^{\lambda\tau}<\infty$ при $\lambda\leqslant\alpha_0$ и $\mathsf{M}e^{\lambda\tau}=\infty$ при $\lambda>\alpha_0$, где

$$\alpha_0=\frac{b}{a+b}\ln\frac{2b}{a+b}+\frac{a}{a+b}\ln\frac{2a}{a+b}.$$

6. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с $\mathsf{M}\xi_i=0$, $\mathsf{D}\xi_i=\sigma_i^2$, $S_n=\xi_1+\ldots+\xi_n$, $\mathscr{F}_n^\xi=\sigma\{\omega:\ \xi_1,\ldots,\xi_n\}$. Доказать справедливость следующих утверждений, обобщающих тождества Вальда (14) и (15): если $\mathsf{M}\sum_{j=1}^{\tau}\mathsf{M}|\xi_j|<\infty$, то $\mathsf{M}S_\tau=0$; если $\mathsf{M}\sum_{j=1}^{\tau}\mathsf{M}\xi_j^2<\infty$, то

$$\mathsf{M}S_\tau^2=\mathsf{M}\sum_{j=1}^{\tau}\xi_j^2=\mathsf{M}\sum_{j=1}^{\tau}\sigma_j^2. \tag{22}$$

§ 3. Основные неравенства

1. Пусть $X=(X_n, \mathscr{F}_n)_{n\geqslant 0}$ — стохастическая последовательность,

$$X_n^*=\max_{0\leqslant j\leqslant n}|X_j|,\quad \|X_n\|_p=(\mathsf{M}|X_n|^p)^{1/p},\quad p>0.$$

Теорема 1 (Дуб). *Пусть $X=(X_n, \mathscr{F}_n)$ — неотрицательный субмартингал. Тогда для всякого $\varepsilon>0$ и любого $n\geqslant 0$*

$$\mathsf{P}\{X_n^*\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon}\int_{\{X_n^*\geqslant\varepsilon\}} X_n\,d\mathsf{P}\leqslant\frac{\mathsf{M}X_n}{\varepsilon}; \tag{1}$$

$$\|X_n\|_p\leqslant\|X_n^*\|_p\leqslant\frac{p}{p-1}\|X_n\|_p, \quad \text{если } p>1; \tag{2}$$

$$\|X_n\|_p\leqslant\|X_n^*\|_p\leqslant\frac{e}{e-1}\{1+\|X_n\ln^+X_n\|_p\}, \text{ если } p=1. \tag{3}$$

Доказательство. Обозначим

$$\tau_n=\min\{j\leqslant n\colon X_j\geqslant\varepsilon\},$$

полагая $\tau_n=n$, если $\max\limits_{0\leqslant j\leqslant n} X_j<\varepsilon$. Тогда, согласно (2.6),

$$\mathsf{M}X_n\geqslant\mathsf{M}X_{\tau_n}=$$
$$=\int_{\{X_n^*\geqslant\varepsilon\}} X_{\tau_n}\,d\mathsf{P}+\int_{\{X_n^*<\varepsilon\}} X_{\tau_n}\,d\mathsf{P}\geqslant\varepsilon\int_{\{X_n^*\geqslant\varepsilon\}} d\mathsf{P}+\int_{\{X_n^*<\varepsilon\}} X_n\,d\mathsf{P}.$$

Поэтому

$$\varepsilon\mathsf{P}\{X_n^*\geqslant\varepsilon\}\leqslant\mathsf{M}X_n-\int_{\{X_n^*<\varepsilon\}} X_n\,d\mathsf{P}=\int_{\{X_n^*\geqslant\varepsilon\}} X_n\,d\mathsf{P}\leqslant\mathsf{M}X_n,$$

что и доказывает (1).

Первые неравенства в (2) и (3) очевидны.

Для доказательства второго неравенства в (2) предположим сначала, что

$$\|X_n^*\|_p<\infty, \tag{4}$$

и воспользуемся тем фактом, что для любой неотрицательной случайной величины ξ и $r>0$

$$\mathsf{M}\xi^r=r\int_0^\infty t^{r-1}\mathsf{P}(\xi\geqslant t)\,dt. \tag{5}$$

Тогда из (1) и теоремы Фубини получаем, что для $p>1$

$$\mathsf{M}(X_n^*)^p=p\int_0^\infty t^{p-1}\mathsf{P}\{X_n^*\geqslant t\}\,dt\leqslant p\int_0^\infty t^{p-2}\left(\int_{\{X_n^*\geqslant t\}} X_n\,d\mathsf{P}\right)dt=$$
$$=p\int_0^\infty t^{p-2}\left[\int_\Omega X_n I\{X_n^*\geqslant t\}\right]dt=p\int_\Omega X_n\left[\int_0^{X_n^*} t^{p-2}\,dt\right]d\mathsf{P}=$$
$$=\frac{p}{p-1}\mathsf{M}\left[X_n(X_n^*)^{p-1}\right]. \tag{6}$$

Отсюда по неравенству Гёльдера

$$\mathsf{M}(X_n^*)^p \leqslant q\|X_n\|_p \cdot \|(X_n^*)^{p-1}\|_q = q\|X_n\|_p [\mathsf{M}(X_n^*)^p]^{1/q}, \tag{7}$$

где $q = \frac{p}{p-1}$.

Если выполнено (4), то из (7) сразу получаем второе неравенство в (2).

Если же условие (4) не выполнено, то следует поступить таким образом. Рассмотрим в (6) вместо X_n^* величину $(X_n^* \wedge L)$, где L — некоторая константа. Тогда получим

$$\mathsf{M}(X_n^* \wedge L)^p \leqslant q\mathsf{M}[X_n (X_n^* \wedge L)^{p-1}] \leqslant q\|X_n\|_p [\mathsf{M}(X_n^* \wedge L)^p]^{1/q},$$

откуда в силу неравенства $\mathsf{M}(X_n^* \wedge L)^p \leqslant L^p < \infty$ следует, что

$$\mathsf{M}(X_n^* \wedge L)^p \leqslant q^p \mathsf{M} X_n^p = q^p \|X_n\|_p^p$$

и, значит,

$$\mathsf{M}(X_n^*)^p = \lim_{L \to \infty} \mathsf{M}(X_n^* \wedge L)^p \leqslant q^p \|X_n\|_p^p.$$

Докажем теперь второе неравенство в (3).

Снова применяя (1), находим, что

$$\mathsf{M}X_n^* - 1 \leqslant \mathsf{M}(X_n^* - 1)^+ = \int_0^\infty \mathsf{P}\{X_n^* - 1 \geqslant t\}\,dt \leqslant$$

$$\leqslant \int_0^\infty \frac{1}{1+t}\left[\int_{\{X_n^* \geqslant 1+t\}} X_n\,d\mathsf{P}\right]dt = \mathsf{M}X_n \int_0^{X_n^*-1} \frac{dt}{1+t} = \mathsf{M}X_n \ln X_n^*.$$

Поскольку для любых $a \geqslant 0$ и $b > 0$

$$a \ln b \leqslant a \ln^+ a + be^{-1}, \tag{8}$$

то

$$\mathsf{M}X_n^* - 1 \leqslant \mathsf{M}X_n \ln X_n^* \leqslant \mathsf{M}X_n \ln^+ X_n + e^{-1}\mathsf{M}X_n^*.$$

Если $\mathsf{M}X_n^* < \infty$, то отсюда сразу получаем второе неравенство (3). Если же $\mathsf{M}X_n^* = \infty$, то следует поступить, как и выше, перейдя от величин X_n^* к $X_n^* \wedge L$.

Теорема доказана.

Следствие 1. Пусть $X = (X_n, \mathscr{F}_n)$ — квадратично интегрируемый мартингал. Тогда $X^2 = (X_n^2, \mathscr{F}_n)$ — субмартингал и из (1) следует, что

$$\mathsf{P}\left\{\max_{j \leqslant n} |X_j| \geqslant \varepsilon\right\} \leqslant \frac{\mathsf{M}X_n^2}{\varepsilon^2}. \tag{9}$$

В частности, если $X_j = \xi_0 + \ldots + \xi_j$, где (ξ_j) — последовательность независимых случайных величин с $\mathsf{M}\xi_j = 0$ и $\mathsf{M}\xi_j^2 < \infty$, то неравенство (9) превращается в неравенство Колмогорова (§ 2 гл. IV).

Следствие 2. Если $X=(X_n, \mathscr{F}_n)$ — квадратично интегрируемый мартингал, то из (2) получаем, что

$$\mathsf{M}\left[\max_{j\leqslant n} X_j^2\right]\leqslant 4\mathsf{M}X_n^2. \tag{10}$$

2. Пусть $X=(X_n, \mathscr{F}_n)$ — субмартингал и

$$X_n=M_n+A_n$$

— его разложение Дуба. Тогда, поскольку $\mathsf{M}M_n=0$, то из (1) следует, что

$$\mathsf{P}\{X_n^*\geqslant\varepsilon\}\leqslant\frac{\mathsf{M}A_n}{\varepsilon}.$$

Нижеследующая теорема 2 показывает, что это неравенство справедливо не только для субмартингалов, но и для более широкого класса последовательностей, обладающих свойством доминируемости в следующем смысле.

Определение. Пусть $X=(X_n, \mathscr{F}_n)$ — некоторая неотрицательная стохастическая последовательность и $A=(A_n, \mathscr{F}_{n-1})$ — возрастающая предсказуемая последовательность. Будем говорить, что X *доминируется* последовательностью A, если

$$\mathsf{M}X_\tau\leqslant\mathsf{M}A_\tau \tag{11}$$

для всякого момента остановки τ.

Теорема 2. *Если $X=(X_n, \mathscr{F}_n)$ — неотрицательная стохастическая последовательность, доминируемая возрастающей предсказуемой последовательностью $A=(A_n, \mathscr{F}_{n-1})$, то для $\varepsilon>0$, $a>0$ и любого момента остановки τ*

$$\mathsf{P}\{X_\tau^*\geqslant\varepsilon\}\leqslant\frac{\mathsf{M}A_\tau}{\varepsilon}, \tag{12}$$

$$\mathsf{P}\{X_\tau^*\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon}\mathsf{M}(A_\tau\wedge a)+\mathsf{P}(A_\tau\geqslant a), \tag{13}$$

$$\|X_\tau^*\|_p\leqslant\left(\frac{2-p}{1-p}\right)^{1/p}\|A_\tau\|_p,\qquad 0<p<1. \tag{14}$$

Доказательство. Положим

$$\sigma_n=\min\{j\leqslant\tau\wedge n:\ X_j\geqslant\varepsilon\},$$

считая $\sigma_n=\tau\wedge n$, если $\{\cdot\}=\varnothing$. Тогда

$$\mathsf{M}A_\tau\geqslant\mathsf{M}A_{\sigma_n}\geqslant\mathsf{M}X_{\sigma_n}\geqslant\int\limits_{\{X_{\tau\wedge n}^*>\varepsilon\}}X_{\sigma_n}\,d\mathsf{P}\geqslant\varepsilon\mathsf{P}\{X_{\tau\wedge n}^*>\varepsilon\},$$

откуда

$$\mathsf{P}\{X^*_{\tau\wedge n} > \varepsilon\} \leqslant \frac{1}{\varepsilon}\mathsf{M}A_\tau,$$

что в силу леммы Фату доказывает неравенство (12).

Для доказательства (13) введем момент

$$\gamma = \inf\{j:\ A_{j+1} \geqslant a\},$$

полагая $\gamma = \infty$, если $\{\cdot\} = \varnothing$. Тогда

$$\begin{aligned}\mathsf{P}\{X^*_\tau \geqslant \varepsilon\} &= \mathsf{P}\{X^*_\tau \geqslant \varepsilon,\ A_\tau < a\} + \mathsf{P}\{X^*_\tau \geqslant \varepsilon,\ A_\tau \geqslant a\} \leqslant \\ &\leqslant \mathsf{P}\left\{I_{\{A_\tau < a\}} X^*_\tau \geqslant \varepsilon\right\} + \mathsf{P}\{A_\tau \geqslant a\} \leqslant \\ &\leqslant \mathsf{P}\{X^*_{\tau\wedge\gamma} \geqslant \varepsilon\} + \mathsf{P}\{A_\tau \geqslant a\} \leqslant \frac{1}{\varepsilon}\mathsf{M}A_{\tau\wedge\gamma} + \mathsf{P}\{A_\tau \geqslant a\} \leqslant \\ &\leqslant \frac{1}{\varepsilon}\mathsf{M}(A_\tau \wedge a) + \mathsf{P}(A_\tau \geqslant a),\end{aligned}$$

где использовано неравенство (12) и то, что $I_{\{A_\tau < a\}} X^*_\tau \leqslant X^*_{\tau\wedge\gamma}$. Наконец, из (13)

$$\begin{aligned}\|X^*_\tau\|_p^p &= \mathsf{M}(X^*_\tau)^p = \int_0^\infty \mathsf{P}\{(X^*_\tau)^p \geqslant t\}\,dt = \int_0^\infty \mathsf{P}\{X^*_\tau \geqslant t^{1/p}\}\,dt \leqslant \\ &\leqslant \int_0^\infty t^{-1/p}\mathsf{M}[A_\tau \wedge t^{1/p}]\,dt + \int_0^\infty \mathsf{P}\{A^p_\tau \geqslant t\}\,dt = \\ &= \mathsf{M}\int_0^{A^p_\tau} dt + \mathsf{M}\int_{A^p_\tau}^\infty (A_\tau t^{-1/p})\,dt + \mathsf{M}A^p_\tau = \frac{2-p}{1-p}\mathsf{M}A^p_\tau.\end{aligned}$$

Теорема доказана.

Следствие. Пусть при каждом $k \geqslant 1$ последовательности X^k и A^k удовлетворяют условиям теоремы 2 и для некоторого момента остановки τ

$$A^k_\tau \xrightarrow{\mathsf{P}} 0, \qquad k \to \infty.$$

Тогда

$$(X^k)^*_\tau \xrightarrow{\mathsf{P}} 0, \qquad k \to \infty,$$

что сразу следует из неравенства

$$\mathsf{P}\{(X^k)^*_\tau \geqslant \varepsilon\} \leqslant \frac{a}{\varepsilon} + \mathsf{P}\left\{A^k_\tau \geqslant a\right\},$$

вытекающего из (13).

3. В этом пункте будет приведен (без доказательства, но с применениями) ряд замечательных неравенств для мартингалов,

являющихся обобщениями неравенств Хинчина и неравенств Марцинкевича и Зигмунда для сумм независимых случайных величин.

Неравенства Хинчина. *Пусть* $\xi_1, \xi_2, \ldots$ — *независимые одинаково распределенные бернуллиевские случайные величины с* $\mathsf{P}(\xi_i=1)=\mathsf{P}(\xi_i=-1)=1/2$ *и* $(c_n)_{n\geqslant 1}$ — *некоторая последовательность чисел.*

Тогда для любого $0<p<\infty$ *существуют такие универсальные константы* A_p *и* B_p *(не зависящие от* (c_n)*), что для любого* $n\geqslant 1$

$$A_p\left(\sum_{j=1}^{n} c_j^2\right)^{1/2} \leqslant \left\|\sum_{j=1}^{n} c_j\xi_j\right\|_p \leqslant B_p\left(\sum_{j=1}^{n} c_j^2\right)^{1/2}. \tag{15}$$

Обобщением этих неравенств (для $p\geqslant 1$) являются.

Неравенства Марцинкевича и Зигмунда. *Если* $\xi_1, \xi_2, \ldots$ — *последовательность независимых интегрируемых случайных величин с* $\mathsf{M}\xi_i=0$, *то для* $p\geqslant 1$ *найдутся такие универсальные константы* A_p *и* B_p *(не зависящие от* (ξ_n)*), что для любого* $n\geqslant 1$

$$A_p\left\|\left(\sum_{j=1}^{n} \xi_j^2\right)^{1/2}\right\|_p \leqslant \left\|\sum_{j=1}^{n} \xi_j\right\|_p \leqslant B_p\left\|\left(\sum_{j=1}^{n} \xi_j^2\right)^{1/2}\right\|_p. \tag{16}$$

В неравенствах (15) и (16) последовательности $X=(X_n)$ с $X_n = \sum_{j=1}^{n} c_j\xi_j$ и $X_n=\sum_{j=1}^{n}\xi_j$ образуют мартингалы. Естественно задаться вопросом о том, нельзя ли обобщить эти неравенства на случай произвольных мартингалов.

Первый результат в этом направлении был получен Буркхольдером.

Неравенства Буркхольдера. *Если* $X=(X_n, \mathscr{F}_n)$ — *мартингал, то для всякого* $p>1$ *существуют такие универсальные константы* A_p *и* B_p *(не зависящие от* X*), что для любого* $n\geqslant 1$

$$A_p\|\sqrt{[X]_n}\|_p \leqslant \|X_n\|_p \leqslant B_p\|\sqrt{[X]_n}\|_p, \tag{17}$$

где $[X]_n$ — *квадратическая вариация* X_n,

$$[X]_n=\sum_{j=1}^{n}(\Delta X_j)^2, \quad X_0=0. \tag{18}$$

В качестве констант A_p *и* B_p *можно взять*

$$A_p=[18p^{3/2}/(p-1)]^{-1}, \quad B_p=18p^{3/2}/(p-1)^{1/2}.$$

С учетом (2) из (17) следует, что

$$A_p\|\sqrt{[X]_n}\|_p \leqslant \|X_n^*\|_p \leqslant B_p^*\|\sqrt{[X]_n}\|_p, \tag{19}$$

Неравенства Буркхольдера (17) справедливы для $p>1$, в то время как неравенства Марцинкевича — Зигмунда (16) верны и для $p=1$. Что можно сказать о справедливости неравенств (17) для $p=1$? Оказывается их обобщение на случай $p=1$ в форме (17) уже несправедливо, что показывает следующий

Пример. Пусть $\xi_1, \xi_2, \ldots$ — независимые бернуллиевские случайные величины с $\mathsf{P}(\xi_i=1)=\mathsf{P}(\xi_i=-1)=1/2$ и

$$X_n=\sum_{j=1}^{n\wedge\tau}\xi_j,$$

где $\tau=\inf\left\{n\geq 1:\ \sum_{j=1}^{n}\xi_j=1\right\}$.

Последовательность $X=(X_n, \mathscr{F}_n^\xi)$ является мартингалом с

$$\|X_n\|_1=\mathsf{M}|X_n|=2\mathsf{M}X_n^+\to 2, \qquad n\to\infty.$$

Но

$$\|\sqrt{[X]_n}\|_1=\mathsf{M}\sqrt{[X]_n}=\mathsf{M}\left(\sum_{j=1}^{\tau\wedge n}1\right)^{1/2}=\mathsf{M}\sqrt{\tau\wedge n}\to\infty.$$

Следовательно, первое неравенство в (17) несправедливо.

Оказалось, что на случай $p=1$ обобщаются не неравенства (17), а неравенства (19) (эквивалентные, если $p>1$).

Неравенства Дэвиса. *Если $X=(X_n, \mathscr{F}_n)$ — мартингал, то существуют такие универсальные константы A и B, $0<A<B<\infty$, что*

$$A\|\sqrt{[X]_n}\|_1\leq\|X_n^*\|_1\leq B\|\sqrt{[X]_n}\|_1, \tag{20}$$

т. е.

$$A\mathsf{M}\sqrt{\sum_{j=1}^{n}(\Delta X_j)^2}\leq\mathsf{M}\left[\max_{1\leq j\leq n}|X_n|\right]\leq B\mathsf{M}\sqrt{\sum_{j=1}^{n}(\Delta X_j)^2}.$$

Следствие 1. Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины, $S_n=\xi_1+\ldots+\xi_n$. Если $\mathsf{M}|\xi_1|<\infty$ и $\mathsf{M}\xi_1=0$, то, согласно тождеству Вальда (2.14), для всякого момента остановки τ (относительно $(\mathscr{F}_n^\xi)$) с $\mathsf{M}\tau<\infty$

$$\mathsf{M}S_\tau=0. \tag{21}$$

Оказывается, что для справедливости (21) предположение $\mathsf{M}\tau<\infty$ можно ослабить, если усилить требования на сами случайные величины. Именно, если

$$\mathsf{M}|\xi_1|^r<\infty,$$

где $1<r\leq 2$, то условие $\mathsf{M}\tau^{1/r}<\infty$ достаточно для справедливости равенства $\mathsf{M}S_\tau=0$.

Для доказательства обозначим $\tau_n = \tau \wedge n$, $Y = \sup\limits_n |S_{\tau_n}|$, и пусть для $t > 0$ $m = [t^r]$ — целая часть числа t^r. В силу следствия 1 к теореме 1 § 1 $\mathsf{M}S_{\tau_n} = 0$. Поэтому для справедливости соотношения $\mathsf{M}S_\tau = 0$ достаточно (согласно теореме о мажорируемой сходимости) проверить, что $\mathsf{M}\sup\limits_n |S_{\tau_n}| < \infty$.

Пользуясь неравенствами (1) и (17), находим

$$\begin{aligned}
\mathsf{P}(Y \geqslant t) &= \mathsf{P}(\tau \geqslant t^r,\ Y \geqslant t) + \mathsf{P}(\tau < t^r,\ Y \geqslant t) \leqslant \\
&\leqslant \mathsf{P}(\tau \geqslant t^r) + \mathsf{P}\left\{\max_{1 \leqslant j \leqslant m} |S_{\tau_j}| \geqslant t\right\} \leqslant \\
&\leqslant \mathsf{P}(\tau \geqslant t^r) + t^{-r}\mathsf{M}|S_{\tau_m}|^r \leqslant \\
&\leqslant \mathsf{P}(\tau \geqslant t^r) + t^{-r}B_r\mathsf{M}\left(\sum_{j=1}^{\tau_m} \xi_j^2\right)^{r/2} \leqslant \\
&\leqslant \mathsf{P}(\tau \geqslant t^r) + t^{-r}B_r\mathsf{M}\sum_{j=1}^{\tau_m} |\xi_j|^r.
\end{aligned}$$

Заметим, что $(\mathscr{F}_0^\xi = \{\varnothing, \Omega\})$

$$\begin{aligned}
\mathsf{M}\sum_{j=1}^{\tau_m} |\xi_j|^r &= \mathsf{M}\sum_{j=1}^{\infty} I(j \leqslant \tau_m)|\xi_j|^r = \\
&= \sum_{j=1}^{\infty} \mathsf{M}\mathsf{M}\left[I(j \leqslant \tau_m)|\xi_j|^r \mid \mathscr{F}_{j-1}^\xi\right] = \\
&= \sum_{j=1}^{\infty} I(j \leqslant \tau_m)\mathsf{M}\left[|\xi|^r \mid \mathscr{F}_{j-1}^\xi\right] = \mathsf{M}\sum_{j=1}^{\tau_m} \mathsf{M}|\xi_j|^r = \mu_r\mathsf{M}\tau_m,
\end{aligned}$$

где $\mu_r = \mathsf{M}|\xi_1|^r$. Поэтому

$$\begin{aligned}
\mathsf{P}(Y \geqslant t) &\leqslant \mathsf{P}(\tau \geqslant t^r) + t^{-r}B_r\mu_r\mathsf{M}\tau_m = \\
&= \mathsf{P}(\tau \geqslant t^r) + B_r\mu_r t^{-r}\left[m\mathsf{P}(\tau \geqslant t^r) + \int_{\{\tau < t^r\}} \tau\, d\mathsf{P}\right] \leqslant \\
&\leqslant (1 + B_r\mu_r)\mathsf{P}(\tau \geqslant t^r) + B_r\mu_r t^{-r}\int_{\{\tau < t^r\}} \tau\, d\mathsf{P}
\end{aligned}$$

и, значит,

$$\begin{aligned}
\mathsf{M}Y &= \int_0^\infty \mathsf{P}(Y \geqslant t)\,dt \leqslant (1 + B_r\mu_r)\mathsf{M}\tau^{1/r} + B_r\mu_r\int_0^\infty t^{-r}\left[\int_{\{\tau < t^r\}} \tau\, d\mathsf{P}\right]dt = \\
&= (1 + B_r\mu_r)\mathsf{M}\tau^{1/r} + B_r\mu_r\int_\Omega \tau\left[\int_{\tau^{1/r}}^\infty t^{-r}\,dt\right]d\mathsf{P} = \\
&= \left(1 + B_r\mu_r + \frac{B_r\mu_r}{r-1}\right)\mathsf{M}\tau^{1/r} < \infty.
\end{aligned}$$

Следствие 2. Пусть $M=(M_n)$ — мартингал с $\mathsf{M}|M_n|^{2r}<\infty$ для некоторого $r\geqslant 1$ и такой, что ($M_0=0$)

$$\sum_{n=1}^{\infty}\frac{\mathsf{M}|\Delta M_n|^{2r}}{n^{1+r}}<\infty, \tag{22}$$

тогда (ср. с теоремой 2 § 3 гл. IV) имеет место усиленный закон больших чисел:

$$\frac{M_n}{n}\to 0\ (\mathsf{P}\text{-п. н.}),\ n\to\infty. \tag{23}$$

В случае $r=1$ доказательство проводится по той же схеме что и доказательство теоремы 2 § 3 гл. IV. А именно, пусть

$$m_n=\sum_{k=1}^{n}\frac{\Delta M_k}{k}.$$

Тогда

$$\frac{M_n}{n}=\frac{\sum\limits_{k=1}^{n}\Delta M_k}{n}=\frac{1}{n}\sum_{k=1}^{n}k\Delta m_k$$

и, согласно лемме Кронекера (§ 3 гл. IV) для сходимости (P-п. н.)

$$\frac{1}{n}\sum_{k=1}^{n}k\Delta m_k\to 0,\qquad n\to\infty,$$

достаточно, чтобы (P-п. н.) существовал конечный предел $\lim\limits_n m_n$, что в свою очередь (теоремы 1 и 4 из § 10 гл. II) имеет место в том и только том случае, когда

$$\mathsf{P}\left\{\sup_{k\geqslant 1}|m_{n+k}-m_n|\geqslant\varepsilon\right\}\to 0,\ n\to\infty. \tag{24}$$

В силу неравенства (1)

$$\mathsf{P}\left\{\sup_{k\geqslant 1}|m_{n+k}-m_n|\geqslant\varepsilon\right\}\leqslant\frac{\sum\limits_{k=n}^{\infty}\frac{\mathsf{M}(\Delta M_k)^2}{k^2}}{\varepsilon^2}.$$

Поэтому требуемый результат следует из (22) и (24).

Пусть теперь $r>1$. Утверждение (23) эквивалентно тому (теорема 1 § 10 гл. II), что для всякого $\varepsilon>0$

$$\varepsilon^{2r}\mathsf{P}\left\{\sup_{j\geqslant n}\frac{|M_j|}{j}\geqslant\varepsilon\right\}\to 0,\ n\to\infty. \tag{25}$$

В силу неравенства (29) из задачи 1

$$\varepsilon^{2r}\mathsf{P}\left\{\sup_{j\geqslant n}\frac{|M_j|}{j}\geqslant\varepsilon\right\}=\varepsilon^{2r}\lim_{m\to\infty}\mathsf{P}\left\{\max_{n\leqslant j\leqslant m}\frac{|M_j|^{2r}}{j^{2r}}\geqslant\varepsilon^{2r}\right\}\leqslant$$
$$\leqslant\frac{1}{n^{2r}}\mathsf{M}|M_n|^{2r}+\sum_{j\geqslant n+1}\frac{1}{j^{2r}}\mathsf{M}(|M_j|^{2r}-|M_{j-1}|^{2r}).$$

Из леммы Кронекера и условия (22) вытекает, что

$$\lim_{n\to\infty}\frac{1}{n^{2r}}\mathsf{M}|M_n|^{2r}=0.$$

Поэтому для доказательства (25) достаточно лишь показать, что

$$\sum_{j\geqslant 2}\frac{1}{j^{2r}}\mathsf{M}(|M_j|^{2r}-|M_{j-1}|^{2r})<\infty. \tag{26}$$

Имеем

$$I_N=\sum_{j=2}^{N}\frac{1}{j^{2r}}[\mathsf{M}|M_j|^{2r}-\mathsf{M}|M_{j-1}|^{2r}]\leqslant$$
$$\leqslant\sum_{j=3}^{N}\left[\frac{1}{(j-1)^{2r}}-\frac{1}{j^{2r}}\right]\mathsf{M}|M_{j-1}|^{2r}+\frac{\mathsf{M}|M_N|^{2r}}{N^{2r}}.$$

В силу неравенства Буркхольдера (17) и неравенства Гёльдера

$$\mathsf{M}|M_j|^{2r}\leqslant\mathsf{M}\left[\sum_{i=1}^{j}(\Delta M_i)^2\right]^r\leqslant\mathsf{M}j^{r-1}\sum_{i=1}^{j}|\Delta M_i|^{2r}.$$

Поэтому

$$I_N\leqslant\sum_{j=2}^{N-1}\left[\frac{1}{j^{2r}}-\frac{1}{(j+1)^{2r}}\right]j^{r-1}\sum_{i=1}^{j}\mathsf{M}|\Delta M_i|^{2r}\leqslant$$
$$\leqslant C_1\sum_{j=2}^{N-1}\frac{1}{j^{r+2}}\sum_{i=1}^{j}\mathsf{M}|\Delta M_i|^{2r}\leqslant C_2\sum_{j=2}^{N}\frac{\mathsf{M}|\Delta M_j|^{2r}}{j^{r+1}}+C_3$$

(C_i — некоторые константы), что в силу (22) доказывает оценку (26).

Последовательность случайных величин $\{X_n\}_{n\geqslant 1}$ имеет с вероятностью единица предел $\lim X_n$ (конечный или бесконечный) тогда и только тогда, когда число «осцилляций между двумя любыми (рациональными) числами a и b, $a<b$», конечно с вероятностью единица. Приводимая ниже теорема 3 дает оценку сверху среднего числа «осцилляций» для субмартингалов, которая в следующем параграфе будет использована для доказательства фундаментального результата о их сходимости.

Зафиксируем два числа a и b, $a<b$, и для стохастической последовательности $X=(X_n, \mathscr{F}_n)$ определим моменты:

$$\begin{aligned} \tau_0 &= 0, \\ \tau_1 &= \min\{n>0\colon X_n \leqslant a\}, \\ \tau_2 &= \min\{n>\tau_1\colon X_n \geqslant b\}, \\ &\dots\dots\dots\dots\dots \\ \tau_{2m-1} &= \min\{n>\tau_{2m-2}\colon X_n \leqslant a\}, \\ \tau_{2m} &= \min\{n>\tau_{2m-1}\colon X_n \geqslant b\}, \end{aligned}$$

полагая $\tau_k=0$, если соответствующее множество $\{\cdot\}$ пусто.

Далее, для каждого $n\geqslant 1$ определим случайные величины

$$\beta_n(a, b) = \begin{cases} 0, & \text{если } \tau_2 > n, \\ \max\{m\colon \tau_{2m}\leqslant n\}, & \text{если } \tau_2\leqslant n. \end{cases}$$

По своему смыслу $\beta_n(a, b)$ есть *число пересечений* (снизу вверх) интервала $[a, b]$ последовательностью $X_1, \dots, X_n$.

