А. Н. Ширяев

ВЕРОЯТНОСТЬ — 1

А. Н. Ширяев

ВЕРОЯТНОСТЬ — 1

Элементарная теория вероятностей.
Математические основания. Предельные теоремы

Издание четвертое,
переработанное и дополненное

*Допущено Министерством образования России в качестве
учебника для студентов высших учебных заведений
по физико-математическим направлениям и специальностям*

Москва
Издательство МЦНМО, 2007

УДК 519.21(075.8)
ББК 22.171
Ш64

Ширяев А. Н.

Ш64 Вероятность: В 2-х кн. — 4-е изд., переработ. и доп. — М.: МЦНМО, 2007.

ISBN 978-5-94057-036-3

Кн. 1. — 552 с. — ISBN 978-5-94057-105-6

Настоящее издание (в двух книгах «Вероятность — 1» и «Вероятность — 2») представляет собой расширенный курс лекций по теории вероятностей.

Первая книга «Вероятность — 1» содержит материал, относящийся к элементарной теории вероятностей, и может служить пособием для первичного ознакомления с предметом. Большой материал отводится математическим основаниям теории вероятностей, базирующимся на аксиоматике Колмогорова, рассматриваются основные вопросы предельных теорем теории вероятностей.

Вторая книга «Вероятность — 2» посвящена случайным процессам с дискретным временем.

Книги рассчитаны на студентов физико-математических специальностей университетов. Могут служить учебным пособием для аспирантов и справочным пособием для специалистов.

Табл. 9. Ил. 42. Библиогр. 137 назв.

Предыдущее издание вышло в 2004 г.

ББК 22.171

ISBN 978-5-94057-105-6

9 785940 571056 >

ISBN 978-5-94057-036-3
ISBN 978-5-94057-105-6 (кн. 1)

К столетию со дня рождения
Андрея Николаевича Колмогорова
(1903—1987)

ОГЛАВЛЕНИЕ

КНИГА ПЕРВАЯ. ВЕРОЯТНОСТЬ — 1

КНИГА ВТОРАЯ. ВЕРОЯТНОСТЬ — 2

Предисловие к четвертому изданию

По сравнению с третьим изданием настоящее пополнилось новым материалом. Особо отметим два новых параграфа в первой главе — § 13 «Производящие функции» и § 14 «Принцип включения-исключения».

В элементарной теории вероятностей, оперирующей с дискретным пространством элементарных исходов, и в дискретной математике вообще метод производящих функций является одним из мощных аналитических средств алгебраического характера, применяемых при решении разнообразных задач. В новом § 13 этот метод иллюстрируется как на ряде вероятностно-комбинаторных задач, так и примерами из дискретной математики типа подсчета числа целочисленных решений в линейных соотношениях при тех или иных ограничениях на искомые величины и примерами типа отыскания элементов последовательностей, подчиняющихся рекуррентным соотношениям.

Материал, относящийся к принципу (формулам) включения-исключения, не столь часто выносится в пособия по теории вероятностей, как того этот принцип заслуживает в связи с его эффективностью при решении разного рода вероятностно-комбинаторных задач. В новом § 14 приводятся основные формулы «включения-исключения» и даются примеры их применения.

Отметим, что после выхода в свет третьего издания в двух книгах «Вероятность — 1» и «Вероятность — 2» нами было опубликовано пособие «Задачи по теории вероятностей», упорядоченное в соответствии с содержанием этих книг. Приводимые в этом пособии задачи являются не только «задачами-упражнениями», полезными для контроля усвоения основных понятий и результатов теории вероятностей, но и во многом носят характер «теории в задачах», что дает большой дополнительный материал для углубленного изучения теории вероятностей.

В заключение отметим также, что в ряд разделов книг «Вероятность — 1» и «Вероятность — 2» была внесена правка редакционного характера.

Москва, ноябрь 2006 *А. Ширяев*

Предисловие к третьему изданию

Если принять во внимание, что первое издание нашей книги «Вероятность» вышло в 1980 г., второе — в 1989 г., а настоящее, третье, — в 2004 г., то можно сказать, что издания выходят примерно с десяти-пятнадцатилетним интервалом. (На английском языке книга издавалась в 1984 и 1996 гг. и на немецком — в 1988 г.)

Время показывает, что отбор материала для первых двух изданий таков, что он сохраняет свою актуальность и сейчас. Это определило то, что мы сохранили структуру предшествующих изданий, внеся, однако, в настоящие книги «Вероятность — 1» и «Вероятность — 2» ряд существенных изменений и дополнений.

Прежде всего это относится к последней, восьмой, главе, посвященной изложению теории марковских цепей с дискретным временем. По существу, эта глава написана заново. Теперь в ней содержится больше материала, приводятся подробные доказательства многих утверждений, ранее изложенные лишь в конспективной форме. Особое внимание уделено строго марковскому свойству и концепциям стационарных и эргодических распределений. Специальный параграф отведен для изложения теории оптимальных правил остановки марковских цепей.

Новый материал добавлен также и в седьмую главу, относящуюся к теории мартингалов с дискретным временем. В § 9 приводится дискретный вариант формулы К. Ито, что может рассматриваться как введение в стохастическое исчисление для броуновского движения, где формула замены переменных Ито играет исключительно важную роль. В § 10 этой главы показывается, как методы теории мартингалов позволяют просто получать оценки вероятностей разорения страховых компаний, эволюция капитала которых описывается моделью Крамера—Лундберга. Материал § 11 относится к «теории арбитража» в стохастической финансовой математике. Здесь приводятся две «фундаментальные теоремы теории арбитража», дающие в мартингальных терминах условия отсутствия арбитражных возможностей и условия, гарантирующие существование такого портфеля ценных бумаг, который позволяет достичь поставленной финансовой цели. В § 12 приводится ряд важных результатов «финансовой инженерии» относительно расчетов рациональных (справедливых) стоимостей опционов и структуры «хеджирующих» портфелей ценных бумаг. Общей теории оптимальных правил остановки для произвольных случайных последовательностей отведен § 13 этой главы. Излагаемый материал служит хорошей иллюстрацией того, как понятия и результаты теории мартингалов «работают» в разнообразных проблемах «стохастической оптимизации».

В другие главы также внесен ряд изменений и дополнений.

Отметим здесь новый материал, относящийся к теоремам о монотонных классах (§ 2 главы II), основанный на детальном изложении понятий и свойств «π-λ-систем», и § 13 главы III, посвященный некоторым фундаментальным теоремам математической статистики.

Новым в настоящем издании является «Очерк истории становления математической теории вероятностей», помещенный в конце книги «Вероятность — 2».

В ряд параграфов внесены новые задачи.

Автор глубоко признателен Т. Б. Толозовой за большой труд по научному редактированию книги и руководству издательства Московского центра непрерывного математического образования как за предложение о новом издании, так и за быстрое и эффективное осуществление проекта по изданию книги.

Москва, 2003 *А. Ширяев*

Математический институт
им. В. А. Стеклова РАН,

Московский государственный
университет им. М. В. Ломоносова

Предисловие ко второму изданию

В предисловии к первому изданию, вышедшему в 1980 г., отмечалось, что основу книги составили лекции, читаемые автором на механико-математическом факультете Московского университета им. М. В. Ломоносова и частично изданные ротапринтным способом под названием «Вероятность, статистика, случайные процессы, I, II» издательством МГУ. Наш первоначальный замысел при написании первого издания настоящей книги состоял в том, чтобы пособие содержало три части: «Вероятность», «Математическая статистика», «Теория случайных процессов», соответствующие программе трехсеместрового курса лекций для математических специализаций университетов. Однако в ходе работы над книгой этот замысел осуществить полностью не удалось, поскольку принятый характер изложения потребовал бы значительно большего объема. В связи с этим в предисловии к первому изданию говорилось, что «достаточно полно здесь представлена лишь теория вероятностей и теория случайных процессов с дискретным временем».

Практически весь текст первого издания входит и в настоящее. Изменения, исправления носят, как правило, редакционный характер с учетом также и тех замечаний, которые были получены мною от советских и зарубежных читателей, знакомых с книгой по русскому изданию и переводам на английский и немецкий языки [98], [99]. Автор искренне признателен всем им за внимание, советы и доброжелательную критику.

В настоящее, второе, издание добавлен новый материал: в главе III это § 5, §§ 7—12, в главе IV — § 5, в главе VII — § 8. Наиболее существенно дополнена третья глава. Здесь читатель найдет изложение ряда вопросов, относящихся к более углубленному изучению таких тем, как меры близости вероятностных мер, метризуемость слабой сходимости, контигуальность вероятностных мер. В эту же главу добавлены доказательства ряда важных результатов о скорости сходимости в центральной предельной теореме и в теореме Пуассона об аппроксимации биномиального распределения пуассоновским, которые в первом издании присутствовали лишь в виде формулировок.

Отметим также новый материал о вероятностях больших уклонений (глава IV, § 5) и центральную предельную теорему для сумм зависимых случайных величин (глава VII, § 8).

За последние несколько лет вероятностная литература, издаваемая Главной редакцией физико-математической литературы издательства «Наука», пополнилась учебниками Б. А. Севастьянова [97], 1982 г., Ю. А. Розанова [95], 1985 г., учебным пособием А. А. Боровкова [7], 1986 г., и учебником Б. В. Гнеденко [15], 1988 г. В 1985 г. и 1986 г. издательство МГУ издало учебное пособие Я. Г. Синая [128]. Представляется, что эти издания вместе с настоящим, во многом отличаясь и дополняя друг друга, охватывают довольно обширный материал, в существенном достаточно полно удовлетворяющий современным требованиям, предъявляемым к преподаванию теоретико-вероятностных дисциплин студентам физико-математических специализаций.

В учебнике Б. В. Гнеденко приводится много хорошо подобранных примеров в том числе и прикладного содержания, дается большой материал методологического характера и обширный очерк истории теории вероятностей. Учебное пособие А. А. Боровкова [7], пожалуй, наиболее сходно с настоящей книгой по стилю изложения. Специального упоминания заслуживают главы 9 («Элементы теории восстановления»), 11 («Факторизационные тождества») и 17 («Функциональные предельные теоремы»), которые отличают пособие [7] от настоящего и [15], [95]. Учебник Ю. А. Розанова содержит большой материал, касающийся разнообразных математических моделей, которые теория вероятностей и математическая статистика предлагают для описания случайных явлений и их эволюции. В основу учебника Б. А. Севастьянова положен двухсеместровый курс его лекций в МГУ. Материал заключительных четырех глав этого учебника охватывает тот необходимый минимум, который входит в университетскую программу годового курса по теории вероятностей и математической статистике. В нашем пособии, возможно в большей степени, нежели чем в отмеченных выше, значительное место уделено теоретико-множественным аспектам и математическим основаниям теории вероятностей.

В учебниках Б. В. Гнеденко и Б. А. Севастьянова в конце каждой главы, а в настоящем пособии — в конце каждого параграфа добавлены упражнения и задачи, которые вместе, например, с задачниками А. В. Прохорова, В. Г. Ушакова, Н. Г. Ушакова «Задачи по теории вероятностей» — М.: Наука, 1986, и А. М. Зубкова, Б. А. Севастьянова, В. П. Чистякова «Сборник задач по теории вероятностей» — М.: Наука, 1988, могут быть использованы читателем для самоконтроля, а преподавателями для проведения семинарских занятий со студентами.

Москва, октябрь 1988 *А. Ширяев*

Предисловие к первому изданию

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 в широком смысле. 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 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.

А. Ширяев

Введение

Предметом теории вероятностей является математический анализ случайных явлений — 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 статистической 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, осуществляющегося в результате экспериментов. Исходя из этого, теория вероятностей постулирует существование у события A определенной числовой характеристики (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 (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 не определенными (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 сейчас хотя бы кратко остановиться на основных исторических вехах теории вероятностей. Подробный «Очерк истории становления математической теории вероятностей» приведен в книге «Вероятность — 2».

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 Aleæ Ludo» («О расчетах в азартных играх»), 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 предположений»), 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 принадлежит заслуга введения в науку «классического» определения понятия «вероятность события» как отношения числа возможных результатов испытаний, благоприятствующих рассматриваемому событию, к числу всех возможных результатов испытаний.

Я. Бернулли был, вероятно, первым, кто делал четкое различие между понятиями «вероятность» события и «частота» его появления и кто осознал важность рассмотрения бесконечных последовательностей повторных испытаний.

Муавру принадлежит заслуга в определении таких понятий, как «независимость», «математическое ожидание», «условная вероятность». В 1812 г. выходит большой трактат Лапласа (1749—1827) «Théorie 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» (А. Н. Колмогоров [30]). До Чебышева основной интерес в теории вероятностей был связан с подсчетом вероятностей случайных событий. Им же впервые была ясно осознана и использована вся сила понятий «случайная величина» и «математическое ожидание 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» (С. Н. Бернштейн [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 не только охватить 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 настоящих книгах «Вероятность — 1» и «Вероятность — 2» 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 («Вероятность — 1»).

Начиная с 20—30 годов прошлого столетия 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, теория информации.

Основное внимание в книге «Вероятность — 2» уделяется случайным процессам с дискретным временем — случайным последовательностям. Однако тот материал, который излагается во второй главе книги

«Вероятность — 1», дает основательную базу (прежде всего логического характера), необходимую при изучении *общей* теории случайных процессов.

Хотя настоящее издание книг «Вероятность — 1» и «Вероятность — 2» посвящено *теории вероятностей*, уместно будет сейчас сказать несколько слов о *математической статистике* и, более общим образом, о *статистике* и их взаимоотношениях с теорией вероятностей.

Во многих странах (например, в Великобритании) *Теория вероятностей* рассматривается как «интегральная» часть *Статистики*, обслуживающая ее математические аспекты. При этом *Статистика* предполагается состоящей из следующих разделов: *описательная статистика* и *математическая статистика*. (Многие энциклопедические издания отмечают, что первоначальное значение слова *статистика* — это «наука о состоянии государства» (по латыни *status* — состояние). На ранних этапах ее называли «политической арифметикой», цель которой состояла в оценивании тех или иных показателей, характеризующих состояние общества, экономики и т. д., и выявлении разного рода количественных свойств массовых явлений по неполным данным.)

Описательная статистика занимается организацией представлений статистических данных («статистического сырья») в удобных для анализа формах. (*Ключевыми словами* здесь являются, например, такие: популяция, выборка, частотные распределения и их гистограммы, относительные частотные распределения и их гистограммы, частотные полигоны и др.).

В настоящее время имеется большое число пакетов статистических программ (MINITAB, SAS, SPSSX и др.), которые позволяют представлять даже очень большие массивы статистических данных в удобных для анализа формах (в виде различных диаграмм, гистограмм и т. п.).

Математическая статистика призвана, собственно говоря, заниматься математической обработкой «статистического сырья», оцениванием выборочных характеристик, выборочных распределений и вынесением статистических выводов с указанием степени их надежности. (*Ключевые слова*: оценивание — точечное и интервальное, различение гипотез, непараметрические тесты, дисперсионный анализ, регрессионный анализ, статистика процессов, ...)

В России традиционным образом *математическая статистика* рассматривается как естественный раздел *теории вероятностей*, занимающийся «обратными вероятностными задачами», т. е. задачами определения той вероятностной модели, которая наиболее адекватным образом отвечает полученным статистическим данным.

Подобный взгляд на математическую статистику (как часть теории вероятностей) дает возможность придать методам и заключениям статистики строгую математическую базу и облечь статистические выводы в форму строгих теоретико-вероятностных утверждений. (См., например, § 13 «Фундаментальные теоремы математической статистики» в гл. III, «Вероятность — 1».) В этой связи уместно будет напомнить, что *первая* предельная теорема теории вероятностей — *закон больших чисел* возникла у Я. Бернулли в «Ars Conjectandi» [134], собственно говоря, именно из желания получить *математическое* обоснование использования «частоты» для оценивания «вероятности успеха» в «схеме Бернулли». (См. по этому поводу «Вероятность — 1», гл. I, § 7.)

В заключение настоящего введения приведем текст Я. Бернулли из «Ars Conjectandi» (глава вторая из части четвертой):

«Относительно того, что твердо известно и не подлежит сомнению, мы говорим, что *знаем* или *понимаем*, относительно всего прочего — что только догадываемся или *предполагаем*.

Делать о какой-либо вещи предположения — все равно, что измерять ее вероятность. Поэтому искусство предположений (Ars conjectandi sive stochastice) у нас определяется как искусство возможно точнее измерять вероятности вещей затем, чтобы в наших суждениях или действиях мы могли всегда выбирать или следовать тому, что будет найдено лучшим, более удовлетворительным, спокойным и разумным».

Латинскому словосочетанию ars conjectandi (искусство предположений) соответствует греческое στοχαστική τέχνη (второе слово часто опускается). Это словосочетание происходит от греческого στόχοζ — цель, догадка, предположение [134, с. 27, 75, 83].

В настоящее время слово «стохастический» широко используется как синоним слова «случайный». Так, выражения «стохастические процессы» и «случайные процессы» рассматриваются как равноправные. Нелишне будет отметить, что *теория случайных процессов* и *статистика случайных процессов* входят ныне в число основных интенсивно развивающихся разделов теории вероятностей и математической статистики.

Глава I

ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Мы называем элементарной теорией вероятностей ту часть теории вероятностей, в которой приходится иметь дело с вероятностями лишь конечного числа событий.

А. Н. Колмогоров. «Основные понятия теории вероятностей» [32]

§ 1. Вероятностная модель эксперимента с конечным числом исходов

1. Рассмотрим некоторый эксперимент, результаты которого (при данном «*комплексе условий*») описываются конечным числом различных *исходов* (явлений) $\omega_1, \ldots, \omega_N$. Для нас несущественна реальная природа этих исходов, важно лишь то, что их число N конечно.

Исходы $\omega_1, \ldots, \omega_N$ будем также 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* (вероятностной «теории») 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 предполагаем, что «комплексом условий» исключаются возможности типа «монета стала на ребро», «монета исчезла», …, но в то же самое время предполагается возможность достоверным образом регистрировать результат подбрасывания.)

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:

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

Замечание. При изложении теории вероятностей «комплекс условий», как правило, не упоминается и «по умолчанию» предполагается данным. Между тем, это понятие важно уже на уровне и элементарной теории вероятностей, поскольку разные «комплексы условий» для одного и того же эксперимента могут приводить к весьма разным вероятностным моделям («теориям»). (В этой связи см. далее текст в начале п. 3, дающий некоторые пояснения, в чем здесь дело.)

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

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 *с возвращением* пространство элементарных событий Ω имеет следующую структуру:

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

и число различных *исходов* (выборок), называемых в комбинаторике *размещениями* из M по *n с повторениями*, равно

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

Если же рассматриваются *неупорядоченные* выборки *с возвращением* (в комбинаторике — *сочетания* из M по *n с повторениями*), то

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

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

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

где $C_k^l \equiv \dfrac{k!}{l!\,(k-l)!}$ — «*число сочетаний из k элементов по l*».