Теорема 3 (Дуб). *Пусть $X=(X_n, \mathscr{F}_n)_{n\geqslant 1}$ — субмартингал. Тогда для любого $n\geqslant 1$*

$$\mathsf{M}\beta_n(a, b) \leqslant \frac{\mathsf{M}\,[X_n-a]^+}{b-a}. \tag{27}$$

Доказательство. Число пересечений субмартингалом $X=(X_n, \mathscr{F}_n)$ интервала $[a, b]$ совпадает с числом пересечений интервала $[0, b-a]$ неотрицательным субмартингалом $X^+=((X_n-a)^+, \mathscr{F}_n)$. Поэтому, считая рассматриваемый субмартингал X неотрицательным и $a=0$, надо доказать, что

$$\mathsf{M}\beta_n(0, b) \leqslant \frac{\mathsf{M}X_n}{b}. \tag{28}$$

Положим $X_0=0$, $\mathscr{F}_0=\{\varnothing, \Omega\}$, и пусть для $i=1, 2, \dots$

$$\varphi_i = \begin{cases} 1, & \text{если } \tau_m < i \leqslant \tau_{m+1} \text{ для некоторого нечетного } m, \\ 0, & \text{если } \tau_m < i \leqslant \tau_{m+1} \text{ для некоторого четного } m. \end{cases}$$

Нетрудно видеть, что

$$b\beta_n(0, b) \leqslant \sum_{i=1}^{n} \varphi_i[X_i - X_{i-1}]$$

и

$$\{\varphi_i = 1\} = \bigcup_{m\text{ — нечетно}} [\{\tau_m < i\} \setminus \{\tau_{m+1} < i\}] \in \mathscr{F}_{i-1}.$$

Поэтому

$$b\mathsf{M}\beta_n(0,\ b)\leqslant \mathsf{M}\sum_{i=1}^{n}\varphi_i[X_i-X_{i-1}]=\sum_{i=1}^{n}\int\limits_{\{\varphi_i=1\}}(X_i-X_{i-1})\,d\mathsf{P}=$$

$$=\sum_{i=1}^{n}\int\limits_{\{\varphi_i=1\}}\mathsf{M}(X_i-X_{i-1}\mid\mathscr{F}_{i-1})\,d\mathsf{P}=$$

$$=\sum_{i=1}^{n}\int\limits_{\{\varphi_i=1\}}[\mathsf{M}(X_i\mid\mathscr{F}_{i-1})-X_{i-1}]\,d\mathsf{P}\leqslant$$

$$\leqslant\sum_{i=1}^{n}\int\limits_{\Omega}[\mathsf{M}(X_i\mid\mathscr{F}_{i-1})-X_{i-1}]\,d\mathsf{P}=\mathsf{M}X_n,$$

что и доказывает неравенство (28).

4. Задачи.

1. Пусть $X=(X_n,\ \mathscr{F}_n)$ — неотрицательный субмартингал и $V=(V_n,\ \mathscr{F}_{n-1})$ — предсказуемая последовательность такая, что $0\leqslant V_{n+1}\leqslant V_n\leqslant C$ (P-п. н.), где C — некоторая константа. Показать, что имеет место следующее обобщение неравенства (1):

$$\varepsilon\mathsf{P}\left\{\max_{1\leqslant j\leqslant n}V_jX_j\right\}+\int\limits_{\left\{\max\limits_{1\leqslant j\leqslant n}V_jX_j<\varepsilon\right\}}V_nX_n\,d\mathsf{P}\leqslant\sum_{j=1}^{n}\mathsf{M}V_j\,\Delta X_j. \tag{29}$$

2. Пусть $X=(X_n,\ \mathscr{F}_n)$ — супермартингал. Показать, что

$$\mathsf{P}\left\{\max_{1\leqslant j\leqslant n}|X_j|\geqslant\varepsilon\right\}\leqslant\frac{C}{\varepsilon}\cdot\max_{1\leqslant j\leqslant n}\mathsf{M}|X_j|,$$

где $C\leqslant 3$ (константа C может быть взята равной единице в случае, когда X — мартингал или когда X не меняет знака).

3. Доказать справедливость *разложения Крикеберга*: всякий мартингал $X=(X_n,\ \mathscr{F}_n)$ с $\sup\mathsf{M}|X_n|<\infty$ может быть представлен как разность двух неотрицательных мартингалов.

4. Пусть $X=(X_n,\ \mathscr{F}_n)$ — субмартингал. Показать, что для всякого $\varepsilon>0$ и $n\geqslant 1$

$$\varepsilon\mathsf{P}\left\{\min_{1\leqslant j\leqslant n}X_j\leqslant-\varepsilon\right\}\leqslant\mathsf{M}(X_n-X_1)-\int\limits_{\left\{\min\limits_{1\leqslant j\leqslant n}X_j\leqslant-\varepsilon\right\}}X_n\,d\mathsf{P}\leqslant$$

$$\leqslant\mathsf{M}X_n^+-\mathsf{M}X_1.$$

5. Пусть $\xi_1,\ \xi_2,\ \ldots$ — последовательность независимых случайных величин, $S_n=\xi_1+\ldots+\xi_n$ и $S_{m,n}=\sum\limits_{j=m+1}^{n}\xi_j$. Доказать

справедливость следующего неравенства Оттавиани:

$$\mathsf{P}\left\{\max_{1\leqslant j\leqslant n}|S_j|>2\varepsilon\right\}\leqslant\frac{\mathsf{P}\{|S_n|>\varepsilon\}}{\min\limits_{1\leqslant j\leqslant n}\mathsf{P}\{|S_{j,n}|\leqslant\varepsilon\}}$$

и вывести из него, что

$$\int_0^\infty \mathsf{P}\left\{\max_{1\leqslant j\leqslant n}|S_j|>2t\right\}dt\leqslant 2\mathsf{M}|S_n|+2\int_{2\mathsf{M}|S_n|}^\infty \mathsf{P}\{|S_n|>t\}\,dt. \qquad (30)$$

6. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с $\mathsf{M}\xi_i=0$. Используя неравенство (30), установить, что для рассматриваемого случая имеет место следующее усиление неравенства (3):

$$\mathsf{M}S_n^*\leqslant 8\mathsf{M}|S_n|.$$

7. Доказать справедливость формулы (5).

8. Доказать неравенство (8).

9. Пусть σ-алгебры $\mathscr{F}_0, \ldots, \mathscr{F}_n$ таковы, что $\mathscr{F}_0\subseteq\mathscr{F}_1\subseteq\ldots\subseteq\mathscr{F}_n$ и события $A_k\in\mathscr{F}_k$, $k=1, \ldots, n$. Используя (13), доказать справедливость следующего *неравенства Дворецкого*: для всякого $\varepsilon>0$

$$\mathsf{P}\left[\bigcup_{k=1}^n A_k|\mathscr{F}_0\right]\leqslant\varepsilon+\mathsf{P}\left[\sum_{k=1}^n \mathsf{P}(A_k|\mathscr{F}_{k-1})>\varepsilon|\mathscr{F}_0\right]\quad (\mathsf{P}\text{-п. н.})$$

§ 4. Основные теоремы о сходимости субмартингалов и мартингалов

1. Следующий результат, являющийся основным во всей проблематике сходимости субмартингалов, можно рассматривать как вероятностный аналог того известного факта из анализа, что ограниченная монотонная числовая последовательность имеет (конечный) предел.

Теорема 1 (Дуб). *Пусть $X=(X_n, \mathscr{F}_n)_{n\geqslant 1}$ — субмартингал с*

$$\sup_n \mathsf{M}|X_n|<\infty. \qquad (1)$$

Тогда с вероятностью единица существует предел $\lim X_n=X_\infty$ *и* $\mathsf{M}|X_\infty|<\infty$.

Доказательство. Предположим, что

$$\mathsf{P}(\overline{\lim}\,X_n>\underline{\lim}\,X_n)>0. \qquad (2)$$

Тогда поскольку

$$\{\overline{\lim}\,X_n>\underline{\lim}\,X_n\}=\bigcup_{a<b}\{\overline{\lim}\,X_n>b>a>\underline{\lim}\,X_n\}$$

(a, b — рациональные числа), то найдутся такие a и b, что

$$\mathsf{P}\{\overline{\lim} X_n > b > a > \underline{\lim} X_n\} > 0. \tag{3}$$

Пусть $\beta_n(a, b)$ — число пересечений снизу вверх последовательностью $X_1, \ldots, X_n$ интервала (a, b) и $\beta_\infty(a, b) = \lim\limits_n \beta_n(a, b)$. Согласно (3.27)

$$\mathsf{M}\beta_n(a, b) \leqslant \frac{\mathsf{M}[X_n - a]^+}{b-a} \leqslant \frac{\mathsf{M}X_n^+ + |a|}{b-a}$$

и, значит,

$$\mathsf{M}\beta_\infty(a, b) = \lim_n \mathsf{M}\beta_n(a, b) \leqslant \frac{\sup\limits_n \mathsf{M}X_n^+ + |a|}{b-a} < \infty,$$

что следует из (1) и того замечания, что для субмартингалов

$$\sup_n \mathsf{M}|X_n| < \infty \Leftrightarrow \sup_n \mathsf{M}X_n^+ < \infty$$

(поскольку $\mathsf{M}X_n^+ \leqslant \mathsf{M}|X_n| = 2\mathsf{M}X_n^+ - \mathsf{M}X_n \leqslant 2\mathsf{M}X_n^+ - \mathsf{M}X_1$). Но условие $\mathsf{M}\beta_\infty(a, b) < \infty$ противоречит допущению (3). Следовательно, с вероятностью единица существует $\lim X_n = X_\infty$, для которого из леммы Фату

$$\mathsf{M}|X_\infty| \leqslant \sup_n \mathsf{M}|X_n| < \infty.$$

Теорема доказана.

Следствие 1. Если X — неположительный субмартингал, то с вероятностью единица существует конечный предел $\lim X_n$.

Следствие 2. Если $X = (X_n, \mathscr{F}_n)_{n \geqslant 1}$ — неположительный субмартингал, то последовательность $\overline{X} = (X_n, \mathscr{F}_n)$ с $1 \leqslant n \leqslant \infty$, $X_\infty = \lim X_n$ и $\mathscr{F}_\infty = \sigma(\bigcup \mathscr{F}_n)$ образует (неположительный) субмартингал.

Действительно, по лемме Фату

$$\mathsf{M}X_\infty = \mathsf{M}\lim X_n \geqslant \overline{\lim}\, \mathsf{M}X_n \geqslant \mathsf{M}X_1 > -\infty$$

и (P-п. н.)

$$\mathsf{M}(X_\infty | \mathscr{F}_m) = \mathsf{M}(\lim X_n | \mathscr{F}_m) \geqslant \overline{\lim}\, \mathsf{M}(X_n | \mathscr{F}_m) \geqslant X_m.$$

Следствие 3. Если $X = (X_n, \mathscr{F}_n)$ — неотрицательный мартингал, то с вероятностью единица существует $\lim X_n$.

В самом деле, тогда

$$\sup_n \mathsf{M}|X_n| = \sup_n \mathsf{M}X_n = \mathsf{M}X_1 < \infty$$

и применима теорема 1.

2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с $\mathsf{P}(\xi_i=0)=\mathsf{P}(\xi_i=2)=1/2$. Тогда $X=(X_n, \mathscr{F}_n^{\xi})$ с $X_n=\prod\limits_{i=1}^{n}\xi_i$ и $\mathscr{F}_n^{\xi}=\sigma\{\omega\colon \xi_1, \ldots, \xi_n\}$ есть мартингал с $\mathsf{M}X_n=1$ и $X_n \to X_\infty \underset{L^1}{\equiv} 0$ (P-п. н.). В то же время ясно, что $\mathsf{M}|X_n-X_\infty|=1$ и, значит, $X_n \not\to X_\infty$. Таким образом, условие (1) не обеспечивает, вообще говоря, сходимость X_n к X_∞ в смысле L^1.

Приводимая далее теорема 2 показывает, что если предположение (1) усилить до предположения равномерной интегрируемости семейства $\{X_n\}$ (из которой условие (1) следует согласно п. 4 § 6 гл. II), то тогда наряду со сходимостью почти наверное будет выполняться и сходимость в смысле L^1.

Теорема 2. *Пусть $X=(X_n, \mathscr{F}_n)$ — субмартингал, для которого семейство случайных величин $\{X_n\}$ равномерно интегрируемо. Тогда существует такая случайная величина X_∞ с $\mathsf{M}|X_\infty|<\infty$, что при $n\to\infty$*

$$X_n \to X_\infty \quad (\mathsf{P}\text{-п. н}), \tag{4}$$

$$X_n \xrightarrow{L^1} X_\infty. \tag{5}$$

При этом последовательность $\bar{X}=(X_n, \mathscr{F}_n)$, $1\leqslant n\leqslant\infty$, с $\mathscr{F}_\infty = \sigma(\bigcup \mathscr{F}_n)$ также образует субмартингал.

Доказательство. Утверждение (4) следует из теоремы 1, а утверждение (5) — из (4) и теоремы 4 § 6 гл. II.

Далее, если $A\in\mathscr{F}_n$ и $m\geqslant n$, то

$$\mathsf{M}I_A|X_m-X_\infty|\to 0, \qquad m\to\infty,$$

и поэтому

$$\lim_{m\to\infty}\int_A X_m\,d\mathsf{P}=\int_A X_\infty\,d\mathsf{P}.$$

Последовательность $\left\{\int_A X_m\,d\mathsf{P}\right\}_{m\geqslant n}$ является неубывающей и, значит,

$$\int_A X_n\,d\mathsf{P}\leqslant\int_A X_m\,d\mathsf{P}\leqslant\int_A X_\infty\,d\mathsf{P},$$

откуда $X_n\leqslant\mathsf{M}(X_\infty|\mathscr{F}_n)$ (P-п. н.) для всех $n\geqslant 1$.

Теорема доказана.

Следствие. Если $X=(X_n, \mathscr{F}_n)$ — субмартингал и для некоторого $p>1$

$$\sup_n \mathsf{M}|X_n|^p<\infty, \tag{6}$$

то существует интегрируемая случайная величина X_∞, для которой выполнены (4) и (5).

Для доказательства достаточно заметить, что, согласно лемме 3 § 6 гл. II, условие (6) обеспечивает равномерную интегрируемость семейства $\{X_n\}$.

3. Приведем теперь теорему о свойствах непрерывности условных математических ожиданий, которая была одним из самых первых результатов относительно сходимости мартингалов.

Теорема 3 (П. Леви). *Пусть* $(\Omega, \mathscr{F}, \mathsf{P})$ — *вероятностное пространство,* $(\mathscr{F}_n)_{n\geqslant 1}$ — *неубывающее семейство σ-алгебр,* $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots \subseteq \mathscr{F}$. *Пусть* ξ — *некоторая случайная величина с* $\mathsf{M}|\xi| < \infty$ *и* $\mathscr{F}_\infty = \sigma\left(\bigcup_n \mathscr{F}_n\right)$. *Тогда* P-п. н. *и в смысле* L^1

$$\mathsf{M}(\xi \mid \mathscr{F}_n) \to \mathsf{M}(\xi \mid \mathscr{F}_\infty), \qquad n \to \infty. \tag{7}$$

Доказательство. Пусть $X_n = \mathsf{M}(\xi \mid \mathscr{F}_n)$, $n \geqslant 1$. Поскольку

$$\int_{\{|X_i| \geqslant a\}} |X_i|\, d\mathsf{P} \leqslant \int_{\{|X_i| \geqslant a\}} \mathsf{M}(|\xi| \mid \mathscr{F}_i)\, d\mathsf{P} = \int_{\{|X_i| \geqslant a\}} |\xi|\, d\mathsf{P}$$

и

$$\sup_i \mathsf{P}\{|X_i| \geqslant a\} \leqslant \sup_i \frac{\mathsf{M}|X_i|}{a} \leqslant \frac{\mathsf{M}|\xi|}{a} \to 0, \qquad a \to \infty,$$

то (лемма 2 § 6 гл. II) семейство $\{X_n\}$ равномерно интегрируемо и, значит, по теореме 2 существует случайная величина X_∞ такая, что $X_n = \mathsf{M}(\xi \mid \mathscr{F}_n) \to X_\infty$ (P-п. н. и в смысле L^1). Поэтому надо лишь показать, что

$$X_\infty = \mathsf{M}(\xi \mid \mathscr{F}_\infty) \quad (\text{P-п. н.}).$$

Пусть $m \geqslant n$ и $A \in \mathscr{F}_n$. Тогда

$$\int_A X_m\, d\mathsf{P} = \int_A X_n\, d\mathsf{P} = \int_A \mathsf{M}(\xi \mid \mathscr{F}_n)\, d\mathsf{P} = \int_A \xi\, d\mathsf{P}.$$

В силу равномерной интегрируемости семейства $\{X_n\}$ и теоремы 5 § 6 гл. II $\mathsf{M} I_A |X_m - X_\infty| \to 0$, $m \to \infty$, и, следовательно,

$$\int_A X_\infty\, d\mathsf{P} = \int_A \xi\, d\mathsf{P}. \tag{8}$$

Это равенство выполнено для любого $A \in \mathscr{F}_n$ и, значит, для любого $A \in \bigcup_{n=1}^{\infty} \mathscr{F}_n$. Поскольку $\mathsf{M}|X_\infty| < \infty$, $\mathsf{M}|\xi| < \infty$, то левая и правая части в (8) представляют σ-аддитивные меры, возможно, принимающие и отрицательные значения, но конечные и совпадающие на алгебре $\bigcup_{n=1}^{\infty} \mathscr{F}_n$. В силу единственности продолжения σ-аддитивной меры с алгебры на наименьшую σ-алгебру ее содержащую (теорема Каратеодори, § 3 гл. II) равенство (9) остается справедливым и для множеств $A \in \mathscr{F}_\infty = \sigma(\bigcup \mathscr{F}_n)$. Итак,

$$\int_A X_\infty \, dP = \int_A \xi \, dP = \int_A M(\xi \mid \mathscr{F}_\infty) \, dP, \qquad A \in \mathscr{F}_\infty. \tag{9}$$

Величины X_∞ и $M(\xi \mid \mathscr{F}_\infty)$ являются $\mathscr{F}_\infty$-измеримыми, поэтому в силу свойства 1 п. 2 § 6 гл. II из (9) следует, что $X_\infty = M(\xi \mid \mathscr{F}_\infty)$ (P-п. н.).

Теорема доказана.

Следствие. Стохастическая последовательность $X = (X_n, \mathscr{F}_n)$ *является равномерно интегрируемым мартингалом* тогда и только тогда, когда существует случайная величина ξ с $M|\xi| < \infty$ такая, что $X_n = M(\xi \mid \mathscr{F}_n)$ для всех $n \geqslant 1$. При этом $X_n \to M(\xi \mid \mathscr{F}_\infty)$ (P-п. н. и в смысле L^1) при $n \to \infty$.

Действительно, если $X = (X_n, \mathscr{F}_n)$ — равномерно интегрируемый мартингал, то по теореме 2 найдется такая интегрируемая случайная величина X_∞, что $X_n \to X_\infty$ (P-п. н. и в смысле L^1) и к тому же $X_n = M(X_\infty \mid \mathscr{F}_n)$. Так что в качестве случайной величины ξ можно взять ($\mathscr{F}_\infty$-измеримую величину) X_∞.

Обратное утверждение следует из теоремы 3.

4. Остановимся на некоторых применениях доказанных теорем.

Пример 1. *Закон «нуля или единицы»*. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, $\mathscr{F}_n^\xi = \sigma\{\omega\colon \xi_1, \ldots, \xi_n\}$ и $\mathscr{X}$ — σ-алгебра «хвостовых» событий. Из теоремы 3 $M(I_A \mid \mathscr{F}_n^\xi) \to M(I_A \mid \mathscr{F}_\infty^\xi) = I_A$ (P-п. н.). Но I_A и $(\xi_1, \ldots, \xi_n)$ независимы. Поэтому $M(I_A \mid \mathscr{F}_n^\xi) = MI_A$ и, значит, (P-п. н.) $I_A = MI_A$, откуда $P(A) = 0$ или $P(A) = 1$.

Следующие два примера иллюстрируют возможности применения приведенных выше теорем о сходимости в математическом анализе.

Пример 2. Если $f = f(x)$ — функция на $[0, 1)$, удовлетворяющая условию Липшица, то она абсолютно непрерывна и, как известно из анализа, найдется такая интегрируемая (по Лебегу) функция $g = g(x)$, что

$$f(x) - f(0) = \int_0^x g(y) \, dy. \tag{10}$$

(В этом смысле $g(x)$ есть «производная» $f(x)$.)

Покажем как этот результат может быть получен из теоремы 1. Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$ и P — лебеговская мера. Положим

$\mathscr{F}_n = \sigma\{x: \xi_1, \ldots, \xi_n\} = \sigma\{x: \xi_n\}$, и пусть

$$X_n = \frac{f(\xi_n + 2^{-n}) - f(\xi_n)}{2^{-n}}.$$

Поскольку при заданном значении ξ_n случайная величина ξ_{n+1} принимает лишь два значения ξ_n и $\xi_n + 2^{-(n+1)}$ с условными вероятностями, равными 1/2, то

$$\mathsf{M}[X_{n+1} \mid \mathscr{F}_n] = \mathsf{M}[X_{n+1} \mid \xi_n] = 2^{n+1}\mathsf{M}[f(\xi_{n+1} + 2^{-(n+1)}) - f(\xi_{n+1}) \mid \xi_n] = 2^{n+1}\left\{\frac{1}{2}[f(\xi_n + 2^{-(n+1)}) - f(\xi_n)\right\} + \frac{1}{2}[f(\xi_n + 2^{-n}) - f(\xi_n + 2^{-(n+1)})]\Big\} = 2^n\{f(\xi_n + 2^{-n}) - f(\xi_n)\} = X_n.$$

Отсюда следует, что $X = (X_n, \mathscr{F}_n)$ есть мартингал, причем равномерно интегрируемый в силу того, что $|X_n| \leqslant L$, где L — константа в условии Липшица: $|f(x) - f(y)| \leqslant L|x - y|$. Заметим, что $\mathscr{F} = \mathscr{B}([0, 1)) = \sigma(\bigcup \mathscr{F}_n)$. Поэтому, согласно следствию к теореме 3, найдется такая $\mathscr{F}$-измеримая функция $g = g(x)$, что $X_n \to g$ (P-п. н.) и

$$X_n = \mathsf{M}[g \mid \mathscr{F}_n]. \tag{11}$$

Возьмем множество $B = [0, k/2^n]$. Тогда из (11)

$$f\left(\frac{k}{2^n}\right) - f(0) = \int_0^{k/2^n} X_n\,dx = \int_0^{k/2^n} g(x)\,dx$$

и в силу произвольности n и k отсюда получаем требуемое равенство (10).

Пример 3. Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$ и P — мера Лебега. Рассмотрим систему функций Хаара $\{H_n(x)\}_{n \geqslant 1}$, определенных в примере 3 § 11 гл. II. Положим $\mathscr{F}_n = \sigma\{x: H_1, \ldots, H_n\}$ и заметим, что $\sigma(\bigcup \mathscr{F}_n) = \mathscr{F}$. Из свойств условных математических ожиданий и структуры функций Хаара нетрудно вывести, что для любой борелевской функции $f \in L$

$$\mathsf{M}[f(x) \mid \mathscr{F}_n] = \sum_{k=1}^{n} a_k H_k(x) \quad (\mathsf{P}\text{-п. н.}), \tag{12}$$

где

$$a_k = (f, H_k) = \int_0^1 f(x) H_k(x)\,dx.$$

Иначе говоря, условное математическое ожидание $\mathsf{M}[f(x) \mid \mathscr{F}_n]$ есть частичная сумма Фурье при разложении функции $f(x)$ по системе Хаара. Тогда, применяя теорему 3 к мартингалу

$(\mathsf{M}(f \mid \mathscr{F}_n), \mathscr{F}_n)$, находим, что при $n \to \infty$

$$\sum_{k=1}^{n} (f, H_k) H_k(x) \to f(x) \quad (\mathsf{P}\text{-п. н.})$$

и

$$\int_0^1 \left| \sum_{k=1}^{n} (f, H_k) H_k(x) - f(x) \right| dx \to 0.$$

5. Задачи.

1. Пусть $\{\mathscr{G}_n\}$ — невозрастающее семейство σ-алгебр $\mathscr{G}_1 \supseteq \supseteq \mathscr{G}_2 \supseteq \ldots$, $\mathscr{G}_\infty = \bigcap \mathscr{G}_n$ и η — некоторая интегрируемая случайная величина. Доказать справедливость следующего аналога теоремы 3: при $n \to \infty$

$$\mathsf{M}(\eta \mid \mathscr{G}_n) \to \mathsf{M}(\eta \mid \mathscr{G}_\infty) \quad (\mathsf{P}\text{-п. н. и в смысле } L^1).$$

2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{M}|\xi_1| < \infty$ и $\mathsf{M}\xi_1 = m$, $S_n = \xi_1 + \ldots + \xi_n$. Показав (см. задачу 2 § 7 гл. II), что

$$\mathsf{M}(\xi_1 \mid S_n, S_{n+1}, \ldots) = \mathsf{M}(\xi_1 \mid S_n) = \frac{S_n}{n} \quad (\mathsf{P}\text{-п. н.}),$$

вывести из результата задачи 1 усиленный закон больших чисел: при $n \to \infty$

$$\frac{S_n}{n} \to m \quad (\mathsf{P}\text{-п. н. и в смысле } L^1).$$

3. Доказать справедливость следующего результата, соединяющего в себе теорему Лебега о мажорируемой сходимости и теорему П. Леви. Пусть $\{\xi_n\}_{n \geq 1}$ — последовательность случайных величин таких, что $\xi_n \to \xi$ (P-п. н.), $|\xi_n| \leq \eta$, $\mathsf{M}\eta < \infty$ и $(\mathscr{F}_m)_{m \geq 1}$ — неубывающее семейство σ-алгебр, $\mathscr{F}_\infty = \sigma(\bigcup \mathscr{F}_n)$. Тогда ($\mathsf{P}$-п. н.)

$$\lim_{\substack{m \to \infty \\ n \to \infty}} \mathsf{M}(\xi_n \mid \mathscr{F}_m) = \mathsf{M}(\xi \mid \mathscr{F}_\infty).$$

4. Доказать справедливость формулы (12).

5. Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$, P — мера Лебега и $f = f(x) \in L^1$. Положим

$$f_n(x) = 2^n \int_{k2^{-n}}^{(k+1)2^{-n}} f(y)\, dy, \quad k2^{-n} \leq x < (k+1)\, 2^{-n}.$$

Показать, что $f_n(x) \to f(x)$ (P-п. н.).

6. Пусть $\Omega = [0, 1)$, $\mathscr{F} = \mathscr{B}([0, 1))$, P — мера Лебега и $f = f(x) \in L^1$. Продолжим эту функцию периодически на $[0, 2)$ и

положим

$$f_n(x)=\sum_{i=1}^{2^n} 2^{-n} f(x+i2^{-n}).$$

Показать, что $f_n(x)\to f(x)$ (Р-п. н.).

7. Доказать, что теорема 1 сохраняет свою силу для обобщенных субмартингалов.

§ 5. О множествах сходимости субмартингалов и мартингалов

1. Пусть $X=(X_n, \mathscr{F}_n)$ — стохастическая последовательность. Будем обозначать через $\{X_n\to\}$, или $\{-\infty<\lim X_n<\infty\}$, множество тех элементарных исходов, для которых $\lim X_n$ *существует* и *конечен*. Будем говорить также, что $A\subseteq B$ (Р-п. н.), если $\mathsf{P}(I_A\leqslant I_B)=1$.

Если X — субмартингал и $\sup \mathsf{M}|X_n|<\infty$ (или, что эквивалентно, $\sup \mathsf{M}X_n^+<\infty$), то в соответствии с теоремой 1 § 4 (Р-п. н.)

$$\{X_n\to\}=\Omega.$$

Рассмотрим вопрос о структуре множеств сходимости $\{X_n\to\}$ для субмартингалов в случае нарушения условия $\sup \mathsf{M}|X_n|<\infty$.

Пусть $a>0$ и $\tau_a=\inf\{n\geqslant 1: X_n>a\}$ с $\tau_a=\infty$, если $\{\cdot\}=\varnothing$.

Определение. Стохастическая последовательность $X=(X_n, \mathscr{F}_n)$ принадлежит *классу* $\boldsymbol{C}^+$ $(X\in\boldsymbol{C}^+)$, если для любого $a>0$

$$\mathsf{M}(\Delta X_{\tau_a})^+ I\{\tau_a<\infty\}<\infty, \tag{1}$$

где $\Delta X_n=X_n-X_{n-1}$, $X_0=0$.

Очевидно, что $X\in\boldsymbol{C}^+$, если

$$\mathsf{M}\sup_n|\Delta X_n|<\infty \tag{2}$$

или, тем более, если (Р-п. н.) для всех $n\geqslant 1$

$$|\Delta X_n|\leqslant C<\infty. \tag{3}$$

Теорема 1. *Если субмартингал* $X\in\boldsymbol{C}^+$, *то* (Р-п. н.)

$$\{\sup X_n<\infty\}=\{X_n\to\}. \tag{4}$$

Доказательство. Включение $\{X_n\to\}\subseteq\{\sup X_n<\infty\}$ очевидно. Для доказательства обратного включения рассмотрим «остановленный» субмартингал $X^{\tau_a}=(X_{\tau_a\wedge n}, \mathscr{F}_n)$. Тогда в силу (1)

$$\sup_n \mathsf{M}X_{\tau_a\wedge n}^+\leqslant a+\mathsf{M}[X_{\tau_a}^+\cdot I\{\tau_a<\infty\}]\leqslant$$
$$\leqslant 2a+\mathsf{M}[(\Delta X_{\tau_a})^+\cdot I\{\tau_a<\infty\}]<\infty, \tag{5}$$

и, значит, по теореме 1 из § 4 (Р-п. н.)

$$\{\tau_a=\infty\}\subseteq\{X_n\to\}.$$

Но $\bigcup\limits_{a>0}\{\tau_a=\infty\}=\{\sup X_n<\infty\}$, поэтому (Р-п. н.) $\{\sup X_n<\infty\}\subseteq$ $\subseteq\{X_n\to\}$.

Теорема доказана.

Следствие. Пусть X — мартингал с $\mathsf{M}\sup|\Delta X_n|<\infty$. Тогда (Р-п. н.)

$$\{X_n\to\}\cup\{\varliminf X_n=-\infty,\ \varlimsup X_n=+\infty\}=\Omega. \qquad (6)$$

В самом деле, применяя теорему 1 к X и $-X$, находим, что (Р-п. н.)

$$\{\varlimsup X_n<\infty\}=\{\sup X_n<\infty\}=\{X_n\to\},$$
$$\{\varliminf X_n>-\infty\}=\{\inf X_n>-\infty\}=\{X_n\to\}.$$

Поэтому (Р-п. н.)

$$\{\varlimsup X_n<\infty,\ \varliminf X_n<\infty\}=\{X_n\to\},$$

что и доказывает (6).

Утверждение (6) означает, что почти все траектории мартингала X, удовлетворяющего условию $\mathsf{M}\sup|\Delta X_n|<\infty$, таковы, что или для них существует конечный предел, или же они устроены «плохо» в том смысле, что для них $\varlimsup X_n=+\infty$, а $\varliminf X_n=-\infty$.

2. Если $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с $\mathsf{M}\xi_i=0$ и $|\xi_i|\leqslant c<\infty$, то, согласно теореме 1 § 2 гл. IV, ряд $\sum\xi_i$ сходится (Р-п. н.) тогда и только тогда, когда $\sum\mathsf{M}\xi_i^2<\infty$. Последовательность $X=(X_n, \mathscr{F}_n)$ с $X_n=\xi_1+\ldots+\xi_n$, $\mathscr{F}_n=\sigma\{\omega\colon \xi_1, \ldots, \xi_n\}_n$ есть квадратично интегрируемый мартингал с $\langle X\rangle_n=\sum\limits_{i=1}^{n}\mathsf{M}\xi_i^2$, и сформулированному утверждению можно придать такую форму: (Р-п. н.)

$$\{\langle X\rangle_\infty<\infty\}=\{X_n\to\}=\Omega,$$

где $\langle X\rangle_\infty=\lim\limits_n\langle X\rangle_n$.

Приводимые далее утверждения обобщают этот результат на случай более общих мартингалов и субмартингалов.

Теорема 2. *Пусть* $X=(X_n, \mathscr{F}_n)$ — *субмартингал и*

$$X_n=m_n+A_n$$

— *его разложение Дуба.*

а) *Если* X — *неотрицательный субмартингал, то* (Р-п. н.)

$$\{A_\infty<\infty\}\subseteq\{X_n\to\}\subseteq\{\sup X_n<\infty\}. \qquad (7)$$

b) *Если* $X \in C^+$, *то* (P-п. н.)

$$\{X_n \to\} = \{\sup X_n < \infty\} \subseteq \{A_\infty < \infty\}. \tag{8}$$

c) *Если* X — *неотрицательный субмартингал и* $X \in C^+$, *то* (P-п. н.)

$$\{X_n \to\} = \{\sup X_n < \infty\} = \{A_\infty < \infty\}. \tag{9}$$

Доказательство. a) Второе включение в (7) очевидно. Для доказательства первого включения введем моменты

$$\sigma_a = \inf\{n \geqslant 1:\ A_{n+1} > a\}, \qquad a > 0,$$

полагая $\sigma_a = +\infty$, если $\{\cdot\} = \varnothing$. Тогда $A_{\sigma_a} \leqslant a$ и в силу следствия 1 к теореме 1 § 2

$$\mathsf{M}X_{n \wedge \sigma_a} = \mathsf{M}A_{n \wedge \sigma_a} \leqslant a.$$

Пусть $Y_n^a = X_{n \wedge \sigma_a}$, тогда $Y^a = (Y_n^a, \mathscr{F}_n)$ — субмартингал с $\sup \mathsf{M}Y_n^a \leqslant a < \infty$ и в силу его неотрицательности из теоремы 1 § 4 следует, что (P-п. н.)

$$\{A_\infty \leqslant a\} = \{\sigma_a = \infty\} \subseteq \{X_n \to\}.$$

Поэтому (P-п. н.)

$$\{A_\infty < \infty\} = \bigcup_{a > 0} \{A_\infty \leqslant a\} \subseteq \{X_n \to\}.$$

b) Первое равенство следует из теоремы 1. Чтобы доказать второе, заметим, что, согласно (5),

$$\mathsf{M}A_{\tau_a \wedge n} = \mathsf{M}X_{\tau_a \wedge n} \leqslant \mathsf{M}X^+_{\tau_a \wedge n} \leqslant 2a + \mathsf{M}\left[(\Delta X_{\tau_a})^+ I\{\tau_a < \infty\}\right]$$

и, значит,

$$\mathsf{M}A_{\tau_a} = \mathsf{M}\lim_n A_{\tau_a \wedge n} < \infty.$$

Поэтому $\{\tau_a = \infty\} \subseteq \{A_\infty < \infty\}$ и требуемое утверждение следует из того, что $\bigcup\limits_{a > 0} \{\tau_a = \infty\} = \{\sup X_n < \infty\}$.

c) Это утверждение есть непосредственное следствие утверждений a) и b).

Теорема доказана.

Замечание. Условие неотрицательности X можно заменить условием $\sup\limits_n \mathsf{M}X_n^- < \infty$.

Следствие 1. Пусть $X_n = \xi_1 + \ldots + \xi_n$, где $\xi_i \geqslant 0$, $\mathsf{M}\xi_i < \infty$, ξ_i — $\mathscr{F}_i$-измеримы и $\mathscr{F}_0 = \{\varnothing, \Omega\}$. Тогда (P-п. н.)

$$\left\{\sum_{n=1}^{\infty} \mathsf{M}(\xi_n \mid \mathscr{F}_{n-1}) < \infty\right\} \subseteq \{X_n \to\}, \tag{10}$$

и если к тому же $\mathsf{M}\sup_n \xi_n < \infty$, то ($\mathsf{P}$-п. н.)

$$\left\{\sum_{n=1}^{\infty} \mathsf{M}(\xi_n \mid \mathscr{F}_{n-1}) < \infty\right\} = \{X_n \to\}. \tag{11}$$

Следствие 2 (лемма Бореля — Кантелли — Леви). Если события $B_n \in \mathscr{F}_n$, то, полагая, в (11) $\xi_n = I_{B_n}$, получаем, что (P-п. н.)

$$\left\{\sum_{n=1}^{\infty} \mathsf{P}(B_n \mid \mathscr{F}_{n-1}) < \infty\right\} = \left\{\sum_{n=1}^{\infty} I_{B_n} < \infty\right\}. \tag{12}$$

3. Теорема 3. *Пусть* $M = (M_n, \mathscr{F}_n)_{n \geqslant 1}$ — *квадратично интегрируемый мартингал. Тогда* (P-п. н.)

$$\{\langle M\rangle_\infty < \infty\} \subseteq \{M_n \to\}. \tag{13}$$

Если к тому же $\mathsf{M}\sup|\Delta M_n|^2 < \infty$, *то* ($\mathsf{P}$-п. н.)

$$\{\langle M\rangle_\infty < \infty\} = \{M_n \to\}, \tag{14}$$

где

$$\langle M\rangle_\infty = \sum_{n=1}^{\infty} \mathsf{M}((\Delta M_n)^2 \mid \mathscr{F}_{n-1}) \tag{15}$$

с $M_0 = 0$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

Доказательство. Рассмотрим два субмартингала $M^2 = (M_n^2, \mathscr{F}_n)$ и $(M+1)^2 = ((M_n+1)^2, \mathscr{F}_n)$. Тогда в их разложениях Дуба

$$M_n^2 = m_n' + A_n', \quad (M_n+1)^2 = m_n'' + A_n''$$

величины A_n' и A_n'' совпадают, поскольку

$$A_n' = \sum_{k=1}^{n} \mathsf{M}(\Delta M_k^2 \mid \mathscr{F}_{k-1}) = \sum_{k=1}^{n} \mathsf{M}((\Delta M_k)^2 \mid \mathscr{F}_{k-1})$$

и

$$A_n'' = \sum_{k=1}^{n} \mathsf{M}(\Delta(M_k+1)^2 \mid \mathscr{F}_{k-1}) = \sum_{k=1}^{n} \mathsf{M}(\Delta M_k^2 \mid \mathscr{F}_{k-1}) = $$
$$= \sum_{k=1}^{n} \mathsf{M}((\Delta M_k)^2 \mid \mathscr{F}_{k-1}).$$

Поэтому из (7) (P-п. н.)

$$\{\langle M\rangle_\infty < \infty\} = \{A_\infty' < \infty\} \subseteq \{M_n^2 \to\} \cap \{(M_n+1)^2 \to\} = \{M_n \to\}.$$

В силу (9) для доказательства (15) достаточно проверить, что условие $\mathsf{M}\sup|\Delta M_n|^2 < \infty$ обеспечивает принадлежность субмартингала M^2 классу $\boldsymbol{C}^+$.

Пусть $\tau_a = \inf\{n \geqslant 1: M_n^2 > 0\}$, $a > 0$. Тогда на множестве $\{\tau_a < \infty\}$

$$|\Delta M_{\tau_a}^2| = |M_{\tau_a}^2 - M_{\tau_a-1}^2| \leqslant |M_{\tau_a} - M_{\tau_a-1}|^2 + \\ + 2|M_{\tau_a-1}| \cdot |M_{\tau_a} - M_{\tau_a-1}| \leqslant (\Delta M_{\tau_a})^2 + 2a^{1/2}|\Delta M_{\tau_a}|,$$

откуда

$$\mathsf{M}|\Delta M_{\tau_a}^2| I\{\tau_a < \infty\} \leqslant \\ \leqslant \mathsf{M}(\Delta M_{\tau_a})^2 I\{\tau_a < \infty\} + 2a^{1/2}\sqrt{\mathsf{M}(\Delta M_{\tau_a})^2 I\{\tau_a < \infty\}} \leqslant \\ \leqslant \mathsf{M}\sup|\Delta M_n|^2 + 2a^{1/2}\sqrt{\mathsf{M}\sup|\Delta M_n|^2} < \infty.$$

Теорема доказана.

В качестве иллюстрации этой теоремы приведем следующий результат, который можно рассматривать как своеобразную форму усиленного закона больших чисел для квадратично интегрируемых мартингалов (ср. с теоремой 2 в § 3 гл. IV и со следствием 2 в п. 3 § 3).

Теорема 4. *Пусть $M = (M_n, \mathscr{F}_n)$ — квадратично интегрируемый мартингал и $A = (A_n, \mathscr{F}_{n-1})$ — предсказуемая возрастающая последовательность с $A_1 \geqslant 1$, $A_\infty = \infty$* (P-п. н.).

Если (P-п. н.)

$$\sum_{i=1}^{\infty} \frac{\mathsf{M}[(\Delta M_i)^2 | \mathscr{F}_{i-1}]}{A_i^2} < \infty, \tag{16}$$

то с вероятностью единица

$$\frac{M_n}{A_n} \to 0, \quad n \to \infty. \tag{17}$$

Доказательство. Рассмотрим квадратично интегрируемый мартингал $m = (m_n, \mathscr{F}_n)$ с

$$m_n = \sum_{i=1}^{n} \frac{\Delta M_i}{A_i}.$$

Тогда

$$\langle m \rangle_n = \sum_{i=1}^{n} \frac{\mathsf{M}[(\Delta M_i)^2 | \mathscr{F}_{i-1}]}{A_i^2}. \tag{18}$$

Поскольку

$$\frac{M_n}{A_n} = \frac{\sum_{k=1}^{n} A_k \Delta m_k}{A_n}, \tag{19}$$

то, согласно лемме Кронекера (§ 3 гл. IV), $\frac{M_n}{A_n} \to 0$ (Р-п. н.), если с вероятностью единица существует конечный предел $\lim m_n$. В силу (13)

$$\{\langle m\rangle_\infty < \infty\} \subseteq \{m_n \to\}, \tag{20}$$

поэтому из (18) следует, что условие (16) достаточно для выполнения (17).

Теорема доказана.

Пример. Рассмотрим последовательность независимых случайных величин $\xi_1, \xi_2, \ldots$ с $\mathsf{M}\xi_i = 0$, $\mathsf{D}\xi_i = D_i > 0$, и пусть последовательность $X = \{X_n\}_{n \geq 0}$ определяется из рекуррентных уравнений

$$X_{n+1} = \theta X_n + \xi_{n+1}, \tag{21}$$

где X_0 не зависит от $\xi_1, \xi_2, \ldots$, а θ — неизвестный параметр, $-\infty < \theta < \infty$.

Будем интерпретировать X_n как результат наблюдения в момент времени n и поставим задачу оценки неизвестного параметра θ. Возьмем в качестве оценки θ по результатам $X_0, X_1, \ldots, X_n$ величину

$$\hat{\theta}_n = \frac{\sum_{k=0}^{n-1} \frac{X_k X_{k+1}}{D_{k+1}}}{\sum_{k=0}^{n-1} \frac{X_k^2}{D_{k+1}}}, \tag{22}$$

полагая ее равной нулю, если знаменатель обращается в нуль. (Величина $\hat{\theta}_n$ есть оценка, полученная по *методу наименьших квадратов*.)

Из (21) и (22) ясно, что

$$\hat{\theta}_n = \theta + \frac{M_n}{A_n},$$

где

$$M_n = \sum_{k=0}^{n-1} \frac{X_k \xi_{k+1}}{D_{k+1}}, \quad A_n = \langle M\rangle_n = \sum_{k=0}^{n-1} \frac{X_k^2}{D_{k+1}}.$$

Поэтому, если истинное значение неизвестного параметра есть θ, то

$$\mathsf{P}(\hat{\theta}_n \to \theta) = 1, \tag{23}$$

когда (Р-п. н.)

$$\frac{M_n}{A_n} \to 0, \quad n \to \infty. \tag{24}$$

Покажем, что условия

$$\sup_n \frac{D_{n+1}}{D_n} < \infty, \quad \sum_{n=1}^{\infty} \mathsf{M}\left(\frac{\xi_n^2}{D_n} \wedge 1\right) = \infty \tag{25}$$

достаточны для (24) и, следовательно, достаточны для (23). Имеем

$$\sum_{n=1}^{\infty} \left(\frac{\xi_n^2}{D_n} \wedge 1\right) \leqslant \sum_{n=1}^{\infty} \frac{\xi_n^2}{D_n} = \sum_{n=1}^{\infty} \frac{(X_n - \theta X_{n-1})^2}{D_n} \leqslant$$

$$\leqslant 2\left[\sum_{n=1}^{\infty} \frac{X_n^2}{D_n} + \theta^2 \sum_{n=1}^{\infty} \frac{X_{n-1}^2}{D_n}\right] \leqslant 2\left[\sup \frac{D_{n+1}}{D_n} + \theta^2\right] \langle M \rangle_{\infty}.$$

Тем самым

$$\left\{\sum_{n=1}^{\infty} \left(\frac{\xi_n^2}{D_n} \wedge 1\right) = \infty\right\} \subseteq \{\langle M \rangle_{\infty} = \infty\}.$$

По теореме о трех рядах (теорема 3 в § 2 гл. IV) расходимость ряда $\sum_{n=1}^{\infty} \mathsf{M}\left(\frac{\xi_n}{D_n} \wedge 1\right)$ обеспечивает расходимость (Р-п. н.) ряда $\sum_{n=1}^{\infty} \left(\frac{\xi_n^2}{D_n} \wedge 1\right)$. Поэтому $\mathsf{P}\{\langle M \rangle_{\infty} = \infty\} = 1$. Далее, если

$$m_n = \sum_{i=1}^{n} \frac{\Delta M_i}{\langle M \rangle_i},$$

то

$$\langle m \rangle_n = \sum_{i=1}^{n} \frac{\Delta \langle M \rangle_i}{\langle M \rangle_i^2}$$

и, поскольку $\mathsf{P}(\langle M \rangle_{\infty} = \infty) = 1$, то $\mathsf{P}(\langle m \rangle_{\infty} < \infty) = 1$. Поэтому требуемсе соотношение (24) следует напосредственно из теоремы 4.

Теорема 5. *Пусть $X = (X_n, \mathscr{F}_n)$ — субмартингал,*

$$X_n = m_n + A_n$$

— его разложение Дуба. Если $|\Delta X_n| \leqslant C$, то (Р-п. н.)

$$\{\langle m \rangle_{\infty} + A_{\infty} < \infty\} = \{X_n \to\}, \tag{26}$$

или, что то же,

$$\left\{\sum_{n=1}^{\infty} \mathsf{M}[\Delta X_n + (\Delta X_n)^2 \mid \mathscr{F}_{n-1}] < \infty\right\} = \{X_n \to\}. \tag{27}$$

Доказательство. Поскольку

$$A_n = \sum_{k=1}^{n} \mathsf{M}(\Delta X_k \mid \mathscr{F}_{k-1}), \tag{28}$$

$$m_n = \sum_{k=1}^{n} [\Delta X_k - \mathsf{M}(\Delta X_k \mid \mathscr{F}_{k-1})], \tag{29}$$

то в силу предположения $|\Delta X_k| \leqslant C$ мартингал $m = (m_n, \mathscr{F}_n)$ является квадратично интегрируемым с $|\Delta m_n| \leqslant 2C$. Тогда из (13)

$$\{\langle m\rangle_\infty + A_\infty < \infty\} \subseteq \{X_n \to\} \tag{30}$$

и согласно (8)

$$\{X_n \to\} \subseteq \{A_\infty < \infty\}.$$

Поэтому из (14) и (20)

$$\begin{aligned}\{X_n \to\} &= \{X_n \to\} \cap \{A_\infty < \infty\} = \{X_n \to\} \cap \{A_\infty < \infty\} \cap \{m_n \to\} = \\ &\doteq \{X_n \to\} \cap \{A_\infty < \infty\} \cap \{\langle m\rangle_\infty < \infty\} \doteq \\ &= \{X_n \to\} \cap \{A_\infty + \langle m\rangle_\infty < \infty\} = \{A_\infty + \langle m\rangle_\infty < \infty\}.\end{aligned}$$

Наконец, эквивалентность утверждений (26) и (27) следует из того, что в силу (29)

$$\langle m\rangle_n = \sum \{\mathsf{M}[(\Delta X_k)^2 \mid \mathscr{F}_{k-1}] - [\mathsf{M}(\Delta X_k \mid \mathscr{F}_{k-1})]^2\},$$

и из сходимости ряда $\sum\limits_{k=1}^{\infty} \mathsf{M}(\Delta X_k \mid \mathscr{F}_{k-1})$, состоящего из неотрицательных членов, следует сходимость ряда $\sum\limits_{k=1}^{\infty} [\mathsf{M}(\Delta X_k \mid \mathscr{F}_{k-1})]^2$.

Теорема доказана.

4. Задачи.

1. Показать, что если субмартингал $X = (X_n, \mathscr{F}_n)$ удовлетворяет условию $\sup\limits_n \mathsf{M}|X_n| < \infty$, то он принадлежит классу C^+.

2. Доказать, что теоремы 1 и 2 остаются справедливыми для обобщенных субмартингалов.

3. Показать, что для обобщенных субмартингалов (Р-п. н.) имеет место включение

$$\{\inf_m \sup_{n \geqslant m} \mathsf{M}(X_n^+ \mid \mathscr{F}_m) < \infty\} \subseteq \{X_n \to\}.$$

4. Показать, что следствие к теореме 1 остается верным и для обобщенных мартингалов.

5. Показать, что всякий обобщенный субмартингал класса C^+ является локальным субмартингалом.

§ 6. Абсолютная непрерывность и сингулярность вероятностных распределений

1. Пусть $(\Omega, \mathscr{F})$ — некоторое измеримое пространство с выделенным на нем семейством σ-алгебр $(\mathscr{F}_n)_{n\geqslant 1}$ таких, что $\mathscr{F}_1 \subseteq \subseteq \mathscr{F}_2 \subseteq \ldots \subseteq \mathscr{F}$ и

$$\mathscr{F} = \sigma\left(\bigcup_{n=1}^{\infty} \mathscr{F}_n\right). \tag{1}$$

Будем предполагать, что на $(\Omega, \mathscr{F})$ заданы две вероятностные меры P и $\tilde{\mathsf{P}}$. Обозначим

$$\mathsf{P}_n = \mathsf{P} \mid \mathscr{F}_n, \quad \tilde{\mathsf{P}}_n = \tilde{\mathsf{P}} \mid \mathscr{F}_n$$

— сужения этих мер на $\mathscr{F}_n$, т. е. пусть P_n и $\tilde{\mathsf{P}}_n$ — меры на $(\Omega, \mathscr{F}_n)$, причем для $B \in \mathscr{F}_n$

$$\mathsf{P}_n(B) = \mathsf{P}(B), \quad \tilde{\mathsf{P}}_n(B) = \tilde{\mathsf{P}}(B).$$

Определение 1. Вероятностная мера $\tilde{\mathsf{P}}$ называется *абсолютно непрерывной* относительно P (обозначение: $\tilde{\mathsf{P}} \ll \mathsf{P}$), если $\tilde{\mathsf{P}}(A) = 0$ всякий раз, когда $\mathsf{P}(A) = 0$, $A \in \mathscr{F}$.

В случае $\tilde{\mathsf{P}} \ll \mathsf{P}$ и $\mathsf{P} \ll \tilde{\mathsf{P}}$ меры $\tilde{\mathsf{P}}$ и P называются *эквивалентными* (обозначение: $\tilde{\mathsf{P}} \sim \mathsf{P}$).

Меры $\tilde{\mathsf{P}}$ и P называются *сингулярными* или *ортогональными*, если существует такое множество $A \in \mathscr{F}$, что $\tilde{\mathsf{P}}(A) = 1$ и $\mathsf{P}(\bar{A}) = 1$ (обозначение: $\tilde{\mathsf{P}} \perp \mathsf{P}$).

Определение 2. Будем говорить, что мера $\tilde{\mathsf{P}}$ *локально абсолютно непрерывна* относительно меры P $\left(\text{обозначение: } \tilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}\right)$, если для любого $n \geqslant 1$

$$\tilde{\mathsf{P}}_n \ll \mathsf{P}_n. \tag{2}$$

Основные вопросы, рассматриваемые в настоящем параграфе, состоят в выяснении условий, при которых из локальной абсолютной непрерывности $\tilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$ следует выполнение свойств $\tilde{\mathsf{P}} \ll \mathsf{P}$, $\tilde{\mathsf{P}} \sim \mathsf{P}$, $\tilde{\mathsf{P}} \perp \mathsf{P}$. Как станет ясно из дальнейшего, теория мартингалов является тем математическим аппаратом, который позволяет исчерпывающим образом ответить на эти вопросы.

Итак, будем предполагать, что $\tilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$. Обозначим

$$z_n = \frac{d\tilde{\mathsf{P}}_n}{d\mathsf{P}_n}$$

производную Радиона — Никодима меры $\tilde{P}_n$ относительно P_n. Ясно, что z_n являются $\mathscr{F}_n$-измеримыми, и если $A \in \mathscr{F}_n$, то

$$\int_A z_{n+1} dP = \int_A \frac{d\tilde{P}_{n+1}}{dP_{n+1}} dP = \tilde{P}_{n+1}(A) = \tilde{P}_n(A) = \\ = \int_A \frac{d\tilde{P}_n}{dP_n} dP = \int_A z_n dP.$$

Отсюда следует, что *относительно меры* P *стохастическая последовательность* $Z = (z_n, \mathscr{F}_n)_{n \geqslant 1}$ *является мартингалом.*

Обозначим

$$z_\infty = \overline{\lim}\, z_n.$$

Поскольку $Mz_n = 1$, то из теоремы 1 § 4 следует, что P-п. н. существует $\lim z_n$ и, значит, $P(z_\infty = \lim z_n) = 1$. (В ходе доказательства теоремы 1 будет установлено, что предел $\lim z_n$ существует и по мере $\tilde{P}$, так что $\tilde{P}(z_\infty = \lim z_n) = 1$.)

Ключевым моментом во всей проблематике «абсолютная непрерывность и сингулярность» является

Теорема 1 (разложение Лебега). *Пусть* $\tilde{P} \overset{\text{loc}}{\ll} P$. *Тогда для всякого* $A \in \mathscr{F}$

$$\tilde{P}(A) = \int_A z_\infty dP + \tilde{P}\{A \cap (z_\infty = \infty)\}, \tag{3}$$

причем меры $\mu(A) = \tilde{P}\{A \cap (z_\infty = \infty)\}$ *и* $P(A)$, $A \in \mathscr{F}$, *сингулярны.*

Доказательство. Прежде всего отметим, что классическое *разложение Лебега* устанавливает, что если P и $\tilde{P}$ — две меры, то найдутся и притом единственные меры λ и μ такие, что $\tilde{P} = \lambda + \mu$, где $\lambda \ll P$ и $\mu \perp P$. Доказываемое утверждение (3) можно рассматривать как конкретизацию этого разложения, связанную с предположением о том, что $\tilde{P}_n \ll P_n$, $n \geqslant 1$.

Введем в рассмотрение вероятностные меры

$$Q = \frac{1}{2}(P + \tilde{P}), \quad Q_n = \frac{1}{2}(P_n + \tilde{P}_n), \quad n \geqslant 1,$$

и обозначим

$$\tilde{\mathfrak{z}} = \frac{d\tilde{P}}{dQ}, \quad \mathfrak{z} = \frac{dP}{dQ}, \quad \tilde{\mathfrak{z}}_n = \frac{d\tilde{P}_n}{dQ_n}, \quad \mathfrak{z}_n = \frac{dP_n}{dQ_n}.$$

Поскольку $\tilde{P}(\tilde{\mathfrak{z}} = 0) = P(\mathfrak{z} = 0) = 0$, то $Q(\tilde{\mathfrak{z}} = 0, \mathfrak{z} = 0) = 0$ и, значит, на множестве $\Omega \setminus \{\tilde{\mathfrak{z}} = 0, \mathfrak{z} = 0\}$ корректно определена величина $\tilde{\mathfrak{z}}\cdot\mathfrak{z}^{-1}$. На множестве $\{\tilde{\mathfrak{z}}=0,\ \mathfrak{z}=0\}$ будем полагать ее равной нулю.

Так как $\tilde{\mathsf{P}}_n \ll \mathsf{P}_n \ll \mathsf{Q}_n$, то (см. (II.7.36))

$$\frac{d\tilde{\mathsf{P}}_n}{d\mathsf{Q}_n}=\frac{d\tilde{\mathsf{P}}_n}{d\mathsf{P}_n}\cdot\frac{d\mathsf{P}_n}{d\mathsf{Q}_n} \quad (\mathsf{Q}\text{-п. н.}), \tag{4}$$

т. е.

$$\tilde{\mathfrak{z}}_n = z_n\mathfrak{z}_n \quad (\mathsf{Q}\text{-п. н.}), \tag{5}$$

откуда

$$z_n=\tilde{\mathfrak{z}}_n\cdot\mathfrak{z}_n^{-1} \quad (\mathsf{Q}\text{-п. н.}),$$

где, как и раньше, на множестве $\{\tilde{\mathfrak{z}}_n=0,\ \mathfrak{z}_n=0\}$, имеющем Q-меру нуль, полагаем $\tilde{\mathfrak{z}}_n\cdot\mathfrak{z}_n^{-1}=0$.

Каждая из последовательностей $(\tilde{\mathfrak{z}}_n, \mathscr{F}_n)$ и $(\mathfrak{z}_n, \mathscr{F}_n)$ образует (относительно меры Q) равномерно интегрируемый мартингал и, следовательно, существуют пределы $\lim\tilde{\mathfrak{z}}_n$ и $\lim\mathfrak{z}_n$. При этом (Q-п. н.)

$$\lim\tilde{\mathfrak{z}}_n=\tilde{\mathfrak{z}}, \quad \lim\mathfrak{z}_n=\mathfrak{z}. \tag{6}$$

Отсюда и из соотношений $z_n=\tilde{\mathfrak{z}}_n\mathfrak{z}_n^{-1}$ (Q-п. н.) и $\mathsf{Q}(\tilde{\mathfrak{z}}=0,\ \mathfrak{z}=0)=0$ вытекает, что Q-п. н. существует $\lim z_n=z_\infty$, равный $\tilde{\mathfrak{z}}\cdot\mathfrak{z}^{-1}$.

Ясно, что $\mathsf{P}\ll\mathsf{Q}$, $\tilde{\mathsf{P}}\ll\mathsf{Q}$. Тем самым $\lim z_n$ существует как по мере P, так и по мере $\tilde{\mathsf{P}}$.

Пусть теперь

$$\lambda(A)=\int_A z_\infty d\mathsf{P}, \qquad \mu(A)=\tilde{\mathsf{P}}\{A\cap(z_\infty=\infty)\}.$$

Для доказательства (3) надо показать, что

$$\tilde{\mathsf{P}}(A)=\lambda(A)+\mu(A),\ \lambda\ll\mathsf{P}, \qquad \mu\perp\mathsf{P}.$$

Имеем

$$\begin{aligned}\tilde{\mathsf{P}}(A)&=\int_A\tilde{\mathfrak{z}}\,d\mathsf{Q}=\int_A\tilde{\mathfrak{z}}\overset{\oplus}{\mathfrak{z}}\mathfrak{z}\,d\mathsf{Q}+\int_A\tilde{\mathfrak{z}}\Big[1-\overset{\oplus}{\mathfrak{z}}\mathfrak{z}\Big]d\mathsf{Q}=\\&=\int_A\tilde{\mathfrak{z}}\overset{\oplus}{\mathfrak{z}}\,d\mathsf{P}+\int_A\Big[1-\overset{\oplus}{\mathfrak{z}}\mathfrak{z}\Big]d\tilde{\mathsf{P}}=\int_A z_\infty\,d\mathsf{P}+\tilde{\mathsf{P}}\{A\cap(\mathfrak{z}=0)\},\end{aligned} \tag{7}$$

где последнее равенство следует из того, что

$$\mathsf{P}\left\{\overset{\oplus}{\mathfrak{z}}=\mathfrak{z}^{-1}\right\}=1, \qquad \tilde{\mathsf{P}}\left\{z_\infty=\tilde{\mathfrak{z}}\cdot\mathfrak{z}^{-1}\right\}=1.$$

Далее,

$$\tilde{\mathsf{P}}\{A\cap(\mathfrak{z}=0)\}=\mathsf{P}\{A\cap(\mathfrak{z}=0)\cap(\tilde{\mathfrak{z}}>0)\}=$$
$$=\tilde{\mathsf{P}}\{A\cap(\tilde{\mathfrak{z}}\cdot\mathfrak{z}^{-1}=\infty)\}=\tilde{\mathsf{P}}\{A\cap(z_\infty\mathfrak{z})=\infty\},$$

что вместе с (7) доказывает разложение (3).

Из конструкции меры λ ясно, что $\lambda\ll\mathsf{P}$, причем $\mathsf{P}(z_\infty<\infty)=1$. И в то же время $\mu(z_\infty<\infty)=\tilde{\mathsf{P}}\{(z_\infty<\infty)\cap(z_\infty=\infty)\}=0$. Тем самым теорема доказана.

Из разложения Лебега (3) вытекают следующие полезные критерии абсолютной непрерывности и сингулярности для локально абсолютно непрерывных вероятностных мер.