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

$$N(k,\, 1) = k = C_k^1.$$

Предположим теперь, что $N(k,\, n) = C_{k+n-1}^n$, $k \leqslant M$, 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,

$$\begin{aligned} N(M,\, n+1) &= N(M,\, n) + N(M-1,\, n) + \ldots + N(1,\, n) = \\ &= C_{M+n-1}^n + C_{M-1+n-1}^n + \ldots + C_n^n = \\ &= (C_{M+n}^{n+1} - C_{M+n-1}^{n+1}) + (C_{M-1+n}^{n+1} - C_{M-1+n-1}^{n+1}) + \ldots \\ &\qquad\qquad \ldots + (C_{n+1}^{n+1} - C_n^n) + C_n^n = C_{M+n}^{n+1}, \end{aligned}$$

где мы воспользовались следующим легко проверяемым свойством *биномиальных коэффициентов C_k^l*:

$$C_k^{l-1} + C_k^l = C_{k+1}^l.$$

(Это свойство лежит в основе подсчета биномиальных коэффициентов из «треугольника Паскаля».)

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* (в комбинаторике — *размещений* из M по *n без повторений*) пространство исходов

$$\Omega = \{\omega\colon\ \omega = (a_1,\, \ldots,\, a_n),\ \ a_k \neq a_l,\ \ k \neq l;\ \ a_i = 1,\, \ldots,\, M\},$$

а число элементов этого множества равно $M(M-1)\ldots(M-n+1)$. Это число обозначается $(M)_n$ или A_M^n и называется «*числом размещений из M по n*».

В случае *неупорядоченных* выборок *без возвращения* (в комбинаторике — *сочетаний* из M по n *без повторений*) пространство исходов

$$\Omega = \{\omega\colon\ \omega = [a_1, \ldots, a_n],\ \ a_k \neq a_l,\ \ k \neq l;\ \ a_i = 1, \ldots, M\}$$

состоит из

$$N(\Omega) = C_M^n \tag{3}$$

элементов. Действительно, из 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, $N(\Omega) = \dfrac{(M)_n}{n!} = C_M^n$.

Результаты о числе исходов в случае n извлечений из урны с M шарами сведем в таблицу 1:

Таблица 1

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

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

$$(\textit{упорядоченные выборки}) \leftrightarrow (\textit{различимые дробинки}),$$
$$(\textit{неупорядоченные выборки}) \leftrightarrow (\textit{неразличимые дробинки}),$$

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 следующие соответствия:

$$(\textit{выбор с возвращением}) \leftrightarrow \begin{pmatrix}\textit{в ячейке может находиться}\\ \textit{любое число дробинок}\end{pmatrix}$$

$$(\textit{выбор без возвращения}) \leftrightarrow \begin{pmatrix}\textit{в ячейке может находиться}\\ \textit{не более одной дробинки}\end{pmatrix}$$

Из этих соответствий можно сконструировать соответствия типа:

$$\begin{pmatrix}\textit{неупорядоченные выборки}\\ \textit{в задаче выбора без воз-}\\ \textit{вращения}\end{pmatrix} \leftrightarrow \begin{pmatrix}\textit{неразличимые дробинки в задаче}\\ \textit{их размещения по ячейкам, когда}\\ \textit{в ячейке не может находиться}\\ \textit{более одной дробинки}\end{pmatrix}$$

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).

Таблица 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 $+$.

Таблица 3

[БЧ][ ][ ] [Б][Ч][ ] [Б][ ][Ч] [Ч][Б][ ] [ ][БЧ][ ] [ ][Б][Ч] [Ч][ ][Б] [ ][Ч][Б] [ ][ ][БЧ]	[++][ ][ ] [ ][++][ ] [ ][ ][++] [+][+][ ] [+][ ][+] [ ][+][+]	Без запрета
[Б][Ч][ ] [Б][ ][Ч] [Ч][Б][ ] [ ][Б][Ч] [Ч][ ][Б] [ ][Ч][Б]	[+][+][ ] [+][ ][+] [ ][+][+]	С запретом
Различимые дробинки	Неразличимые дробинки	Размещение / Тип дробинки

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 таблицы 1, сведены в таблицу 4.

Таблица 4

$N(\Omega)$ в задаче размещения n дробинок по M ячейкам			
Тип дробинки / Размещение	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		

В статистической физике говорят, что различимые (неразличимые) частицы, не подчиняющиеся принципу запрета Паули («не больше одной частицы в каждой ячейке»), 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.

3. Наряду с понятием пространства *элементарных событий* введем теперь важное понятие *события*, лежащее в основе построения всякой вероятностной модели («теории») рассматриваемого эксперимента.

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:

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

и если «комплекс условий» позволяет записать (зафиксировать, «измерить» и т. п.) результаты всех трех подбрасываний, то, скажем, множество

$$A = \{\text{ГГГ, ГГР, ГРГ, РГГ}\}$$

является событием, состоящим в том, что выпадет по крайней мере два «герба». Однако если «комплекс условий» позволяет зафиксировать лишь только результат первого подбрасывания, то рассматриваемое множество A уже нельзя 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» и «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:

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

На языке теории вероятностей $A \cup B$ — событие, состоящее в том, что *произошло хотя бы одно* из событий A или 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:

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

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

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

$$\begin{aligned} &A \cup B = \{\text{ГГ, ГР, РГ, РР}\} \quad (= \Omega), \\ &A \cap B = \{\text{РГ, ГР}\}. \end{aligned}$$

Если A — некоторое подмножество Ω, то под его *дополнением*, обозначаемым в дальнейшем $\overline{A}$, 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 $\overline{A} = \Omega \setminus A$. На языке теории вероятностей $\overline{A}$ — это событие, состоящее *в ненаступлении* события A. Так, если событие $A = \{\text{ГГ, ГР, РГ}\}$, то $\overline{A} = \{\text{РР}\}$ — событие, состоящее в том, что подряд выпадут две «решетки».

Множества A и $\overline{A}$ не имеют общих точек, и, следовательно, множество $A \cap \overline{A}$ 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 принадлежат $\mathscr{A}$.

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

b) $\mathscr{A} = \{A, \overline{A}, \Omega, \varnothing\}$ — система, порожденная событием A;

c) $\mathscr{A} = \{A : A \subseteq \Omega\}$ — 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:

$$D_1 + \ldots + D_n = \Omega.$$

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

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

(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 c), т. е. алгебру всех подмножеств Ω.

Пусть $\mathscr{D}_1$ и $\mathscr{D}_2$ — два разбиения. Будем говорить, что разбиение $\mathscr{D}_2$ «мельче» разбиения $\mathscr{D}_1$, и записывать это в виде $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2$, если $\alpha(\mathscr{D}_1) \subseteq \subseteq \alpha(\mathscr{D}_2)$.

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 $\mathscr{A}$ из примера c), равно 2^N. Действительно, каждое непустое множество $A \in \mathscr{A}$ может быть представлено в виде

$$A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}, \quad \text{где } \omega_{i_j} \in \Omega, \quad 1 \leqslant k \leqslant N.$$

Поставим в соответствие этому множеству последовательность, состоящую из нулей и единиц:

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

где на местах с номерами $i_1, \ldots, i_k$ 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 таблице 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 построению вероятностной модели («теории») эксперимента с конечным числом исходов: выделили пространство исходов Ω и некоторую систему $\mathscr{A}$ его подмножеств, образующих алгебру и называемых событиями. (Пару $\mathscr{E} = (\Omega, \mathscr{A})$ иногда идентифицируют с *экспериментом*.) Сделаем теперь следующий шаг, а именно припишем каждому элементарному событию (исходу, явлению) $\omega_i \in \Omega$, $i = 1, \ldots, N$, некоторый «вес», обозначаемый $p(\omega_i)$ (или p_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* (A) любого события $A \in \mathscr{A}$ по формуле

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

Определение. Принято говорить, что «вероятностное пространство»

$$(\Omega, \mathscr{A}, \quad),$$

где $\Omega = \{\omega_1, \ldots, \omega_N\}$, $\mathscr{A}$ — некоторая алгебра подмножеств Ω и $\quad = \{\quad (A);\, A \in \mathscr{A}\}$, определяет (задает) *вероятностную модель* («теорию») эксперимента с (конечным) пространством исходов (элементарных событий) Ω и алгеброй событий $\mathscr{A}$. (Ясно, что $\quad (\{\omega_i\}) = p(\omega_i)$, $i = 1, \ldots, N$.)

Из определения (4) вытекают следующие свойства вероятностей:

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

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

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

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

$$(A + B) = \quad (A) + \quad (B), \tag{8}$$

и

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

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 приписывания исходам значений их вероятностей, отметим, что существует много практических ситуаций, в которых из соображений *симметрии* или *однородности* 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 $A \in \mathscr{A}$

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

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

Такой способ задания вероятностей носит название *классического*. Ясно, что в этом случае подсчет вероятностей (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 7. *Задача о совпадениях*. Пусть урна содержит M шаров, занумерованных числами $1, 2, \ldots, M$. Производится выбор с возвращением объема n, 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 следующий вопрос: какова вероятность события

$$A = \{\omega\colon\ a_i \neq a_j,\ \ i \neq j\},$$

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

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

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 $P_{365}(n)$ того, что в этом классе с n учениками 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),

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

Следующая таблица дает значения вероятностей $P_{365}(n)$ для некоторых n:

n	4	16	22	23	40	64
$P_{365}(n)$	0,01636	0,28360	0,47569	0,50730	0,89123	0,99711
n	70	100				
$P_{365}(n)$	0,99916	$1-3{,}07249\cdot 10^{-7}$				

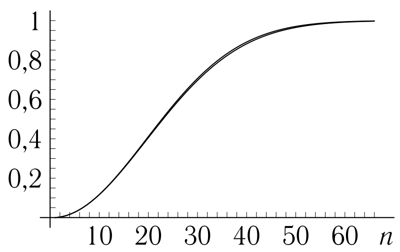
При достаточно больших M

$$\ln \frac{(M)_n}{M^n} = \sum_{k=1}^{n-1} \ln\left(1 - \frac{k}{M}\right) \sim -\frac{1}{M} \sum_{k=1}^{n-1} k = -\frac{1}{M}\frac{n(n-1)}{2},$$

и, значит,

$$P_M(n) \equiv 1 - \frac{(M)_n}{M^n} \sim 1 - e^{-\frac{n(n-1)}{2M}} \quad (\equiv \widetilde{P}_M(n)), \qquad M \to \infty.$$

Ниже приводится график функции $P_{365}(n)$. В том же самом масштабе график приближения $\widetilde{P}_{365}(n)$ практически совпадает с графиком функции $P_{365}(n)$. На интервале [0, 60] максимальное их отличие равно примерно 0,01 (в окрестности значения $n = 30$).



Графики $P_{365}(n)$ и $\widetilde{P}_{365}(n)$

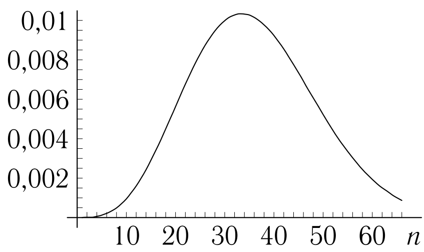


График $P_{365}(n) - \widetilde{P}_{365}(n)$

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 8. *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 такую структуру:

$$\Omega = \{\omega\colon\ \omega = [a_1, \ldots, a_n],\ \ a_k \neq a_l,\ \ k \neq l;\ \ a_i = 1, \ldots, M\}.$$

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

$$A_0 = \{\omega\colon\ \omega = [a_1, \ldots, a_n],\ \ a_k \neq a_l,\ \ k \neq l;\ \ a_i = n+1, \ldots, M\}$$

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

$$(A_0) = \frac{C_{M-n}^n}{C_M^n} = \frac{(M-n)_n}{(M)_n} = \left(1 - \frac{n}{M}\right)\left(1 - \frac{n}{M-1}\right)\ldots\left(1 - \frac{n}{M-n+1}\right)$$

и, значит,

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

Если $M = n^2$ и $n \to \infty$, то $\quad (A_0) \to e^{-1}$ 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 $\cap$ и $\cup$:

$A \cup B = B \cup A$, $A \cap B = B \cap A$ (коммутативность),

$A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$ (ассоциативность),

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (дистрибутивность),

$A \cup A = A$, $A \cap A = A$ (идемпотентность).

Показать также, что справедливы следующие формулы, называемые *законами Моргана* (De Morgan):

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

2. Пусть множество Ω состоит из N элементов. Показать, что число Белла B_N различных разбиений множества Ω определяется формулой

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

(*Указание*. Доказать, что

$$B_N = \sum_{k=0}^{N-1} C_{N-1}^k B_k, \qquad \text{где } B_0 = 1,$$

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

3. Доказать, что для любой конечной системы множеств $A_1, \ldots, A_n$

$$(A_1 \cup \ldots \cup A_n) \leqslant \quad (A_1) + \ldots + \quad (A_n).$$

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

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

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

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

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

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

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

В частности, для $m = 0$

$$(B_0) = 1 - S_1 + S_2 - \ldots \pm S_n.$$

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

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

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

$$(B_1) + \ldots + \quad (B_n) = S_1 - S_2 + \ldots \mp S_n.$$

Доказать справедливость следующих свойств:

(a) *неравенства Бонферрони*: для всякого $k = 1, 2, \ldots$ такого, что $2k \leqslant n$,

$$S_1 - S_2 + \ldots - S_{2k} \leqslant \quad \left(\bigcup_{i=1}^{n} A_i\right) \leqslant S_1 - S_2 + \ldots + S_{2k-1};$$

(b) *неравенства Гумбеля*:

$$\left(\bigcup_{r=1}^{n} A_i\right) \leqslant \frac{\tilde{S}_m}{C_{n-1}^{m-1}}, \qquad m=1, \ldots, n,$$

где

$$\tilde{S}_m = \sum_{1\leqslant i_1<\ldots<i_m\leqslant n} (A_{i_1}\cup\ldots\cup A_{i_m});$$

(c) *неравенства Фреше*:

$$\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \frac{\tilde{S}_{m+1}}{C_{n-1}^{m}} \leqslant \frac{\tilde{S}_m}{C_{n-1}^{m-1}}, \qquad m=1, \ldots, n-1$$

6. Показать, что $(A\cap B\cap C)\geqslant \ (A)+\ (B)+\ (C)-2$ и, по индукции,

$$\left(\bigcap_{i=1}^{n} A_i\right) \geqslant \sum_{i=1}^{n} \ (A_i)-(n-1).$$

7. Исследуйте асимптотику вероятностей $P_M(n)$ из примера 7 при разных предположениях относительно n и M (типа: $n=xM$, $M\to\infty$, или $n=x\sqrt{M}$, $M\to\infty$, где x — фиксировано). Сравните результаты с локальной предельной теоремой из § 6.

§ 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:

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

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

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

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 $\sum\limits_{\omega\in\Omega} p(\omega)=1$.

Рассмотрим все те исходы $\omega=(a_1, \ldots, a_n)$, для которых $\sum\limits_i a_i=k$, где $k=0, 1, \ldots, n$. Согласно таблице 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 $(A)=\sum\limits_{\omega\in A} p(\omega)$, $A\in\mathscr{A}$ (в частности, $(\{\omega\})=p(\omega)$, $\omega\in\Omega$) определяет некоторую вероятностную модель. Естественно назвать ее *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

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

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

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

причем $\sum\limits_{k=0}^{n} (A_k)=1$.

Набор вероятностей $((A_0), \ldots, (A_n))$ 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 $P_n(k)=(A_k)$, $k=0, 1, \ldots, n$. На рис. 1 воспроизведены биномиальные распределения для случая $p=\dfrac{1}{2}$ («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

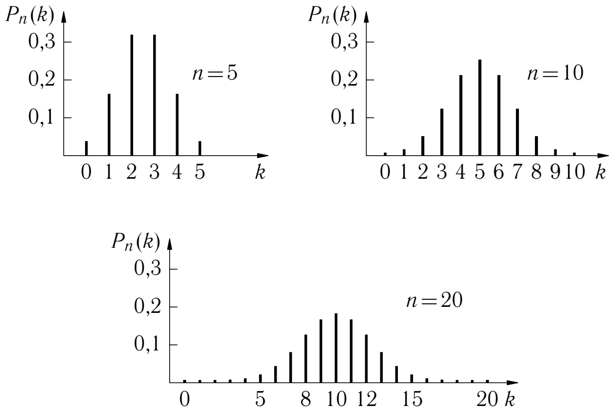


Рис. 1. Графики биномиальных вероятностей $P_n(k)$ для $n=5$, 10, 20

движения частицы может быть полностью 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 $p+q=1$.

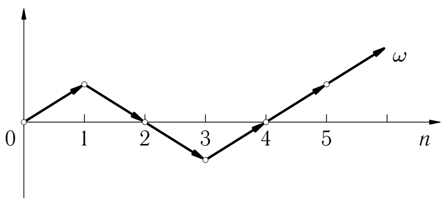


Рис. 2.

Поскольку $\sum\limits_{\omega\in\Omega} p(\omega)=1$, 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, состоящего в том, 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 $C_n^{(n+k)/2}$, и, значит,

$$(A_k)=C_n^{(n+k)/2}p^{(n+k)/2}q^{(n-k)/2}.$$

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

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

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

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)

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

Поэтому

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

и, значит, при больших n

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

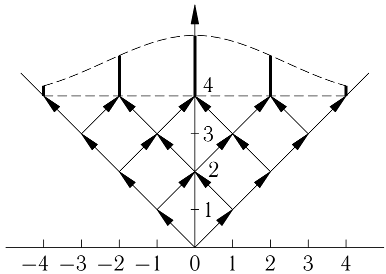


Рис. 3. Возникновение биномиального распределения

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

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

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

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

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

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

где через $C_n(n_1, \ldots, n_r)$ обозначено 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, которыми 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

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

Поэтому

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

и, следовательно, рассматриваемый способ задания вероятностей является корректным.

Пусть

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

Тогда

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

Набор вероятностей $\{\ (A_{n_1,\ldots,n_r})\}$ 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 шаров имеют «цвет» b_1, …, 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

$$\Omega = \{\omega\colon\ \omega = (a_1, \ldots, a_n),\ \ a_k \neq a_l,\ \ k\neq l;\ \ a_i = 1, \ldots, M\}$$

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 шаров имеют цвет b_1, …, n_r шаров имеют цвет b_r, $n_1+\ldots+n_r = n$. Нетрудно показать, что

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

и, значит,

$$(B_{n_1,\dots,n_r}) = \frac{N(B_{n_1,\dots,n_r})}{N(\Omega)} = \frac{C_{M_1}^{n_1} \dots C_{M_r}^{n_r}}{C_M^n}. \tag{3}$$

Набор вероятностей $\{\ (B_{n_1,\dots,n_r})\}$ носит название *многомерного гипергеометрического распределения*. В случае $r = 2$ это распределение называют просто *гипергеометрическим* в связи с тем, что так называемая производящая функция (см. § 13) этого распределения есть гипергеометрическая функция.

Структура многомерного гипергеометрического распределения довольно сложна. Так, вероятность

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

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

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

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

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

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 с 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 задачу 1 в § 8 гл. VIII).

5. Задачи.

1. Доказать утверждение (5).

2. Показать, что для полиномиального распределения $\{\ (A_{n_1,\ldots,n_r})\}$ максимальное значение вероятности достигается в точке $(k_1, \ldots, k_r)$, удовлетворяющей неравенствам:

$$np_i - 1 < k_i \leqslant (n + r - 1)\,p_i, \quad i = 1, \ldots, r.$$

3. *Одномерная модель Изинга*. Пусть имеется n частиц, расположенных в точках $1, 2, \ldots, n$. 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 и комбинаторные соображения, доказать справедливость следующих тождеств:

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

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

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

$$kC_n^k = nC_{n-1}^{k-1},$$

$$C_n^m = \sum_{j=0}^{m} C_k^j C_{n-k}^{m-j}, \quad \text{где } 0 \leqslant m \leqslant n,\ 0 \leqslant k \leqslant n \text{ и полагаем } C_l^j = 0 \text{ при } j < 0 \text{ или } j > l.$$

5. Пусть N — размер некоторой популяции, который требуется оценить «минимальными средствами» без простого пересчета всех элементов этой совокупности. Подобного рода вопрос интересен, например, при оценке числа жителей в той или иной стране, городе и т. д.

В 1786 г. Лаплас для оценки числа N жителей во Франции предложил следующий метод.

Выберем некоторое число, скажем, M, элементов популяции и пометим их. Затем возвратим их в основную совокупность и предположим, что они «хорошо перемешаны» с немаркированными элементами. После этого возьмем из «перемешанной» популяции n элементов. Пусть среди них X элементов оказались маркированными.

Показать, что соответствующая вероятность $_{N,M;n}\{X=m\}$ задается формулой гипергеометрического распределения (ср. с (4)):

$$_{N,M;n}\{X=m\}=\frac{C_M^n C_{N-M}^{n-m}}{C_N^n}.$$

Считая M, n и m фиксированными, найдем максимум этой вероятности по N, т. е. найдем «наиболее правдоподобный» объем всей популяции, приводящий (при заданных M и n) к тому, что число X маркированных элементов оказалось равным m.

Показать, что наиболее правдоподобное значение (обозначим его $\widehat{N}$) определяется формулой ($[\,\cdot\,]$ — целая часть):

$$\widehat{N}=[Mnm^{-1}].$$

Так полученная оценка $\widehat{N}$ для N называется оценкой *максимального правдоподобия*.

(Продолжение этой задачи см. в § 7 (задача 4).)

6. (Ср. с задачей 2 в § 1.) Пусть Ω содержит N элементов и $\tilde{d}(N)$ есть число различных разбиений Ω, обладающих тем свойством, что каждое подмножество разбиения имеет нечетное число элементов. Показать, что

$$\tilde{d}(1)=1,\quad \tilde{d}(2)=1,\quad \tilde{d}(3)=2,$$
$$\tilde{d}(4)=5,\quad \tilde{d}(5)=12,\quad \tilde{d}(6)=37$$

и, вообще,

$$\sum_{n=1}^{\infty}\frac{\tilde{d}(n)x^n}{n!}=e^{\operatorname{sh} x}-1,\qquad |x|<1.$$

§ 3. Условные вероятности. Независимость

1. Понятие *вероятности* события 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 (A) события A, состоящего в том, что вытащенный шар имеет белый цвет? В случае классического подхода $(A)=M_1/M$.

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 интуитивно представляется целесообразным считать, что интересующая нас (условная) вероятность равна $\dfrac{M_1-1}{M-1}$.

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

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

Определение 1. *Условной вероятностью* события B при условии события A с $\ (A) > 0$ (обозначение: $\ (B \mid A)$) называется величина

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

В случае *классического* способа (§ 1, п. 4) задания вероятностей $\ (A) = \dfrac{N(A)}{N(\Omega)}$, $\quad \ (AB) = \dfrac{N(AB)}{N(\Omega)}$ и, значит,

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

Следующие свойства условных вероятностей непосредственно вытекают из определения 1:

$$\begin{aligned} \ (A \mid A) &= 1, \\ \ (\varnothing \mid A) &= 0, \\ \ (B \mid A) &= 1, \quad B \supseteq A, \\ \ (B_1 + B_2 \mid A) &= \ (B_1 \mid A) + \ (B_2 \mid A). \end{aligned}$$

Из этих свойств следует, что при «закрепленном» множестве A условная вероятность $\ (\cdot \mid A)$ 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 $\ (\cdot)$ на $(\Omega, \mathscr{A})$.

Отметим, что

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

однако, вообще говоря,

$$\begin{aligned} \ (B \mid A) + \ (B \mid \bar{A}) &\neq 1, \\ \ (B \mid A) + \ (\bar{B} \mid \bar{A}) &\neq 1. \end{aligned}$$

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 (исходов) здесь, очевидно, таково:

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

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

Будем считать, что каждый исход равновозможен:

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

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 $\quad (AB \mid A)$, а в вопросе b) — условная вероятность $\quad (AB \mid A \cup B)$.

Легко находим, что

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

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

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 $\quad (A_i) > 0$, $i = 1, \ldots, n$ (часто такое разбиение называют также *полной группой* несовместимых событий). Ясно, что

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

и, значит,

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

Но

$$(BA_i) = \quad (B \mid A_i) \quad (A_i).$$

Тем самым имеет место **формула полной вероятности**

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

В частности, если $0 < \quad (A) < 1$, то

$$(B) = \quad (B \mid A) \quad (A) + \quad (B \mid \bar{A}) \quad (\bar{A}). \tag{4}$$

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

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

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

и

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

Интересно отметить, что вероятность $\quad (A)$ 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 $(\quad (A) > 0)$

$$(AB) = \quad (B \mid A) \quad (A). \tag{5}$$

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

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

(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 $(A)>0$ и $(B)>0$. Тогда наряду с (5) справедлива также формула

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

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

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

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

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

В статистических применениях события $A_1, \ldots, A_n$, образующие «полную группу событий» ($A_1+\ldots+A_n=\Omega$), часто называют «гипотезами», а (A_i) — *априорной* *) вероятностью гипотезы A_i. Условная вероятность $(A_i\mid B)$ 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. Предположим, что выпал герб. Спрашивается, какова вероятность того, что выбранная монета симметрична.

Построим соответствующую вероятностную модель. В качестве пространства элементарных событий (исходов) естественно здесь взять множество $\Omega=\{A_1\Gamma, A_1\text{Р}, A_2\Gamma, A_2\text{Р}\}$, 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 рассматриваемых исходов должны быть заданы так, чтобы, согласно условиям задачи,

$$(A_1)=(A_2)=1/2$$

и

$$(\Gamma\mid A_1)=1/2,\qquad (\Gamma\mid A_2)=1/3.$$

Этими условиями вероятности исходов определяются однозначно:

$$(A_1\Gamma)=1/4,\qquad (A_1\text{Р})=1/4,\qquad (A_2\Gamma)=1/6,\qquad (A_2\text{Р})=1/3.$$

Тогда, согласно *формуле Байеса*, интересующая нас вероятность

$$(A_1\mid \Gamma)=\frac{(A_1)\,(\Gamma\mid A_1)}{(A_1)\,(\Gamma\mid A_1)+(A_2)\,(\Gamma\mid A_2)}=\frac{3}{5}$$

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

и, значит,

$$(A_2\,|\,\Gamma) = 2/5.$$

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, будем говорить, что «B не зависит от A», если

$$(B\,|\,A) = \quad (B) \tag{10}$$

(здесь мы предполагаем, что $\quad (A) > 0$).

Поскольку

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

то из (10) находим, что

$$(AB) = \quad (A) \quad (B). \tag{11}$$

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

$$(A\,|\,B) = \quad (A).$$

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

Исходя из этого, примем следующее

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

$$(AB) = \quad (A) \quad (B).$$

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

Приведем соответствующее

Определение 3. *Алгебры* (и, более общо, *системы* подмножеств Ω) $\mathscr{A}_1$ и $\mathscr{A}_2$ называются *независимыми* или *статистически независимыми* (относительно вероятности), 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 пар 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

$$(A_1\bar{A}_2) = \ (A_1) - \ (A_1A_2) = \ (A_1) - \ (A_1)\ (A_2) =$$
$$= \ (A_1)(1 - \ (A_2)) = \ (A_1)\ (\bar{A}_2).$$

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.

Определение 4. Говорят, что *множества* $A_1, \ldots, A_n$ *независимы* или *статистически независимы* в совокупности (относительно вероятности), если для любых $k = 1, \ldots, n$ и $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$

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

Определение 5. *Алгебры* множеств $\mathscr{A}_1, \ldots, \mathscr{A}_n$ называются *независимыми* или *статистически* независимыми в совокупности (относительно вероятности), если независимы любые множества $A_1, \ldots, A_n$, принадлежащие соответственно $\mathscr{A}_1, \ldots, \mathscr{A}_n$.

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

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

попарно независимы, но в то же время

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

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

$$(ABC) = \ (A)\ (B)\ (C),$$

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 $A = \{(i, j) : j = 1, 2$

или 5}, $B=\{(i,\ j)\colon j=4,\ 5$ или $6\}$, $C=\{(i,\ j)\colon i+j=9\}$ имеем

$$(AB)=\frac{1}{6}\neq\frac{1}{4}=\quad(A)\quad(B),$$
$$(AC)=\frac{1}{36}\neq\frac{1}{18}=\quad(A)\quad(C),$$
$$(BC)=\frac{1}{12}\neq\frac{1}{18}=\quad(B)\quad(C),$$

но в то же время

$$(ABC)=\frac{1}{36}=\quad(A)\quad(B)\quad(C).$$

6. С точки зрения понятия *независимости* рассмотрим подробнее классическую модель $(\Omega,\ \mathscr{A},\quad)$, 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 $\quad(\{\omega\})=p(\omega)$ с

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

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

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

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

Ясно, что

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

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

$$(A_kA_l)=p^2,\qquad (A_k\bar{A}_l)=pq,\qquad (\bar{A}_k\bar{A}_l)=q^2.$$

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

Аналогично показывается, что независимы алгебры $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$. Это дает основание назвать рассматриваемую модель $(\Omega, \mathscr{A}, \quad)$ 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 $(\Omega_1, \mathscr{B}_1, \quad_1), \ldots, (\Omega_n, \mathscr{B}_n, \quad_n)$ конечных вероятностных пространств. Образуем пространство $\Omega = \Omega_1 \times \Omega_2 \times \times \ldots \times \Omega_n$ точек $\omega = (a_1, \ldots, a_n)$, где $a_i \in \Omega_i$. Обозначим $\mathscr{A} = \mathscr{B}_1 \otimes \mathscr{B}_2 \otimes \otimes \ldots \otimes \mathscr{B}_n$ — 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 $\quad(A)$ для множеств $A = B_1 \times B_2 \times \ldots \ldots \times B_n$ формулой:

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

Нетрудно проверить, что $\quad(\Omega) = 1$ и, следовательно, тройка $(\Omega, \mathscr{A}, \quad)$ определяет некоторое вероятностное пространство. Это пространство называют *прямым произведением вероятностных пространств*

$$(\Omega_1, \mathscr{B}_1, \quad_1), \ \ldots, \ (\Omega_n, \mathscr{B}_n, \quad_n).$$

Отметим одно легко проверяемое свойство прямого произведения вероятностных пространств: относительно вероятности $\quad$ события

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

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

$$\begin{gathered} \mathscr{A}_1 = \{A_1 : A_1 = \{\omega : a_1 \in B_1\}, \ B_1 \in \mathscr{B}_1\}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mathscr{A}_n = \{A_n : A_n = \{\omega : a_n \in B_n\}, \ B_n \in \mathscr{B}_n\} \end{gathered}$$

являются *независимыми*.

Из приведенных конструкций видно, что *схема Бернулли*

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

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

$$\Omega_i=\{0, 1\}, \qquad \mathscr{B}_i=\{\{0\}, \{1\}, \varnothing, \Omega_i\},$$
$$\ \ _i(\{1\})=p, \qquad \ \ _i(\{0\})=q.$$

7. Задачи.

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

$$\ \ (B\,|\,A)+\ \ (B\,|\,\bar{A})=1,$$
$$\ \ (B\,|\,A)+\ \ (\bar{B}\,|\,\bar{A})=1$$

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

$$\ \ (B_j\,|\,A_k)=k/n.$$

3. Пусть $A_1, \ldots, A_n$ — независимые события. Показать, что тогда

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

4. Пусть $A_1, \ldots, A_n$ — независимые события с $\ \ (A_i)=p_i$. Показать, что 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 $\ \ (A)$ и $\ \ (B)$ выразить вероятности событий, состоящих в том, что произойдет *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 $\ \ (A)$ равно 0 или 1.

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

8. Рассматривается электрическая схема, изображенная на рис. 4. Каждое из реле A, B, C, D и E, работающих независимо, находится в *открытом* состоянии (и, значит, не пропускает электрический сигнал)

или в *закрытом* состоянии (и тогда сигнал пропускается) 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?

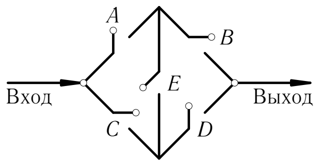


Рис. 4.

9. Пусть $(A+B)>0$. Показать, что

$$(A\,|\,A+B)=\frac{(A)}{(A)+\ \ (B)}.$$

10. Пусть событие A не зависит от событий B_n, $n\geqslant 1$, при этом $B_i\cap B_j=\varnothing$, $i\neq j$. Тогда события A и $\bigcup\limits_{n=1}^{\infty} B_n$ являются независимыми.

11. Показать, что если $(A\,|\,C)>\ \ (B\,|\,C)$ и $(A\,|\,\bar C)>\ \ (B\,|\,\bar C)$, то $(A)>\ \ (B)$.

12. Показать, что

$$(A\,|\,B)=\ \ (A\,|\,BC)\ \ (C\,|\,B)+\ \ (A\,|\,B\bar C)\ \ (\bar C\,|\,B).$$

13. Пусть X и Y — независимые биномиальные величины с параметрами n и p. Показать, что

$$(X=k\,|\,X+Y=m)=\frac{C_n^k C_n^{m-k}}{C_{2n}^m},\quad k=0,\ 1,\ \ldots,\ \min(m,\ n).$$

14. Пусть A, B, C — попарно независимые равновероятные события, причем $A\cap B\cap C=\varnothing$. Найти максимально возможное значение для вероятности (A).

15. В урну, где находился один белый шар, добавили либо белый, либо черный шар (с одинаковыми вероятностями). После этого случайным образом вытащили один шар. Он оказался белым. Какова условная вероятность того, что оставшийся в урне шар тоже белый?

§ 4. Случайные величины и их характеристики

1. Пусть $(\Omega,\ \mathscr{A},\ \ )$ — вероятностная модель некоторого эксперимента с *конечным* числом исходов $N(\Omega)$ и алгеброй $\mathscr{A}$ всех подмножеств Ω. Можно заметить, что в рассмотренных выше примерах, связанных с подсчетом вероятностей тех или иных 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, характеризующие результаты «измерений» 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», отвечающее 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 на $(\Omega, \mathscr{A})$, а распределение вероятностей на *множестве значений* случайной величины. Поскольку в рассматриваемом 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.

*) Для индикатора $I_A(\omega)$ используются также обозначения $I(A)$, I_A. По поводу часто используемых далее свойств индикаторов см. задачу 1.

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

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

Ясно, что значения этих вероятностей полностью определяются вероятностями

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

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 $0,\ 1,\ \ldots,\ n$ 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 приводимых далее примерах мы не конкретизируем структуру основного вероятностного пространства $(\Omega,\ \mathscr{A},\ \ )$, 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

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

называется *функцией распределения* случайной величины ξ.

Ясно, что

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

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

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

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

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

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

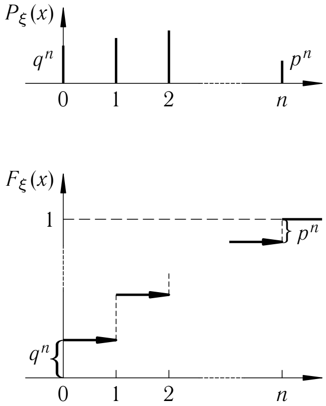


Рис. 5.

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

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

Наряду со случайными величинам 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, $i = 1, \ldots, r$.

Набор вероятностей

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

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

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

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* (в совокупности), если для любых $x_1, \ldots, x_r \in X$

$$\{\xi_1 = x_1, \ldots, \xi_r = x_r\} = \quad \{\xi_1 = x_1\} \ldots \quad \{\xi_r = x_r\},$$

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

$$\{\xi_1 \in B_1, \ldots, \xi_r \in B_r\} = \quad \{\xi_1 \in B_1\} \ldots \quad \{\xi_r \in B_r\}.$$

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

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

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

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

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 $\zeta = \eta + \xi$.

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

$$Z = \{z\colon\ z = x_i + y_j,\ i = 1, \ldots, k,\ j = 1, \ldots, l\},$$

и ясно, что

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

Особо важен случай *независимых* случайных величин ξ и η. Тогда

$$\{\xi = x_i,\ \eta = y_j\} = \quad \{\xi = x_i\} \quad \{\eta = y_j\},$$

и, значит, для любого $z \in Z$

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

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 $\{\xi_i=1\}=p$, $\{\xi_i=0\}=q$, то случайная величина $\zeta=\xi_1+\ldots+\xi_n$ имеет *биномиальное* распределение

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

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

Пусть $(\Omega,\,\mathscr{A},\quad)$ — (конечное) вероятностное пространство и $\xi=\xi(\omega)$ — некоторая случайная величина, принимающая значения в множестве $X=\{x_1,\,\ldots,\,x_k\}$. Если положить $A_i=\{\omega\colon \xi=x_i\}$, $i=1,\,\ldots,\,k$, 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 $p_i=\quad\{\xi=x_i\}$. Интуитивно ясно, что если наблюдать за значениями случайной величины ξ в «n повторных независимых экспериментах», то значение x_i должно встретиться примерно np_i раз, $i=1,\,\ldots,\,k$. (Полезно сопоставить это высказывание с утверждениями законов больших чисел, формулируемых далее в §§ 5, 12.) Таким образом, «среднее значение» этой случайной величины, подсчитанное по результатам n экспериментов, есть примерно

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

Это замечание делает понятным следующее

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

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

Поскольку $A_i = \{\omega\colon \xi(\omega) = x_i\}$ и $P_\xi(x_i) = \quad (A_i)$, то

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

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,

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

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 $\quad \xi$ можно подсчитывать по формуле $\sum\limits_{j=1}^{l} x'_j \quad (B_j)$, не переходя к представлению (5), где все x_i различны. Действительно,

$$\sum_{\{j\,:\,x'_j = x_i\}} x'_j \quad (B_j) = x_i \sum_{\{j\,:\,x'_j = x_i\}} \quad (B_j) = x_i \quad (A_i)$$

и, значит,

$$\sum_{j=1}^{l} x'_j \quad (B_j) = \sum_{i=1}^{k} x_i \quad (A_i).$$

5. Сформулируем основные свойства математических ожиданий:

1) *Если* $\xi \geqslant 0$, *то* $\quad \xi \geqslant 0$.