Теорема 2. *Пусть* $\tilde{\mathsf{P}}\overset{\text{loc}}{\ll}\mathsf{P}$, *т. е.* $\tilde{\mathsf{P}}_n\ll\mathsf{P}_n$, $n\geqslant1$. *Тогда*

$$\tilde{\mathsf{P}}\ll\mathsf{P}\Leftrightarrow\mathsf{M}z_\infty=1\Leftrightarrow\tilde{\mathsf{P}}(z_\infty<\infty)=1, \tag{8}$$

$$\tilde{\mathsf{P}}\perp\mathsf{P}\Leftrightarrow\mathsf{M}z_\infty=0\Leftrightarrow\tilde{\mathsf{P}}(z_\infty=\infty)=1, \tag{9}$$

где M — *усреднение по мере* P.

Доказательство. Полагая в (3) $A=\Omega$, находим, что

$$\mathsf{M}z_\infty=1\Leftrightarrow\tilde{\mathsf{P}}(z_\infty=\infty)=0, \tag{10}$$

$$\mathsf{M}z_\infty=0\Leftrightarrow\tilde{\mathsf{P}}(z_\infty=\infty)=1. \tag{11}$$

Если $\tilde{\mathsf{P}}(z_\infty=\infty)=0$, то снова из (3) следует, что $\tilde{\mathsf{P}}\ll\mathsf{P}$.

Обратно, пусть $\tilde{\mathsf{P}}\ll\mathsf{P}$. Тогда поскольку $\mathsf{P}(z_\infty=\infty)=0$, то $\tilde{\mathsf{P}}(z_\infty=\infty)=0$.

Далее, если $\tilde{\mathsf{P}}\perp\mathsf{P}$, то существует множество $B\in\mathscr{F}$ с $\tilde{\mathsf{P}}(B)=1$ и $\mathsf{P}(B)=0$. Тогда из (3) $\tilde{\mathsf{P}}(B\cap(z_\infty=\infty))=1$ и, значит, $\tilde{\mathsf{P}}(z_\infty=\infty)=1$. Если же $\tilde{\mathsf{P}}(z_\infty=\infty)=1$, то свойство $\tilde{\mathsf{P}}\perp\mathsf{P}$ очевидно, поскольку $\mathsf{P}(z_\infty=\infty)=0$.

Теорема доказана.

2. Из теоремы 2 ясно, что критерии абсолютной непрерывности и сингулярности можно выражать или в терминах меры P (и проверять равенства $\mathsf{M}z_\infty=1$ или $\mathsf{M}z_\infty=0$) или же в терминах меры $\tilde{\mathsf{P}}$ (и тогда проверять, что $\tilde{\mathsf{P}}(z_\infty<\infty)=1$ или $\tilde{\mathsf{P}}(z_\infty=\infty)=1$).

В силу теоремы 5 § 6 гл. II условие $\mathsf{M}z_\infty=1$ равносильно условию равномерной интегрируемости (по мере P) семейства $\{z_n\}_{n\geqslant1}$. Это обстоятельство позволяет давать простые *достаточные условия для абсолютной непрерывности* $\tilde{\mathsf{P}}\ll\mathsf{P}$. Например, если

$$\sup_n \mathsf{M}[z_n\ln^+z_n]<\infty \tag{12}$$

или если

$$\sup_n \mathsf{M} z_n^{1+\varepsilon} < \infty, \qquad \varepsilon > 0, \tag{13}$$

то, согласно лемме 3 § 6 гл. II, семейство случайных величин $\{z_n\}_{n\geqslant 1}$ будет равномерно интегрируемым и, значит, $\tilde{\mathsf{P}} \ll \mathsf{P}$.

Во многих же случаях при проверке свойств абсолютной непрерывности или сингулярности предпочтительнее использовать критерии, выраженные в терминах меры $\tilde{\mathsf{P}}$, поскольку тогда дело сводится к исследованию $\tilde{\mathsf{P}}$-вероятности «хвостового» события $\{z_\infty < \infty\}$, а для этого можно использовать утверждения типа закона «нуля или единицы».

В качестве иллюстрации покажем, как из теоремы 2 выводится альтернатива Какутани.

Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — некоторое вероятностное пространство, $(R^\infty, \mathscr{B}_\infty)$ — измеримое пространство числовых последовательностей $x = (x_1, x_2, \ldots)$ с $\mathscr{B}_\infty = \mathscr{B}(R^\infty)$, и пусть $\mathscr{B}_n = \sigma\{x\colon (x_1, \ldots, x_n)\}$. Предположим, что $\xi = (\xi_1, \xi_2, \ldots)$ и $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots)$ — две последовательности, состоящие из независимых случайных величин.

Обозначим через P и $\tilde{P}$ распределения вероятностей на $(R^\infty, \mathscr{B}_\infty)$ для ξ и $\tilde{\xi}$ соответственно, т. е.

$$P(B) = \mathsf{P}\{\xi \in B\}, \qquad \tilde{P}(B) = \mathsf{P}\{\tilde{\xi} \in B\}, \qquad B \in \mathscr{B}_\infty.$$

Пусть также

$$P_n = P \mid \mathscr{B}_n, \qquad \tilde{P}_n = \tilde{P} \mid \mathscr{B}_n$$

— сужения мер P и $\tilde{P}$ на $\mathscr{B}_n$ и

$$P_{\xi_n}(A) = \mathsf{P}(\xi_n \in A),$$
$$P_{\tilde{\xi}_n}(A) = \mathsf{P}(\tilde{\xi}_n \in A), \qquad A \in \mathscr{B}(R^1).$$

Теорема 3 (альтернатива Какутани). *Пусть $\xi = (\xi_1, \xi_2, \ldots)$ и $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots)$ — последовательности из независимых случайных величин, для которых*

$$P_{\tilde{\xi}_n} \ll P_{\xi_n}, \qquad n \geqslant 1. \tag{14}$$

Тогда или $\tilde{P} \ll P$, или $\tilde{P} \perp P$.

Доказательство. Условие (14), очевидно, равносильно условию, что $\tilde{P}_n \ll P_n$, $n \leqslant 1$, т. е. $\tilde{P} \overset{\text{loc}}{\ll} P$. Ясно, что

$$z_n = \frac{d\tilde{P}_n}{dP_n} = q_1(x_1) \ldots q_n(x_n),$$

где

$$q_i(x_i)=\frac{dP_{\tilde{\xi}_i}}{dP_{\xi_i}}(x_i). \tag{15}$$

Следовательно,

$$\{x\colon z_\infty<\infty\}=\{x\colon \ln z_\infty<\infty\}=\left\{x\colon \sum_{i=1}^{\infty}\ln q_i(x_i)<\infty\right\}.$$

Событие $\left\{x\colon \sum_{i=1}^{\infty}\ln q_i(x_i)<\infty\right\}$ является «хвостовым». Поэтому в силу закона «нуля или единицы» Колмогорова (теорема 1 § 1 гл. IV) вероятность $\tilde{P}\{x\colon z_\infty<\infty\}$ принимает только два значения (0 или 1) и, значит, по теореме 2 или $\tilde{P}\perp P$, или $\tilde{P}\ll P$.

Теорема доказана.

3. Следующая теорема дает критерий абсолютной непрерывности и сингулярности, выраженный в «предсказуемых» терминах.

Теорема 4. *Пусть* $\tilde{\mathsf{P}}\overset{\mathrm{loc}}{\ll}\mathsf{P}$,

$$\alpha_n=z_n z_{n-1}^{\oplus},\qquad n\geqslant 1,$$

с $z_0=1$. *Тогда* ($\mathscr{F}_0=\{\varnothing,\ \Omega\}$)

$$\tilde{\mathsf{P}}\ll\mathsf{P}\Leftrightarrow\tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty}\left[1-\mathsf{M}\left(\sqrt{\alpha_n}\,|\,\mathscr{F}_{n-1}\right)\right]<\infty\right\}=1, \tag{16}$$

$$\tilde{\mathsf{P}}\perp\mathsf{P}\Leftrightarrow\tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty}\left[1-\mathsf{M}\left(\sqrt{\alpha_n}\,|\,\mathscr{F}_{n-1}\right)\right]=\infty\right\}=1. \tag{17}$$

Доказательство. Поскольку

$$\tilde{\mathsf{P}}_n\{z_n=0\}=\int_{\{z_n=0\}}z_n\,d\mathsf{P}=0,$$

то ($\tilde{\mathsf{P}}$-п. н.)

$$z_n=\prod_{k=1}^{n}\alpha_k=\exp\left\{\sum_{k=1}^{n}\ln\alpha_k\right\}. \tag{18}$$

Полагая в (3) $A=\{z_\infty=0\}$, находим, что $\tilde{\mathsf{P}}\{z_\infty=0\}=0$. Поэтому из (18) ($\tilde{\mathsf{P}}$-п. н.)

$$\{z_\infty<\infty\}=\{0<z_\infty<\infty\}=\{0<\lim z_n<\infty\}=$$
$$=\left\{-\infty<\lim\sum_{k=1}^{n}\ln\alpha_k<\infty\right\}. \tag{19}$$

Введем функцию

$$u(x)=\begin{cases} x, & |x|\leqslant 1,\\ \operatorname{sign} x, & |x|>1.\end{cases}$$

Тогда

$$\left\{-\infty<\lim\sum_{k=1}^{n}\ln\alpha_k<\infty\right\}=\left\{-\infty<\lim\sum_{k=1}^{n}u(\ln\alpha_k)<\infty\right\}. \quad (20)$$

Пусть $\tilde{\mathsf M}$ означает усреднение по мере $\tilde{\mathsf P}$ и η — $\mathscr F_n$-измеримая интегрируемая случайная величина. Из свойств условных математических ожиданий следует (задача 4), что

$$z_{n-1}\tilde{\mathsf M}(\eta\mid\mathscr F_{n-1})=\mathsf M(\eta z_n\mid\mathscr F_{n-1})\ (\mathsf P\text{- и }\tilde{\mathsf P}\text{-п. н.}), \quad (21)$$

$$\tilde{\mathsf M}(\eta\mid\mathscr F_{n-1})=z_{n-1}^{\oplus}\mathsf M(\eta z_n\mid\mathscr F_{n-1})\ (\tilde{\mathsf P}\text{-п. н.}). \quad (22)$$

Вспоминая, что $\alpha_n=z_{n-1}^{\oplus}z_n$, из (22) получаем следующую полезную формулу:

$$\tilde{\mathsf M}(\eta\mid\mathscr F_{n-1})=\mathsf M(\alpha_n\eta\mid\mathscr F_{n-1})\ (\tilde{\mathsf P}\text{-п. н.}), \quad (23)$$

из которой, в частности, вытекает, что

$$\mathsf M(\alpha_n\mid\mathscr F_{n-1})=1\ (\tilde{\mathsf P}\text{-п. н.}). \quad (24)$$

Из (23)

$$\tilde{\mathsf M}[u(\ln\alpha_n)\mid\mathscr F_{n-1}]=\mathsf M[\alpha_n u(\ln\alpha_n)\mid\mathscr F_{n-1}]\ (\tilde{\mathsf P}\text{-п. н.}).$$

Поскольку $xu(\ln x)\geqslant x-1$ для всех $x\geqslant 0$, то в силу (24)

$$\tilde{\mathsf M}[u(\ln\alpha_n)\mid\mathscr F_{n-1}]\geqslant 0\ (\tilde{\mathsf P}\text{-п. н.}).$$

Отсюда следует, что стохастическая последовательность $X=(X_n,\mathscr F_n)$ с

$$X_n=\sum_{k=1}^{n}u(\ln\alpha_k)$$

относительно меры $\tilde{\mathsf P}$ является субмартингалом, причем $|\Delta X_n|=|u(\ln\alpha_n)|\leqslant 1$.

Тогда по теореме 5 из § 5 ($\tilde{\mathsf P}$-п. н.)

$$\left\{-\infty<\lim\sum_{k=1}^{n}u(\ln\alpha_k)<\infty\right\}=$$

$$=\left\{\sum_{k=1}^{\infty}\tilde{\mathsf M}[u(\ln\alpha_k)+u^2(\ln\alpha_k)\mid\mathscr F_{k-1}]<\infty\right\}. \quad (25)$$

Тем самым из (19), (20), (22) и (25) находим, что ($\tilde{\mathsf{P}}$-п. н.)

$$\{z_\infty < \infty\} = \left\{\sum_{k=1}^{\infty} \tilde{\mathsf{M}}\,[u(\ln\alpha_k) + u^2(\ln\alpha_k) \mid \mathscr{F}_{k-1}] < \infty\right\} =$$
$$= \left\{\sum_{k=1}^{\infty} \mathsf{M}\,[\alpha_k u(\ln\alpha_k) + \alpha_k u^2(\ln\alpha_k) \mid \mathscr{F}_{k-1}] < \infty\right\}$$

и, следовательно, в силу теоремы 2

$$\tilde{\mathsf{P}} \ll \mathsf{P} \Leftrightarrow \tilde{\mathsf{P}}\left\{\sum_{k=1}^{\infty} \mathsf{M}\,[\alpha_k u(\ln\alpha_k) + \alpha_k u^2(\ln\alpha_k) \mid \mathscr{F}_{k-1}] < \infty\right\} = 1, \quad (26)$$

$$\tilde{\mathsf{P}} \perp \mathsf{P} \Leftrightarrow \tilde{\mathsf{P}}\left\{\sum_{k=1}^{\infty} \mathsf{M}\,[\alpha_k u(\ln\alpha_k) + \alpha_k u^2(\ln\alpha_k) \mid \mathscr{F}_{k-1}] = \infty\right\} = 1. \quad (27)$$

Заметим теперь, что в силу (24)

$$\mathsf{M}\,[(1-\sqrt{\alpha_n})^2 \mid \mathscr{F}_{n-1}] = 2\mathsf{M}\,[1-\sqrt{\alpha_n} \mid \mathscr{F}_{n-1}] \quad (\tilde{\mathsf{P}}\text{-п. н.})$$

и для всех $x \geqslant 0$ найдутся такие константы A и B $(0 < A < B < \infty)$, что

$$A(1-\sqrt{x})^2 \leqslant xu(\ln x) + xu^2(\ln x) + 1 - x \leqslant B(1-\sqrt{x})^2. \quad (28)$$

Поэтому утверждения (16) и (17) следуют из (26), (27) и (24), (28).

Теорема доказана.

Следствие 1. Если для любого $n \geqslant 1$ σ-алгебры $\sigma(\alpha_n)$ и $\mathscr{F}_{n-1}$ независимы по мере P (или $\tilde{\mathsf{P}}$) и $\tilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, то имеет место альтернатива: *либо* $\tilde{\mathsf{P}} \ll \mathsf{P}$, *либо* $\tilde{\mathsf{P}} \perp \mathsf{P}$. При этом

$$\tilde{\mathsf{P}} \ll \mathsf{P} \Leftrightarrow \sum_{n=1}^{\infty} [1 - \mathsf{M}\sqrt{\alpha_n}] < \infty,$$

$$\tilde{\mathsf{P}} \perp \mathsf{P} \Leftrightarrow \sum_{n=1}^{\infty} [1 - \mathsf{M}\sqrt{\alpha_n}] = \infty.$$

В частности, в ситуации Какутани (см. теорему 3) $\alpha_n = q_n$ и

$$\tilde{\mathsf{P}} \ll \mathsf{P} \Leftrightarrow \sum_{n=1}^{\infty} [1 - \mathsf{M}\sqrt{q_n(x_n)}] < \infty,$$

$$\tilde{\mathsf{P}} \perp \mathsf{P} \Leftrightarrow \sum_{n=1}^{\infty} [1 - \mathsf{M}\sqrt{q_n(x_n)}] = \infty.$$

Следствие 2. Пусть $\tilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$. Тогда

$$\tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty} \mathsf{M}(\alpha_n \ln \alpha_n \mid \mathscr{F}_{n-1}) < \infty\right\} = 1 \Rightarrow \tilde{\mathsf{P}} \ll \mathsf{P}.$$

Для доказательства достаточно заметить, что для любого $x \geqslant 0$

$$x \ln x + \frac{3}{2}(1-x) \geqslant 1 - x^{1/2}, \tag{29}$$

и воспользоваться (16) и (24).

Следствие 3. Поскольку ряд $\sum_{n=1}^{\infty}[1-\mathsf{M}(\sqrt{\alpha_n} \mid \mathscr{F}_{n-1})]$, состоящий из неотрицательных ($\tilde{\mathsf{P}}$-п. н.) членов, сходится или расходится одновременно с рядом $\sum |\ln \mathsf{M}(\sqrt{\alpha_n} \mid \mathscr{F}_{n-1})|$, то утверждениям (16) и (17) теоремы 4 можно придать следующую форму:

$$\tilde{\mathsf{P}} \ll \mathsf{P} \Leftrightarrow \tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty} |\ln \mathsf{M}(\sqrt{\alpha_n} \mid \mathscr{F}_{n-1})| < \infty\right\} = 1, \tag{30}$$

$$\tilde{\mathsf{P}} \perp \mathsf{P} \Leftrightarrow \tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty} |\ln \mathsf{M}(\sqrt{\alpha_n} \mid \mathscr{F}_{n-1})| = \infty\right\} = 1. \tag{31}$$

Следствие 4. Пусть существуют константы A и B такие, что $0 \leqslant A < 1$, $B \geqslant 0$ и

$$\mathsf{P}\{1 - A \leqslant \alpha_n \leqslant 1 + B\} = 1, \quad n \geqslant 1.$$

Тогда, если $\tilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$, то

$$\tilde{\mathsf{P}} \ll \mathsf{P} \Leftrightarrow \tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty} \mathsf{M}[(1-\alpha_n)^2 \mid \mathscr{F}_{n-1}] < \infty\right\} = 1,$$

$$\tilde{\mathsf{P}} \perp \mathsf{P} \Leftrightarrow \tilde{\mathsf{P}}\left\{\sum_{n=1}^{\infty} \mathsf{M}[(1-\alpha_n)^2 \mid \mathscr{F}_{n-1}] = \infty\right\} = 1.$$

Для доказательства достаточно заметить, что для $x \in [1-A, 1+B]$, $0 \leqslant A < 1$, $B \geqslant 0$, найдутся такие константы c и C $(0 < c < C < \infty)$, что

$$c(1-x)^2 \leqslant (1-\sqrt{x})^2 \leqslant C(1-x)^2. \tag{32}$$

4. В обозначениях п. 2 предположим, что $\xi = (\xi_1, \xi_2, \ldots)$ и $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2 \ldots)$ — две гауссовские последовательности, $\tilde{P}_n \sim P_n$, $n \geqslant 1$. Покажем, как для таких последовательностей из полученных выше «предсказуемых» критериев следует альтернатива Гаека — Фельдмана: *либо* $\tilde{P} \sim P$, *либо* $\tilde{P} \perp P$

По теореме о нормальной корреляции (теорема 2 § 13 гл. II) условные математические ожидания $M(x_n \mid \mathscr{B}_{n-1})$ и $\tilde{M}(x_n \mid \mathscr{B}_{n-1})$, где M и $\tilde{M}$ — усреднения по мерам P и $\tilde{P}$ соответственно, являются линейными функциями от $x_1, \ldots, x_{n-1}$. Обозначим эти (линейные) функции через $a_{n-1}(x)$ и $\tilde{a}_{n-1}(x)$ соответственно и положим

$$b_{n-1} = (M[x_n - a_{n-1}(x)]^2)^{1/2},$$
$$\tilde{b}_{n-1} = (\tilde{M}[x_n - \tilde{a}_{n-1}(x)]^2)^{1/2}.$$

По той же самой теореме о нормальной корреляции найдутся последовательности $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots)$ и $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots)$, состоящие из независимых гауссовских случайных величин с нулевым средним и единичной дисперсией такие, что

$$\begin{aligned} x_n &= a_{n-1}(x) + b_{n-1}\varepsilon_n \quad (P\text{-п. н.}),\\ x_n &= \tilde{a}_{n-1}(x) + \tilde{b}_{n-1}\tilde{\varepsilon}_n \quad (\tilde{P}\text{-п. н.}). \end{aligned} \tag{33}$$

Заметим, что в случае $b_{n-1} = 0$ $(\tilde{b}_{n-1} = 0)$ для построения величин ε_n $(\tilde{\varepsilon}_n)$ приходится, вообще говоря, расширять вероятностное пространство. Однако если $b_{n-1} = 0$, то распределение вектора $(x_1, \ldots, x_n)$ сосредоточено (P-п. н.) на линейном многообразии $x_n = a_{n-1}(x)$, и поскольку по предположению $\tilde{P}_n \sim P_n$, то $\tilde{b}_{n-1} = 0$, $a_{n-1}(x) = \tilde{a}_{n-1}(x)$ и $\alpha_n(x) = 1$ (P- и $\tilde{P}$-п. н.). Поэтому без ограничения общности можно считать, что $b_n^2 > 0$, $\tilde{b}_n^2 > 0$ при всех $n \geqslant 1$, поскольку в противном случае вклад соответствующих членов в сумму $\sum_{n=1}^{\infty} \left[1 - M\sqrt{\alpha_n} \mid \mathscr{B}_{n-1}\right]$ (см. (16) и (17)) равен нулю.

Используя предположения о гауссовости, из (33) находим, что для $n \geqslant 1$

$$\alpha_n = d_{n-1}^{-1} \exp\left\{-\frac{(x_n - a_{n-1}(x))^2}{2b_{n-1}^2} + \frac{(x_n - \tilde{a}_{n-1}(x))^2}{2\tilde{b}_{n-1}^2}\right\}, \tag{34}$$

где $d_n = |\tilde{b}_n \cdot b_n^{-1}|$ и

$$a_0(x) = \mathsf{M}\xi_1, \quad \tilde{a}_0(x) = \tilde{\mathsf{M}}\tilde{\xi}_1,$$

$$b_0^2 = \mathsf{D}\xi_1, \qquad \tilde{b}_0^2 = \tilde{\mathsf{D}}\tilde{\xi}_1.$$

Из (34)

$$\ln M\left(\alpha_n^{1/2} \mid \mathscr{B}_{n-1}\right) = \frac{1}{2}\ln\frac{2d_{n-1}}{1+d_{n-1}^2} - \frac{d_{n-1}^2}{1+d_{n-1}^2}\left(\frac{a_{n-1}(x) - \tilde{a}_{n-1}(x)}{b_{n-1}}\right)^2.$$

Поскольку $\ln\frac{2d_{n-1}}{1+d_{n-1}^2} \leqslant 0$, то утверждение (30) принимает следующую форму:

$$\tilde{P} \ll P \Leftrightarrow \tilde{P}\left\{\sum_{n=1}^{\infty}\left[\frac{1}{2}\ln\frac{1+d_{n-1}^2}{2d_{n-1}} + \right.\right.$$

$$\left.\left.+\frac{d_{n-1}^2}{1+d_{n-1}^2}\cdot\left(\frac{a_{n-1}(x)-\tilde{a}_{n-1}(x)}{b_{n-1}}\right)^2\right] < \infty\right\} = 1. \quad (35)$$

Ряды $\sum_{n=1}^{\infty}\ln\frac{1+d_{n-1}^2}{2d_{n-1}}$ и $\sum_{n=1}^{\infty}(d_{n-1}^2-1)$ сходятся или расходятся одновременно, поэтому из (35) следует, что

$$\tilde{P} \ll P \Leftrightarrow \tilde{P}\left\{\sum_{n=0}^{\infty}\left[\left(\frac{\tilde{b}_n^2}{b_n^2}-1\right)^2 + \frac{\Delta_n^2(x)}{b_n^2}\right] < \infty\right\} = 1, \quad (36)$$

где $\Delta_n(x) = a_n(x) - \tilde{a}_n(x)$.

В силу линейности $a_n(x)$ и $\tilde{a}_n(x)$ последовательность случайных величин $\left\{\frac{\Delta_n(x)}{b_n}\right\}_{n \geqslant 0}$ образует гауссовскую систему (как по мере $\tilde{P}$, так и по мере P). Как следует из приводимой далее леммы, для таких последовательностей имеет место следующий аналог закона «нуля или единицы»:

$$\tilde{P}\left\{\sum\left(\frac{\Delta_n(x)}{b_n}\right)^2 < \infty\right\} = 1 \Leftrightarrow \sum \tilde{M}\left(\frac{\Delta_n(x)}{b_n}\right)^2 < \infty. \quad (37)$$

Поэтому из (36) следует, что

$$\tilde{P} \ll P \Leftrightarrow \sum_{n=0}^{\infty}\left[\tilde{M}\left(\frac{\Delta_n(x)}{b_n}\right)^2 + \left(\frac{\tilde{b}_n^2}{b_n^2}-1\right)^2\right] < \infty$$

и аналогичным образом

$$\tilde{P} \perp P \Leftrightarrow \sum_{n=0}^{\infty}\left[\tilde{M}\left(\frac{\Delta_n(x)}{b_n}\right)^2 + \left(\frac{\tilde{b}_n^2}{b_n^2}-1\right)^2\right] = \infty.$$

Отсюда ясно, что если меры $\tilde{P}$ и P не сингулярны, то $\tilde{P} \ll P$. Но по предположению $\tilde{P}_n \sim P_n$, $n \geqslant 1$; поэтому в силу симметрии $P \ll \tilde{P}$. Тем самым имеет место следующая

Теорема 5 (альтернатива Гаека — Фельдмана). *Пусть* $\xi = (\xi_1, \xi_2, \ldots)$ *и* $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots)$ — *две гауссовские последовательности, конечномерные распределения которых эквивалентны:* $\tilde{P}_n \sim P_n$,

$n \geqslant 1$. *Тогда либо* $\tilde{P} \sim P$, *либо* $\tilde{P} \perp P$. *При этом*

$$
\begin{aligned}
\tilde{P} \sim P &\Leftrightarrow \sum_{n=0}^{\infty}\left[\tilde{M}\left(\frac{\Delta_n(x)}{b_n}\right)^2 + \left(\frac{\tilde{b}_n^2}{b_n^2} - 1\right)^2\right] < \infty, \\
\tilde{P} \perp P &\Leftrightarrow \sum_{n=0}^{\infty}\left[\tilde{M}\left(\frac{\Delta_n(x)}{b_n}\right)^2 + \left(\frac{\tilde{b}_n^2}{b_n^2} - 1\right)^2\right] = \infty.
\end{aligned} \tag{38}
$$

Докажем теперь закон «нуля или единицы» для гауссовских последовательностей, использованный при доказательстве теоремы 5.

Лемма. *Пусть* $\beta = (\beta_n)_{n \geqslant 1}$ — *гауссовская последовательность, заданная на* $(\Omega, \mathscr{F}; \mathsf{P})$. *Тогда*

$$
\mathsf{P}\left\{\sum_{n=1}^{\infty} \beta_n^2 < \infty\right\} = 1 \Leftrightarrow \sum_{n=1}^{\infty} \mathsf{M}\beta_n^2 < \infty. \tag{39}
$$

Доказательство. Импликация $\Leftarrow$ следует из теоремы Фубини. Установим обратное утверждение, предположив сначала, что $\mathsf{M}\beta_n = 0$, $n \geqslant 1$. С этой целью достаточно показать, что

$$
\mathsf{M}\sum_{n=1}^{\infty} \beta_n^2 \leqslant \left[\mathsf{M}\exp\left(-\sum_{n=1}^{\infty} \beta_n^2\right)\right]^{-2}, \tag{40}
$$

поскольку тогда из условия $\mathsf{P}\{\sum \beta_n^2 < \infty\} = 1$ будет следовать, что правая часть в (40) меньше бесконечности.

Зафиксируем некоторое $n \geqslant 1$. Тогда из §§ 11 и 13 гл. II следует, что можно найти такие независимые гауссовские случайные величины $\beta_{k,n}$, $k = 1, \ldots, r \leqslant n$, с $\mathsf{M}\beta_{k,n} = 0$, что

$$
\sum_{k=1}^{n} \beta_k^2 = \sum_{k=1}^{r} \beta_{k,n}^2.
$$

Если обозначить $\mathsf{M}\beta_{k,n}^2 = \lambda_{k,n}$, то легко найдем, что

$$
\mathsf{M}\sum_{k=1}^{r} \beta_{k,n}^2 = \sum_{k=1}^{r} \lambda_{k,n} \tag{41}
$$

и

$$
\mathsf{M}\exp\left(-\sum_{k=1}^{r} \beta_{k,n}^2\right) = \prod_{k=1}^{r} (1 + 2\lambda_{k,n})^{-1/2}. \tag{42}
$$

Сравнивая правые части в (41) и (42), получаем

$$
\mathsf{M}\sum_{k=1}^{n} \beta_k^2 = \mathsf{M}\sum_{k=1}^{r} \beta_{k,n}^2 \leqslant \left[\mathsf{M}e^{-\sum_{k=1}^{r} \beta_{k,n}^2}\right]^{-2} = \left[\mathsf{M}e^{-\sum_{k=1}^{n} \beta_k^2}\right]^{-2},
$$

откуда предельным переходом (при $n\to\infty$) получаем требуемое неравенство (40).

Предположим теперь, что $\mathsf{M}\beta_n\not\equiv 0$.

Рассмотрим новую последовательность $\tilde\beta=(\tilde\beta_n)_{n\geqslant 1}$ с тем же распределением, что и у последовательности $\beta=(\beta_n)_{n\geqslant 1}$, и не зависящую от нее (в случае необходимости расширяя исходное вероятностное пространство). Тогда, если $\mathsf{P}\left\{\sum_{n=1}^{\infty}\beta_n^2<\infty\right\}=1$, то $\mathsf{P}\left\{\sum_{n=1}^{\infty}(\beta_n-\tilde\beta_n)^2<\infty\right\}=1$, и по доказанному

$$2\sum_{n=1}^{\infty}\mathsf{M}(\beta_n-\mathsf{M}\beta_n)^2=\sum_{n=1}^{\infty}\mathsf{M}(\beta_n-\tilde\beta_n)^2<\infty.$$

Так как

$$(\mathsf{M}\beta_n)^2\leqslant 2\beta_n^2+2(\beta_n-\mathsf{M}\beta_n)^2,$$

то $\sum_{n=1}^{\infty}(\mathsf{M}\beta_n)^2<\infty$ и, значит,

$$\sum_{n=1}^{\infty}\mathsf{M}\beta_n^2=\sum_{n=1}^{\infty}(\mathsf{M}\beta_n)^2+\sum_{n=1}^{\infty}\mathsf{M}(\beta_n-M\beta_n)^2<\infty.$$

Лемма доказана.

5. Задачи.

1. Доказать справедливость утверждений (6).

2. Пусть $\tilde{\mathsf{P}}_n\sim\mathsf{P}_n$, $n\geqslant 1$. Показать, что

$$\tilde{\mathsf{P}}\sim\mathsf{P}\Leftrightarrow\tilde{\mathsf{P}}\{z_\infty<\infty\}=\mathsf{P}\{z_\infty>0\}=1,$$
$$\tilde{\mathsf{P}}\perp\mathsf{P}\Leftrightarrow\tilde{\mathsf{P}}\{z_\infty=\infty\}=1 \text{ или } \mathsf{P}\{z_\infty=0\}=1.$$

3. Пусть $\tilde{\mathsf{P}}_n\ll\mathsf{P}_n$, $n\geqslant 1$, τ — момент остановки (относительно $(\mathscr{F}_n)$), $\tilde{\mathsf{P}}_\tau=\tilde{\mathsf{P}}\,|\,\mathscr{F}_\tau$ и $\mathsf{P}_\tau=\mathsf{P}\,|\,\mathscr{F}_\tau$ — сужения мер $\tilde{\mathsf{P}}$ и P на σ-алгебру $\mathscr{F}_\tau$. Показать, что $\tilde{\mathsf{P}}_\tau\ll\mathsf{P}_\tau$, если и только если $\{\tau=\infty\}=\{z_\infty<\infty\}$ ($\tilde{\mathsf{P}}$-п. н.). (В частности, если $\tilde{\mathsf{P}}\{\tau<\infty\}=1$, то $\tilde{\mathsf{P}}_\tau\ll\mathsf{P}_\tau$.)