2) $\quad (a\xi + b\eta) = a \quad \xi + b \quad \eta$, a, b — *постоянные.*

3) *Если* $\xi \geqslant \eta$, *то* $\quad \xi \geqslant \quad \eta$.

4) $| \quad \xi| \leqslant \quad |\xi|$.

5) *Если* ξ *и* η *независимы, то* $\quad \xi\eta = \quad \xi \cdot \quad \eta$.

6) $(\quad |\xi\eta|)^2 \leqslant \quad \xi^2 \cdot \quad \eta^2$ (*неравенство Коши—Буняковского*, называемое также *неравенством Коши—Шварца* или *неравенством Шварца*).

7) *Если* $\xi = I(A)$, *то* $\quad \xi = \quad (A)$.

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

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

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

$$| \quad \xi| = \left| \sum_i x_i \quad (A_i) \right| \leqslant \sum_i |x_i| \quad (A_i) = \quad |\xi|.$$

Для доказательства свойства 5) достаточно заметить, что

$$\xi\eta = \quad \left(\sum_i x_i I(A_i) \right) \left(\sum_j y_j I(B_j) \right) =$$
$$= \quad \sum_{i,j} x_i y_j I(A_i \cap B_j) = \sum_{i,j} x_i y_j \quad (A_i \cap B_j) =$$
$$= \sum_{i,j} x_i y_j \quad (A_i) \quad (B_j) = \left(\sum_i x_i \quad (A_i) \right) \left(\sum_j y_j \quad (B_j) \right) = \quad \xi \cdot \quad \eta,$$

где мы воспользовались тем, что для *независимых* случайных величин ξ и η события

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

являются независимыми: $\quad (A_i \cap B_j) = \quad (A_i) \quad (B_j)$.

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

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

Пусть $\xi^2 > 0$, $\eta^2 > 0$. Положим

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

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

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

$$\xi_1 \ldots \xi_r = \ \xi_1 \ldots \ \xi_r.$$

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

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

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

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

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

находим, что

$$S_n = np.$$

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

$$\{S_n = k\} = C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

Поэтому

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

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

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

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

6. Пусть $\xi = \sum x_i I(A_i)$, 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 $(I_{B_j}(\omega) = I(B_j))$

$$\varphi(\xi(\omega)) = \sum_j y_j I_{B_j}(\omega),$$

и, следовательно,

$$\varphi = \sum_j y_j \quad (B_j) = \sum_j y_j P_\varphi(y_j). \tag{9}$$

Но ясно также, что $(I_{A_j}(\omega) = I(A_j))$

$$\varphi(\xi(\omega)) = \sum_i \varphi(x_i) I_{A_j}(\omega).$$

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

$$\varphi(\xi) = \sum_j \varphi(x_i) P_\xi(x_i).$$

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

Определение 5. *Дисперсией* величины ξ (обозначается $\ \ \xi$) называется величина

$$\xi = \ \ (\xi - \ \ \xi)^2.$$

Величина $\sigma = +\sqrt{\ \ \xi}$ называется *стандартным отклонением* значений случайной величины ξ от ее среднего значения $\ \ \xi$.

Поскольку

$$(\xi - \ \ \xi)^2 = \ \ (\xi^2 - 2\xi \cdot \ \ \xi + (\ \ \xi)^2) = \ \ \xi^2 - (\ \ \xi)^2,$$

то

$$\xi = \ \ \xi^2 - (\ \ \xi)^2.$$

Ясно, что $\ \ \xi \geqslant 0$. Из определения дисперсии также следует, что

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

В частности, $\ \ a = 0$, $\ \ (b\xi) = b^2 \ \ \xi$.

Пусть ξ и η — две случайные величины. Тогда

$$(\xi + \eta) = \ \ ((\xi - \ \ \xi) + (\eta - \ \ \eta))^2 = \ \ \xi + \ \ \eta + 2 \ \ (\xi - \ \ \xi)(\eta - \ \ \eta).$$

Обозначим

$$(\xi,\, \eta) = \quad (\xi - \quad \xi)(\eta - \quad \eta).$$

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

$$\rho(\xi,\, \eta) = \frac{(\xi,\, \eta)}{\sqrt{\quad \xi \cdot \quad \eta}}$$

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 $\xi - \quad \xi$ и $\eta - \quad \eta$, а значит, по свойству 5) математических ожиданий

$$(\xi,\, \eta) = \quad (\xi - \quad \xi) \cdot \quad (\eta - \quad \eta) = 0.$$

С учетом введенного обозначения для ковариации находим, что

$$(\xi + \eta) = \quad \xi + \quad \eta + 2 \qquad (\xi,\, \eta). \tag{10}$$

Если ξ и η *независимы*, то *дисперсия суммы* $\xi + \eta$ *равна сумме дисперсий*:

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

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. $\qquad (\xi,\, \eta) = 0$.

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 зависимы (относительно вероятности $\quad$):

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

но и, более того, *функционально зависимы*: $\xi^2 + \eta^2 = 1$.

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

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

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

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

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

$$\xi = \quad (\xi - \quad \xi)^2 = \quad (\xi - p)^2 = (1-p)^2 p + p^2 q = pq.$$

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

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

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

$$(\eta - f^*(\xi))^2 = \inf_f \quad (\eta - f(\xi))^2.$$

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

$$\begin{aligned} \frac{\partial g(a, b)}{\partial a} &= -2 \quad [\eta - (a + b\xi)], \\ \frac{\partial g(a, b)}{\partial b} &= -2 \quad [(\eta - (a + b\xi))\xi], \end{aligned}$$

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

$$a^* = \quad \eta - b^* \quad \xi, \qquad b^* = \frac{(\xi, \eta)}{\xi}. \tag{15}$$

Иначе говоря,

$$\lambda^*(\xi) = \quad \eta + \frac{(\xi, \eta)}{\xi}(\xi - \quad \xi). \tag{16}$$

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

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

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 (т. е. $\rho(\xi,\, \eta) = 0$), то $\lambda^*(\xi) = \eta$. Таким образом, в случае отсутствия корреляции между ξ и η лучшей оценкой η по ξ является просто η (ср. с задачей 4).

9. Задачи.

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

$$I_\varnothing = 0, \quad I_\Omega = 1, \quad I_{\bar{A}} = 1 - I_A,$$
$$I_{AB} = I_A \cdot I_B, \quad I_{A \cup B} = I_A + I_B - I_{AB},$$
$$I_{A \setminus B} = I_A(1 - I_B), \quad I_{A \triangle B} = (I_A - I_B)^2 = I_A + I_B \pmod 2,$$
$$I_{\bigcup_{i=1}^n A_i} = 1 - \prod_{i=1}^n (1 - I_{A_i}), \quad I_{\overline{\bigcup_{i=1}^n A_i}} = \prod_{i=1}^n (1 - I_{A_i}), \quad I_{\sum_{i=1}^n A_i} = \sum_{i=1}^n I_{A_i},$$

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

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

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

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

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

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

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

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

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

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

6. Пусть $P_\xi(x) = \ \{\xi=x\}$ и $F_\xi(x) = \ \{\xi \leqslant x\}$. 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

$$F_{\xi^2}(y) = F_\xi(+\sqrt{y}) - F_\xi(-\sqrt{y}) + P_\xi(-\sqrt{y}).$$

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 $\ \xi>0$, $\ \eta>0$ 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

$$\frac{\eta- \ \eta}{\sqrt{\ \eta}} = \frac{\xi- \ \xi}{\sqrt{\ \xi}}$$

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

$$\frac{\eta- \ \eta}{\sqrt{\ \eta}} = -\frac{\xi- \ \xi}{\sqrt{\ \xi}}$$

(и, значит, $a<0$).

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

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

9. Используя равенство

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

доказать формулу

$$(B_0) = 1 - S_1 + S_2 - \ldots \pm S_n$$

из задачи 5 § 1.

10. Если $\xi_1, \ldots, \xi_n$ — независимые случайные величины, $\varphi_1 = \varphi_1(\xi_1, \ldots$ $\ldots, \xi_k)$ 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, то 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 $F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \quad \{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\}$.

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 $\quad \{\xi\eta \leqslant z\}$ и $\quad \left\{\dfrac{\xi}{\eta} \leqslant z\right\}$ через вероятности $P_\xi(x)$ и $P_\eta(y)$.

15. Пусть ξ, η, ζ — случайные величины, $|\xi| \leqslant 1$, $|\eta| \leqslant 1$, $|\zeta| \leqslant 1$. Доказать справедливость *неравенства Белла*:

$$| \quad \xi\zeta - \quad \eta\zeta| \leqslant 1 - \quad \xi\eta.$$

(См., например, [136].)

16. В n урн независимым образом бросаются k шаров. (Для каждого шара вероятность его попадания в каждую конкретную урну равна $1/n$.) Найти математическое ожидание числа непустых урн.

§ 5. Схема Бернулли. I. Закон больших чисел

1. В соответствии с данными выше определениями тройка

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

называется вероятностной моделью, отвечающей n независимым испытаниям с двумя исходами, или *схемой Бернулли*.

В этом и следующем параграфах мы изучим некоторые предельные (в указываемом ниже смысле) свойства схем Бернулли, которые оказывается удобным ввести 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 $\xi_i(\omega)$, $i = 1, \ldots, n$, независимы и одинаково распределены:

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

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

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

Как было найдено выше, $S_n = np$ и, следовательно,

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

Иначе говоря, *среднее* значение частоты появления «успеха», т. е. величины 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

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

Действительно, при $0 < p < 1$

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

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

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

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

$$\xi \geqslant \varepsilon \quad I(\xi \geqslant \varepsilon) = \varepsilon \quad \{\xi \geqslant \varepsilon\},$$

что и доказывает (3). □

Следствия. *Пусть* ξ — *произвольная случайная величина. Тогда для* $\varepsilon > 0$

$$\begin{aligned} &\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\ |\xi|}{\varepsilon}, \\ &\{|\xi| \geqslant \varepsilon\} = \ \{\xi^2 \geqslant \varepsilon^2\} \leqslant \frac{\ \xi^2}{\varepsilon^2}, \\ &\{|\xi - \ \xi| \geqslant \varepsilon\} \leqslant \frac{\ \xi}{\varepsilon^2}. \end{aligned} \tag{4}$$

Воспользуемся последним неравенством, взяв $\xi = S_n/n$. Тогда с учетом (14) § 4 получим

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

Итак,

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

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

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

и, в сущности, в (5) 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 и *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 (при $p=1/2$).

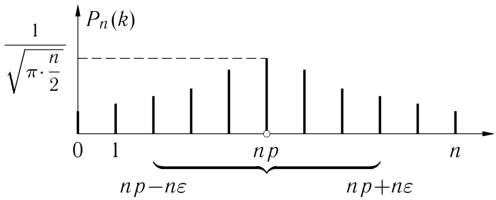


Рис. 6.

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:

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

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

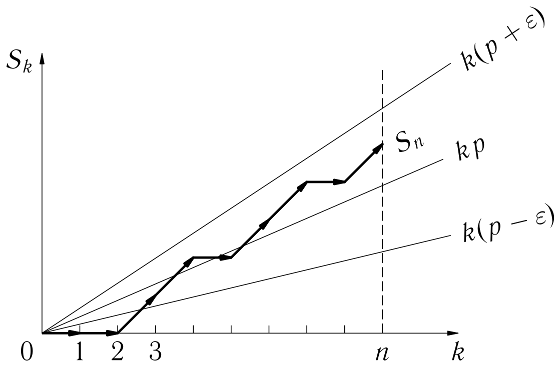


Рис. 7.

бесконечная последовательность независимых бернуллиевских случайных величин $\xi_1, \xi_2, \ldots$ 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 *элементарной* теории вероятностей, то можно утверждать лишь следующее.

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

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

и

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

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

Тогда

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

Утверждения типа (7)—(9) носят название

«Закон 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 и 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

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

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 из закона больших чисел можно заключить, что

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

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

Важно при этом 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 .

3. Рассмотрим полученную выше оценку

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

для ответа на следующий, типичный для *математической статистики* вопрос: каково наименьшее число наблюдений n, гарантирующее выполнение (для любого $0<p<1$) неравенства

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

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 $\left\{\left|\frac{S_n}{n}-p\right|\geqslant\varepsilon\right\}$.

4. Обозначим

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

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 $(n,\,\varepsilon)$-типичными).

Поставим следующий вопрос: каково число $N(C(n,\,\varepsilon))$ *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 соответственно 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 траекторий типа $(0,\,0,\,\ldots,\,0)$ или $(1,\,1,\,\ldots,\,1)$) будут типичными, 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 («энтропией»).

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 ln — натуральный логарифм и $0 \cdot \ln 0 = 0$. 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 $r = 2$).

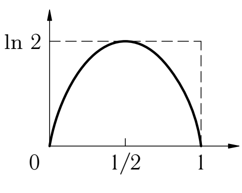


Рис. 8. График функции $H(p) = -p \ln p - (1-p) \ln(1-p)$

Если рассматривать распределение вероятностей $(p_1, p_2, \ldots, p_r)$ как вероятности появления некоторых событий, скажем, $A_1, A_2, \ldots, A_r$, 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 было бы иметь *количественную* характеристику *меры неопределенности* различных распределений вероятностей, что позволяло бы их сравнивать с этой стороны. Такой удачной характеристикой меры неопределенности как раз и оказалась *энтропия* H, играющая существенную роль в статистической механике, теории кодирования и теории связи, что видно из следующих рассмотрений.

Предположим, что пространство исходов

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

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

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

Ясно, что

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

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 $\left\{\left|\dfrac{\nu_i(\omega)}{n} - p_i\right| \geqslant \varepsilon\right\}$ 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 $\omega \in \Omega$ веса

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

Поэтому, если ω — *типичная* траектория, то в силу (14)

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

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

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

где

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

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

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

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

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

то

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

и аналогично

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

Поскольку $\quad (C(n, \varepsilon_1)) \to 1$, $n \to \infty$, то найдется n_1 такое, что для $n > n_1$ $\ (C(n, \varepsilon_1)) > 1 - \varepsilon$ 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 $n \geqslant n_0 = \max(n_1, n_2)$

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

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

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

$$B_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k p^k (1-p)^{n-k}, \quad 0 \leqslant p \leqslant 1, \ n \geqslant 0, \tag{15}$$

называемые *полиномами Бернштейна* по имени автора (С. Н. Бернштейна) приводимого доказательства теоремы Вейерштрасса.

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

$$f(S_n/n) = B_n(p).$$

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, $|f(x)| \leqslant \leqslant M < \infty$.

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 следует, что для полиномов Бернштейна (15)

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

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:

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

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

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

$$\ \{|\xi| \geqslant \varepsilon\} \geqslant \frac{\ f(\xi) - f(\varepsilon)}{f(C)}.$$

В частности, для $f(x) = x^2$

$$\frac{\ \xi^2 - \varepsilon^2}{C^2} \leqslant \ \{|\xi - \ \xi| \geqslant \varepsilon\} \leqslant \frac{\ \xi}{\varepsilon^2}.$$

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

$$\ \left\{\left|\frac{\xi_1+\ldots+\xi_n}{n} - \frac{\ (\xi_1+\ldots+\xi_n)}{n}\right| \geqslant \varepsilon\right\} \leqslant \frac{C}{n\varepsilon^2}. \tag{16}$$

(С теми же оговорками, какие были сделаны к соотношению (8), из неравенства (16) 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 $\ \{\xi_i = 1\} = p > 0$, $\ \{\xi_i = -1\} = 1 - p$. Имеет место следующая *оценка Бернштейна*: существует $a > 0$ такое, что

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

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

5. Пусть ξ — неотрицательная случайная величина, $a > 0$. Найти точную верхнюю грань для $\ \{x \geqslant a\}$, если известно, что

(i) $\ \xi = 20$;

(ii) $\ \xi = 20$, $\ \xi = 25$;

(iii) $\ \xi = 20$, $\ \xi = 25$ и ξ симметрична относительно своего среднего значения.

§ 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

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

и в силу (14) § 4

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

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 $\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\}$. Естественно думать, что напрашивающемуся 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

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

или, что то же, вероятности

$$\left\{\left|\frac{S_n - \ S_n}{\sqrt{\ S_n}}\right| \leqslant x\right\}$$

(поскольку $\ S_n = np$ и $\ S_n = npq$).

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

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

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 $n\to\infty$

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

где $\varphi(n)$ — *любая неотрицательная функция такая, что* $\varphi(n)=$ $=o(npq)^{2/3}$.

Доказательство существенно использует формулу Стирлинга (6) § 2

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

где $R(n)\to 0$, $n\to\infty$.

Согласно этой формуле, если $n\to\infty$, $k\to\infty$, $n-k\to\infty$, то

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

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

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

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

$$
\begin{aligned}
P_n(k) &= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}}\left(\frac{p}{\hat{p}}\right)^k\left(\frac{1-p}{1-\hat{p}}\right)^{n-k}(1+\varepsilon)= \\
&= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}}\ \exp\left\{k\ \ln\ \frac{p}{\hat{p}}+(n-k)\ \ln\ \frac{1-p}{1-\hat{p}}\right\}\cdot(1+\varepsilon)= \\
&= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}}\ \exp\left\{n\left[\frac{k}{n}\ \ln\ \frac{p}{\hat{p}}+\left(1-\frac{k}{n}\right)\ \ln\ \frac{1-p}{1-\hat{p}}\right]\right\}\cdot(1+\varepsilon)= \\
&= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}}\ \exp\{-nH(\hat{p})\}\cdot(1+\varepsilon),
\end{aligned}
$$

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 $0<x<1$

$$
\begin{aligned}
H'(x) &= \ln\ \frac{x}{p}-\ln\ \frac{1-x}{1-p}, \\
H''(x) &= \frac{1}{x}+\frac{1}{1-x}, \\
H'''(x) &= -\frac{1}{x^2}+\frac{1}{(1-x)^2},
\end{aligned}
$$

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

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

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,

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

если sup брать по тем k, для которых

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

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* $\{0,\,1,\,\ldots,\,n\}$,

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

т. е. при $n\to\infty$

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

где $\psi(n)=o(npq)^{1/6}$.

С учетом замечаний, сделанных по поводу формулы (8) § 5, полученные результаты на вероятностном языке можно переформулировать следующим образом:

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

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

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

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

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

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

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

Перейдем к точным формулировкам.

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

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

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 $|t_k| \leqslant \leqslant T < \infty$,

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

где

$$\sup_{|t_k| \leqslant T} |\varepsilon(t_k,\, n)| \to 0, \qquad n \to \infty. \tag{13}$$

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

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

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

где

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

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

Из известных свойств интегральных сумм

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

Ясно также, что

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

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

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

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

Обозначим

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

Тогда из (14)—(16) вытекает, что

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

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

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

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

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

Отсюда и из (19) следует, что

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

и, значит,

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

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 $-\infty \leqslant -T \leqslant a \leqslant b \leqslant T \leqslant \infty$

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

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

$$\leqslant \frac{\varepsilon}{4} + P_n(-\infty,\, -T] + \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-T} e^{-x^2/2}\, dx + P_n(T,\, \infty) + \frac{1}{\sqrt{2\pi}} \int\limits_T^{\infty} e^{-x^2/2}\, dx \leqslant$$

$$\leqslant \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{8} + \frac{\varepsilon}{8} = \varepsilon.$$

С учетом (18) отсюда легко выводится, что $P_n(a,\, b]$ стремится к $\Phi(b) - \Phi(a)$ *равномерно* по всем $-\infty \leqslant a < b \leqslant \infty$.

Итак, доказана

Интегральная теорема Муавра—Лапласа. *Пусть* $0 < p < 1$,

$$P_n(k) = C_n^k p^k q^{n-k}, \qquad P_n(a,\, b] = \sum_{a < x \leqslant b} P_n(np + x\sqrt{npq}).$$

Тогда

$$\sup_{-\infty \leqslant a < b \leqslant \infty} \left|P_n(a,\, b] - \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2}\, dx\right| \to 0, \quad n \to \infty. \tag{21}$$

С точностью до тех же самых замечаний, которые были сделаны по поводу соотношения (8) § 5, результат (21) можно на вероятностном языке сформулировать следующим образом:

$$\sup_{-\infty \leqslant a < b \leqslant \infty} \left| \quad \left\{a < \frac{S_n - \quad S_n}{\sqrt{\quad S_n}} \leqslant b\right\} - \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2}\, dx\right| \to 0, \quad n \to \infty.$$

Из этой формулы сразу следует, что для любых $-\infty \leqslant A < B \leqslant \infty$

$$\{A < S_n \leqslant B\} - \left[\Phi\left(\frac{B - np}{\sqrt{npq}}\right) - \Phi\left(\frac{A - np}{\sqrt{npq}}\right)\right] \to 0, \quad n \to \infty. \tag{22}$$

Пример. 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 «на руках» представляет весьма трудоемкую задачу. Если же воспользоваться интегральной теоремой, то найдем, что интересующая нас вероятность P примерно равна ($n = 12\,000$, $p = \dfrac{1}{6}$, $A = 1800$, $B = 2100$)

$$\Phi\left(\frac{2100-2000}{\sqrt{12000\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right)-\Phi\left(\frac{1800-2000}{\sqrt{12000\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right)=$$
$$=\Phi(\sqrt{6})-\Phi(-2\sqrt{6})\approx\Phi(2{,}449)-\Phi(-4{,}898)\approx 0{,}992,$$

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 график (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 $\dfrac{1}{\sqrt{2\pi npq}}e^{-x^2/2}$. Интегральная же теорема говорит о том, что вероятность

$$P_n(a,\, b] = \quad \{a\sqrt{npq} < S_n - np \leqslant b\sqrt{npq}\} =$$
$$= \quad \{np + a\sqrt{npq} < S_n \leqslant np + b\sqrt{npq}\}$$

хорошо аппроксимируется интегралом $(1/\sqrt{2\pi})\int\limits_a^b e^{-x^2/2}\,dx$.

Обозначим

$$F_n(x) = P_n(-\infty,\, x] \quad \left(= \quad \left\{\frac{S_n - np}{\sqrt{npq}} \leqslant x\right\}\right).$$

Тогда из (21) следует, что

$$\sup_{-\infty\leqslant x\leqslant\infty} |F_n(x) - \Phi(x)| \to 0, \qquad n \to \infty. \tag{23}$$

Интересно было бы понять, насколько *быстро* с ростом n происходит стремление к нулю в (21) и (23). Приведем результат, относящийся сюда и

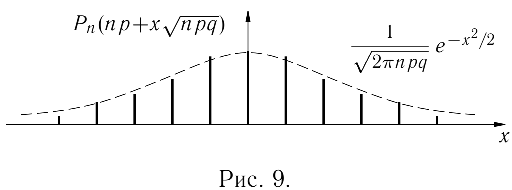


Рис. 9.

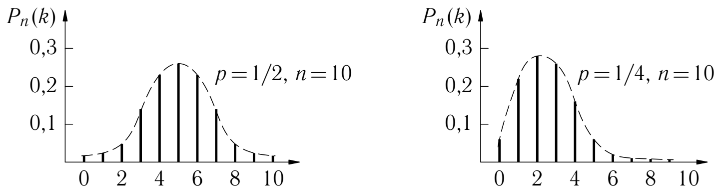


Рис. 10.

являющийся частным случаем так называемой *теоремы Берри—Эссеена* (§ 11 гл. III):

$$\sup_{-\infty \leqslant x \leqslant \infty} |F_n(x) - \Phi(x)| \leqslant \frac{p^2 + q^2}{\sqrt{npq}}. \tag{24}$$

Важно подчеркнуть, что порядок оценки $1/\sqrt{npq}$ 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, слева). Однако при *малых* значениях p биномиальное распределение приобретает *асимметричную* форму (см. рис. 10, справа), 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

$$P_n(k) = \begin{cases} C_n^k p^k q^{n-k}, & k = 0,\ 1,\ \ldots,\ n, \\ 0, & k = n+1,\ n+2,\ \ldots, \end{cases}$$

и предположим сейчас, что p является функцией от n, $p = p(n)$.

Теорема Пуассона. *Пусть* $p(n) \to 0$, $n \to \infty$, *причем так, что* $np(n) \to \lambda$, *где* $\lambda > 0$. *Тогда для любого* $k = 0,\ 1,\ \ldots$

$$P_n(k) \to \pi_k, \qquad n \to \infty, \tag{25}$$

где

$$\pi_k = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0,\ 1,\ \ldots \tag{26}$$

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 ($\pi_k > 0$, $\sum_{k=0}^{\infty} \pi_k = 1$). 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 следующий результат (Ю. В. Прохоров), показывающий, с какой скоростью величины $P_n(k)$ сходятся к π_k при $n \to \infty$: *если* $np(n) = \lambda$, *то* *)

$$\sum_{k=0}^{\infty} |P_n(k) - \pi_k| \leqslant \frac{2\lambda}{n} \cdot \min(2, \lambda). \tag{27}$$

5. Вернемся к предельной теореме Муавра—Лапласа. Покажем, как из нее следует закон больших чисел (с оговорками, сделанными в § 5 к формуле (8)). Поскольку

$$\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} = \left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant \varepsilon\sqrt{\frac{n}{pq}}\right\},$$

то из (21) ясно, что для $\varepsilon > 0$

$$\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} - \frac{1}{\sqrt{2\pi}} \int\limits_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} e^{-x^2/2}\, dx \to 0, \quad n \to \infty, \tag{28}$$

*) Доказательство несколько более слабого результата дано в § 12 гл. III.

откуда

$$\left\{\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right\}\to 1,\quad n\to\infty,$$

что и составляет утверждение закона больших чисел.

Из (28)

$$\left\{\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right\}\sim\frac{1}{\sqrt{2\pi}}\int\limits_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}}e^{-x^2/2}\,dx,\quad n\to\infty,\tag{29}$$

в то время как неравенство Чебышева давало лишь оценку

$$\left\{\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right\}\geqslant 1-\frac{pq}{n\varepsilon^2}.$$

В п. 3 § 5 было показано, что для числа наблюдений, необходимого для справедливости соотношения

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

неравенство Чебышева дает следующую оценку:

$$n\geqslant\left[\frac{1}{4\varepsilon^2\alpha}\right]\quad(=n_1(\alpha)),$$

где $[x]$ — целая часть числа x. 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

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-k(\alpha)}^{k(\alpha)}e^{-x^2/2}\,dx=1-\alpha.$$

Поскольку $\varepsilon\sqrt{\dfrac{n}{pq}}\geqslant 2\varepsilon\sqrt{n}$, то, определяя (наименьшее целое) n из неравенства

$$2\varepsilon\sqrt{n}\geqslant k(\alpha),\tag{30}$$

получим, что

$$\left\{\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right\}\geqslant 1-\alpha.\tag{31}$$

Из (30) находим, что $n=n_2(\alpha)$ с

$$n_2(\alpha)=\left[\frac{k^2(\alpha)}{4\varepsilon^2}\right],$$

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	0,50	0,3173	0,10	0,05	0,0454	0,01	0,0027
$k(\alpha)$	0,675	1,000	1,645	1,960	2,000	2,576	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 (непрерывных) 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). Остановимся сейчас

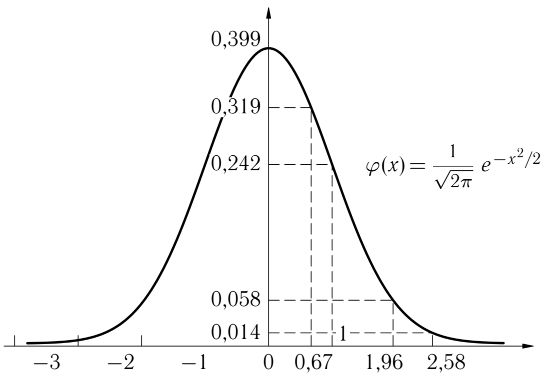


Рис. 11. График плотности $\varphi(x)$ нормального распределения

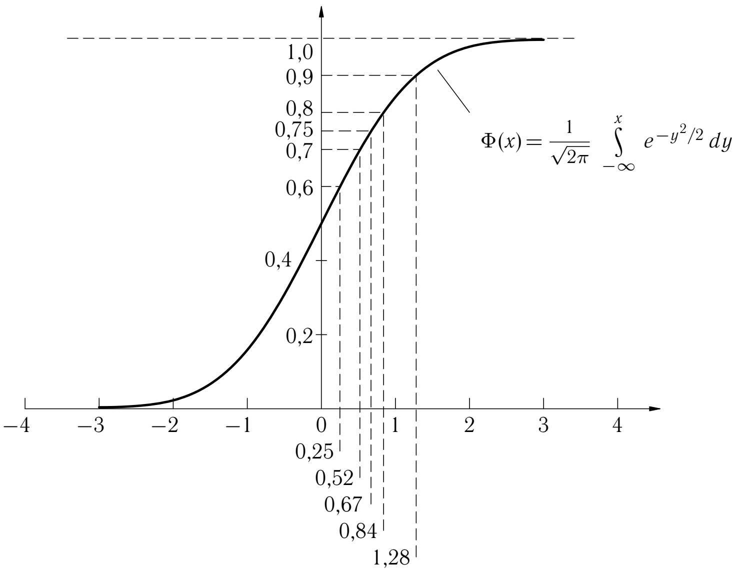


Рис. 12. График функции нормального распределения $\Phi(x)$

на некоторых простейших свойствах функций $\varphi(x)$ и $\Phi(x)$, графики которых приведены рис. 11 и 12.

Функция $\varphi(x)$ является симметричной колоколообразной кривой, убывающей с ростом $|x|$ очень быстро: $\varphi(1) = 0{,}24197$, $\varphi(2) = 0{,}053991$, $\varphi(3) = 0{,}004432$, $\varphi(4) = 0{,}000134$, $\varphi(5) = 0{,}000016$. Максимум этой кривой достигается в точке $x = 0$ и равен $(2\pi)^{-1/2} \approx 0{,}399$.

Кривая

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

быстро приближается с ростом x к единице:

$$\Phi(1) = 0{,}841345, \qquad \Phi(2) = 0{,}977250, \qquad \Phi(3) = 0{,}998650,$$
$$\Phi(4) = 0{,}999968, \qquad \Phi(5) = 0{,}999997.$$

По поводу *таблиц* функций $\varphi(x)$ и $\Phi(x)$, а также других основных функций, используемых в теории вероятностей и математической статистике, см. [6].

Полезно отметить, что при расчетах, наряду с функцией $\Phi(x)$, часто используется родственная ей *функция ошибок*

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt, \qquad x > 0.$$

Очевидно, что $(x > 0)$

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \qquad \operatorname{erf}(x) = 2\Phi(\sqrt{2}x) - 1.$$

7. В конце п. 3 § 5 было отмечено, что оценка сверху для вероятности события $\left\{\omega\colon \left|\dfrac{S_n}{n} - p\right| \geqslant \varepsilon\right\}$, даваемая неравенством Чебышева, является достаточно грубой. Получение этой оценки основывалось на неравенстве Чебышева

$$\{X \geqslant \varepsilon\} \leqslant \frac{X^2}{\varepsilon^2}$$

для неотрицательных случайных величин $X \geqslant 0$. Можно попытаться, однако, воспользоваться неравенством Чебышева в форме

$$\{X \geqslant \varepsilon\} = \quad \{X^{2k} \geqslant \varepsilon^{2k}\} \leqslant \frac{X^{2k}}{\varepsilon^{2k}}. \tag{33}$$

Но можно пойти и дальше, а именно воспользоваться «экспоненциальной» формой неравенства Чебышева: если $X \geqslant 0$ и $\lambda > 0$, то

$$\{X \geqslant \varepsilon\} = \quad \{e^{\lambda X} \geqslant e^{\lambda\varepsilon}\} \leqslant \quad e^{\lambda(X-\varepsilon)}. \tag{34}$$

В силу произвольности $\lambda > 0$ ясно, что

$$\{X \geqslant \varepsilon\} \leqslant \inf_{\lambda>0} \quad e^{\lambda(X-\varepsilon)}. \tag{35}$$

Посмотрим, к чему приводит этот путь в случае $X = S_n/n$,

$$S_n = \xi_1 + \ldots + \xi_n, \qquad \{\xi_i = 1\} = p, \qquad \{\xi_i = 0\} = q, \quad i \geqslant 1.$$

Обозначим $\varphi(\lambda) = \quad e^{\lambda\xi_1}$. Тогда

$$\varphi(\lambda) = 1 - p + pe^{\lambda}$$

и в предположении независимости величин $\xi_1, \xi_2, \ldots, \xi_n$

$$e^{\lambda S_n} = [\varphi(\lambda)]^n.$$

Поэтому $(0 < a < 1)$

$$\left\{\frac{S_n}{n} \geqslant a\right\} \leqslant \inf_{\lambda>0} \quad e^{\lambda\left(\frac{S_n}{n} - a\right)} = \inf_{\lambda>0} e^{-n\left[\frac{\lambda}{n}a - \ln\varphi\left(\frac{\lambda}{n}\right)\right]} =$$
$$= \inf_{s>0} e^{-n[as - \ln\varphi(s)]} = e^{-n \sup\limits_{s>0}[as - \ln\varphi(s)]}. \tag{36}$$

Аналогично

$$\left\{\frac{S_n}{n} \leqslant a\right\} \leqslant e^{-n \sup\limits_{s<0}[as - \ln \varphi(s)]}. \tag{37}$$

Функция $f(s) = as - \ln[1 - p + pe^s]$ при $p \leqslant a \leqslant 1$ достигает максимума в точке s_0 ($f'(s_0) = 0$), определяемой из равенства

$$e^{s_0} = \frac{a(1-p)}{p(1-a)}.$$

Поэтому

$$\sup_{s>0} f(s) = H(a),$$

где

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

— функция, уже рассмотренная выше при доказательстве локальной теоремы (п. 1).

Итак, при $p \leqslant a \leqslant 1$

$$\left\{\frac{S_n}{n} \geqslant a\right\} \leqslant e^{-nH(a)}, \tag{38}$$

и поскольку $H(p+x) \geqslant 2x^2$, $0 \leqslant p + x \leqslant 1$, то при $\varepsilon > 0$, $0 \leqslant p \leqslant 1$

$$\left\{\frac{S_n}{n} - p \geqslant \varepsilon\right\} \leqslant e^{-2n\varepsilon^2}. \tag{39}$$

Аналогичным образом устанавливается, что для $a \leqslant p \leqslant 1$

$$\left\{\frac{S_n}{n} \leqslant a\right\} \leqslant e^{-nH(a)}, \tag{40}$$

и, следовательно, для всякого $\varepsilon > 0$, $0 \leqslant p \leqslant 1$

$$\left\{\frac{S_n}{n} - p \leqslant -\varepsilon\right\} \leqslant e^{-2n\varepsilon^2}. \tag{41}$$

Тем самым

$$\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant 2e^{-2n\varepsilon^2}. \tag{42}$$

Отсюда вытекает, что число наблюдений $n_3(\alpha)$, гарантирующее выполнение при любых $0 \leqslant p \leqslant 1$ неравенства

$$\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha, \tag{43}$$

определяется формулой

$$n_3(\alpha) = \left[\frac{\ln(2/\alpha)}{2\varepsilon^2}\right], \tag{44}$$

где $[x]$ — целая часть числа x. Пренебрегая «целыми частями» и сравнивая $n_3(\alpha)$ с $n_1(\alpha) = \left[\frac{1}{4\alpha\varepsilon^2}\right]$, находим, что

$$\frac{n_1(\alpha)}{n_3(\alpha)} = \frac{1}{2\alpha \ln \frac{2}{\alpha}} \uparrow \infty, \quad \alpha \downarrow 0.$$

Отсюда видно, что при $\alpha \downarrow 0$ оценка *минимального* необходимого числа наблюдений, полученная с помощью экспоненциального неравенства Чебышева, является более точной, чем оценка, найденная из обычного неравенства Чебышева, особенно при малых α. Воспользовавшись без труда («правило Лопиталя») устанавливаемым соотношением

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\, dy \sim \frac{1}{\sqrt{2\pi}\, x} e^{-x^2/2}, \quad x \to \infty,$$

можно показать, что $k^2(\alpha) \sim 2 \ln \frac{2}{\alpha}$, $\alpha \downarrow 0$. Тем самым

$$\frac{n_2(\alpha)}{n_3(\alpha)} \to 1, \quad \alpha \downarrow 0.$$

Неравенства типа (38)—(42) носят название *неравенств для вероятностей больших уклонений*. Объяснение этому названию следующее.