4. Доказать формулы (21) и (22).

5. Проверить справедливость неравенств (28), (29), (32).

6. Доказать формулу (34).

7. Пусть в п. 2 последовательности $\xi=(\xi_1,\xi_2,\ldots)$ и $\tilde\xi=(\tilde\xi_1,\tilde\xi_2,\ldots)$ состоят из независимых одинаково распределенных случайных величин. Показать, что если $P_{\tilde\xi_1}\ll P_{\xi_1}$, то $\tilde P\ll P$ в том и только том случае, когда меры $P_{\tilde\xi_1}$ и P_{ξ_1} совпадают. Если же $P_{\tilde\xi_1}\ll P_{\xi_1}$ и $P_{\tilde\xi_1}\neq P_{\xi_1}$, то $\tilde P\perp P$.

§ 7. Об асимптотике вероятности выхода случайного блуждания за криволинейную границу

1. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, $S_n=\xi_1+\ldots+\xi_n$, $g=g(n)$ — некоторая «граница», $n\geqslant 1$, и

$$\tau=\inf\{n\geqslant 1:\ S_n<g(n)\}$$

— тот первый момент, когда случайное блуждание $(S_n)_{n\geqslant 1}$ окажется ниже границы $g=g(n)$. (Как обычно, $\tau=\infty$, если $\{\cdot\}=\varnothing$.)

Отыскание точного вида распределения для момента τ является весьма трудной задачей. В настоящем параграфе находится асимптотика вероятности $\mathsf{P}(\tau>n)$ при $n\to\infty$ для широкого класса границ $g=g(n)$ и в предположении, что величины ξ_i нормально распределены. Применяемый метод доказательства основан на идее «абсолютно непрерывной замены меры» с использованием ряда изложенных выше свойств мартингалов и марковских моментов.

Теорема 1. *Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные, $\xi_i\sim\mathcal{N}(0, 1)$, случайные величины. Предположим, что граница $g=g(n)$ такова, что $g(1)<0$ и для $n\geqslant 2$*

$$0\leqslant\Delta g(n+1)\leqslant\Delta g(n), \tag{1}$$

где $\Delta g(n)=g(n)-g(n-1)$ и

$$\ln n=o\left(\sum_{k=2}^{n}[\Delta g(k)]^2\right),\qquad n\to\infty. \tag{2}$$

Тогда

$$\mathsf{P}(\tau>n)=\exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2(1+o(1))\right\},\qquad n\to\infty. \tag{3}$$

Прежде чем переходить к доказательству, отметим, что условия (1) и (2) выполнены, если, скажем,

$$g(n)=an^{\nu}+b,\quad 1/2<\nu\leqslant 1,\quad a+b<0,$$

или (при достаточно больших n)

$$g(n)=n^{\nu}L(n),\quad 1/2\leqslant\nu\leqslant 1,$$

где $L(n)$ — некоторая медленно меняющаяся функция (например, $L(n)=C(\ln n)^{\beta}$ с любым β при $1/2<\nu<1$ и с $\beta>0$ при $\nu=1/2$).

2. Следующие два вспомогательных предложения будут использоваться при доказательстве теоремы 1.

Будем предполагать, что $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, $\xi_i\sim\mathcal{N}(0, 1)$. Обозначим $\mathcal{F}_0=\{\varnothing, \Omega\}$, $\mathcal{F}_n=\sigma\{\omega:\ \xi_1, \ldots, \xi_n\}$, и пусть $\alpha=(\alpha_n, \mathscr{F}_{n-1})$ — некоторая предсказуемая последовательность с $\mathsf{P}(|\alpha_n|\leqslant C)=1$, $n\geqslant 1$, где C — некоторая константа. Образуем последовательность $z=(z_n, \mathscr{F}_n)$ с

$$z_n=\exp\left\{\sum_{k=1}^{n}\alpha_k\xi_k-\frac{1}{2}\sum_{k=1}^{n}\alpha_k^2\right\},\qquad n\geqslant 1. \tag{4}$$

Нетрудно проверить, что (относительно меры P) последовательность $z=(z_n, \mathscr{F}_n)$ образует мартингал с $\mathsf{M}z_n=1$, $n\geqslant 1$.

Зафиксируем некоторое $n\geqslant 1$ и на измеримом пространстве $(\Omega, \mathscr{F}_n)$ введем вероятностную меру $\tilde{\mathsf{P}}_n$, полагая

$$\tilde{\mathsf{P}}_n(A)=\mathsf{M}I(A)\,z_n,\qquad A\in\mathscr{F}_n. \tag{5}$$

Лемма 1. *Относительно меры $\tilde{\mathsf{P}}_n$ случайные величины $\tilde{\xi}_k=\xi_k-\alpha_k$, $1\leqslant k\leqslant n$, являются независимыми и нормально распределенными, $\tilde{\xi}_k\sim\mathscr{N}(0, 1)$.*

Доказательство. Пусть символ $\tilde{\mathsf{M}}_n$ означает усреднение по мере $\tilde{\mathsf{P}}_n$. Тогда для $\lambda_k\in R$, $1\leqslant k\leqslant n$,

$$\tilde{\mathsf{M}}_n\exp\left\{i\sum_{k=1}^{n}\lambda_k\tilde{\xi}_k\right\}=\mathsf{M}\exp\left\{i\sum_{k=1}^{n}\lambda_k\tilde{\xi}_k\right\}z_n=$$

$$=\mathsf{M}\left[\exp\left\{i\sum_{k=1}^{n-1}\lambda_k\tilde{\xi}_k\right\}z_{n-1}\cdot\mathsf{M}\left\{\exp\left(i\lambda_n(\xi_n-\alpha_n)+\right.\right.\right.$$

$$\left.\left.\left.+\alpha_n\xi_n-\frac{\alpha_n^2}{2}\right)\,\middle|\,\mathscr{F}_{n-1}\right\}\right]=$$

$$=\mathsf{M}\left[\exp\left\{i\sum_{k=1}^{n-1}\lambda_k\tilde{\xi}_k\right\}z_{n-1}\right]\exp\left\{\frac{\lambda_n^2}{2}\right\}=\ldots=\exp\left\{\frac{1}{2}\sum_{k=1}^{n}\lambda_k^2\right\}.$$

Теперь требуемое утверждение вытекает из теоремы 4 § 12 гл. II.

Лемма 2. *Пусть $X=(X_n, \mathscr{F}_n)_{n\geqslant 1}$ — квадратично интегрируемый мартингал с нулевым средним и*

$$\sigma=\inf\{n\geqslant 1:\ X_n\leqslant -b\},$$

где константа $b>0$. Предположим, что

$$\mathsf{P}(X_1<-b)>0.$$

Тогда существует константа $C>0$ такая, что для всех $n\geqslant 1$

$$\mathsf{P}(\sigma>n)\geqslant\frac{C}{\mathsf{M}X_n^2}. \tag{6}$$

Доказательство. По следствию 1 к теореме VII.2.1 $\mathsf{M}X_{\sigma\wedge n}=0$, откуда

$$-\mathsf{M}I(\sigma\leqslant n)X_\sigma=\mathsf{M}I(\sigma>n)X_n. \tag{7}$$

На множестве $\{\sigma\leqslant n\}$

$$-X_\sigma\geqslant b>0.$$

Поэтому при $n\geqslant 1$

$$-\mathsf{M}I(\sigma\leqslant n)X_\sigma\geqslant b\mathsf{P}(\sigma\leqslant n)\geqslant b\mathsf{P}(\sigma=1)=b\mathsf{P}(X_1<-b)>0. \tag{8}$$

С другой стороны, в силу неравенства Коши — Буняковского

$$\mathsf{M}I(\sigma>n)X_n\leqslant[\mathsf{P}(\sigma>n)\cdot\mathsf{M}X_n^2]^{1/2}, \tag{9}$$

что вместе с (7) и (8) приводит к требуемому неравенству.

Доказательство теоремы 1. Достаточно показать, что

$$\varliminf_{n\to\infty}\ln\mathsf{P}(\tau>n)\Big/\sum_{k=2}^{n}[\Delta g(k)]^2\geqslant-\frac{1}{2} \tag{10}$$

и

$$\varlimsup_{n\to\infty}\ln\mathsf{P}(\tau>n)\Big/\sum_{k=2}^{n}[\Delta g(k)]^2\leqslant-\frac{1}{2}. \tag{11}$$

С этой целью рассмотрим (неслучайную) последовательность $(\alpha_n)_{n\geqslant 1}$ с

$$\alpha_1=0,\ \alpha_n=\Delta g(n),\qquad n\geqslant 2,$$

и вероятностные меры $(\tilde{\mathsf{P}}_n)_{n\geqslant 1}$, определенные формулой (5). Тогда в силу неравенства Гёльдера

$$\tilde{\mathsf{P}}_n(\tau>n)=\mathsf{M}I(\tau>n)z_n\leqslant(\mathsf{P}(\tau>n))^{1/q}(\mathsf{M}z_n^p)^{1/p}, \tag{12}$$

где $p>1$ и $q=\frac{p}{p-1}$.

Последний сомножитель легко вычисляется в явном виде:

$$(\mathsf{M}z_n^p)^{1/p}=\exp\left\{\frac{p-1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2\right\}. \tag{13}$$

Оценим теперь вероятность $\tilde{\mathsf{P}}_n(\tau>n)$, входящую в левую часть (12). Имеем

$$\tilde{\mathsf{P}}_n(\tau>n)=\tilde{\mathsf{P}}_n(S_k\geqslant g(k),\ 1\leqslant k\leqslant n)=$$
$$=\tilde{\mathsf{P}}_n(\tilde{S}_k\geqslant g(1),\quad 1\leqslant k\leqslant n),$$

где $\tilde{S}_k = \sum_{i=1}^{k} \tilde{\xi}_i$, $\tilde{\xi}_i = \xi_i - \alpha_i$. Согласно лемме 1 величины $\tilde{\xi}_1, \ldots, \tilde{\xi}_n$ по мере $\tilde{\mathsf{P}}_n$ являются независимыми и нормально распределенными, $\tilde{\xi}_i \sim \mathcal{N}(0, 1)$. Тогда по лемме 2 (примененной к $b = -g(1)$, $\mathsf{P} = \tilde{\mathsf{P}}_n$, $X_n = \tilde{S}_n$) находим, что

$$\tilde{\mathsf{P}}(\tau > n) \geqslant \frac{c}{n}, \tag{14}$$

где C — некоторая константа.

Тогда из (12) — (14) следует, что для любого $p > 1$

$$\mathsf{P}(\tau > n) \geqslant C_p \exp\left\{-\frac{p}{2} \sum_{k=2}^{n} [\Delta g(k)]^2 - \frac{p}{p-1} \ln n\right\}, \tag{15}$$

где C_p — некоторая константа. Из условий теоремы и в силу произвольности $p > 1$ из (15) получаем оценку снизу (10).

Для получения оценки сверху (11) прежде всего заметим, что, поскольку $z_n > 0$ (P- и $\tilde{\mathsf{P}}_n$-п. н.), то в силу (5)

$$\mathsf{P}(\tau > n) = \tilde{\mathsf{M}}_n I(\tau > n) z_n^{-1}, \tag{16}$$

где $\tilde{\mathsf{M}}_n$ — усреднение по мере $\tilde{\mathsf{P}}_n$.

В рассматриваемом нами случае $\alpha_1 = 0$, $\alpha_n = \Delta g(n)$, $n \geqslant 2$, поэтому для $n \geqslant 2$

$$z_n^{-1} = \exp\left\{-\sum_{k=2}^{n} \Delta g(k) \cdot \xi_k + \frac{1}{2} \sum_{k=2}^{n} [\Delta g(k)]^2\right\}.$$

По формуле суммирования по частям (см. доказательство леммы 2 в § 3 гл. IV)

$$\sum_{k=2}^{n} \Delta g(k) \cdot \xi_k = \Delta g(n) \cdot S_n - \sum_{k=2}^{n} S_{k-1} \Delta(\Delta g(k)),$$

откуда с учетом того, что по условиям теоремы $\Delta g(k) \geqslant 0$, $\Delta(\Delta g(k)) \leqslant 0$, находим, что на множестве $\{\tau > n\} = \{S_k \geqslant g(k),\ 1 \leqslant k \leqslant n\}$

$$\sum_{k=2}^{n} \Delta g(k) \cdot \xi_k \geqslant \Delta g(n) \cdot g(n) - \sum_{k=3}^{n} g(k-1) \Delta(\Delta g(k)) - \xi_1 \Delta g(2) =$$

$$= \sum_{k=2}^{n} [\Delta g(k)]^2 + g(1) \Delta g(2) - \xi_1 \Delta g(2),$$

Итак, из (16)

$$\mathsf{P}(\tau > n) \leqslant$$
$$\leqslant \exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2 - g(1)\,\Delta g(2)\right\}\tilde{\mathsf{M}}_n I(\tau > n)\, e^{-\xi_1 \Delta g(2)} \leqslant$$
$$\leqslant \exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2\right\}\tilde{\mathsf{M}}_n I(\tau > n)\, e^{-\xi_1 \Delta g(2)},$$

где

$$\tilde{\mathsf{M}}_n I(\tau > n)\, e^{-\xi_1 \Delta g(2)} \leqslant \mathsf{M} z_n e^{-\xi_1 \Delta g(2)} = \mathsf{M} e^{-\xi_1 \Delta g(2)} < \infty.$$

Поэтому

$$\mathsf{P}(\tau > n) \leqslant C \exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2\right\},$$

где C — некоторая положительная константа, что и доказывает оценку сверху (11).

Теорема доказана.

3. Идеи абсолютно непрерывной замены меры позволяют исследовать аналогичную задачу и для случая двусторонних границ. Приведем (без доказательства) один из результатов в этом направлении.

Теорема 2. *Пусть* $\xi_1, \xi_2, \ldots$ — *независимые одинаково распределенные,* $\xi_i \sim \mathcal{N}(0, 1)$, *случайные величины. Предположим, что* $f = f(n)$ — *положительная функция такая, что*

$$f(n) \to \infty, \qquad n \to \infty,$$

и

$$\sum_{k=2}^{n}[\Delta f(k)]^2 = o\left(\sum_{k=1}^{n} f^{-2}(k)\right), \qquad n \to \infty.$$

Тогда, если

$$\sigma = \inf\{n \geqslant 1:\ |S_n| \geqslant f(n)\},$$

то

$$\mathsf{P}(\sigma > n) = \exp\left\{-\frac{\pi^2}{8}\sum_{k=1}^{n} f^{-2}(k)\,(1 + o(1))\right\}, \qquad n \to \infty. \tag{17}$$

4. Задачи.

1. Показать, что последовательность, определенная в (4), является мартингалом.
2. Установить справедливость формулы (13).
3. Доказать формулу (17).

ГЛАВА VIII

ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН, ОБРАЗУЮЩИЕ МАРКОВСКУЮ ЦЕПЬ

§ 1. Определения и основные свойства

1. В главе I (§ 12) для случая конечных вероятностных пространств были изложены основные принципы, положенные в основу понятия *марковской зависимости* между случайными величинами. Там же были приведены разнообразные примеры и рассмотрены простейшие закономерности, которыми обладают случайные величины, связанные в цепь Маркова.

В настоящей главе дается общее определение стохастической последовательности случайных величин, связанных марковской зависимостью, и основное внимание уделяется изучению асимптотических свойств марковских цепей со счетным множеством состояний.

2. Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство с выделенным на нем неубывающим семейством σ-алгебр $(\mathscr{F}_n)$, $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \subseteq \ldots \subseteq \mathscr{F}$.

Определение. Стохастическая последовательность $X = (X_n, \mathscr{F}_n)$ называется *марковской цепью* или *цепью Маркова* (по отношению к мере P), если для любых $n \geqslant m \geqslant 0$ и любого $B \in \mathscr{B}(R)$

$$\mathsf{P}\{X_n \in B \mid \mathscr{F}_m\} = \mathsf{P}\{X_n \in B \mid X_m\} \quad (\mathsf{P}\text{-п. н.}). \tag{1}$$

Свойство (1), называемое *марковским свойством*, допускает различные эквивалентные формулировки.

Так, (1) равносильно тому, что для любой ограниченной борелевской функции $g = g(x)$

$$\mathsf{M}[g(X_n) \mid \mathscr{F}_m] = \mathsf{M}[g(X_n) \mid X_m] \quad (\mathsf{P}\text{-п. н.}). \tag{2}$$

Свойство (1) эквивалентно также тому, что при фиксированном «настоящем» X_m «будущее» Б и «прошлое» П независимы, т. е.

$$\mathsf{P}(\text{БП} \mid X_m) = \mathsf{P}(\text{Б} \mid X_m)\,\mathsf{P}(\text{П} \mid X_m), \tag{3}$$

где событие $\text{Б} \in \sigma\{\omega\colon X_i,\ i \geqslant m\}$, а событие $\text{П} \in \mathscr{F}_m$, $m \leqslant n$.

В том частном случае, когда

$$\mathscr{F}_n=\mathscr{F}_n^X=\sigma\{\omega: X_0, \ldots, X_n\}$$

и стохастическая последовательность $X=(X_n, \mathscr{F}_n^X)$ образует марковскую цепь, принято говорить, что сама *последовательность* (X_n) *является марковской цепью.* В этой связи полезно отметить, что если $X=(X_n, \mathscr{F}_n)$ — цепь Маркова, то (X_n) также есть марковская цепь.

Замечание. В данном выше определении предполагалось, что величины X_n принимают действительные значения. Аналогичным образом дается определение марковской цепи и в том случае, когда величины X_n принимают значения в некотором измеримом пространстве $(E, \mathscr{E})$. При этом, если все одноточечные множества измеримы, то это пространство называют *фазовым* и говорят, что $X=(X_n, \mathscr{F}_n)$ — марковская цепь со значениями в фазовом пространстве $(E, \mathscr{E})$. В том случае, когда E — конечное или счетное множество (и $\mathscr{E}$ — σ-алгебра всех его подмножеств), марковские цепи называют *дискретными.* В свою очередь дискретные цепи с конечным фазовым пространством называют *конечными* цепями.

Изложение теории конечных марковских цепей, данное в § 12 гл. I, показывает, что в их исследовании особо важную роль играют переходные вероятности $\mathsf{P}(X_{n+1}\in B\,|\,X_n)$ за один шаг. В силу теоремы 3 из § 7 гл. II существуют функции $P_{n+1}(x; B)$ — *регулярные условные вероятности,* являющиеся при фиксированном x мерами на $(R, \mathscr{B}(R))$ и при фиксированном B измеримыми функциями по x, такие, что

$$\mathsf{P}(X_{n+1}\in B\,|\,X_n)=P_{n+1}(X_n; B) \quad (\mathsf{P}\text{-п. н.}). \tag{4}$$

Функции $P_n=P_n(x; B)$, $n\geqslant 0$, называют *переходными функциями* и в том случае, когда они совпадают ($P_1=P_2=\ldots$), соответствующую марковскую цепь X принято называть *однородной* (по времени).

Все дальнейшие рассмотрения будут вестись лишь для однородных марковских цепей, а переходная функция $P_1=P_1(x; B)$ будет обозначаться просто $P=P(x; B)$.

Наряду с переходной функцией важной вероятностной характеристикой марковской цепи является *начальное распределение* $\pi=\pi(B)$, т. е. распределение вероятностей, определяемое равенством $\pi(B)=\mathsf{P}\{X_0\in B\}$.

Набор объектов (π, P), где π — начальное распределение, а P — переходная функция, полностью определяет вероятностные свойства последовательности X, поскольку все конечномерные распределения выражаются (задача 2) через π и P: для любого $n \geqslant 0$ и $A \in \mathscr{B}(R^{n+1})$

$$\mathsf{P}\{(X_0, \ldots, X_n) \in A\} = \\ = \int\limits_R \pi(dx_0) \int\limits_R P(x_0; dx_1) \ldots \int\limits_R I_A(x_0, \ldots, x_n) P(x_{n-1}; dx_n). \quad (5)$$

Отсюда стандартным предельным переходом выводится, что для любой $\mathscr{B}(R^{n+1})$-измеримой функции (одного знака или ограниченной) $g(x_0, \ldots, x_n)$

$$\mathsf{M}g(X_0, \ldots, X_n) = \\ = \int\limits_R \pi(dx_0) \int\limits_R P(x_0; dx_1) \ldots \int\limits_R g(x_0, \ldots, x_n) P(x_{n-1}; dx_n). \quad (6)$$

3. Обозначим через $P^{(n)} = P^{(n)}(x; B)$ — регулярный вариант *переходной вероятности за n шагов*:

$$\mathsf{P}(X_n \in B \mid X_0) = P^n(X_0; B) \quad (\mathsf{P}\text{-п. н.}). \quad (7)$$

Из марковского свойства непосредственно выводится, что для любых k, $l \geqslant 1$ (P-п. н.)

$$P^{(k+l)}(X_0; B) = \int\limits_R P^{(k)}(X_0; dy) P^{(l)}(y; B). \quad (8)$$

Отсюда *не* следует, конечно, что тогда для *всех* $x \in R$

$$P^{(k+l)}(x; B) = \int\limits_R P^{(k)}(x; dy) P^{(l)}(y; B). \quad (9)$$

Оказывается, однако, что регулярные варианты переходных вероятностей можно выбрать так (см. по этому поводу соответствующее место в историко-библиографической справке), что свойство (9) будет выполнено для *всех* $x \in R$.

Уравнение (9) носит название *уравнения Колмогорова — Чэпмена* (ср. с (I.12.13)) и служит отправным моментом при исследовании вероятностных свойств марковских цепей.

4. Как следует из вышеизложенного, каждой марковской цепи $X = (X_n, \mathscr{F}_n)$, заданной на $(\Omega, \mathscr{F}, \mathsf{P})$, сопоставляется набор (π, P). Естественно поставить вопрос о том, каким условиям должен удовлетворять набор (π, P), где $\pi = \pi(B)$ есть распределение вероятностей на $(R, \mathscr{B}(R))$, а $P = P(x; B)$ — функция, являющаяся измеримой по x при фиксированном B и вероятностной мерой по B при каждом x, чтобы π было начальным распределением, а P — переходной функцией некоторой марковской цепи. Как сейчас будет показано, для этого никаких дополнительных условий налагать не требуется.

В самом деле, возьмем в качестве $(\Omega, \mathscr{F})$ измеримое пространство $(R^\infty, \mathscr{B}(R^\infty))$ и определим на множествах $A \in \mathscr{B}(R^{n+1})$

вероятностную меру с помощью выражения, стоящего в правой части формулы (5). Как следует из § 9 гл. II, на $(R^\infty, \mathscr{B}(R^\infty))$ существует вероятностная мера P такая, что

$$\mathsf{P}\{\omega\colon (x_0, \ldots, x_n) \in A\} = \\ = \int_R \pi(dx_0) \int_R P(x_0;\ dx_1) \ldots \int_R I_A(x_0, \ldots, x_n) P(x_{n-1};\ dx_n). \tag{10}$$

Покажем, что если для $\omega = (x_0, x_1, \ldots)$ положить $X_n(\omega) = x_n$, то последовательность $X = (X_n)_{n \geqslant 0}$ будет образовывать (по отношению к построенной мере P) марковскую цепь.

Действительно, если $B \in \mathscr{B}(R)$, $C \in \mathscr{B}(R^{n+1})$, то

$$\mathsf{P}\{X_{n+1} \in B,\ (X_0, \ldots, X_n) \in C\} = \int_R \pi(dx_0) \int_R P(x_0;\ dx_1) \ldots \\ \ldots \int_R I_B(x_{n+1}) I_C(x_0, \ldots, x_n) P(x_n;\ dx_{n+1}) = \\ = \int_R \pi(dx_0) \int_R P(x_0;\ dx_1) \ldots \int_R P(x_n;\ B) I_C(x_0, \ldots, x_n) P(x_{n-1};\ dx_n) = \\ = \int_{\{\omega\colon (X_0, \ldots, X_n) \in C\}} P(X_n;\ B)\, d\mathsf{P},$$

откуда (P-п. н.)

$$\mathsf{P}\{X_{n+1} \in B \mid X_0, \ldots, X_n\} = P(X_n;\ B). \tag{11}$$

Аналогичным образом проверяется, что (P-п. н.)

$$\mathsf{P}\{X_{n+1} \in B \mid X_n\} = P(X_n;\ B). \tag{12}$$

Из (11) и (12) вытекает требуемое равенство (1). Точно так же доказывается, что (P-п. н.) для любого $k \geqslant 1$ и $n \geqslant 0$

$$\mathsf{P}\{X_{n+k} \in B \mid X_0, \ldots, X_n\} = \mathsf{P}\{X_{n+k} \in B \mid X_n\}.$$

Отсюда следует однородность марковской цепи.

Построенная марковская цепь $X = (X_n)$ называется марковской цепью, порожденной парой (π, P). При этом, чтобы подчеркнуть, что построенная на $(R^\infty, \mathscr{B}(R^\infty))$ мера P отвечает именно начальному распределению π, ее часто обозначают через P_π.

Если мера π сосредоточена в одной точке $\{x\}$, то вместо P_π пишут P_x и соответствующую марковскую цепь называют цепью, *выходящей из точки* x (поскольку $\mathsf{P}_x\{X_0 = x\} = 1$).

Таким образом, с каждой переходной функцией $P = P(x;\ B)$ связывается на самом деле *целое семейство вероятностных мер* $\{\mathsf{P}_x,\ x \in R\}$, а, значит, и целое семейство марковских цепей, возникающих, когда последовательность $(X_n)_{n \geqslant 0}$ рассматривается относительно мер P_x, $x \in R$. В дальнейшем под словами «марковкая цепь с заданной переходной функцией» будем понимать именно семейство марковских цепей в указанном смысле.

Заметим, что построенные по переходной функции $P=P(x;\ B)$ меры P_π и P_x согласованы в том смысле, что для $A\in\mathscr{B}(R^\infty)$

$$\mathsf{P}_\pi\{(X_0, X_1, \ldots)\in A\,|\,X_0=x\}=\mathsf{P}_x\{(X_0,\ X_1,\ \ldots)\in A\}\ (\pi\text{-п. н.}) \tag{13}$$

и

$$\mathsf{P}_\pi\{(X_0,\ X_1, \ldots)\in A\}=\int\limits_R \mathsf{P}_x\{(X_0,\ X_1,\ \ldots)\in A\}\,\pi(dx). \tag{14}$$

5. Будем предполагать, что $(\Omega,\ \mathscr{F})=(R^\infty,\ \mathscr{B}(R^\infty))$ и что рассматриваемые последовательности $X=(X_n)$ заданы координатным образом, т. е. $X_n(\omega)=x_n$ для $\omega=(x_0,\ x_1,\ \ldots)$. Пусть также $\mathscr{F}_n=\sigma\{\omega\colon X_0, \ldots,\ X_n\}$, $n\geqslant 0$.

Определим на Ω *операторы сдвига* θ_n, $n\geqslant 0$, с помощью равенства

$$\theta_n(x_0,\ x_1,\ \ldots)=(x_n,\ x_{n+1},\ \ldots),$$

и для каждой случайной величины $\eta=\eta(\omega)$ определим случайные величины $\theta_n\eta$, полагая

$$(\theta_n\eta)(\omega)=\eta(\theta_n\omega).$$

Используя эти обозначения, марковскому свойству однородных цепей можно придать (задача 1) следующую форму: для любой $\mathscr{F}$-измеримой случайной величины $\eta=\eta(\omega)$, любого $n\geqslant 0$ и $B\in\mathscr{B}(R)$

$$\mathsf{P}\{\theta_n\eta\in B\,|\,\mathscr{F}_n\}=\mathsf{P}_{X_n}\{\eta\in B\}\quad(\mathsf{P}\text{-п. н.}). \tag{15}$$

Именно эта форма марковского свойства допускает важное обобщение, состоящее в том, что соотношение (15) останется справедливым, если вместо n рассмотреть моменты остановки τ.

Теорема. *Пусть $X=(X_n)$ — однородная марковская цепь, заданная на $(R^\infty,\ \mathscr{B}(R^\infty),\ \mathsf{P})$ и τ — момент остановки. Тогда справедливо следующее строго марковское свойство:*

$$\mathsf{P}\{\theta_\tau\eta\in B\,|\,\mathscr{F}_\tau\}=\mathsf{P}_{X_\tau}\{\eta\in B\}\quad(\mathsf{P}\text{-п. н.}). \tag{16}$$

Доказательство. Если $A\in\mathscr{F}_\tau$, то

$$\mathsf{P}\{\theta_\tau\eta\in B,\ A\}=\sum_{n=0}^{\infty}\mathsf{P}\{\theta_\tau\eta\in B,\ A,\ \tau=n\}=$$
$$=\sum_{n=0}^{\infty}\mathsf{P}\{\theta_n\eta\in B,\ A,\ \tau=n\}. \tag{17}$$

Событие $A\cap\{\tau=n\}\in\mathscr{F}_n$ и, значит,

$$\mathsf{P}\{\theta_n\eta\in B,\ A\cap\{\tau=n\}\}=\int\limits_{A\cap\{\tau=n\}}\mathsf{P}\{\theta_n\eta\in B\,|\,\mathscr{F}_n\}\,d\mathsf{P}=$$
$$=\int\limits_{A\cap\{\tau=n\}}\mathsf{P}_{X_n}\{\eta\in B\}\,d\mathsf{P}=\int\limits_{A\cap\{\tau=n\}}\mathsf{P}_{X_\tau}\{\eta\in B\}\,d\mathsf{P},$$

что вместе с (17) доказывает (16).

Следствие. Если σ — момент остановки такой, что $\mathrm{P}(\sigma \geqslant \tau)=1$ и σ — $\mathscr{F}_\tau$-измерим, то

$$\mathrm{P}\{X_\sigma \in B,\ \sigma<\infty \mid \mathscr{F}_\tau\}=\mathrm{P}_{X_\tau}(B) \quad (\{\sigma<\infty\};\ \mathrm{P}\text{-п. н.}). \tag{18}$$

6. Как уже отмечалось, в дальнейшем будут рассматриваться лишь дискретные цепи Маркова (с фазовым пространством состояний $E=\{\ldots, i, j, k, \ldots\}$). Для простоты записи будем в этом случае обозначать переходные функции $P(i; \{j\})$ через p_{ij} и называть их переходными вероятностями, а вероятности перехода из i в j за n шагов обозначать через $p_{ij}^{(n)}$.