Интегральная теорема Муавра—Лапласа дает возможность просто оценивать вероятности событий $\{|S_n - np| \leqslant x\sqrt{n}\}$, характеризующих «стандартное» уклонение (на величину порядка $\sqrt{n}$) S_n от np. Неравенства же (39), (41) и (42) дают оценку для вероятностей событий $\{\omega\colon |S_n - np| \leqslant \leqslant xn\}$, описывающих уклонения порядка большего, нежели $\sqrt{n}$, а именно порядка n.

Мы продолжим рассмотрение вопроса о вероятностях больших уклонений в более общих случаях в § 5 гл. IV.

8. 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

$$\{10 < S_{100} \leqslant 12\}, \qquad \{20 < S_{100} \leqslant 22\},$$
$$\{33 < S_{100} \leqslant 35\}, \qquad \{40 < S_{100} \leqslant 42\},$$
$$\{50 < S_{100} \leqslant 52\}$$

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

$$\sup_j \left|\sqrt{\pi n}\quad\{Z_{2n}=j\}-e^{-j^2/4^n}\right|\to 0, \quad n\to\infty.$$

3. Доказать, что в теореме Пуассона (с $p=\lambda/n$) имеет место следующая скорость сходимости:

$$\sup_k \left|P_n(k)-\frac{\lambda^k e^{-\lambda}}{k!}\right|\leqslant \frac{\lambda^2}{n}.$$

(Для доказательства полезно ознакомиться с § 12 гл. III.)

§ 7. Оценка вероятности «успеха» в схеме Бернулли

1. В рассмотренной выше схеме Бернулли $(\Omega, \mathscr{A}, \quad)$ с $\Omega=\{\omega: \omega= =(x_1, \ldots, x_n),\ x_i=0,\ 1\}$, $\mathscr{A}=\{A: A\subseteq\Omega\}$, $\quad(\{\omega\})=p(\omega)$, где

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

предполагалось, что число p (вероятность «успеха») *известно*.

Представим теперь, что p заранее не известно 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 $\Theta=[0, 1]$. Часто говорят, что набор

$$\mathscr{E}=(\Omega, \mathscr{A}, \quad_\theta;\ \theta\in\Theta) \quad \text{с} \quad _\theta(\{\omega\})=\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

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, называют *оценкой*.

Если $S_n=\xi_1+\ldots+\xi_n$ и $T_n^*=\dfrac{S_n}{n}$, то из закона больших чисел следует, что оценка T_n^* является *состоятельной* в том смысле, что $(\varepsilon>0)$

$$_\theta\{|T_n^*-\theta|\geqslant\varepsilon\}\to 0, \quad n\to\infty. \tag{1}$$

Кроме того, эта оценка является *несмещенной*: для всякого $\theta\in\Theta$

$$_\theta T_n^*=\theta, \tag{2}$$

где $\quad_\theta$ — математическое ожидание, отвечающее вероятности $\quad_\theta$.

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

$$T_n = \frac{b_1\xi_1 + \ldots + b_n\xi_n}{n},$$

где $b_1 + \ldots + b_n = n$. При этом для таких оценок также будет выполняться закон больших чисел (1) (по крайней мере, если $|b_i| \leqslant K < \infty$), 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 $\widetilde{T}_n$ *эффективной* (в классе несмещенных оценок T_n), если

$$ {}_\theta \widetilde{T}_n = \inf_{T_n} \quad {}_\theta T_n, \qquad \theta \in \Theta, \tag{3}$$

где ${}_\theta T_n$ — дисперсия оценки T_n, т. е. величина ${}_\theta (T_n - \theta)^2$.

Покажем, что рассмотренная выше оценка T_n^* является эффективной. Имеем

$$ {}_\theta T_n^* = {}_\theta \left(\frac{S_n}{n} \right) = \frac{{}_\theta S_n}{n^2} = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}. \tag{4}$$

Поэтому, для того чтобы установить, что оценка T_n^* эффективна, достаточно показать, что

$$\inf_{T_n} \quad {}_\theta T_n \geqslant \frac{\theta(1-\theta)}{n}. \tag{5}$$

При $\theta = 0$ или 1 это неравенство очевидно. Пусть теперь $\theta \in (0, 1)$ и

$$p_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}.$$

Ясно, что ${}_\theta(\{\omega\}) = p_\theta(\omega)$, где

$$p_\theta(\omega) = \prod_{i=1}^{n} p_\theta(x_i).$$

Обозначим

$$L_\theta(\omega) = \ln\, p_\theta(\omega).$$

Тогда

$$L_\theta(\omega) = \ln\theta \cdot \sum x_i + \ln(1-\theta) \cdot \sum (1 - x_i)$$

и

$$\frac{\partial L_\theta(\omega)}{\partial\theta} = \frac{\sum(x_i-\theta)}{\theta(1-\theta)}.$$

Поскольку

$$1 \equiv \quad {}_\theta 1 = \sum_\omega p_\theta(\omega)$$

и в силу несмещенности оценки T_n

$$\theta \equiv \quad {}_\theta T_n = \sum_\omega T_n(\omega)\, p_\theta(\omega),$$

то после дифференцирования по θ получим, что

$$0 = \sum_\omega \frac{\partial p_\theta(\omega)}{\partial\theta} = \sum_\omega \frac{\dfrac{\partial p_\theta(\omega)}{\partial\theta}}{p_\theta(\omega)}\, p_\theta(\omega) = \quad {}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right],$$

$$1 = \sum_\omega T_n \frac{\dfrac{\partial p_\theta(\omega)}{\partial\theta}}{p_\theta(\omega)}\, p_\theta(\omega) = \quad {}_\theta\left[T_n \frac{\partial L_\theta(\omega)}{\partial\theta}\right].$$

Значит,

$$1 = \quad {}_\theta\left[(T_n-\theta)\frac{\partial L_\theta(\omega)}{\partial\theta}\right]$$

и, согласно неравенству Коши—Буняковского,

$$1 \leqslant \quad {}_\theta[T_n-\theta]^2 \cdot \quad {}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2,$$

откуда

$$ {}_\theta[T_n-\theta]^2 \geqslant \frac{1}{I_n(\theta)}, \tag{6}$$

где величина

$$I_n(\theta) = \quad {}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2$$

носит название *информации Фишера*.

Из (6) получаем частный случай так называемого *неравенства Рао—Крамера* для несмещенных оценок T_n

$$\inf_{T_n} \quad {}_\theta T_n \geqslant \frac{1}{I_n(\theta)}. \tag{7}$$

В рассматриваемом случае

$$I_n(\theta) = \quad {}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2 = \quad {}_\theta\left[\frac{\sum(\xi_i-\theta)}{\theta(1-\theta)}\right]^2 = \frac{n\theta(1-\theta)}{[\theta(1-\theta)]^2} = \frac{n}{\theta(1-\theta)},$$

что и доказывает неравенство (5), из которого, как уже отмечалось, следует эффективность несмещенной оценки $T_n^* = S_n/n$ 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 ω величины $T_n^* = T_n^*(\omega)$ 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 $\{|\theta - T_n^*| > \delta\}$ будет достаточно мала.

Согласно неравенству Чебышева,

$$ {}_{\theta}\{|\theta - T_n^*| > \delta\} \leqslant \frac{{}_{\theta} T_n^*}{\delta^2} = \frac{\theta(1-\theta)}{n\delta^2} $$

и, значит, для всякого $\lambda > 0$

$$ {}_{\theta}\left\{|\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} \geqslant 1 - \frac{1}{\lambda^2}. $$

Если взять, к примеру, $\lambda = 3$, то найдем, что с ${}_{\theta}$-вероятностью, большей, чем 0,8888 ($1 - \frac{1}{3^2} = \frac{8}{9} \approx 0{,}8888$), 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,

$$ {}_{\theta}\left\{|\theta - T_n^*| \leqslant \frac{3}{2\sqrt{n}}\right\} = {}_{\theta}\left\{T_n^* - \frac{3}{2\sqrt{n}} \leqslant \theta \leqslant T_n^* + \frac{3}{2\sqrt{n}}\right\} \geqslant 0{,}8888. $$

Иначе говоря, можно утверждать, что истинное значение параметра θ принадлежит интервалу $\left[T_n^* - \frac{3}{2\sqrt{n}},\ T_n^* + \frac{3}{2\sqrt{n}}\right]$ с вероятностью, большей, чем 0,8888. 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 $\left[T_n^* - \dfrac{3}{2\sqrt{n}},\ T_n^* + \dfrac{3}{2\sqrt{2}}\right]$ является примером так называемых доверительных интервалов для неизвестного параметра.

Определение. Множество вида

$$[\psi_1(\omega),\ \psi_2(\omega)],$$

где $\psi_1(\omega)$ и $\psi_2(\omega)$ — две функции элементарных событий, назовем *доверительным интервалом с коэффициентом надежности* $1-\delta$ (или с *уровнем значимости* δ), если для всех $\theta \in \Theta$

$$_\theta\{\psi_1(\omega) \leqslant \theta \leqslant \psi_2(\omega)\} \geqslant 1-\delta.$$

Приведенные выше рассуждения показывают, что интервал

$$\left[T_n^* - \frac{\lambda}{2\sqrt{n}},\ T_n^* + \frac{\lambda}{2\sqrt{n}}\right]$$

имеет надежность $1 - \dfrac{1}{\lambda^2}$. 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

$$\left\{\omega\colon\ |\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} = \{\omega\colon\ \psi_1(T_n^*,\ n) \leqslant \theta \leqslant \psi_2(T_n^*,\ n)\},$$

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

$$F_\theta^n(x) = \ _\theta\left\{\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leqslant x\right\}.$$

Тогда в силу (24) § 6

$$\sup_x |F_\theta^n(x) - \Phi(x)| \leqslant \frac{1}{\sqrt{n\theta(1-\theta)}}.$$

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,

$$
\begin{aligned}
{}_{\theta}\{\psi_1(T_n^*,\, n) \leqslant \theta \leqslant \psi_2(T_n^*,\, n)\} = {}&{}_{\theta}\left\{|\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} = \\
= {}&{}_{\theta}\left\{\frac{|S_n - n\theta|}{\sqrt{n\theta(1-\theta)}} \leqslant \lambda\right\} \geqslant (2\Phi(\lambda) - 1) - \frac{2}{\Delta\sqrt{n}}.
\end{aligned}
$$

Пусть λ^* — то наименьшее λ, для которого

$$
(2\Phi(\lambda) - 1) - \frac{2}{\Delta\sqrt{n}} \leqslant 1 - \delta^*,
$$

где δ^* — заданный уровень значимости. Обозначая $\delta = \delta^* - \dfrac{2}{\Delta\sqrt{n}}$, находим, что λ^* есть корень уравнения

$$
\Phi(\lambda) = 1 - \frac{\delta}{2}.
$$

В случае больших n можно пренебречь членом $2/\Delta\sqrt{n}$ и считать, что λ^* удовлетворяет соотношению

$$
\Phi(\lambda^*) = 1 - \frac{\delta^*}{2}.
$$

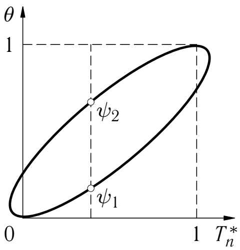


Рис. 13.

В частности, если $\lambda^* = 3$, то $1 - \delta^* = 0{,}9973\ldots$ Так что с вероятностью, *примерно* равной 0,9973,

$$
T_n^* - 3\sqrt{\frac{\theta(1-\theta)}{n}} \leqslant \theta \leqslant T_n^* + 3\sqrt{\frac{\theta(1-\theta)}{n}}, \tag{8}
$$

или, после итерирования и отбрасывания членов порядка $O(n^{-3/4})$,

$$
T_n^* - 3\sqrt{\frac{T_n^*(1-T_n^*)}{n}} \leqslant \theta \leqslant T_n^* + 3\sqrt{\frac{T_n^*(1-T_n^*)}{n}}. \tag{9}
$$

Отсюда следует, что доверительный интервал

$$
\left[T_n^* - \frac{3}{2\sqrt{n}},\; T_n^* + \frac{3}{2\sqrt{n}}\right] \tag{10}
$$

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 $\dfrac{3}{2\sqrt{n}}$. (См. по этому поводу также конец § 5.)

3. Задачи.

1. Пусть *a priori* известно, что параметр θ принимает значения в $\Theta_0 \subseteq \subseteq [0, 1]$. Выяснить, когда существует несмещенная оценка для параметра θ, принимающая значения лишь в 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.

4. В дополнение к задаче 5 в § 2 исследовать вопрос о несмещенности и эффективности оценки $\widehat{N}$, считая N достаточно большим, $N \gg M$, $N \gg n$. Построить, по аналогии с доверительными интервалами для параметра θ (см. формулы (8) и (9)), доверительные интервалы $[\widehat{N} - a(\widehat{N}),\ \widehat{N} + b(\widehat{N})]$ для N такие, что

$$_{N,M;n}\{\widehat{N} - a(\widehat{N}) \leqslant N \leqslant \widehat{N} + b(\widehat{N})\} \approx 1 - \alpha,$$

где α — некоторое малое число.

§ 8. Условные вероятности и математические ожидания относительно разбиений

1. Пусть $(\Omega, \mathscr{A},\ \ )$ — конечное вероятностное пространство и

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

— некоторое *разбиение* Ω ($D_i \in \mathscr{A}$, $\ (D_i) > 0$, $i = 1, \ldots, k$, и $D_1 + \ldots + D_k = \Omega$). Пусть, далее, A — событие из $\mathscr{A}$ и $\ (A \mid D_i)$ — условная вероятность события A относительно события D_i.

С набором условных вероятностей $\{\ (A \mid D_i),\ i = 1, \ldots, k\}$ можно связать случайную величину

$$\pi(\omega) = \sum_{i=1}^{k} \ (A \mid D_i) I_{D_i}(\omega) \tag{1}$$

(ср. с (5) § 4), принимающую на атомах разбиения D_i значения $\ (A \mid D_i)$. Чтобы подчеркнуть, что эта *случайная величина* связана именно с разбиением $\mathscr{D}$, ее обозначают

$$\ (A \mid \mathscr{D}) \text{ или } \ (A \mid \mathscr{D})(\omega)$$

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.

Следующие два свойства условных вероятностей очевидны:

$$(A+B\,|\,\mathscr{D}) = \quad (A\,|\,\mathscr{D}) + \quad (B\,|\,\mathscr{D}); \tag{2}$$

если $\mathscr{D}$ — тривиальное разбиение, состоящее из одного множества Ω, то

$$(A\,|\,\mathscr{D}) = \quad (A). \tag{3}$$

Определение условной вероятности $\quad (A\,|\,\mathscr{D})$ как *случайной величины* дает возможность говорить о ее математическом ожидании, используя которое можно следующим компактным образом записать *формулу полной вероятности* (3) § 3:

$$(A\,|\,\mathscr{D}) = \quad (A). \tag{4}$$

Действительно, поскольку

$$(A\,|\,\mathscr{D})(\omega) = \sum_{i=1}^{k} \quad (A\,|\,D_i) I_{D_i}(\omega),$$

то по определению математического ожидания (см. (5) и (6) § 4)

$$(A\,|\,\mathscr{D}) = \sum_{i=1}^{k} \quad (A\,|\,D_i) \quad (D_i) = \sum_{i=1}^{k} \quad (AD_i) = \quad (A).$$

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 $\quad (A\,|\,\mathscr{D}_\eta)$ будем в дальнейшем обозначать $\quad (A\,|\,\eta)$ или $\quad (A\,|\,\eta)(\omega)$ и называть *условной вероятностью события A относительно случайной величины η*. Условимся также под $\quad (A\,|\,\eta = y_j)$ понимать условную вероятность $\quad (A\,|\,D_j)$, 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

$$D_{y_1, y_2, \ldots, y_m} = \{\omega\colon \eta_1(\omega) = y_1, \eta_2(\omega) = y_2, \ldots, \eta_m(\omega) = y_m\},$$

то $\quad (A\,|\,\mathscr{D}_{\eta_1, \eta_2, \ldots, \eta_m})$ обозначается $\quad (A\,|\,\eta_1, \eta_2, \ldots, \eta_m)$ 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 $(A\,|\,\eta)$ 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, то при $\{\eta=y\}>0$ имеем

$$(\xi+\eta=z\,|\,\eta=y)=\ \{\xi+y=z\}. \tag{5}$$

В самом деле,

$$\begin{aligned}(\xi+\eta=z\,|\,\eta=y)&=\frac{\{\xi+\eta=z,\ \eta=y\}}{\{\eta=y\}}=\frac{\{\xi+y=z,\ \eta=y\}}{\{\eta=y\}}=\\&=\frac{\{\xi+y=z\}\ \ \{\eta=y\}}{\{\eta=y\}}=\ \{\xi+y=z\}.\end{aligned}$$

Используя эту формулу, находим, что

$$\begin{aligned}(A\,|\,\eta)(\omega)&=\ (\xi+\eta=k\,|\,\eta)(\omega)=\\&=\ (\xi+\eta=k\,|\,\eta=0)I_{\{\eta=0\}}(\omega)+\ (\xi+\eta=k\,|\,\eta=1)I_{\{\eta=1\}}(\omega)=\\&=\ \{\xi=k\}I_{\{\eta=0\}}(\omega)+\ \{\xi=k-1\}I_{\{\eta=1\}}(\omega).\end{aligned}$$

Итак,

$$(\xi+\eta=k\,|\,\eta)(\omega)=\begin{cases}qI_{\{\eta=0\}}(\omega), & k=0,\\ pI_{\{\eta=0\}}(\omega)+qI_{\{\eta=1\}}(\omega), & k=1,\\ pI_{\{\eta=1\}}(\omega), & k=2,\end{cases} \tag{6}$$

или, что то же самое,

$$(\xi+\eta=k\,|\,\eta)=\begin{cases}q(1-\eta), & k=0,\\ p(1-\eta)+q\eta, & k=1,\\ p\eta, & k=2.\end{cases} \tag{7}$$

2. Пусть $\xi=\xi(\omega)$ — случайная величина, принимающая значения в множестве $X=\{x_1,\ \ldots,\ x_l\}$:

$$\xi=\sum_{j=1}^{l}x_jI_{A_j},\qquad A_j=\{\omega\colon\ \xi=x_j\},$$

и $\mathscr{D}=\{D_1,\ \ldots,\ D_k\}$ — некоторое разбиение. Подобно тому как для ξ по вероятностям (A_j), $j=1,\ \ldots,\ l$, было определено математическое ожидание

$$\xi=\sum_{j=1}^{l}x_j\ \ (A_j), \tag{8}$$

так и с помощью условных вероятностей $(A_j \,|\, \mathscr{D})$, $j = 1, \ldots, l$, естественно определить *условное математическое ожидание случайной величины ξ относительно разбиения $\mathscr{D}$*, обозначаемое $(\xi \,|\, \mathscr{D})$ или $(\xi \,|\, \mathscr{D})(\omega)$, формулой

$$(\xi \,|\, \mathscr{D}) = \sum_{j=1}^{l} x_j \quad (A_j \,|\, \mathscr{D}). \tag{9}$$

$$\begin{tikzcd}
\mathsf{P}(\cdot) \arrow[r, "(8)"] \arrow[d, "(3.1)"] & \mathsf{E}\xi \\
\mathsf{P}(\cdot \,|\, D) \arrow[r, "(10)"] \arrow[d, "(1)"] & \mathsf{E}(\xi \,|\, D) \arrow[d, "(11)"] \\
\mathsf{P}(\cdot \,|\, \mathscr{D}) \arrow[r, "(9)"] & \mathsf{E}(\xi \,|\, \mathscr{D})
\end{tikzcd}$$

Рис. 14.

Согласно этому определению, условное математическое ожидание $(\xi \,|\, \mathscr{D})(\omega)$ является *случайной величиной*, принимающей для всех элементарных событий ω, принадлежащих одному и тому же атому D_i, одно и то же значение $\sum\limits_{j=1}^{l} x_j \quad (A_j \,|\, D_i)$. Это замечание показывает, что к определению условного математического ожидания $(\xi \,|\, \mathscr{D})$ можно было бы подойти иначе. А именно, сначала определить $(\xi \,|\, D_i)$ — условное математическое ожидание ξ относительно события D_i формулой

$$(\xi \,|\, D_i) = \sum_{j=1}^{l} x_j \quad (A_j \,|\, D_i) \quad \left(= \frac{[\xi I_{D_i}]}{(D_i)} \right), \tag{10}$$

а затем положить по определению

$$(\xi \,|\, \mathscr{D})(\omega) = \sum_{i=1}^{l} \quad (\xi \,|\, D_i) I_{D_i}(\omega) \tag{11}$$

(см. диаграмму на рис. 14).

Полезно отметить, что значения $(\xi \,|\, D)$ и $(\xi \,|\, \mathscr{D})$ *не зависят* от способа представления случайной величины ξ.

Приводимые далее свойства условных математических ожиданий непосредственно вытекают из их определения:

$$(a\xi + b\eta \,|\, \mathscr{D}) = a \quad (\xi \,|\, \mathscr{D}) + b \quad (\eta \,|\, \mathscr{D}), \quad a,\ b \text{ — константы;} \tag{12}$$

$$(\xi \,|\, \Omega) = \quad \xi; \tag{13}$$

$$(C \,|\, \mathscr{D}) = C, \quad C \text{ — константа;} \tag{14}$$

если $\xi = I_A(\omega)$, то

$$(\xi \,|\, \mathscr{D}) = \quad (A \,|\, \mathscr{D}). \tag{15}$$

Последнее равенство показывает, в частности, что свойства условных вероятностей можно получать непосредственно из свойств условных математических ожиданий.

Следующее важное свойство обобщает *формулу полной вероятности* (4):

$$(\xi \mid \mathscr{D}) = \quad \xi. \tag{16}$$

Для доказательства достаточно заметить, что, согласно (4),

$$(\xi \mid \mathscr{D}) = \quad \sum_{j=1}^{l} x_j \quad (A_j \mid \mathscr{D}) = \sum_{j=1}^{l} x_j \quad (A_j \mid \mathscr{D}) = \sum_{j=1}^{l} x_j \quad (A_j) = \quad \xi.$$

Пусть $\mathscr{D} = \{D_1, \ldots, D_k\}$ — разбиение и $\eta = \eta(\omega)$ — некоторая случайная величина. Будем говорить, что случайная величина η является *измеримой* относительно этого разбиения, или $\mathscr{D}$-измерима, если $\mathscr{D}_\eta \preccurlyeq \mathscr{D}$, 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 величина η будет $\mathscr{D}$-измеримой 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

$$(\xi\eta \mid \mathscr{D}) = \eta \quad (\xi \mid \mathscr{D}) \tag{17}$$

и, в частности,

$$(\eta \mid \mathscr{D}) = \eta \quad (\quad (\eta \mid \mathscr{D}_\eta) = \eta). \tag{18}$$

Для доказательства (17) заметим, что если $\xi = \sum_{j=1}^{l} x_j I_{A_j}$, то

$$\xi\eta = \sum_{j=1}^{l} \sum_{i=1}^{k} x_j y_i I_{A_j D_i}$$

и, значит,

$$(\xi\eta \mid \mathscr{D}) = \sum_{j=1}^{l} \sum_{i=1}^{k} x_j y_i \quad (A_j D_i \mid \mathscr{D}) = \sum_{j=1}^{l} \sum_{i=1}^{k} x_j y_i \sum_{m=1}^{k} \quad (A_j D_i \mid D_m) I_{D_m}(\omega) =$$

$$= \sum_{j=1}^{l} \sum_{i=1}^{k} x_j y_i \quad (A_j D_i \mid D_i) I_{D_i}(\omega) = \sum_{j=1}^{l} \sum_{i=1}^{k} x_j y_i \quad (A_j \mid D_i) I_{D_i}(\omega). \tag{19}$$

С другой стороны, учитывая, что $I_{D_i}^2 = I_{D_i}$ и $I_{D_i} I_{D_m} = 0$, $i \neq m$, получаем,

$$\eta\quad (\xi \,|\, \mathscr{D}) = \left[\sum_{i=1}^{k} y_i I_{D_i}(\omega)\right] \cdot \left[\sum_{j=1}^{l} x_j \quad (A_j \,|\, \mathscr{D})\right] =$$

$$= \left[\sum_{i=1}^{k} y_i I_{D_i}(\omega)\right] \cdot \sum_{m=1}^{k} \left[\sum_{j=1}^{l} x_j \quad (A_j \,|\, D_m)\right] \cdot I_{D_m}(\omega) =$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{l} y_i x_j \quad (A_j \,|\, D_i) I_{D_i}(\omega),$$

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 $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2$ (разбиение $\mathscr{D}_2$ «мельче» разбиения $\mathscr{D}_1$). Тогда справедливо «телескопическое» свойство:

$$[\quad (\xi \,|\, \mathscr{D}_2) \,|\, \mathscr{D}_1] = \quad (\xi \,|\, \mathscr{D}_1). \tag{20}$$

Для доказательства предположим, что

$$\mathscr{D}_1 = \{D_{11}, \ldots, D_{1m}\}, \qquad \mathscr{D}_2 = \{D_{21}, \ldots, D_{2n}\}.$$

Тогда, если $\xi = \sum\limits_{j=1}^{l} x_j I_{A_j}$, то

$$(\xi \,|\, \mathscr{D}_2) = \sum_{j=1}^{l} x_j \quad (A_j \,|\, \mathscr{D}_2),$$

и достаточно лишь установить, что

$$[\quad (A_j \,|\, \mathscr{D}_2) \,|\, \mathscr{D}_1] = \quad (A_j \,|\, \mathscr{D}_1). \tag{21}$$

Поскольку

$$(A_j \,|\, \mathscr{D}_2) = \sum_{q=1}^{n} \quad (A_j \,|\, D_{2q}) I_{D_{2q}},$$

то

$$
\begin{aligned}
[\ (A_j \mid \mathscr{D}_2) \mid \mathscr{D}_1] &= \sum_{q=1}^{n} \ (A_j \mid D_{2q}) \ (D_{2q} \mid \mathscr{D}_1) = \\
&= \sum_{q=1}^{n} \ (A_j \mid D_{2q}) \left[\sum_{p=1}^{m} \ (D_{2q} \mid D_{1p}) I_{D_{1p}} \right] = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \sum_{q=1}^{n} \ (A_j \mid D_{2q}) \ (D_{2q} \mid D_{1p}) = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \sum_{\{q \,:\, D_{2q} \subseteq D_{1p}\}} \ (A_j \mid D_{2q}) \ (D_{2q} \mid D_{1p}) = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \sum_{\{q \,:\, D_{2q} \subseteq D_{1p}\}} \frac{(A_j D_{2q})}{(D_{2q})} \cdot \frac{(D_{2q})}{(D_{1p})} = \\
&= \sum_{p=1}^{m} I_{D_{1p}} \ (A_j \mid D_{1p}) = \ (A_j \mid \mathscr{D}_1),
\end{aligned}
$$

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 $(\xi \mid \mathscr{D}_{\eta_1, \ldots, \eta_k})$ будет обозначаться $(\xi \mid \eta_1, \ldots, \eta_k)$ (или $(\xi \mid \eta_1, \ldots, \eta_k)(\omega)$) и называться *условным математическим ожиданием случайной величины ξ относительно $\eta_1, \ldots, \eta_k$*.

Непосредственно из определения $(\xi \mid \eta)$ следует, что если ξ и η *независимы*, то

$$(\xi \mid \eta) = \ \xi. \tag{22}$$

Из (18) следует также, что

$$(\eta \mid \eta) = \eta. \tag{23}$$

С учетом обозначения $(\xi \mid \eta)$ для $(\xi \mid \mathscr{D}_\eta)$ формуле (16), являющейся обозначением формулы полной вероятности (4), можно придать следующий, широко испльзуемый вид

$$(\xi \mid \eta) = \ \xi. \tag{24}$$

(См. также свойство (27) в задаче 3.)

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

$$(\xi \,|\, \mathscr{D}) = \quad \xi.$$

Из (20) в качестве частного случая получаем следующую полезную формулу:

$$[\quad (\xi \,|\, \eta_1, \eta_2) \,|\, \eta_1] = \quad (\xi \,|\, \eta_1). \tag{25}$$

Пример 3. Для случайных величин ξ и η, рассмотренных в примере 1, найдем $\quad (\xi + \eta \,|\, \eta)$. В силу (22) и (23)

$$(\xi + \eta \,|\, \eta) = \quad \xi + \eta = p + \eta.$$

Этот результат можно получить и отправляясь от (8):

$$(\xi + \eta \,|\, \eta) = \sum_{k=0}^{2} k \quad (\xi + \eta = k \,|\, \eta) = p(1 - \eta) + q\eta + 2p\eta = p + \eta.$$

Пример 4. Пусть ξ и η — независимые одинаково распределенные случайные величины. Тогда

$$(\xi \,|\, \xi + \eta) = \quad (\eta \,|\, \xi + \eta) = \frac{\xi + \eta}{2}. \tag{26}$$

Действительно, считая для простоты, что ξ и η принимают значения $1, 2, \ldots, m$, находим, что $(1 \leqslant k \leqslant m,\ 2 \leqslant l \leqslant 2m)$

$$(\xi = k \,|\, \xi + \eta = l) = \frac{\{\xi = k,\ \xi + \eta = l\}}{\{\xi + \eta = l\}} = \frac{\{\xi = k,\ \eta = l - k\}}{\{\xi + \eta = l\}} =$$
$$= \frac{\{\xi = k\} \quad \{\eta = l - k\}}{\{\xi + \eta = l\}} = \frac{\{\eta = k\} \quad \{\xi = l - k\}}{\{\xi + \eta = l\}} = \quad (\eta = k \,|\, \xi + \eta = l).$$

Этим доказано первое равенство в (26). Для доказательства второго достаточно заметить, что

$$2 \quad (\xi \,|\, \xi + \eta) = \quad (\xi \,|\, \xi + \eta) + \quad (\eta \,|\, \xi + \eta) = \quad (\xi + \eta \,|\, \xi + \eta) = \xi + \eta.$$

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 $(\xi\,|\,\mathscr{B})$ случайной величины ξ относительно алгебры $\mathscr{B}$ просто совпадает с $(\xi\,|\,\mathscr{D})$ — 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 $(\xi\,|\,\mathscr{B})$ и $(\xi\,|\,\mathscr{D})$, понимая всякий раз, что $(\xi\,|\,\mathscr{B})$ есть по определению просто $(\xi\,|\,\mathscr{D})$.

4. Задачи.

1. Привести пример двух случайных величин ξ и η, которые не являются независимыми, но для которых

$$(\xi\,|\,\eta) = \quad \xi.$$

(Ср. с утверждением (22).)

2. *Условной дисперсией* ξ относительно разбиения $\mathscr{D}$ называется случайная величина

$$(\xi\,|\,\mathscr{D}) = \quad [(\xi - \quad (\xi\,|\,\mathscr{D}))^2|\mathscr{D}].$$

Показать, что дисперсия

$$\xi = \quad (\xi\,|\,\mathscr{D}) + \quad (\xi\,|\,\mathscr{D}).$$

3. Отправляясь от (17), доказать, что для всякой функции $f=f(\eta)$ условное математическое ожидание $(\xi\,|\,\eta)$ обладает следующим свойством:

$$[f(\eta) \quad (\xi\,|\,\eta)] = \quad [\xi f(\eta)]. \tag{27}$$

4. Пусть ξ и η — случайные величины. Показать, что $\inf_f \quad (\eta - f(\xi))^2$ достигается на функции $f^*(\xi) = \quad (\eta\,|\,\xi)$. (Таким образом, оптимальной в среднеквадратическом смысле *оценкой* η по ξ является условное математическое ожидание $(\eta\,|\,\xi)$.)

5. Пусть $\xi_1, \ldots, \xi_n$, τ — *независимые* случайные величины, причем величины $\xi_1, \ldots, \xi_n$ одинаково распределены, а τ принимает значения $1, 2, \ldots, n$. Показать, что если $S_\tau = \xi_1 + \ldots + \xi_\tau$ — *сумма случайного числа случайных величин*, то

$$(S_\tau\,|\,\tau) = \tau \quad \xi_1, \qquad (S_\tau|\tau) = \tau \quad \xi_1$$

и

$$S_\tau = \quad \tau \cdot \quad \xi_1, \qquad S_\tau = \quad \tau \cdot \quad \xi_1 + \quad \tau \cdot (\quad \xi_1)^2.$$

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 $\{S_n = k\}$ и $\{A < S_n \leqslant B\}$. Роль этих теорем состоит также и в том, что они носят *универсальный* характер, т. е. остаются справедливыми не только для независимых бернуллиевских случайных величин $\xi_1, \xi_2, \ldots$, 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 блужданий, описываемых схемой Бернулли, хотя многие выводы остаются справедливыми и для блужданий более общего вида.

2. Рассмотрим схему Бернулли $(\Omega, \mathscr{A}, \ )$, где $\Omega = \{\omega\colon \omega = (x_1, \ldots, x_n),\ x_i = \pm 1\}$, $\mathscr{A}$ — система всех подмножеств Ω и $(\{\omega\}) = p^{\nu(\omega)} q^{n-\nu(\omega)}$, $\nu(\omega) = \dfrac{\sum x_i + n}{2}$. 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,

$$\{\xi_i = 1\} = p, \qquad \{\xi_i = -1\} = q, \qquad p + q = 1.$$

Положим $S_0 = 0$, $S_k = \xi_1 + \ldots + \xi_k$, $1 \leqslant k \leqslant n$. Последовательность $(S_k)_{k \leqslant n}$ можно рассматривать как *траекторию* случайного блуждания некоторой «частицы», выходящей из нуля. При этом $S_{k+1} = S_k + \xi_{k+1}$, 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 в точке 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 $S_k^x = x + S_k$,

$$\tau_k^x = \min\{0 \leqslant l \leqslant k\colon\ S_l^x = A \text{ или } B\}, \tag{1}$$

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 $\{\omega\colon \tau_k^x = l\}$ 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 $\{\omega\colon \tau_k^x = l,\ S_l^x = A\}$ и $\{\omega\colon \tau_k^x = l,\ S_l^x = B\}$ 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 $0 \leqslant k \leqslant n$

$$\begin{aligned} \mathscr{A}_k^x &= \sum_{0 \leqslant l \leqslant k} \{\omega\colon\ \tau_k^x = l,\ S_l^x = A\}, \\ \mathscr{B}_k^x &= \sum_{0 \leqslant l \leqslant k} \{\omega\colon\ \tau_k^x = l,\ S_l^x = B\}, \end{aligned} \tag{2}$$

и пусть

$$\alpha_k(x) = \quad (\mathscr{A}_k^x), \qquad \beta_k(x) = \quad (\mathscr{B}_k^x)$$

— 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 (3) § 3

$$\begin{aligned}\beta_k(x) = \quad & (\mathscr{B}_k^x) = \\ = \quad & (\mathscr{B}_k^x \mid S_1^x = x+1) \quad \{\xi_1 = 1\} + \quad (\mathscr{B}_k^x \mid S_1^x = x-1) \quad \{\xi_1 = -1\} = \\ & = p \quad (\mathscr{B}_k^x \mid S_1^x = x+1) + q \quad (\mathscr{B}_k^x \mid S_1^x = x-1). \qquad (3)\end{aligned}$$

Покажем, что

$$(\mathscr{B}_k^x \mid S_1^x = x+1) = \quad (\mathscr{B}_{k-1}^{x+1}), \qquad (\mathscr{B}_k^x \mid S_1^x = x-1) = \quad (\mathscr{B}_{k-1}^{x-1}).$$

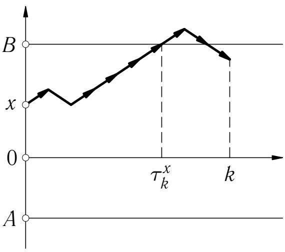


Рис. 15. Пример траектории из множества B_k^x

С этой целью заметим, что множество $\mathscr{B}_k^x$ можно представить в виде

$$\mathscr{B}_k^x = \{\omega\colon\ (x,\ x+\xi_1,\ \ldots,\ x+\xi_1+\ldots+\xi_k) \in B_k^x\},$$

где B_k^x — множество траекторий вида

$$(x,\ x+x_1,\ \ldots,\ x+x_1+\ldots+x_k)$$

с $x_i = \pm 1$, 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 $(x,\ x+1,\ x+1+x_2,\ \ldots,\ x+1+x_2+\ldots+x_k)$ 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 (6) § 8, находим, что

$$\begin{aligned}(\mathscr{B}_k^x \mid S_1^x = x+1) &= \quad (\mathscr{B}_k^x \mid \xi_1 = 1) = \\ &= \quad ((x,\ x+\xi_1,\ \ldots,\ x+\xi_1+\ldots+\xi_k) \in B_k^k \mid \xi_1 = 1) = \\ &= \quad \{(x+1,\ x+1+\xi_2,\ \ldots,\ x+1+\xi_2+\ldots+\xi_k) \in B_{k-1}^{x+1}\} = \\ &= \quad \{(x+1,\ x+1+\xi_1,\ \ldots,\ x+1+\xi_1+\ldots+\xi_{k-1}) \in B_{k-1}^{x+1}\} = \quad (\mathscr{B}_{k-1}^{x+1}).\end{aligned}$$

Точно так же

$$(\mathscr{B}_k^x \mid S_1^x = x-1) = \quad (\mathscr{B}_{k-1}^{x-1}).$$

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

$$\beta_l(B) = 1, \quad \beta_l(A) = 0, \quad 0 \leqslant l \leqslant n. \tag{5}$$

Аналогично

$$\alpha_k(x) = p\alpha_{k-1}(x+1) + q\alpha_{k-1}(x-1) \tag{6}$$

с

$$\alpha_l(A) = 1, \quad \alpha_l(B) = 0, \quad 0 \leqslant l \leqslant n.$$

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. далее 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

$$\beta(B) = 1, \quad \beta(A) = 0, \tag{8}$$

получаемыми 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 общее 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

$$\alpha(x)=p\alpha(x+1)+q\alpha(x-1), \quad x\in(A,\,B), \tag{11}$$

с граничными условиями

$$\alpha(A)=1, \ \ \alpha(B)=0 \tag{12}$$

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 $0\leqslant p\leqslant 1$

$$\alpha(x)+\beta(x)=1. \tag{16}$$

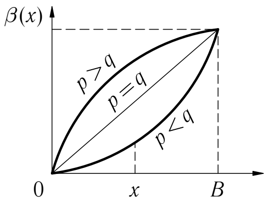


Рис. 16. График $\beta(x)$ — вероятности достижения точки B раньше точки 0, когда частица выходит из точки x

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 $(\Omega,\, \mathscr{A},\ \ )$ оказывается слишком «бедным» для того, чтобы на нем существовала такая *бесконечная* последовательность независимых случайных величин. В дальнейшем мы увидим (§ 9 гл. II), что такую последовательность *действительно* можно построить и что величины $\beta(x)$ и $\alpha(x)$ в самом деле являются вероятностями разорения при неограниченном числе шагов.