Основные вопросы, которые будут изучаться в §§ 2—4, связаны с выяснением условий, при которых ($E=\{1, 2, \ldots\}$):

A) *Существуют пределы* $\pi_j=\lim\limits_n p_{ij}^{(n)}$, не зависящие от i;

B) Пределы $(\pi_1, \pi_2, \ldots)$ образуют *распределение вероятностей*, т. е. $\pi_i \geqslant 0$, $\sum\limits_{i=1}^{\infty} \pi_i=1$;

C) Цепь является *эргодической*, т. е. пределы $(\pi_1, \pi_2, \ldots)$ таковы, что $\pi_i>0$, $\sum\limits_{i=1}^{\infty} \pi_i=1$;

D) Существует и притом единственное *стационарное распределение вероятностей* $\mathbb{Q}=(q_1, q_2, \ldots)$, т. е. такое, что $q_i \geqslant 0$, $\sum\limits_{i=1}^{\infty} q_i=1$ и $q_j=\sum\limits_i q_i p_{ij}$, $j \in E$.

Для получения ответа на эти вопросы проведем классификацию состояний марковской цепи в зависимости от арифметических и асимптотических свойств вероятностей $p_{ij}^{(n)}$ и $p_{ii}^{(n)}$.

7. Задачи.

1. Доказать эквивалентность определений марковости (1), (2), (3) и (15).
2. Доказать справедливость формулы (5).
3. Доказать соотношение (18).
4. Пусть $(X_n)_{n \geqslant 0}$ — марковская цепь. Показать, что обращенная последовательность $(\ldots X_n, X_{n-1}, \ldots, X_0)$ также образует цепь Маркова.

§ 2. Классификация состояний марковской цепи по арифметическим свойствам переходных вероятностей $p_{ij}^{(n)}$

1. Будем называть состояние $i \in E=\{1, 2, \ldots\}$ *несущественным*, если из него с положительной вероятностью можно за конечное число шагов выйти, но нельзя в него вернуться, т. е. существуют такие m и j, что $p_{ij}^{(m)}>0$, но для всех n и j $p_{ji}^{(n)}=0$.

Выделим из множества E все несущественные состояния. Тогда оставшееся множество *существенных* состояний обладает тем свойством, что, попав в него, блуждающая «частица» никогда из него не выйдет (рис. 36). Как станет ясно из дальнейшего, основной интерес представляют именно существенные состояния.

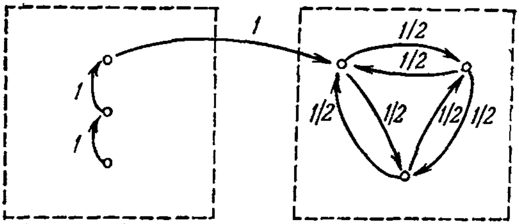


Рис. 36.

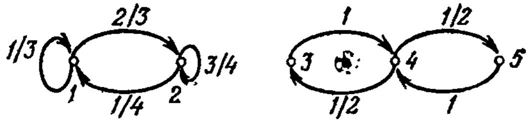
Рассмотрим сейчас множество существенных состояний. Назовем состояние j *достижимым* из точки i ($i \to j$), если существует такое $m \geqslant 0$, что $p_{ij}^{(m)} > 0$ ($p_{ij}^{(0)} = 1$, если $i = j$, и 0, если $i \neq j$). Состояния i и j назовем *сообщающимися* ($i \leftrightarrow j$), если j достижимо из i и i достижимо из j.

По самому определению отношение «$\leftrightarrow$» является симметричным и рефлексивным. Нетрудно убедиться, что оно транзитивно ($i \leftrightarrow j$, $j \leftrightarrow k \Rightarrow i \leftrightarrow k$). Следовательно, множество существенных состояний разбивается на конечное или счетное число непересекающихся множеств E_1, E_2, ..., состоящих из сообщающихся состояний и характеризующихся тем, что переходы между различными множествами невозможны.

Для краткости множества E_1, E_2, ... будем называть *классами* или *неразложимыми классами* (существенных сообщающихся состояний), а марковскую цепь, состояния которой образуют один неразложимый класс, назовем *неразложимой*.

Для иллюстрации введенных понятий рассмотрим цепь с матрицей

Граф этой цепи с множеством состояний $E=\{1, 2, 3, 4, 5\}$ имеет следующий вид:



Ясно, что у рассматриваемой цепи есть два неразложимых класса $E_1=\{1, 2\}$, $E_2=\{3, 4, 5\}$, и исследование ее свойств сводится к исследованию свойств каждой из двух цепей, состояниями которых являются множества E_1 и E_2, а матрицы переходных вероятностей равны соответственно $\mathbb{P}_1$ и $\mathbb{P}_2$.

Рассмотрим теперь какой-нибудь неразложимый класс E. Для примера пусть им будет класс, изображенный на рис. 37.

Заметим, что здесь возвращение в каждое состояние возможно лишь за четное число шагов, переход в соседнее состояние — за нечетное число шагов, а матрица переходных вероятностей имеет блочную структуру:

$$\mathbb{P}=\begin{pmatrix} 0 & 0 & \vdots & ^1/_2 & ^1/_2 \\ 0 & 0 & \vdots & ^1/_2 & ^1/_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ ^1/_2 & ^1/_2 & \vdots & 0 & 0 \\ ^1/_2 & ^1/_2 & \vdots & 0 & 0 \end{pmatrix}.$$

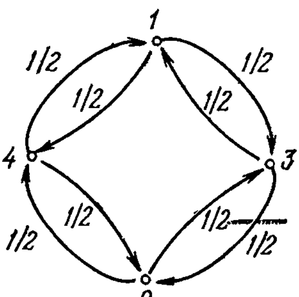


Рис. 37. Пример марковский цепи с периодом $d=2$.

Отсюда видно, что класс $E=\{1, 2, 3, 4\}$ разбивается на два подкласса $C_0=\{1, 2\}$ и $C_1=\{3, 4\}$, обладающих следующим свойством *цикличности*: за один шаг из C_0 частица непременно переходит в C_1, а из C_1 — в C_0.

Этот пример подсказывает классификацию неразложимых классов на *циклические подклассы*.

2. Будем говорить, что состояние j имеет период $d=d(j)$, если выполнены следующие два условия:

1) $p_{jj}^{(n)}>0$ только для тех n, которые имеют вид $n=dm$;

2) d есть наибольшее из чисел, обладающих свойством 1).

Иначе говоря, d есть *общий наибольший делитель* чисел n таких, что $p_{jj}^{(n)}>0$. (Если $p_{jj}^{(n)}=0$ для всех $n\geqslant 1$, то полагаем $d(j)=0$.)

Покажем, что все состояния одного неразложимого класса E имеют один и тот же период d, который поэтому естественно назвать периодом этого класса, $d=d(E)$.

Пусть $i, j\in E$. Тогда найдутся такие k и l, что $p_{ij}^{(k)}>0$, $p_{ji}^{(l)}>0$. Поэтому $p_{ii}^{(k+l)}\geqslant p_{ij}^{(k)}p_{ji}^{(l)}>0$ и, значит, $k+l$ делится на $d(i)$. Предположим, что $n>0$ и n не делится на $d(i)$. Тогда $n+k+l$ также не делится на $d(i)$ и, следовательно, $p_{ii}^{(n+k+l)}=0$.

Но

$$p_{ii}^{(n+k+l)} \geqslant p_{ij}^{(k)} p_{jj}^{(n)} p_{ji}^{(l)}$$

и, значит, $p_{jj}^{(n)}=0$. Отсюда вытекает, что если $p_{jj}^{(n)}>0$, то n должно делиться на $d(i)$, а поэтому $d(i) \leqslant d(j)$. В силу симметрии $d(j) \leqslant d(i)$. Следовательно, $d(i)=d(j)$.

Если $d(j)=1$ $(d(E)=1)$, то состояние j (класс E) будем называть *апериодическим*.

Пусть $d=d(E)$ — период неразложимого класса E. Переходы внутри такого класса могут осуществляться весьма причудливым образом, однако (как и в рассмотренном выше примере) имеет место некоторая цикличность в переходах из одной группы состояний в другую. Чтобы это показать, зафиксируем некоторое состояние i_0 и введем (для $d \geqslant 1$) следующие подклассы:

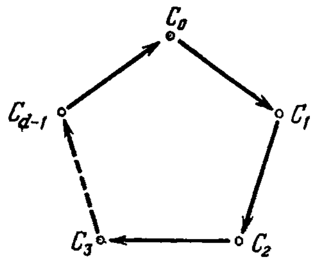


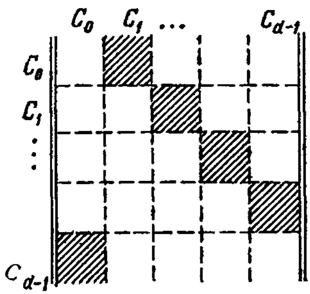
Рис. 38. Движение по циклическим подклассам.

$$C_0=\{j \in E \;:\; p_{i_0 j}^{(n)}>0 \Rightarrow n \equiv 0 \pmod d\};$$
$$C_1=\{j \in E \;:\; p_{i_0 j}^{(n)}>0 \Rightarrow n \equiv 1 \pmod d\};$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$C_{d-1}=\{j \in E \;:\; p_{i_0 j}^{(n)}>0 \Rightarrow n \equiv d-1 \pmod d\}.$$

Ясно, что $E=C_0+C_1+\ldots+C_{d-1}$. Покажем, что движение из подкласса в подкласс осуществляется так, как это изображено на рис. 38.

В самом деле, пусть состояние $i \in C_p$ и $p_{ij}>0$. Покажем, что тогда непременно $j \in C_{p+1 \pmod d}$. Пусть n таково, что $p_{i_0 i}^{(n)}>0$. Тогда $n=ad+p$, а значит, $n \equiv p \pmod d$ и $n+1 \equiv p+1 \pmod d$. Отсюда $p_{i_0 j}^{(n+1)}>0$ и $j \in C_{p+1 \pmod d}$.

Заметим, что из приведенных рассуждений следует, что матрица $\mathbb{P}$ переходных вероятностей неразложимой цепи имеет блочную структуру:



Рассмотрим некоторый подкласс C_p. Если считать, что в начальный момент частица находится во множестве C_0, то в моменты $s=p+dt$, $t=0, 1, \ldots$, она будет находиться в подклассе C_p. Следовательно, с каждым подклассом C_p можно связать новую марковскую цепь с матрицей переходных вероятностей $(p_{ij}^d)_{i,j\in C_p}$, которая будет неразложимой и апериодической. Тем самым, принимая во внимание проведенную классификацию (см. сводный

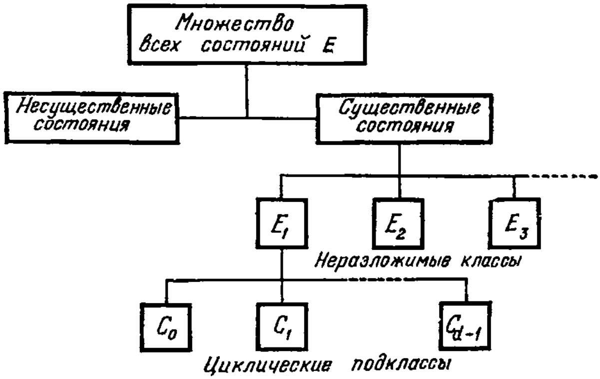


Рис. 39. Классификация состояний марковской цепи по арифметическим свойствам вероятностей $p_{ij}^{(n)}$.

рис. 39), заключаем, что при исследовании вопросов о предельных свойствах вероятностей $p_{ij}^{(n)}$ можно ограничиваться рассмотрением лишь *апериодических неразложимых цепей*.

3. Задачи.

1. Показать, что отношение «$\leftrightarrow$» является транзитивным.

2. Для примера 1, рассмотренного в § 5, показать, что для $0<p<1$ все состояния образуют один класс с периодом $d=2$.

3. Показать, что марковские цепи, рассмотренные в примерах 4 и 5 в § 5, являются апериодическими.

§ 3. Классификация состояний марковской цепи по асимптотическим свойствам вероятностей $p_{ii}^{(n)}$

1. Пусть $\mathbb{P}=\|p_{ij}\|$ — матрица переходных вероятностей марковской цепи $(X_n)_{n\geq 0}$,

$$f_{ii}^{(k)}=\mathrm{P}_i\{X_k=i,\ X_l\neq i,\ 1\leq l\leq k-1\} \tag{1}$$

и для $i\neq j$

$$f_{ij}^{(k)}=\mathrm{P}_i\{X_k=j,\ X_l\neq j,\ 1\leq l\leq k-1\} \tag{2}$$

— вероятность *первого возвращения в состояние* i и вероятность *первого попадания в состояние* j в момент времени k, когда $X_0=i$.

Используя строго марковское свойство (1.16), аналогично (I.12.38) показывается, что

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}. \tag{3}$$

Введем для каждого $i \in E$ величину

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}, \tag{4}$$

по своему смыслу являющуюся вероятностью того, что частица, выходящая из состояния i, рано или поздно вернется в это состояние. Иначе говоря, $f_{ii} = \mathsf{P}_i\{\sigma_i < \infty\}$, где $\sigma_i = \inf\{n \geqslant 1\colon X_n = i\}$ с $\sigma_i = \infty$, когда $\{\cdot\} = \varnothing$.

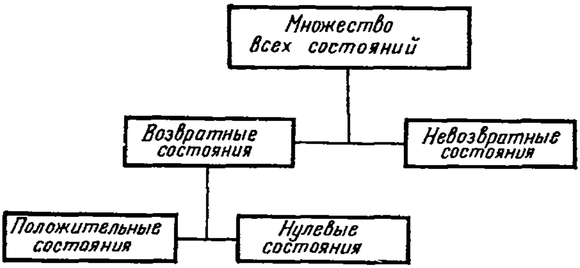


Рис. 40. Классификация состояний марковской цепи по асимптотическим свойствам вероятностей $p_{ii}^{(n)}$.

Назовем состояние i *возвратным*, если

$$f_{ii} = 1,$$

и *невозвратным*, если

$$f_{ii} < 1.$$

Каждое возвратное состояние можно в свою очередь отнести к одному из двух типов в зависимости от конечности или бесконечности *среднего времени возвращения*.

А именно, будем называть возвратное состояние i *положительным*, если

$$\mu_i^{-1} \equiv \left(\sum_{n=1}^{\infty} n f_{ii}^{(n)} \right)^{-1} > 0,$$

и *нулевым*, если

$$\mu_i^{-1} \equiv \left(\sum_{n=1}^{\infty} n f_{ii}^{(n)} \right)^{-1} = 0.$$

Итак, в зависимости от свойств вероятностей $p_{ii}^{(n)}$ получаем классификацию состояний цепи, изображенную на рис. 40.

2. Поскольку отыскание функций $f_{ii}^{(n)}$ довольно-таки сложно, то для определения того, является состояние i возвратным или невозвратным, полезен следующий критерий.

Лемма 1. a) *Состояние i возвратно тогда и только тогда, когда*

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty. \tag{5}$$

b) *Если состояние j возвратно и $i \leftrightarrow j$, то состояние i также возвратно.*

Доказательство. a) В силу (3)

$$p_{ii}^{(n)} = \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)},$$

и, значит, ($p_{ii}^{(0)} = 1$)

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} = \sum_{k=1}^{\infty} f_{ii}^{(k)} \sum_{n=k}^{\infty} p_{ii}^{(n-k)} =$$
$$= f_{ii} \sum_{n=0}^{\infty} p_{ii}^{(n)} = f_{ii} \left(1 + \sum_{n=1}^{\infty} p_{ii}^{(n)} \right).$$

Поэтому, если $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$, то $f_{ii} < 1$ и, значит, состояние i невозвратно. Далее, пусть $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$. Тогда

$$\sum_{n=1}^{N} p_{ii}^{(n)} = \sum_{n=1}^{N} \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} = \sum_{k=1}^{N} f_{ii}^{(k)} \sum_{n=1}^{N} p_{ii}^{(n-k)} \leqslant \sum_{k=1}^{N} f_{ii}^{(k)} \sum_{l=0}^{N} p_{ii}^{(l)},$$

и поэтому

$$f_{ii} = \sum_{k=1}^{\infty} f_{ii}^{(k)} \geqslant \sum_{k=1}^{N} f_{ii}^{(k)} \geqslant \frac{\sum_{n=1}^{N} p_{ii}^{(n)}}{\sum_{l=0}^{N} p_{ii}^{(l)}} \to 1, \qquad N \to \infty.$$

Итак, если $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, то $f_{ii} = 1$, т. е. состояние i возвратно.

b) Пусть $p_{ij}^{(s)} > 0$, $p_{ji}^{(t)} > 0$. Тогда

$$p_{ii}^{(n+s+t)} \geqslant p_{ij}^{(s)} p_{jj}^{(n)} p_{ji}^{(t)},$$

и если $\sum_{j} p_{jj}^{(n)} = \infty$, то и $\sum_{i} p_{ii}^{(n)} = \infty$, т. е. состояние i возвратно.

3. Из критерия (5) легко выводится следующий первый результат об асимптотическом поведении вероятностей $p_{ij}^{(n)}$.

Лемма 2. *Если состояние j невозвратно, то для любого i*

$$\sum p_{ij}^{(n)} < \infty \tag{6}$$

и, значит,

$$p_{ij}^{(n)} \to \infty, \quad n \to \infty. \tag{7}$$

Доказательство. Из (3) и леммы 1

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \sum_{n=0}^{\infty} p_{jj}^{(n)} =$$
$$= f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)} \leqslant \sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty,$$

где мы воспользовались тем, что $f_{ij} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \leqslant 1$, поскольку это есть вероятность того, что частица, вышедшая из i, рано или поздно попадет в j. Итак, (6), а значит, и (7) доказаны.

Перейдем теперь к случаю возвратных состояний.

Лемма 3. *Пусть j является возвратным состоянием с $d(j) = 1$.*

a) *Если i сообщается с j, то*

$$p_{ij}^{(n)} \to \frac{1}{\mu_j}, \quad n \to \infty. \tag{8}$$

Если к тому же j является положительным состоянием, то

$$p_{ij}^{(n)} \to \frac{1}{\mu_j} > 0, \quad n \to \infty. \tag{9}$$

Если же j является нулевым, то

$$p_{ij}^{(n)} \to 0, \quad n \to \infty. \tag{10}$$

b) *Если i и j принадлежат разным классам сообщающихся состояний, то*

$$p_{ij}^{(n)} \to \frac{f_{ij}}{\mu_i}, \quad n \to \infty. \tag{11}$$

Доказательство леммы будет опираться на следующий результат из анализа.

Пусть $f_1, f_2, \ldots$ — последовательность неотрицательных чисел с $\sum_{i=1}^{\infty} f_i = 1$ такая, что общий наибольший делитель тех чисел j, для которых $f_j > 0$, равен единице. Пусть $u_0 = 1$, $u_n = \sum_{k=1}^{n} f_k u_{n-k}$, $n = 1, 2, \ldots$, и $\mu = \sum_{n=1}^{\infty} n f_n$. Тогда $u_n \to \frac{1}{\mu}$, $n \to \infty$ (доказательство см., например, в [69] § 10 гл. XIII).

Учитывая соотношения (3), применим этот результат к $u_n = p_{jj}^{(n)}$, $f_k = f_{jj}^{(k)}$. Тогда сразу получаем, что

$$p_{jj}^{(n)} \to \frac{1}{\mu_j},$$

где $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$.

Перепишем теперь соотношение (3) в виде $(p_{jj}^{(s)} = 0,\ s < 0)$

$$p_{ij}^{(n)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} p_{jj}^{(n-k)}. \tag{12}$$

Согласно доказанному для каждого фиксированного k $p_{jj}^{(n-k)} \to \mu_j^{-1}$, $n \to \infty$. Поэтому, если предположить, что

$$\lim_n \sum_{k=1}^{\infty} f_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \lim_n p_{jj}^{(n-k)}, \tag{13}$$

то тогда сразу получим

$$p_{ij}^{(n)} \to \frac{1}{\mu_j} \left(\sum_{k=1}^{\infty} f_{ij}^{(k)} \right) = \frac{1}{\mu_j} f_{ij}, \tag{14}$$

что и доказывает (11).

По своему смыслу f_{ij} есть вероятность того, что частица, вышедшая из состояния i, рано или поздно попадет в состояние j. Состояние j возвратно, и если i сообщается с j, то естественно ожидать, что тогда $f_{ij} = 1$. Покажем, что это действительно так.

Пусть f'_{ij} — вероятность того, что частица, вышедшая из состояния i, бесконечно много раз побывает в состоянии j. Понятно, что $f_{ij} \geqslant f'_{ij}$. Поэтому, если показать, что для возвратного состояния j и сообщающегося с ним состояния i вероятность $f'_{ij} = 1$, то требуемое равенство $f_{ij} = 1$ будет установлено.

Согласно утверждению b) леммы 1 состояние i также возвратно и, значит,

$$f_{ii} = \sum f_{ii}^{(n)} = 1. \tag{15}$$

Пусть

$$\sigma_i = \inf\{n \geqslant 1:\ X_n = i\}$$

— момент первого (в моменты $n \geqslant 1$) попадания частицы в состояние i; полагаем $\sigma_i = \infty$, если такого момента не существует.

Тогда

$$1 = f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = \sum_{n=1}^{\infty} \mathsf{P}_i(\sigma_i = n) = \mathsf{P}_i(\sigma_i < \infty), \tag{16}$$

и, следовательно, возвратность состояния i означает, что частица, вышедшая из i, рано или поздно снова вернется в это состояние (в случайный момент σ_i). Но после возвращения в это состояние «жизнь» частицы как бы начинается сначала (в силу справедливости строго марковского свойства). Отсюда напрашивается вывод, что если состояние i возвратно, то частица будет попадать в него бесконечно часто:

$$\mathsf{P}_i\{X_n = i \text{ для бесконечно многих } n\} = 1. \tag{17}$$

Дадим теперь формальное доказательство этого утверждения.

Пусть i — какое-то состояние (возвратное или невозвратное). Покажем, что вероятность возвращения в него по крайней мере r раз равна $(f_{ii})^r$.

Для $r = 1$ это следует из определения f_{ii}. Пусть утверждение доказано для $r = m - 1$. Тогда, используя строго марковское свойство и (16), находим, что

$$\begin{aligned}
&\mathsf{P}_i\,(\text{число возвращений в } i \text{ больше или равно } m) = \\
&= \sum_{k=1}^{\infty} \mathsf{P}_i\begin{pmatrix}\sigma_i = k, & \text{число возвращений в } i \text{ после момента } k\\ & \text{больше или равно } m-1\end{pmatrix} = \\
&= \sum_{k=1}^{\infty} \mathsf{P}_i(\sigma_i = k)\, \mathsf{P}_i\left(\begin{matrix}\text{по крайней мере } (m-1) \text{ значение}\\ \text{из } X_{\sigma_i+1}, X_{\sigma_i+2}, \ldots \text{ равно } i\end{matrix}\;\middle|\; \sigma_i = k\right) = \\
&= \sum_{k=1}^{\infty} \mathsf{P}_i(\sigma_i = k)\, \mathsf{P}_i\begin{pmatrix}\text{по крайней мере } (m-1) \text{ значение}\\ \text{из } X_1, X_2, \ldots \text{ равно } i\end{pmatrix} = \\
&= \sum_{k=1}^{\infty} f_{ii}^{(k)} (f_{ii})^{m-1} = f_{ii}^m.
\end{aligned}$$

Отсюда, в частности, следует, что для возвратного состояния i справедлива формула (17). Если же состояние невозвратно, то

$$\mathsf{P}_i\{X_n = i \text{ для бесконечно многих } n\} = 0. \tag{18}$$

Перейдем теперь к доказательству того, что $f_{ij} = 1$. Поскольку состояние i возвратно, то в силу (17) и строго марковского свойства

$$1=\sum_{k=1}^{\infty}\mathsf{P}_i(\sigma_j=k)+\mathsf{P}_i(\sigma_j=\infty)=$$

$$=\sum_{k=1}^{\infty}\mathsf{P}_i\begin{pmatrix}\sigma_j=k, & \text{число возвращений в } i\\ & \text{после момента } k \text{ равно}\\ & \text{бесконечности}\end{pmatrix}+\mathsf{P}_i(\sigma_j=\infty)=$$

$$=\sum_{k=1}^{\infty}\mathsf{P}_i\begin{pmatrix}\sigma_j=k, & \text{бесконечно много значений}\\ & \text{из } X_{\sigma_j+1},\ X_{\sigma_j+2},\ \ldots \text{ равно } i\end{pmatrix}+\mathsf{P}_i(\sigma_j=\infty)=$$

$$=\sum_{k=1}^{\infty}P_i(\sigma_j=k)\cdot P_i\left(\begin{matrix}\text{бесконечно много зна-}\\ \text{чений из } X_{\sigma_j+1},\ X_{\sigma_j+2},\ldots\\ \ldots\text{равно } i\end{matrix}\,\middle|\,\begin{matrix}\sigma_j=k\\ X_{\sigma_j}=\\ =j\end{matrix}\right)+\mathsf{P}_i(\sigma_j=\infty)=$$

$$=\sum_{k=1}^{\infty}f_{ij}^{(k)}\cdot\mathsf{P}_j\begin{pmatrix}\text{бесконечно много значений}\\ \text{из } X_1,\ X_2,\ \ldots \text{ равно } i\end{pmatrix}+(1-f_{ij})=$$

$$=\sum_{k=1}^{\infty}f_{ij}^{(k)}f'_{ij}+(1-f_{ij})=f'_{ij}f_{ij}+(1-f_{ij}).$$

Итак,

$$1=f'_{ij}f_{ij}+1-f_{ij}$$

и, значит,

$$f_{ij}=f'_{ij}\cdot f_{ij}.$$

Поскольку $i\leftrightarrow j$, то $f_{ij}>0$, а следовательно, $f'_{ij}=1$ и $f_{ij}=1$.

Таким образом, в предположении (13) из (14) и равенства $f_{ij}=1$ следует, что для сообщающихся состояний i и j

$$p_{ij}^{(n)}\to\frac{1}{\mu_j},\qquad n\to\infty.$$

Что же касается равенства (13), то его справедливость следует из теоремы о мажорируемой сходимости и того замечания, что $p_{jj}^{(n-k)}\to\frac{1}{\mu_j}$, $n\to\infty$, $\sum\limits_{k=1}^{\infty}f_{ij}^{(k)}=f_{ij}\leqslant 1$.

Лемма доказана.

Перейдем теперь к рассмотрению случая периодических состояний.

Лемма 4. *Пусть j — возвратное состояние и $d(j)>1$.*

a) *Если состояния i и j принадлежат одному классу состояний, при этом i принадлежит циклическому подклассу C_r, а $j \in C_{r+a}$, то*

$$p_{ij}^{(nd+a)} \to \frac{d}{\mu_j}. \tag{19}$$

b) *Если же i произвольно, то*

$$p_{ij}^{(nd+a)} \to \left[\sum_{r=0}^{\infty} f_{ij}^{(rd+a)}\right] \cdot \frac{d}{\mu_j}, \quad a=0,\ 1, \ldots, d-1. \tag{20}$$

Доказательство. a) Пусть сначала $a=0$. Относительно матрицы переходных вероятностей $\mathbb{P}^d$ состояние j возвратно и апериодично. Следовательно, согласно (8)

$$p_{ij}^{(nd)} \to \frac{1}{\sum\limits_{k=1}^{\infty} kf_{jj}^{(kd)}} = \frac{d}{\sum\limits_{k=1}^{\infty} kdf_{jj}^{(kd)}} = \frac{d}{\mu_j}.$$

Предположим, что (19) доказано для $a=r$. Тогда

$$p_{ij}^{(nd+r+1)} = \sum_{k=1}^{\infty} p_{ik} p_{kj}^{(nd+r)} \to \sum_{k=1}^{\infty} p_{ik} \cdot \frac{d}{\mu_j} = \frac{d}{\mu_j}.$$

b) Ясно, что

$$p_{ij}^{(nd+a)} = \sum_{k=1}^{nd+a} f_{ij}^{(k)} p_{jj}^{(nd+a-k)}, \quad a=0,\ 1, \ldots, d-1.$$

Период состояния j равен d, поэтому $p_{jj}^{(nd+a-k)}=0$, за исключением лишь случаев, когда $k-a$ имеет вид $r \cdot d$. Значит,

$$p_{ij}^{(nd+a)} = \sum_{r=0}^{n} f_{ij}^{(rd+a)} p_{jj}^{((n-r)d)}$$

и требуемый результат (20) следует из (19).

Лемма доказана.

Из лемм 2—4 вытекает, в частности, следующий результат о предельном поведении вероятностей $p_{ij}^{(n)}$.

Теорема 1. *Пусть марковская цепь неразложима (т. е. ее состояния образуют один класс существенных сообщающихся состояний) и апериодична.*

Тогда:

a) *если все состояния нулевые или невозвратные, то для всех i и j*

$$p_{ij}^{(n)} \to 0, \qquad n \to \infty; \tag{21}$$

b) *если все состояния j положительны, то для всех i*

$$p_{ij}^{(n)} \to \frac{1}{\mu_j} > 0, \qquad n \to \infty; \tag{22}$$

4. Обсудим результат этой теоремы для случая марковской цепи с конечным числом состояний $E = \{1, 2, \ldots, r\}$. Будем предполагать, что рассматриваемая цепь неразложима и апериодична. Оказывается, что тогда она автоматически *возвратна и положительна*:

$$\begin{pmatrix}\text{неразложимость}\\ d=1\end{pmatrix} \Rightarrow \begin{pmatrix}\text{неразложимость}\\ \text{возвратность}\\ \text{положительность}\\ d=1\end{pmatrix}. \tag{23}$$

Для доказательства предположим, что все состояния невозвратны. Тогда в силу (21) и конечности множества состояний цепи

$$1 = \lim_n \sum_{j=1}^{r} p_{ij}^{(n)} = \sum_{j=1}^{r} \lim_n p_{ij}^{(n)} = 0. \tag{24}$$

Полученное противоречие показывает, что все состояния не могут быть невозвратными. Пусть i_0 — возвратное состояние, а j — произвольное состояние. Поскольку $i_0 \leftrightarrow j$, то по лемме 1 состояние j также возвратно.

Итак, все состояния у апериодической неразложимой цепи возвратны.

Теперь установим, что все возвратные состояния положительны.

Если предположить, что все они нулевые, то тогда снова придем к противоречивому равенству (24). Следовательно, существует по крайней мере одно положительное состояние, скажем i_0. Пусть i — какое-то другое состояние. Поскольку $i \leftrightarrow i_0$, то найдутся s и t такие, что $p_{i_0 i}^{(s)} > 0$, $p_{i i_0}^{(t)} > 0$ и, значит,

$$p_{ii}^{(n+s+t)} \geqslant p_{ii_0}^{(s)} p_{i_0 i_0}^{(n)} p_{i_0 i}^{(t)} \to p_{ii_0}^{(s)} \frac{1}{\mu_{i_0}} \cdot p_{i_0 i}^{(t)} > 0. \tag{25}$$

Поэтому найдется $\varepsilon > 0$ такое, что для всех достаточно больших n $p_{ii}^{(n)} \geqslant \varepsilon > 0$. Но $p_{ii}^{(n)} \to \frac{1}{\mu_i}$, а значит, $\mu_i > 0$. Тем самым импликация (23) доказана.

Обозначим $\pi_j = 1/\mu_j$. Тогда в силу (22) $\pi_j > 0$ и, поскольку $1 = \lim_n \sum_{j=1}^{r} p_{ij}^{(n)} = \sum_{j=1}^{r} \pi_j$, то (апериодическая неразложимая) цепь является эргодической. Понятно, что для всякой эргодической конечной цепи

$$\textit{найдется } n_0 \textit{ такое, что для всех } n \geqslant n_0 \quad \min_{i,j} p_{ij}^{(n)} > 0. \tag{26}$$

В § 12 гл. I было показано, что верно и обратное: из (26) следует эргодичность.

Таким образом, имеют место следующие импликации:

$$\begin{pmatrix}\text{неразложимость}\\ d=1\end{pmatrix} \Leftrightarrow \begin{pmatrix}\text{неразложимость}\\ \text{возвратность}\\ \text{положительность}\\ d=1\end{pmatrix} \Rightarrow \begin{pmatrix}\text{эргодич-}\\ \text{ность}\end{pmatrix} \Leftrightarrow (26)$$

Можно доказать, однако, больше.

Теорема 2. *В случае конечной марковской цепи*

$$\begin{pmatrix}\textit{неразложимость}\\ d=1\end{pmatrix} \Leftrightarrow \begin{pmatrix}\textit{неразложимость}\\ \textit{возвратность}\\ \textit{положительность}\\ d=1\end{pmatrix} \Leftrightarrow \begin{pmatrix}\textit{эргодич-}\\ \textit{ность}\end{pmatrix} \Leftrightarrow (26)$$

Доказательство. Достаточно доказать импликацию

$$(\text{эргодичность}) \Rightarrow \begin{pmatrix}\text{неразложимость}\\ \text{возвратность}\\ \text{положительность}\\ d=1\end{pmatrix}.$$

Неразложимость следует из (26). Что же касается апериодичности, возвратности и положительности, то они справедливы в более общей ситуации (достаточно лишь существования предельного распределения), что доказывается в теореме 2 § 4.

5. Задачи.

1. Рассмотрим неразложимую цепь с множеством состояний 0, 1, 2, ... Для того чтобы она была невозвратной, необходимо и достаточно, чтобы система уравнений $u_j = \sum_i u_i p_{ij}$, $j = 0, 1, \ldots$, имела ограниченное решение такое, что $u_i \not\equiv c$, $i = 0, 1, \ldots$

2. Для того чтобы неразложимая цепь со множеством состояний 0, 1, ... была возвратной, достаточно существования такой последовательности $(u_0, u_1, \ldots)$ с $u_i \to \infty$, $i \to \infty$, чтобы для всех $j \neq 0$ $u_j \geqslant \sum_i u_i p_{ij}$.

3. Для того чтобы неразложимая цепь с состояниями 0, 1, ... была возвратной и положительной, необходимо и достаточно,

чтобы система уравнений $u_j = \sum_i u_i p_{ij}$, $j = 0, 1, \ldots$ имела не тождественно равное нулю решение, для которого $\sum_i |u_i| < \infty$.

4. Рассматривается марковская цепь с состояниями 0, 1, ... и переходными вероятностями

$$p_{00} = r_0, \quad p_{01} = p_0 > 0,$$

$$p_{ij} = \begin{cases} p_i > 0, & j = i+1, \\ r_i \geqslant 0, & j = i, \\ q_i > 0, & j = i-1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Пусть $\rho_0 = 1$, $\rho_m = \frac{q_1 \ldots q_m}{p_1 \ldots p_m}$. Доказать справедливость следующих утверждений:

$$\text{цепь возвратна} \Leftrightarrow \sum \rho_m = \infty,$$

$$\text{цепь невозвратна} \Leftrightarrow \sum \rho_m < \infty,$$

$$\text{цепь положительная} \Leftrightarrow \sum \frac{1}{p_m \rho_m} < \infty,$$

$$\text{цепь нулевая} \Leftrightarrow \sum \rho_m = \infty, \ \sum \frac{1}{p_m \rho_m} = \infty.$$

5. Показать, что

$$f_{ik} \geqslant f_{ij} f_{jk},$$

$$\sup_n p_{ij}^{(n)} \leqslant f_{ij} \leqslant \sum_{n=1}^{\infty} p_{ij}^{(n)}.$$

6. Показать, что для любой марковской цепи со счетным множеством состояний всегда существуют пределы для $p_{ij}^{(n)}$ *в смысле Чезаро*:

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \frac{f_{ij}}{\mu_j}.$$

7. Рассматривается марковская цепь $\xi_0, \xi_1, \ldots$ с $\xi_{k+1} = (\xi_k)^+ + \eta_{k+1}$, $k \geqslant 0$, где $\eta_1, \eta_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\mathsf{P}(\eta_k = j) = p_j$, $j = 0, 1, \ldots$ Выпишите матрицу переходных вероятностей и покажите, что если $p_0 > 0$, $p_0 + p_1 < 1$, то цепь возвратна тогда и только тогда, когда $\sum_k k p_k \leqslant 1$.

§ 4. О существовании предельных и стационарных распределений

1. Начнем с некоторых необходимых условий существования стационарных распределений.

Теорема 1. *Пусть марковская цепь со счетным множеством состояний $E=\{1, 2, \ldots\}$ и матрицей переходных вероятностей $\mathbb{P}=\|p_{ij}\|$ такова, что для всех i и j существуют пределы*

$$\lim_n p_{ij}^{(n)}=\pi_j,$$

не зависящие от i.

Тогда

a) $\sum\limits_j \pi_j \leqslant 1, \quad \sum\limits_i \pi_i p_{ij}=\pi_j;$

b) *или все* $\pi_j=0$, *или же* $\sum\limits_j \pi_j=1;$

c) *если все* $\pi_j=0$, *то стационарного распределения не существует; если же* $\sum\limits_j \pi_j=1$, *то* $\Pi=(\pi_1, \pi_2, \ldots)$ *образует единственное стационарное распределение.*

Доказательство. По лемме Фату

$$\sum_j \pi_j=\sum_j \lim_n p_{ij}^{(n)} \leqslant \varliminf_n \sum_j p_{ij}^{(n)}=1.$$

Далее,

$$\sum_i \pi_i p_{ij}=\sum_i \left(\lim_n p_{ki}^{(n)}\right) p_{ij} \leqslant \varliminf_n \sum_i p_{ki}^{(n)} p_{ij}=\varliminf_n p_{kj}^{(n+1)}=\pi_j,$$

т. е. для любого j

$$\sum_i \pi_i p_{ij} \leqslant \pi_j.$$

Предположим, что для некоторого j_0

$$\sum_i \pi_i p_{ij_0}<\pi_{j_0}.$$

Тогда

$$\sum_j \pi_j>\sum_j\left(\sum_i \pi_i p_{ij}\right)=\sum_i \pi_i \sum_j p_{ij}=\sum_i \pi_i.$$

Полученное противоречие доказывает, что

$$\sum_i \pi_i p_{ij}=\pi_j \tag{1}$$

для любых j.

Из (1) следует, что

$$\sum_i \pi_i p_{ij}^{(n)} = \pi_j.$$

Поэтому

$$\pi_j = \lim_n \sum_i \pi_i p_{ij}^{(n)} = \sum_i \pi_i \lim_n p_{ij}^{(n)} = \left(\sum_i \pi_i\right) \pi_j,$$

т. е. для всех j

$$\pi_j \left(1 - \sum_i \pi_i\right) = 0,$$

откуда следует утверждение b).

Пусть теперь $\mathbb{Q} = (q_1, q_2, \ldots)$ — какое-то стационарное распределение. Тогда, поскольку $\sum_i q_i p_{ij}^{(n)} = q_j$ и, значит, $\sum_i q_i \pi_j = q_j$, т. е. $\pi_j = q_j$ для всех j, то это стационарное распределение должно совпадать с $\Pi = (\pi_1, \pi_2, \ldots)$. Поэтому, если все $\pi_j = 0$, то стационарного распределения нет. Если же $\sum_j \pi_j = 1$, то $\Pi = (\pi_1, \pi_2, \ldots)$ будет единственным стационарным распределением.

Теорема доказана.

Сформулируем и докажем основной результат о существовании единственного стационарного распределения.

Теорема 2. *Для марковских цепей со счетным множеством состояний единственное стационарное распределение существует тогда и только тогда, когда во множестве состояний существует в точности один положительный возвратный класс (существенных сообщающихся состояний).*

Доказательство. Обозначим через N число положительных возвратных классов.

Пусть $N = 0$. Тогда все состояния невозвратные или возвратные нулевые и в силу (3.10) и (3.20) $\lim_n p_{ij}^{(n)} = 0$ для любых i и j. Следовательно, по теореме 1 стационарного распределения не существует.

Пусть $N = 1$ и C — единственный положительный возвратный класс. Если $d(C) = 1$, то, согласно (3.8),

$$p_{ij}^{(n)} \to \frac{1}{\mu_j} > 0, \qquad i, j \in C.$$

Если $j \notin C$, то j невозвратно и в силу (3.7) для всех i $p_{ij}^{(n)} \to 0$, $n \to \infty$.

Положим

Тогда по теореме 1 набор $\mathbb{Q}=(q_1, q_2, \ldots)$ образует единственное стационарное распределение.

Пусть теперь $d=d(C)>1$. Обозначим $C_0, \ldots, C_{d-1}$ циклические подклассы. Каждый из этих подклассов C_k относительно матрицы $\mathbb{P}^d$ образует возвратный апериодический класс. Тогда, если $i, j \in C_k$, то согласно (3.19),

$$p_{ij}^{(nd)} \to \frac{d}{\mu_j} > 0.$$

Поэтому на каждом из множеств C_k набор $\frac{d}{\mu_j}$, $j \in C_k$, образует (относительно матрицы $\mathbb{P}^d$) единственное стационарное распределение. Отсюда, в частности, следует, что $\sum_{j \in C_k} \frac{d}{\mu_j}=1$, т. е. $\sum_{j \in C_k} \frac{1}{\mu_j}=\frac{1}{d}$.

Положим

$$q_j=\begin{cases} \frac{1}{\mu_j}, & j \in C=C_0+\ldots+C_{d-1}, \\ 0, & j \notin C, \end{cases}$$

и покажем, что для исходной цепи набор $\mathbb{Q}=(q_1, q_2, \ldots)$ образует единственное стационарное распределение.

В самом деле, для $i \in C$

$$p_{ii}^{(nd)}=\sum_{j \in C} p_{ij}^{(nd-1)} p_{ji}.$$

Тогда по лемме Фату

$$\frac{d}{\mu_i}=\lim_n p_{ii}^{(nd)} \geqslant \sum_{j \in C} \underline{\lim_n}\, p_{ij}^{(nd-1)} p_{ji}=\sum_{j \in C} \frac{1}{\mu_j} p_{ji}$$

и, значит,

$$\frac{1}{\mu_i} \geqslant \sum_{j \in C} \frac{1}{\mu_j} p_{ji}.$$

Но

$$\sum_{i \in C} \frac{1}{\mu_i}=\sum_{k=0}^{d-1}\left(\sum_{i \in C_k} \frac{1}{\mu_i}\right)=\sum_{k=0}^{d-1} \frac{1}{d}=1.$$

Так же, как и в теореме 1, отсюда показывается, что на самом деле

$$\frac{1}{\mu_i}=\sum_{j \in C} \frac{1}{\mu_j} p_{ji}.$$

Это доказывает, что набор $\mathbb{Q}=(q_1, q_2, \ldots)$ образует стационарное распределение, которое единственно в силу теоремы 1.

Пусть теперь число положительных возвратных классов $N \geqslant 2$. Обозначим их $C^1, \ldots, C^N$, и пусть $\mathbb{Q}^i=(q_1^i, q_2^i, \ldots)$ — стационарное распределение, соответствующее классу C^l и построенное по формуле

$$q_j^l = \begin{cases} \frac{1}{\mu_j} > 0, & j \in C^l, \\ 0, & j \notin C^l. \end{cases}$$

Тогда для любых неотрицательных чисел $a_1, \ldots, a_N$ таких, что $a_1 + \ldots + a_N = 1$, набор $a_1\mathbb{Q}^1 + \ldots + a_N\mathbb{Q}^N$ будет также образовывать стационарное распределение, поскольку $(a_1\mathbb{Q}^1 + \ldots + a_N\mathbb{Q}^N)\mathbb{P} = a_1\mathbb{Q}^1\mathbb{P} + \ldots + a_N\mathbb{Q}^N\mathbb{P} = a_1\mathbb{Q}^1 + \ldots + a_N\mathbb{Q}^N$. Отсюда следует, что для $N \geqslant 2$ существует континуум стационарных распределений. Таким образом, единственное стационарное распределение существует лишь в случае $N = 1$.

Теорема доказана.

2. Следующая теорема дает ответ на вопрос об условиях, при которых существует предельное распределение для марковских цепей со счетным множеством состояний E.

Теорема 3. *Для того чтобы существовало предельное распределение, необходимо и достаточно, чтобы во множестве E всех состояний цепи нашелся в точности один апериодический положительный возвратный класс C такой, что $f_{ij} = 1$ для всех $j \in C$ и $i \in E$.*

Доказательство. Необходимость. Пусть $q_j = \lim\limits_n p_{ij}^{(n)}$ и набор $\mathbb{Q} = (q_1, q_2, \ldots)$ образует распределение $\left(q_i \geqslant 0,\ \sum\limits_i q_i = 1\right)$. Тогда по теореме 1 это предельное распределение будет единственным стационарным распределением, а значит, по теореме 2 существует один и только один возвратный положительный класс C. Покажем, что период этого класса $d = 1$. Предположим противное, т. е. пусть $d > 1$. Обозначим $C_0, C_1, \ldots, C_{d-1}$ — циклические подклассы. Если $i \in C_0$ и $j \in C_1$, то, согласно (19), $p_{ij}^{(nd+1)} \to \frac{d}{\mu_j}$ и $p_{ij}^{(nd)} = 0$ для всех n. Но $\frac{d}{\mu_j} > 0$, поэтому $p_{ij}^{(n)}$ не имеет предела при $n \to \infty$, что противоречит исходному предположению о существовании $\lim\limits_n p_{ij}^{(n)}$. Пусть теперь $j \in C$ и $i \in E$. Тогда, согласно (3.11), $p_{ij}^{(n)} \to \frac{f_{ij}}{\mu_j}$. Следовательно, $\pi_j = \frac{f_{ij}}{\mu_j}$. Но π_j не зависит от i. Значит, $f_{ij} = f_{jj} = 1$.

Достаточность. В силу (3.11), (3.10) и (3.7)

Поэтому, если $f_{ij}=1$ для всех $j\in C$ и $i\in E$, то $q_j=\lim\limits_n p_{ij}^{(n)}$ не зависит от i. Класс C положительный, значит, $q_j>0$ для $j\in C$. Тогда по теореме 1 $\sum\limits_j q_j=1$ и набор $\mathbb{Q}=(q_1,\ q_2,\ \ldots)$ образует предельное распределение.

3. Резюмируем полученные выше результаты о существовании предельного распределения, единственного стационарного распределения и эргодичности для случая конечных цепей.

Теорема 4. *Для конечных марковских цепей имеют место следующие импликации:*

$$\begin{array}{ccc}
(\textit{эргодичность}) & \overset{\{1\}}{\Longleftrightarrow} & \left(\begin{array}{c}\textit{цепь неразложима воз-}\\ \textit{вратна, положительна}\\ \textit{с } d=1\end{array}\right)\\
\Downarrow & & \Downarrow\\
\left(\begin{array}{c}\textit{существует пре-}\\ \textit{дельное распреде-}\\ \textit{ление}\end{array}\right) & \overset{\{2\}}{\Longleftrightarrow} & \left(\begin{array}{c}\textit{существует в точности}\\ \textit{один возвратный поло-}\\ \textit{жительный класс с } d=1\end{array}\right)\\
\Downarrow & & \Downarrow\\
\left(\begin{array}{c}\textit{существует един-}\\ \textit{ственное стацио-}\\ \textit{нарное распределе-}\\ \textit{ние}\end{array}\right) & \overset{\{3\}}{\Longleftrightarrow} & \left(\begin{array}{c}\textit{существует в точности}\\ \textit{один возвратный поло-}\\ \textit{жительный класс}\end{array}\right)
\end{array}$$

Доказательство. Все «вертикальные» импликации $\Downarrow$ очевидны. Импликации $\{1\}$ установлены в теореме 2 § 3, импликации $\{2\}$ — в теореме 3, импликации $\{3\}$ — в теореме 2.

4. Задачи.

1. Показать, что в примере 1 из § 5 стационарные и предельные распределения отсутствуют.

2. Рассмотреть вопрос о стационарных и предельных распределениях для марковской цепи с матрицей переходных вероятностей

$$\mathbb{P}=\begin{pmatrix}1/2 & 0 & 1/2 & 0\\ 0 & 0 & 0 & 1\\ 1/4 & 1/2 & 1/4 & 0\\ 0 & 1/2 & 1/2 & 0\end{pmatrix}.$$

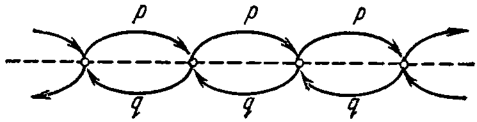
3. Пусть $\mathbb{P}=\|p_{ij}\|$ — конечная *дважды стохастическая* матрица, т. е. $\sum\limits_{i=1}^{m} p_{ij}=1$, $j=1,\ \ldots,\ m$. Показать, что для соответствующей марковской цепи стационарным распределением является вектор $\mathbb{Q}=(1/m,\ \ldots,\ 1/m)$.

§ 5. Примеры

1. Приведем несколько примеров, иллюстрирующих введенные понятия и полученные выше результаты относительно классификации и предельного поведения переходных вероятностей.

Пример 1. Будем называть *простым случайным блужданием* марковскую цепь, в которой частица с некоторой вероятностью остается в каждом состоянии и с некоторыми вероятностями переходит в соседние.

Простое случайное блуждание, соответствующее следующему графу:



описывает блуждание частицы по состояниям $E=\{0, \pm 1, \ldots\}$ с переходом на единицу вправо с вероятностью p и влево с вероятностью q. Понятно, что вероятности перехода равны

$$p_{ij}=\begin{cases} p, & j=i+1, \\ q, & j=i-1, \quad p+q=1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Если $p=0$, то частица детерминированным образом движется влево, если же $p=1$, то вправо. Эти случаи мало интересны, поскольку тогда все состояния несущественны. Будем поэтому предполагать, что $0<p<1$.

В этом предположении, состояния цепи образуют один класс (существенных сообщающихся состояний). В каждое состояние можно вернуться за 2, 4, 6, ... шагов. Поэтому цепь имеет период $d=2$.

Поскольку для любого $i \in E$

$$p_{ii}^{(2n)}=C_{2n}^{n}(pq)^n=\frac{(2n)!}{(n!)^2}(pq)^n,$$

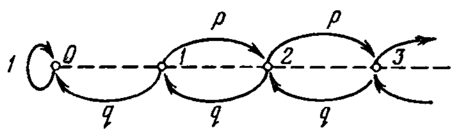
то по формуле Стирлинга $(n! \sim \sqrt{2\pi n}\, n^n e^{-n})$

$$p_{ii}^{(2n)} \sim \frac{(4pq)^n}{\sqrt{\pi n}}.$$

Поэтому, если $p=q$, то $\sum_n p_{ii}^{(2n)}=\infty$, и если $p \neq q$, то $\sum_n p_{ii}^{(2n)}<\infty$. Иначе говоря, если $p=q$, то цепь возвратна, если же $p \neq q$, то невозвратна. В § 10 гл. I было показано, что в случае $p=q=1/2$ $f_{ii}^{(2n)} \sim \frac{1}{2\sqrt{\pi} n^{3/2}}$, $n \to \infty$. Значит, $\mu_i = \sum_n (2n) f_{ii}^{(2n)} = \infty$, т. е. все возвратные состояния являются нулевыми. Поэтому в силу теоремы 1 из § 3 для всех $0<p<1$ $p_{ij}^{(n)} \to 0$, $n \to \infty$, для любых i и j.

Стационарные, предельные и эргодические распределения отсутствуют.

Пример 2. Рассмотрим простое случайное блуждание с $E = \{0, 1, 2, \ldots\}$, где 0 является *поглощающим экраном*:



Состояние 0 образует единственный положительный возвратный класс с $d=1$. Все же остальные состояния невозвратны. Поэтому, согласно теореме 2 из § 4, существует единственное стационарное распределение

$$\Pi = (\pi_0, \pi_1, \pi_2, \ldots)$$

с $\pi_0 = 1$ и $\pi_i = 0$, $i \geqslant 1$.

Рассмотрим теперь вопрос о предельном распределении. Ясно, что $p_{00}^{(n)} = 1$, $p_{ij}^{(n)} \to 0$, $j \geqslant 1$, $i \geqslant 0$. Покажем теперь, что для всякого $i \geqslant 1$ величины $\alpha(i) = \lim_n p_{i0}^{(n)}$ определяются формулами

$$\alpha(i) = \begin{cases} \left(\frac{q}{p}\right)^i, & p>q, \\ 1, & p \leqslant q. \end{cases} \tag{1}$$

С этой целью прежде всего заметим, что поскольку состояние 0 является поглощающим, то $p_{i0}^{(n)} = \sum_{k \leqslant n} f_{i0}^{(k)}$ и, следовательно, $\alpha(i) = f_{i0}$, т. е. интересующая нас вероятность $\alpha(i)$ есть вероятность того, что частица, выходящая из состояния i, рано или поздно достигнет нулевого состояния. Для этих вероятностей тем же методом, что и в § 12 гл. I (см. также § 2 гл. VII), выводятся рекуррентные соотношения

$$\alpha(i) = p\alpha(i+1) + q\alpha(i-1), \tag{2}$$

при этом $\alpha(0)=1$. Общее решение этого уравнения имеет вид

$$\alpha(i)=a+b\,(q/p)^i, \tag{3}$$

и условие $\alpha(0)=1$ дает одно условие на константы a и b: $a+b=1$.

Если предположить, что $q>p$, то тогда в силу ограниченности $\alpha(i)$ сразу получаем, что $b=0$, а значит, $\alpha(i)=1$. Этот результат вполне понятен, поскольку в случае $q>p$ частица имеет тенденцию двигаться по направлению к нулевому состоянию.

Если же $p>q$, то ситуация обратная — имеется тенденция хода вправо, и естественно поэтому ожидать, что тогда

$$\alpha(i)\to 0,\quad i\to\infty, \tag{4}$$

а, значит, $a=0$ и

$$\alpha(i)=\left(\frac{q}{p}\right)^i. \tag{5}$$

Чтобы доказать это равенство, мы не будем устанавливать (4), а поступим иначе.

Наряду с поглощающим экраном в точке 0 введем в рассмотрение поглощающий экран в целочисленной точке N. Вероятность того, что частица, выходящая из точки i, достигнет нулевого состояния раньше, чем состояния N, обозначим $\alpha_N(i)$. Для вероятностей $\alpha_N(i)$ справедливы уравнения (2) с граничными условиями

$$\alpha_N(0)=1,\quad \alpha_N(N)=0,$$

и, как это уже было показано в § 9 гл. I,

$$\alpha_N(i)=\frac{\left(\frac{q}{p}\right)^i-\left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)^N},\qquad 0\leqslant i\leqslant N \tag{6}$$

Отсюда

$$\lim_N \alpha_N(i)=\left(\frac{q}{p}\right)^i$$

и, следовательно, для доказательства требуемого результата (5) надо лишь показать, что

$$\alpha(i)=\lim_N \alpha_N(i). \tag{7}$$

Интуитивно это понятно. Строгое же доказательство можно получить на следующем пути.

Будем предполагать, что частица выходит из фиксированного состояния i. Тогда

$$\alpha(i)=\mathsf{P}_i(A), \tag{8}$$

где A — событие, состоящее в том, что найдется такое N, что частица, выходящая из точки i, достигнет нулевого состояния раньше, чем состояния N. Если

$$A_N = \{\text{частица достигнет } 0 \text{ раньше чем } N\},$$

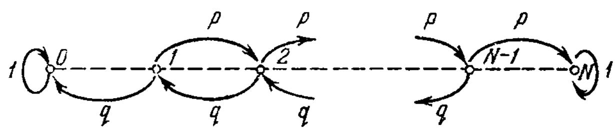
то $A = \bigcup\limits_{N=i+1}^{\infty} A_N$. Ясно, что $A_N \subseteq A_{N+1}$ и

$$\mathsf{P}_i\left(\bigcup_{N=i+1}^{\infty} A_N\right) = \lim_{N\to\infty} \mathsf{P}_i(A_N). \tag{9}$$

Но $\alpha_N(i) = \mathsf{P}_i(A_N)$, так что (7) сразу следует из (8) и (9).

Итак, если $p > q$, то предельные значения $\lim\limits_n p_{i0}^{(n)}$ зависят от i и, следовательно, в этом случае предельного распределения не существует. Если же $p \leqslant q$, то для любого $\lim\limits_n p_{i0}^{(n)} = 1$ и $\lim\limits_n p_{ij}^{(n)} = 0$, $j \geqslant 1$. Таким образом, в этом случае предельное распределение Π имеет вид $\Pi = (1, 0, 0, \dots)$.

Пример 3. Рассмотрим простое случайное блуждание с *поглощающими экранами в точках* 0 *и* N:

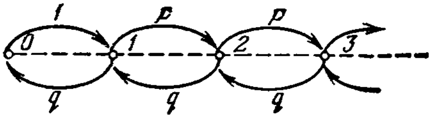


Здесь существуют два положительных возвратных класса $\{0\}$ и $\{N\}$. Все остальные состояния $\{1, \dots, N-1\}$ невозвратны. Из теоремы 1 § 3 следует, что существует бесконечно много стационарных распределений $\Pi = (\pi_0, \pi_1, \dots, \pi_N)$ с $\pi_0 = a$, $\pi_N = b$, $\pi_1 = \dots = \pi_{N-1} = 0$, где $a \geqslant 0$, $b \geqslant 0$, $a + b = 1$. Из теоремы 4 § 4 вытекает также, что предельного распределения не существует. Это следует и из того, что, согласно результатам п. 2 § 9 гл. I,

$$\lim_{n\to\infty} p_{i0}^{(n)} = \begin{cases} \dfrac{\left(\dfrac{q}{p}\right)^i - \left(\dfrac{q}{p}\right)^N}{1 - \left(\dfrac{q}{p}\right)^N}, & p \neq q, \\[2ex] 1 - \dfrac{i}{N}, & p = q, \end{cases} \tag{10}$$

$$\lim_n p_{iN}^{(n)} = 1 - \lim_n p_{i0}^{(n)} \quad \text{и} \quad \lim_n p_{ij}^{(n)} = 0, \; 1 \leqslant j \leqslant N-1.$$

Пример 4. Рассмотрим простое случайное блуждание с $E = \{0, 1, \dots\}$ и *отражающим экраном в нуле*:



Нетрудно понять, что цепь является апериодической. Предположим, что $p > q$ (блуждающая частица имеет тенденцию ухода вправо). Пусть $i > 1$; для отыскания вероятностей f_{i1} можно воспользоваться формулой (1), из которой следует, что

$$f_{i1} = \left(\frac{q}{p}\right)^{i-1} < 1, \quad i > 1.$$

Все состояния рассматриваемой цепи сообщаются между собой. Поэтому, если состояние i было бы возвратным, то тогда бы и состояние 1 также было бы возвратным. Но (см. доказательство леммы 3 в § 3) тогда f_{i1} было бы равно единице. Следовательно, все состояния рассматриваемой цепи в случае $p > q$ невозвратны, а значит, $p_{ij}^{(n)} \to 0$, $n \to \infty$, $i, j \in E$, и предельного или стационарного распределения не существует.

Пусть теперь $p \leqslant q$. Тогда из (1) $f_{i1} = 1$ для $i > 1$ и $f_{11} = q + pf_{21} = 1$. Поэтому цепь возвратна.

Рассмотрим систему уравнений, определяющую стационарное распределение $\Pi = (\pi_0, \pi_1, \dots)$:

$$\pi_0 = \pi_1 q,$$
$$\pi_1 = \pi_0 + \pi_2 q,$$
$$\pi_2 = \pi_1 p + \pi_3 q,$$

т. е.

$$\pi_1 = \pi_1 q + \pi_2 q,$$
$$\pi_2 = \pi_2 q + \pi_3 q,$$
$$\dots\dots\dots$$

откуда

$$\pi_j = \left(\frac{p}{q}\right)\pi_{j-1}, \qquad j = 2, 3, \dots$$

Если $p = q$, то тогда $\pi_1 = \pi_2 = \dots$ и, следовательно, $\pi_0 = \pi_1 = \pi_2 = \dots = 0$. Иначе говоря, если $p = q$, то не существует стационарного, а значит, и предельного распределения. Отсюда и из теоремы 3 § 4, в частности, следует, что в этом случае все состояния цепи нулевые.

Осталось рассмотреть случай $p<q$. Из условия $\sum_{j=0}^{\infty}\pi_j=1$ находим, что

$$\pi_1\left[q+1+\left(\frac{p}{q}\right)+\left(\frac{p}{q}\right)^2+\ldots\right]=1,$$

т. е.

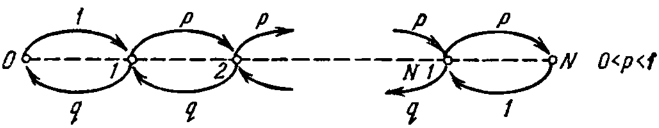
$$\pi_1=\frac{q-p}{2q}$$

и

$$\pi_j=\frac{q-p}{2q}\cdot\left(\frac{p}{q}\right)^{j-1};\qquad j\geqslant 2.$$

Тем самым это распределение $\prod$ является единственным стационарным распределением. Поэтому в случае $p<q$ цепь является апериодической возвратной и положительной (теорема 2 § 4). Распределение $\prod$ является также предельным и эргодическим.

Пример 5. Снова рассматривается простое случайное блуждание с *двумя отражающими экранами в точках* 0 и N:



Все состояния цепи являются апериодическими, возвратными и положительными. Согласно теореме 4 § 4 цепь является эргодической. Решая систему $\pi_j=\sum_{i=0}^{N}\pi_i p_{ij}$ с условием $\sum_{i=0}^{N}\pi_i=1$, находим эргодическое распределение:

$$\pi_i=\frac{\left(\frac{p}{q}\right)^{i-1}}{1+\sum_{j=1}^{N-1}\left(\frac{p}{q}\right)^{j-1}},\qquad 2\leqslant j\leqslant N-1,$$

и

$$\pi_0=\pi_1 q,\qquad \pi_N=\pi_{N-1}p.$$

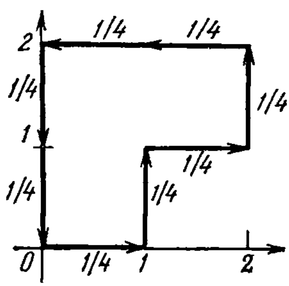


Рис. 41. Блуждание на плоскости.

2. Пример 6. Из примера 1 следует, что рассмотренное в нем простое случайное блуждание по целочисленным точкам прямой будет возвратно, если $p=q$, и невозвратно, если $p\neq q$. Рассмотрим теперь с точки зрения возвратности и невозвратности симметричные случайные блуждания на плоскости и в пространстве.

В случае плоскости будем предполагать, что частица из каждого состояния (i, j) с вероятностью 1/4 сдвигается вверх, вниз, вправо или влево (рис. 41).

Рассмотрим для определенности состояние (0, 0). Тогда вероятность $P_k = p^{(k)}_{(0,0),(0,0)}$ перехода из состояния (0, 0) за k шагов в (0, 0) задается формулами

$$P_{2n+1} = 0, \qquad n = 0, 1, 2, \ldots$$

$$P_{2n} = \sum_{\{(i,j):\, i+j=n\}} \frac{(2n)!}{i!\,i!\,j!\,j!}\left(\frac{1}{4}\right)^{2n}, \qquad n = 1, 2, \ldots$$

Умножая числитель и знаменатель каждого члена суммы на $(n!)^2$, получим

$$P_{2n} = \left(\frac{1}{4}\right)^{2n} C_{2n}^n \sum_{i=0}^{n} C_n^i C_n^{n-i} = \left(\frac{1}{4}\right)^{2n} (C_{2n}^n)^2,$$

поскольку

$$\sum_{i=0}^{n} C_n^i C_n^{n-i} = C_{2n}^n.$$

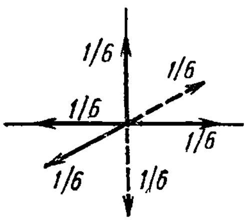
Применяя теперь формулу Стирлинга, найдем, что

$$P_{2n} \sim \frac{1}{\pi n}$$

и, значит, $\sum P_{2n} = \infty$. Следовательно, состояние (0, 0) (также как и любое другое) является *возвратным*.

Оказывается, однако, что в случае размерности *три и больше* симметричное случайное блуждание *невозвратно*. Покажем это для блуждания по целочисленным точкам (i, j, k) в пространстве.

Будем предполагать, что из точки (i, j, k) частица с вероятностью 1/6 сдвигается на единицу вдоль одного из направлений координатных осей:



Тогда, если P_k — вероятность возвращения за k шагов из состояния (0, 0, 0) в (0, 0, 0), то

$$P_{2n+1} = 0, \qquad n = 0, 1, \ldots,$$

$$P_{2n} = \sum_{\{(i,j):\, 0 \leqslant i+j \leqslant n\}} \frac{(2n)!}{(i!)^2 (j!)^2 ((n-i-j)!)^2}\left(\frac{1}{6}\right)^{2n} =$$

$$= \frac{1}{2^{2n}} C_{2n}^n \sum_{\{(i,j):\, 0 \leqslant i+j \leqslant n\}} \left[\frac{n!}{i!\,j!\,(n-i-j)!}\right]^2 \left(\frac{1}{3}\right)^{2n} \leqslant$$

$$\leqslant C_n \frac{1}{2^{2n}} C_{2n}^n \frac{1}{3^n} \sum_{\{(i,j):\, 0 \leqslant i+j \leqslant n\}} \frac{n!}{i!\,j!\,(n-i-j)!}\left(\frac{1}{3}\right)^n =$$

$$= C_n \frac{1}{2^{2n}} C_{2n}^n \frac{1}{3^n}, \qquad n = 1, 2, \ldots, \qquad (11)$$

где

$$C_n = \max_{\{(i, j):\, 0 \leqslant i+j \leqslant n\}} \left[\frac{n!}{i!\, j!\, (n-i-j)!}\right]. \qquad (12)$$

Докажем, что при больших n max в (12) достигается при $i \sim n/3$, $j \sim n/3$. Для этого сбозначим через i_0 и j_0 значения, для которых достигается max. Тогда, очевидно, будут справедливы следующие неравенства:

$$\frac{n!}{j_0!\,(i_0-1)!\,(n-j_0-i_0+1)!} \leqslant \frac{n!}{j_0!\, i_0!\,(n-j_0-i_0)!},$$

$$\frac{n!}{j_0!\,(i_0+1)!\,(n-j_0-i_0-1)!} \leqslant \frac{n!}{(i_0-1)!\, i_0!\,(n-j_0-i_0+1)!} \leqslant$$

$$\leqslant \frac{n!}{(j_0+1)!\, i_0!\,(n-j_0-i_0-1)!},$$

откуда

$$n - i_0 - 1 \leqslant 2j_0 \leqslant n - i_0 + 1,$$

$$n - j_0 - 1 \leqslant 2i_0 \leqslant n - j_0 + 1,$$

и, значит, для больших n $i_0 \sim n/3$, а $j_0 \sim n/3$ и

$$C_n \sim \frac{n!}{\left[\left(\frac{n}{3}\right)!\right]^3}.$$

По формуле Стирлинга

$$C_n \frac{1}{2^{2n}} C_{2n}^n \frac{1}{3^n} \sim \frac{3\sqrt{3}}{2\pi^{3/2} n^{3/2}},$$

и поскольку

$$\sum_{n=1}^{\infty} \frac{3\sqrt{3}}{2\pi^{3/2} n^{3/2}} < \infty,$$

то $\sum_n P_{2n} < \infty$. Следовательно, состояние (0, 0, 0), а также любое другое состояние, являются невозвратными. Аналогичный результат остается верным и для размерностей, больших трех.

Итак, справедлив следующий результат (Пойа): *для пространств R^1 и R^2 симметричное случайное блуждание возвратно, а для пространств R^n, $n \geqslant 3$, является невозвратным.*

3. Задачи.

1. Вывести рекуррентные соотношения (1).

2. Установить справедливость соотношений (4).

3. Показать, что в примере 5 все состояния являются апериодическими, возвратными и положительными.

4. Дать классификацию состояний марковской цепи с матрицей переходных вероятностей

$$\mathbb{P} = \begin{pmatrix} p & q & 0 & 0 \\ 0 & 0 & p & q \\ p & q & 0 & 0 \\ 0 & 0 & p & q \end{pmatrix},$$

где $p+q=1$, $p \geqslant 0$, $q \geqslant 0$.

ИСТОРИКО-БИБЛИОГРАФИЧЕСКАЯ СПРАВКА

Введение

История теории вероятностей до Лапласа изложена в монографии Тодхантер [68]. Период от Лапласа до конца XIX в. освещен в статье Б. В. Гнеденко и О. В. Шейнина, опубликованной в сборнике [45]. В книге Д. Е. Майстрова [44] история теории вероятностей изложена от ее возникновения до 30-х годов текущего столетия. Краткий очерк теории вероятностей имеется в учебнике Б. В. Гнеденко [15]. О происхождении многих вероятностных терминов см. книгу Н. В. Александровой [2].

По поводу основных понятий теории вероятностей см. книги А. Н. Колмогорова [32], Б. В. Гнеденко [15], А. А. Боровкова [7], Б. В. Гнеденко и А. Я. Хинчина [17], А. М. Яглома и И. М. Яглома [84], справочное пособие Ю. В. Прохорова и Ю. А. Розанова [56], справочник [65] и книги В. Феллера [69], [70], Ю. Неймана [51], М. Лоэва [42], Дж. Л. Дуба [20], переведенные с английского. Укажем также на сборники [46] и [67], содержащие большое количество задач по теории вероятностей.

При составлении настоящего учебного пособия автор использовал разнообразную литературу. Из учебных руководств на английском языке особо отметим книги Л. Бреймана [8], Р. Эша [81], [82] и Р. Эша и М. Гарднера [83], являющихся (по мнению автора) образцами удачной подачи материала.

Полезный справочный материал по теории вероятностей и математической статистике читатель может найти в Большой Советской Энциклопедии, Малой Советской Энциклопедии и в Математической Энциклопедии (изд. «Советская Энциклопедия»).

Основным научным журналом по теории вероятностей и математической статистике, издаваемым в нашей стране, является журнал «Теория вероятностей и ее применения» (изд-во «Наука»), выходящий с 1956 г.

«Реферативный журнал», выпускаемый ВИНИТИ — Всесоюзным институтом научной и технической информации (Москва), печатает рефераты на статьи по теории вероятностей и математической статистике, публикуемые как у нас, так и за рубежом.

Для большинства вероятностно-статистических приложений, требующих обращения к таблицам, полезными являются «Таблицы математической статистики» Л. Н. Большева и Н. В. Смирнова [6].

Глава I

§ 1. О построении вероятностных моделей см. также статью А. Н. Колмогорова [31], книгу Б. В. Гнеденко [15]. Большой материал, касающийся вопросов типа «размещение дробинок по ячейкам», см. в книге В. Ф. Колчина, Б. А. Севастьянова и В. П. Чистякова [34].

§ 2. По поводу различных вероятностных моделей (в частности, одномерной модели Изинга), возникающих в статистической физике, см., например, книгу Исихара [25].

§ 3. Формула и теорема Байеса лежат в основе так называемого «байесовского подхода» в математической статистике. См., например, книги Де Гроота [18] и Закса [22].

§ 4. Различные задачи, касающиеся случайных величин и их вероятностных характеристик, можно найти в сборниках задач [46] и [67].

§ 5. Комбинаторное доказательство закона больших чисел, восходящее к Я. Бернулли, можно найти, например, в [69]. По поводу эмпирической интерпретации закона больших чисел см. статью А. Н. Колмогорова [31].

§ 6. По поводу уточнений в локальной и интегральной теоремах, а также в теореме Пуассона см. книгу А. А. Боровкова [7] и статью Ю. В. Прохорова [54].

§ 7. Излагаемый здесь материал на примере схемы Бернулли иллюстрирует некоторые основные понятия и методы математической статистики. Подробнее см., например, монографии Г. Крамера [35] и Ван дер Вардена [10].

§ 8. Рассмотрение условных вероятностей и условных математических ожиданий относительно разбиений поможет лучше освоиться с вводимыми далее более сложными понятиями условных вероятностей и условных математических ожиданий относительно σ-алгебр.

§ 9. Задача о разорении рассматривалась в приводимой здесь форме, в сущности, еще Лапласом. См. по этому поводу статью Б. В. Гнеденко и О. В. Шейнина [45]. Обширный материал на эту тему содержится в книге В. Феллера [69].

§ 10. Принятое здесь изложение следует в основном книге В Феллера [69]. Метод доказательства соотношений (10) и (11) дан в статье [19].

§ 11. Теория мартингалов подробно изложена в книге Дж. Дуба [20]. Иное доказательство теоремы о баллотировке можно найти, например, в книге В. Феллера [69].

§ 12. Обширный материал по марковским цепям содержится в книгах В. Феллера [69], Е. Б. Дынкина [21], Дж. Кемени и Дж. Снелла [27], Т. А. Сарымсакова [61], С. Х. Сираждинова [64]. Теории ветвящихся процессов посвящена монография Б. А. Севастьянова [62].

Глава II

§ 1. Аксиоматика Колмогорова изложена в его книге [32].

§ 2. Дополнительный материал об алгебрах и σ-алгебрах можно найти, например, в книгах А. Н. Колмогорова и С. В. Фомина [33], Ж. Нёве [49], Л. Бреймана [8], Р. Эша [82].

§ 3. Доказательство теоремы Каратеодори см. в [42], [71].

§§ 4—5. Большой материал об измеримых функциях можно найти в книге П. Халмоша [71].

§ 6. См. также книги А. Н. Колмогорова и С. В. Фомина [33], П. Халмоша [71], Р. Эша [82]. В этих книгах содержится и доказательство теоремы Радона—Никодима. Иногда неравенством Чебышева называют неравенство

$$\mathsf{P}(|\xi|\geqslant\varepsilon)\leqslant\frac{\mathsf{M}\xi^2}{\varepsilon^2},$$

а неравенство

$$\mathsf{P}(|\xi|\geqslant\varepsilon)\leqslant\frac{\mathsf{M}|\xi|^r}{\varepsilon^r},\qquad r>0,$$

называют *неравенством Маркова*.

§ 7. Определение условной вероятности и условного математического ожидания относительно σ-алгебр было дано А. Н. Колмогоровым [32]. Обширный материал по рассматриваемым вопросам содержится в книгах Л. Бреймана [8] и Р. Эша [82].

§ 8. См. также книги А. А. Боровкова [7], Р. Эша [82], Г. Крамера [35], Б. В. Гнеденко [15].

§ 9. Теорема Колмогорова о существовании процесса с заданными конечномерными распределениями содержится в его книге [32]. По поводу теоремы Ионеску Тулчи см. также книги Ж. Невё [49] и Р. Эша [82]. Приводимое здесь доказательство следует [82].

§§ 10—11. См. также книги А. Н. Колмогорова и С. В. Фомина [33], Р. Эша [82], Дж. Дуба [20], М. Лоэва [42].

§ 12. Теория характеристических функций излагается во многих книгах. См., например, Б. В. Гнеденко [15], Б. В. Гнеденко и А. Н. Колмогоров [16], Рамачандран [57]. Изложение формул связи моментов и семиинвариантов следует статье В. П. Леонова и А. Н. Ширяева [40].

§ 13. См. также книги И. А. Ибрагимова и Ю. А. Розанова [24], Л. Бреймана [8], Р. Ш. Липцера и А. Н. Ширяева [41].

Глава III

§ 1. Подробное изложение вопросов слабой сходимости вероятностных мер и распределений содержится в книгах Б. В. Гнеденко и А. Н. Колмогорова [16] и П. Биллингсли [5].

§ 2. Теорема Ю. В. Прохорова содержится в его статье [55].

§ 3. Методу характеристических функций в доказательстве предельных теорем теории вероятностей посвящена монография Б. В. Гнеденко и А. Н. Колмогорова [16]. См. также П. Биллингсли [5]. Приводимая задача 2 охватывает как закон больших чисел Я. Бернулли, так и закон больших чисел Пуассона, который предполагал, что $\xi_1, \xi_2, \ldots$ независимы, принимают два значения (1 и 0), но, вообще говоря, разнораспределенны: $\mathsf{P}(\xi_i = 1) = p_i$, $\mathsf{P}(\xi_i = 0) = 1 - p_i$, $i \geqslant 1$.

§§ 4—5. Изложение рассматриваемых здесь вопросов следует книгам Б. В. Гнеденко и А. Н. Колмогорова [16] и Р. Эша [82]. Теорема 2 в § 4 носит название теоремы Линдеберга—Феллера См. также В. Феллер [70].

Глава IV

§ 1. Закон «нуля или единицы» Колмогорова содержится в его книге [32]. По поводу закона «нуля или единицы» Хьюитта и Сэвиджа см. также А. А. Боровков [7], Л. Брейман [8], Р. Эш [82].

§§ 2—4. Основные результаты здесь получены А. Н. Колмогоровым и А. Я. Хинчиным (см. [32] и литературу там). См. также книги В. В. Петрова [53] и Стоута [66]. По поводу вероятностных методов в теории чисел см. книгу И. Кубилюса [36].

Глава V

§§ 1—3. При изложении теории стационарных (в узком смысле) случайных последовательностей использованы книги Л. Бреймана [8], Я. Г. Синая [63] и Дж. Ламперти [38]. Простое доказательство максимальной эргодической теоремы дано А. Гарсиа [12].

Глава VI

§ 1. Теории стационарных (в широком смысле) случайных последовательностей посвящены книги Ю. А. Розанова [60], И. И. Гихмана и А. В. Скорохода [13], [14]. Пример 6 часто приводился в лекциях А. Н. Колмогорова.

§ 2. По поводу ортогональных стохастических мер и стохастических интегралов см. также Дж. Дуб [20], И. И. Гихман и А. В. Скороход [14], Ю. А. Розанов [60], Р. Эш и М. Гарднер [83].

§ 3. Спектральное представление (2) получено Г. Крамером и М. Лоэвом (см., например, [42]). В других терминах такое представление содержится в работе А. Н. Колмогорова [29]. См. также книги Дж. Дуба [20], Ю. А. Розанова [60], Р. Эша и М. Гарднера [83].

§ 4. Подробное изложение вопросов статистического оценивания ковариационной функции и спектральной плотности содержится в книгах Э. Хеннана [72], [73].

§§ 5—6. См. также книги Ю. А. Розанова [60], Дж. Ламперти [38], И. И. Гихмана и А. В. Скорохода [13], [14].

§ 7. Изложение здесь следует книге Р. Ш. Липцера и А. Н. Ширяева [41].

Глава VII

§ 1. Большинство основных результатов теории мартингалов получено Дж. Дубом [20]. Теорема 1 содержится у П. Мейера [47]. См. также книги П. Мейера [48], Р. Ш. Липцера и А. Н. Ширяева [41], И. И. Гихмана и А. В. Скорохода [14].

§ 2. Теорема 1 часто называется теоремой «О преобразовании свободного выбора», [20]. По поводу тождеств (14), (15) и фундаментального тождества Вальда см. книгу [9].

§ 3. Подробное освещение излагаемых здесь результатов, включая доказательство неравенств Хинчина, Марцинкевича и Зигмунда, Буркхольдера, Дэвиса содержится в книге Чао и Тейчера [74]. Теорема 2 принадлежит Ленгляру [39].

§ 4. См. монографию Дж. Дуба [20].

§ 5. Излагаемый здесь материал следует статьям Ю. М. Кабанова, Р. Ш. Липцера и А. Н Ширяева [26], Г. Ю. Энгельберта и А. Н. Ширяева [80] и книге Ж. Невё [50]. Теорема 4 и пример даны Р. Ш. Липцером.

§ 6. Приводимый здесь подход к проблематике «абсолютная непрерывность и сингулярность» и излагаемые результаты содержатся в работе Ю. М. Кабанова, Р. Ш. Липцера и А. Н. Ширяева [26].

§ 7. Теоремы 1 и 2 принадлежат А. А. Новикову [52]. Лемма 1 является «дискретным» аналогом известной «леммы Гирсанова» (см [26]).

Глава VIII

§ 1. По поводу основных определений см. книги Е. Б. Дынкина [21], А. Д. Вентцеля [11], Дж. Дуба [20], И. И. Гихмана и А. В. Скорохода [14]. Существование регулярных переходных вероятностей, для которых при всех $x \in R$ выполнены уравнения Колмогорова — Чэпмена (9), доказано в [49] (следствие к предложению V.2.1) и в [14] (том I, гл. II, § 4). С. Е. Кузнецовым доказана (см. Abstracts of 12 European Meeting of Statisticians, Varna, 1979) справедливость (далеко нетривиального) аналогичного результата для марковских процессов с непрерывным временем и со значениями в универсально измеримых пространствах.

§§ 2—5. Изложение здесь следует статье А. Н. Колмогорова [28] и книгам А. А. Боровкова [7] и Р. Эша [81].

ЛИТЕРАТУРА

[1] Александров П. С. Введение в общую теорию множеств и функций. — М.: Гостехиздат, 1948.
[2] Александрова Н. В. Математические термины. — М.: Высшая школа, 1978.
[3] Бернштейн С. Н. О работах П. Л. Чебышева по теории вероятностей. — В кн.: Научное наследие П. Л. Чебышева. Вып. 1, Математика, 1945, с 59—60.
[4] Бернштейн С. Н. Теория вероятностей. — 4-е изд. — М.: Гостехиздат, 1946.
[5] Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977.
[6] Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. — М.: Наука, 1965.
[7] Боровков А. А. Теория вероятностей. — М.: Наука, 1976.
[8] Брейман (Breiman L.) Probability: Addison-Wesley, 1968.
[9] Вальд А. Последовательный анализ. — М.: Физматгиз, 1960.
[10] Ван дер Варден Б. Л. Математическая статистика. — М.: ИЛ, 1960.
[11] Вентцель А. Д. Курс теории случайных процессов. — М.: Наука, 1976.
[12] Гарсиа (Garsia A.) A simple proof of Eberhard Hopf's maximal ergodic theorem. — J. Math. and Mech. 1965, **14**, 381—382.
[13] Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. — М.: Наука, 1977.
[14] Гихман И. И., Скороход А. В. Теория случайных процессов. — М.: Наука, 1971, 1973, 1975. — Т. I, II, III.
[15] Гнеденко Б. В. Курс теории вероятностей —5-е изд.—М.: Наука, 1969.
[16] Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — Л. — М.: Гостехиздат, 1949.
[17] Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей. — М.: Наука, 1976.
[18] Де Гроот, Оптимальные статистические решения. — М.: Мир, 1974.
[19] Дохерти (Doherty). An amusing proof in fluctuation theory. — Lecture Notes in Mathematics, 1975, **452**, 101—104.
[20] Дуб Дж. Л. Вероятностные процессы. — М.: ИЛ, 1956.
[21] Дынкин Е. Б. Марковские процессы. — М.: Физматгиз, 1963.
[22] Закс Ш. Теория статистических выводов. — М.: Мир, 1975.
[23] Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. — М.: Наука, 1965.
[24] Ибрагимов И. А., Розанов Ю. А. Гауссовские случайные процессы. М.: Наука, 1970.
[25] Исихара А. Статистическая физика. — М.: Мир, 1973.
[26] Кабанов Ю. М., Липцер Р. Ш., Ширяев А. Н. К вопросу об абсолютной непрерывности и сингулярности вероятностных мер. — Матем. сб., 1977, **104 (146)**, 227—247.
[27] Кемени Дж., Снелл Дж. Конечные цепи Маркова. — М.: Наука, 1970.

[28] Колмогоров А. Н. Цепи Маркова со счетным числом возможных состояний. — Бюлл. МГУ, 1937, 1, № 3, 1—16.
[29] Колмогоров А. Н. Стационарные последовательности в гильбертовском пространстве. — Бюлл. МГУ, 1941, 2, № 6, 1—40.
[30] Колмогоров А. Н. Роль русской науки в развитии теории вероятностей. — Учен. зап. МГУ, 1947, вып. 91, с. 56.
[31] Колмогоров А. Н. Теория вероятностей. — В сб.: Математика, ее содержание, методы и значение. — М.: Изд-во Ан СССР, 1956. — Т. II.
[32] Колмогоров А. Н. Основные понятия теории вероятностей. — М.: Наука, 1974.
[33] Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Наука, 1968.
[34] Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. Случайные размещения. — М.: Наука, 1976.
[35] Крамер Г. Математические методы статистики. — М.: Мир, 1976.
[36] Кубилюс И. Вероятностные методы в теории чисел. — Вильнюс: Гос. изд.-во полит. и научн. литер. Лит. ССР, 1959.
[37] Ламперти Дж. Вероятность. — М.: Наука, 1973.
[38] Ламперти (Lamperti J). Stochastic Processes. — Lecture Notes Series Aarhus Univ., 1974, 38.
[39] Ленгляр (Lenglart E.) Relation de domination entre deux processus. — Annales de l'institut H. Poincare, 1977, Sect. B. XIII, 2, 171—179.
[40] Леонов В. П., Ширяев А. Н. К технике вычисления семиинвариантов. — Теория вероятн. и ее прим., 1959, IV, 2, 342—355.
[41] Липцер Р. Ш., Ширяев А. Н, Статистика случайных процессов. — М.: Наука, 1974.
[42] Лоэв М. Теория вероятностей. — М.: ИЛ, 1962.
[43] Марков А. А. Исчисление вероятностей. — 3-е изд. — СПб, 1913.
[44] Майстров Д. Е. Теория вероятностей (исторический очерк). — М.: Наука, 1967.
[45] Математика XIX века / под ред. А. Н, Колмогорова и А. П, Юшкевича. — М.: Наука, 1978.
[46] Мешалкин Л. Д. Сборник задач по теории вероятностей. — М.: изд-во МГУ, 1963.
[47] Мейер (Meyer P. A.) Martingales and Stochastic Integrals I. — Lectures Notes in Mathematics, 284. Heidelberg, New York: Springer-Verlag, 1972.
[48] Мейер П. А. Вероятность и потенциалы. — М.: Мир, 1973.
[49] Невё Ж. Математические основы теории вероятностей. — М.: Мир, 1969.
[50] Невё (Neveu J.) Discrete parameter martingales. — Amsterdam: North Holland Publishing Co., 1975.
[51] Нейман Ю. Вводный курс теории вероятностей и математической статистики. — М.: Наука, 1968.
[52] Новиков А. А. Об оценках и асимптотическом поведении вероятностей непересечения подвижных границ суммами независимых случайных величин. — Изв. АН СССР. Серия математ., 1980, 40, 4, 868—885.
[53] Петров В. В, Суммы независимых случайных величин. — М.: Наука, 1972.
[54] Прохоров Ю. В. Асимптотическое поведение биномиального распределения. — УМН, 1953, VIII, 3 (55), 135—142.
[55] Прохоров Ю. В. Сходимость случайных процессов и предельные теоремы теории вероятностей. — Теория вероятн. и ее примен, 1956, I, 2, 177—238.
[56] Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. — М.: Наука, 1973.
[57] Рамачандран Б. Теория характеристических функций. — М.: Наука, 1975.
[58] Реньи (Renyi A), Probability Theory, Budapest: Akadémiai Kiado, 1970.

[59] Роббинс Г., Сигмунд Д., Чао И. Теория оптимальных правил остановки.—М.: Наука, 1977.
[60] Розанов Ю. А. Стационарные случайные процессы.—М.: Физматгиз, 1963.
[61] Сарымсаков Т. А. Основы теории процессов Маркова.—М.: Гостехиздат, 1954.
[62] Севастьянов Б. А. Ветвящиеся процессы.—М.: Наука, 1971.
[63] Синай Я. Г. Введение в эргодическую теорию.—Ереван: Изд-во Ереванского ун-та, 1973.
[64] Сираждинов С. Х. Предельные теоремы для однородных цепей Маркова.—Ташкент: Изд-во Ан УзССР, 1955.
[65] Справочник по теории вероятностей и математической статистике / под ред. В. С. Королюка.—Киев: Наукова думка, 1978.
[66] Стоут (Stout W. F.). Almost Sure Convergence.—New York: Academic Press, 1974.
[67] Теорія ймовірностей.—Київ: Вища школа, 1976.
[68] Тодхантер (Todhunter I.). A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.—London; Mac Millan, 1865.
[69] Феллер В. Введение в теорию вероятностей и ее приложения.—М.: Мир, 1964.—Т. 1.
[70] Феллер В. Введение в теорию вероятностей и ее приложения.—М.: Мир, 1967.—Т. 2.
[71] Халмош П. Теория меры.—М.: ИЛ, 1953.
[72] Хеннан Э. Анализ временных рядов.—М.: Наука, 1964.
[73] Хеннан Э. Многомерные временные ряды.—М.: Мир. 1974.
[74] Чао, Тейчер (Chow Y. S., Teiher H.). Probability Theory.—New York; Springer-Verlag, 1978.
[75] Чебышев П. Л. Теория вероятностей: Лекции акад. П. Л. Чебышева, читанные в 1879, 1880 гг. /Издано А. Н. Крыловым по записи А. М. Ляпунова/.—М.—Л., 1936.
[76] Чжун Кай-лай. Однородные цепи Маркова.—М.: Мир, 1964.
[77] Ширяев А. Н. Случайные процессы.—М.: Изд-во МГУ, 1972.
[78] Ширяев А. Н. Вероятность, статистика, случайные процессы.—М.: Изд-во МГУ, 1973, 1974.—Т. I и II.
[79] Ширяев А. Н. Статистический последовательный анализ.—М.: Наука, 1976.
[80] Энгельберт, Ширяев (Engelbert H.-J., Shiryaev A. N.). On the Sets of Convergence of Generalized Submartingales —Stochastics, 1979, 2, 3, 155—166.
[81] Эш (Ash R.). Basic Probability theory —New York: John Wiley and Sons, 1970.
[82] Эш (Ash R.). Real Analysis and Probability.—New York: Academic Press, 1972.
[83] Эш, Гарднер (Ash R., Gardner M. F.). Topics in Stochastic Processes.—New York: Academic Press, 1975.
[84] Яглом А. М., Яглом И. М. Вероятность и информация.—М.: Наука, 1973.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

УКАЗАТЕЛЬ ОБОЗНАЧЕНИЙ

$\equiv$ — равно по определению, тождественно равно

$A \Rightarrow B$ — импликация, т. е. высказывание «если A, то B», или «A есть достаточное условие для B», или «B есть необходимое условие для A»

$A \Leftrightarrow B$ — эквиваленция, т. е. единая запись двух импликаций $A \Rightarrow B$ и $B \Rightarrow A$

$[t_1, \ldots, t_n]$ — неупорядоченный набор

$(t_1, \ldots, t_n)$ — упорядоченный набор

ЛАТИНСКИЙ АЛФАВИТ

Aa	а	*Jj*	йот	*Ss*	эс
Bb	бе	*Kk*	ка	*Tt*	тэ
Cc	це	*Ll*	эль	*Uu*	у
Dd	де	*Mm*	эм	*Vv*	ве
Ee	е	*Nn*	эн	*Ww*	дубль-ве
Ff	эф	*Oo*	о	*Xx*	икс
Gg	ге	*Pp*	пэ	*Yy*	игрек
Hh	аш	*Qq*	ку	*Zz*	зет
Ii	и	*Rr*	эр		

ГОТИЧЕСКИЙ АЛФАВИТ

𝔄𝔞	а	𝔍𝔧	йот	𝔖𝔰	эс
𝔅𝔟	бэ	𝔎𝔨	ка	𝔗𝔱	тэ
ℭ𝔠	цэ	𝔏𝔩	эль	𝔘𝔲	у
𝔇𝔡	дэ	𝔐𝔪	эм	𝔙𝔳	фау
𝔈𝔢	э	𝔑𝔫	эн	𝔚𝔴	вэ
𝔉𝔣	эф	𝔒𝔬	о	𝔛𝔵	икс
𝔊𝔤	гэ	𝔓𝔭	пэ	𝔜𝔶	ипсилон
ℌ𝔥	ха	𝔔𝔮	ку	ℨ𝔷	цэт
ℑ𝔦	и	ℜ𝔯	эр		

ГРЕЧЕСКИЙ АЛФАВИТ

Αα	альфа	Ιι	иота	Ρρ	ро
Ββ	бета	Κϰ	каппа	Σσ	сигма
Γγ	гамма	Λλ	ламбда	Ττ	тау
Δδ	дельта	Μμ	мю	Υυ	ипсилон
Εε	эпсилон	Νν	ню	Φφ	фи
Ζζ	дзета	Ξξ	кси	Χχ	хи
Ηη	эта	Οο	омикрон	Ψψ	пси
Θθϑ	тэта	Ππ	пи	Ωω	омега