Обратимся к ряду следствий, вытекающих 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 выполняется неравенство $q>p$, 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 $\{\xi_i=1/2\}=p$, $\{\xi_i=$ $=-1/2\}=q$. 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

$$\alpha_n=\alpha_n(0),\qquad \beta_n=\beta_n(0),\qquad \gamma_n=1-(\alpha_n+\beta_n).$$

Ясно, что

$$\gamma_n=\ \{A<S_k<B,\ 0\leqslant k\leqslant n\},$$

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

$$\begin{aligned} \zeta_1 &= \xi_1 + \ldots + \xi_m, \\ \zeta_2 &= \xi_{m+1} + \ldots + \xi_{2m}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \zeta_r &= \xi_{m(r-1)+1} + \ldots + \xi_{rm}. \end{aligned}$$

Тогда, если $C = |A| + B$, то нетрудно убедиться в том, что

$$\{A < S_k < B,\ 1 \leqslant k \leqslant rm\} \subseteq \{|\zeta_1| < C,\ \ldots,\ |\zeta_r| < C\},$$

и, значит, в силу независимости величин $\zeta_1, \ldots, \zeta_r$ и их одинаковой распределенности

$$\gamma_n \leqslant \quad \{|\zeta_1| < C,\ \ldots,\ |\zeta_r| < C\} = \prod_{i=1}^{r} \quad \{|\zeta_i| < C\} = (\quad \{|\zeta_1| < C\})^r. \tag{18}$$

Заметим, что $\quad \zeta_1 = m[1 - (p-q)^2]$. Поэтому при $0 < p < 1$ для достаточно больших m

$$\{|\zeta_1| < C\} \leqslant \varepsilon_1, \tag{19}$$

где $\varepsilon_1 < 1$, поскольку если $\quad \{|\zeta_1| \leqslant C\} = 1$, то $\quad \zeta_1 \leqslant C^2$.

Если же $p = 0$ или $p = 1$, то для достаточно больших m вероятность $\quad \{|\zeta_1| < C\} = 0$, и, следовательно, (19) выполнено при всех $0 \leqslant p \leqslant 1$.

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

$$\begin{aligned} 0 \leqslant \alpha - \alpha_n \leqslant \gamma_n \leqslant \varepsilon^n, \\ 0 \leqslant \beta - \beta_n \leqslant \gamma_n \leqslant \varepsilon^n \end{aligned}$$

с $\varepsilon < 1$. Аналогичные оценки справедливы и для разностей $\alpha(x) - \alpha_n(x)$ и $\beta(x) - \beta_n(x)$.

4. Обратимся теперь к вопросу о *средней длительности* случайного блуждания.

Пусть $m_k(x) = \quad \tau_k^x$ — математическое ожидание момента остановки τ_k^x, $k \leqslant n$. Поступая, как и при выводе рекуррентных соотношений для $\beta_k(x)$, получаем, что для $x \in (A,\, B)$

$$\begin{aligned} m_k(x) = \ \ \tau_k^x &= \sum_{1 \leqslant l \leqslant k} l \ \ \{\tau_k^x = l\} = \\ &= \sum_{1 \leqslant l \leqslant k} l[p \ \ (\tau_k^x = l \,|\, \xi_1 = 1) + q \ \ (\tau_k^x = l \,|\, \xi_1 = -1)] = \\ &= \sum_{1 \leqslant l < k} l[p \ \ \{\tau_{k-1}^{x+1} = l-1\} + q \ \ \{\tau_{k-1}^{x-1} = l-1\}] = \\ &= \sum_{0 \leqslant l \leqslant k-1} (l+1)[p \ \ \{\tau_{k-1}^{x+1} = l\} + q \ \ \{\tau_{k-1}^{x-1} = l\}] = \\ &= pm_{k-1}(x+1) + qm_{k-1}(x-1) + \\ &\qquad\qquad + \sum_{0 \leqslant l \leqslant k-1} [p \ \ \{\tau_{k-1}^{x+1} = l\} + q \ \ \{\tau_{k-1}^{x-1} = l\}] = \\ &= pm_{k-1}(x+1) + qm_{k-1}(x-1) + 1. \end{aligned}$$

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

$$m(x) < \infty, \qquad x \in (A,\, B). \tag{25}$$

Тогда, если $p \neq q$, то частное решение имеет вид $\dfrac{x}{q-p}$ и общее решение (см. (9)) записывается в виде

$$m(x) = \frac{x}{p-q} + a + b\left(\frac{q}{p}\right)^x.$$

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

$$m(x) = (B - x)(x - A). \tag{27}$$

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 секунду, 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 $S_{\tau_n} = S_{\tau_n}(\omega)$, определенную следующим равенством:

$$S_{\tau_n}(\omega) = \sum_{k=0}^{n} S_k(\omega) I_{\{\tau_n = k\}}(\omega). \tag{28}$$

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 $p = q = 1/2$

$$S_{\tau_n} = 0, \tag{29}$$

$$S_{\tau_n}^2 = \quad \tau_n. \tag{30}$$

Для доказательства первого равенства заметим, что

$$S_{\tau_n} = \sum_{k=0}^{n} \quad [S_k I_{\{\tau_n = k\}}(\omega)] = \sum_{k=0}^{n} \quad [S_n I_{\{\tau_n = k\}}(\omega)] + \\ + \sum_{k=0}^{n} \quad [(S_k - S_n) I_{\{\tau_n = k\}}(\omega)] = \quad S_n + \sum_{k=0}^{n} \quad [(S_k - S_n) I_{\{\tau_n = k\}}(\omega)], \tag{31}$$

где, очевидно, $S_n = 0$. Покажем, что

$$\sum_{k=0}^{n} [(S_k - S_n) I_{\{\tau_n = k\}}(\omega)] = 0.$$

Для $0 \leqslant k < n$ имеем $\{\tau_n > k\} = \{A < S_1 < B, \ldots, A < S_k < B\}$. Событие $\{\omega : A < S_1 < B, \ldots, A < S_k < B\}$ может быть, очевидно, представлено в виде

$$\{\omega : (\xi_1, \ldots, \xi_k) \in A_k\}, \tag{32}$$

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 10 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,

$$[(S_n - S_k) I_{\{\tau_n = k\}}] = [S_n - S_k] \cdot I_{\{\tau_n = k\}} = 0,$$

что и доказывает формулу (29).

Тем же методом доказывается и формула (30):

$$S_{\tau_n}^2 = \sum_{k=0}^{n} S_k^2 I_{\{\tau_n = k\}} = \sum_{k=0}^{n} ([S_n + (S_k - S_n)]^2 I_{\{\tau_n = k\}}) =$$

$$= \sum_{k=0}^{n} [S_n^2 I_{\{\tau_n = k\}} + 2 S_n (S_k - S_n) I_{\{\tau_n = k\}} + (S_n - S_k)^2 I_{\{\tau_n = k\}}] =$$

$$= S_n^2 - \sum_{k=0}^{n} (S_n - S_k)^2 I_{\{\tau_n = k\}} = n - \sum_{k=0}^{n} (n - k) \{\tau_n = k\} =$$

$$= \sum_{k=0}^{n} k \{\tau_n = k\} = \tau_n.$$

Итак, для $p = q = 1/2$ имеют место формулы (29), (30). В случае же произвольных p и q $(p + q = 1)$ аналогично устанавливается, что

$$S_{\tau_n} = (p - q) \tau_n, \tag{33}$$

$$[S_{\tau_n} - \tau_n \xi_1]^2 = \xi_1 \cdot \tau_n, \tag{34}$$

где $\xi_1 = p - q$, $\xi_1 = 1 - (p - q)^2$.

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)

$$\tau_n \leqslant \max(A^2,\, B^2). \tag{35}$$

Если же $p \neq q$, то из (33)

$$\tau_n \leqslant \frac{\max(|A|,\, B)}{|p-q|}, \tag{36}$$

откуда ясно, что $m(0) < \infty$.

Заметим также, что в случае $p = q = 1/2$

$$\tau_n = \quad S^2_{\tau_n} = A^2\alpha_n + B^2\beta_n + \quad [S_n^2 I_{\{A<S_n<B\}} I_{\{\tau_n=n\}}]$$

и, значит,

$$A^2\alpha_n + B^2\beta_n \leqslant \quad \tau_n \leqslant A^2\alpha_n + B^2\beta_n + \max(A^2,\, B^2)\cdot\gamma_n.$$

Отсюда и из неравенств (20) следует, что математические ожидания τ_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 $p\neq q$:

$$\textit{экспоненциально быстро} \quad \tau_n \to m(0) = \frac{\alpha A + \beta B}{p-q}.$$

5. Задачи.

1. Показать, что в обобщение (33) и (34) справедливы следующие формулы:

$$S^x_{\tau^x_n} = x + (p-q) \quad \tau^x_n,$$
$$[S^x_{\tau^x_n} - \tau^x_n \quad \xi_1]^2 = \quad \xi_1 \cdot \quad \tau^x_n + x^2.$$

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 $|S_n|$ при больших n?

4. Два игрока независимым образом подбрасывают (каждый свою) симметричные монеты. 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 $\sum\limits_{k=0}^{n}(C_n^k)^2 = C_{2n}^n$ (см. также задачу 4 в § 2).

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 математические ожидания $\quad \min(\sigma_n,\, n)$.

§ 10. Случайное блуждание. II. Принцип отражения. Закон арксинуса

1. Как и в предыдущем параграфе, будем предполагать, что $\xi_1, \ldots$ $\ldots, \xi_{2n}$ — последовательность независимых одинаково распределенных бернуллиевских случайных величин с

$$\{\xi_i = 1\} = p, \qquad \{\xi_i = -1\} = q,$$
$$S_k = \xi_1 + \ldots + \xi_k, \quad 1 \leqslant k \leqslant 2n; \quad S_0 = 0.$$

Обозначим

$$\sigma_{2n} = \min\{1 \leqslant k \leqslant 2n\colon\ S_k = 0\},$$

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 $0 \leqslant k \leqslant n$

$$u_{2k} = \quad \{S_{2k} = 0\}, \qquad f_{2k} = \quad \{\sigma_{2n} = 2k\}. \tag{1}$$

Ясно, что $u_0 = 1$ и

$$u_{2k} = C_{2k}^k \cdot 2^{-2k}.$$

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

$$\begin{aligned} f_{2k} = \quad &\{S_1 \neq 0, \ldots, S_{2k-1} \neq 0, S_{2k} = 0\} = \\ &= 2 \quad \{S_1 > 0, \ldots, S_{2k-1} > 0, S_{2k} = 0\}. \end{aligned} \tag{3}$$

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

$$f_{2k} = 2 \sum_{(a_{2k+1},\ldots,a_{2n})} L_{2n}(S_1 > 0, \ldots, S_{2k-1} > 0, S_{2k} = 0, S_{2k+1} = a_{2k+1}, \ldots \\ \ldots, S_{2n} = a_{2k+1} + \ldots + a_{2n}) \cdot 2^{-2n} = \\ = 2L_{2k}(S_1 > 0, \ldots, S_{2k-1} > 0, S_{2k} = 0) \cdot 2^{-2k}, \quad (4)$$

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*

$$L_k(S_1 > 0, \ldots, S_{k-1} > 0, S_k = a - b) = \frac{a-b}{k} C_k^a. \quad (5)$$

Доказательство. Действительно,

$$L_k(S_1 > 0, \ldots, S_{k-1} > 0, S_k = a - b) = \\ = L_k(S_1 = 1, S_2 > 0, \ldots, S_{k-1} > 0, S_k = a - b) = L_k(S_1 = 1, S_k = a - b) - \\ - L_k(S_1 = 1, S_k = a - b; \exists i, 2 \leqslant i \leqslant k - 1, \text{ такое, что } S_i \leqslant 0). \quad (6)$$

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 временну́ю ось. *)

Заметим теперь, что

$$L_k(S_1 = 1, S_k = a - b; \exists i, 2 \leqslant i \leqslant k - 1, \text{ такое, что } S_i \leqslant 0) = \\ = L_k(S_1 = -1, S_k = a - b), \quad (7)$$

т. е. *число путей, идущих из точки $\alpha = (1, 1)$ в точку $\beta = (k, a - b)$ и касающихся или пересекающих временну́ю 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,

*) Путь $(S_1, \ldots, S_k)$ называется *положительным* (неотрицательным), если все $S_i > 0$ ($S_i \geqslant 0$); путь называется *касающимся* временно́й оси, если $S_j \geqslant 0$ или $S_j \leqslant 0$ для всех $1 \leqslant j \leqslant k$ и найдется такое $1 \leqslant i \leqslant k$, что $S_i = 0$, и называется *пересекающим* временну́ю ось, если найдутся такие два момента времени i и j, что $S_i > 0$, а $S_j < 0$.

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.

Из (6) и (7) находим

$$
\begin{aligned}
L_k(S_1 > 0, \ldots, S_{k-1} > 0, S_k = a - b) = \\
= L_k(S_1 = 1, S_k = a - b) - \\
- L_k(S_k = -1, S_k = a - b) = \\
= C_{k-1}^{a-1} - C_{k-1}^{a} = \frac{a-b}{k} C_k^a,
\end{aligned}
$$

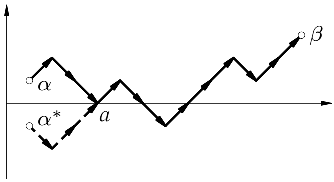


Рис. 17. К принципу отражения

что и доказывает утверждение (5). $\square$

Возвращаясь к подсчету вероятности f_{2k}, находим, что, согласно (4) и (5) (с $a = k$, $b = k - 1$),

$$
\begin{aligned}
f_{2k} &= 2L_{2k}(S_1 > 0, \ldots, S_{2k-1} > 0, S_{2k} = 0) \cdot 2^{-2k} = \\
&= 2L_{2k-1}(S_1 > 0, \ldots, S_{2k-1} = 1) \cdot 2^{-2k} = 2 \cdot 2^{-2k} \cdot \frac{1}{2k-1} C_{2k-1}^{k} = \frac{1}{2k} u_{2(k-1)}.
\end{aligned}
$$

Итак, формула (2) доказана.

Приведем еще одно доказательство этой формулы, основанное на следующем замечании. Непосредственная проверка показывает, что

$$
\frac{1}{2k} u_{2(k-1)} = u_{2(k-1)} - u_{2k}. \tag{8}
$$

В то же самое время ясно, что

$$
\begin{gathered}
\{\sigma_{2n} = 2k\} = \{\sigma_{2n} > 2(k-1)\} \setminus \{\sigma_{2n} > 2k\}, \\
\{\sigma_{2n} > 2l\} = \{S_1 \neq 0, \ldots, S_{2l} \neq 0\}
\end{gathered}
$$

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

$$
f_{2k} = \quad \{S_1 \neq 0, \ldots, S_{2(k-1)} \neq 0\} - \quad \{S_1 \neq 0, \ldots, S_{2k} \neq 0\},
$$

и, следовательно, в силу (8) для доказательства равенства $f_{2k} = \frac{1}{2k} u_{2(k-1)}$ достаточно 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

$$
2L_{2k}(S_1 > 0, \ldots, S_{2k} > 0) = L_{2k}(S_1 \geqslant 0, \ldots, S_{2k} \geqslant 0) \tag{10}
$$

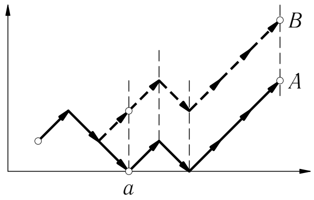


Рис. 18.

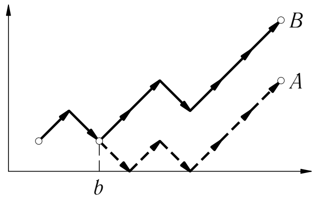


Рис. 19.

и

$$L_{2k}(S_1 \geqslant 0, \ldots, S_{2k} \geqslant 0) = L_{2k}(S_{2k} = 0). \qquad (11)$$

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.

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, пусть $B = (S_1, \ldots, S_{2k})$ — 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 неотрицательными 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

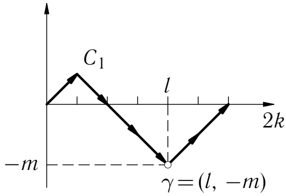


Рис. 20.

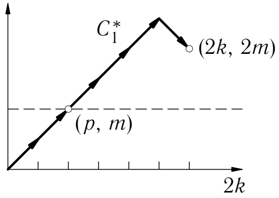


Рис. 21.

начало координат в точку $(l,\,-m)$. Полученная траектория C_1^* будет положительным путем.

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 единственный 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

$$\min(\sigma_{2n},\,2n)=\sum_{k=1}^{n}2k\quad\{\sigma_{2n}=2k\}+2nu_{2n}=\sum_{k=1}^{n}u_{2(k-1)}+2nu_{2n}$$

является довольно-таки большим.

Более того, $\sum\limits_{k=1}^{\infty} u_{2(k-1)} = \infty$. 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 $2n$ *среднее 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 среднее число ничьих имеет порядок $\sqrt{2n}$ (см. (17) в § 9 гл. VII). Отсюда вытекает, в частности, что, вопреки ожидаемому, «типичные» реализации блуждания $(S_0, S_1, \ldots, S_n)$ должны иметь *не синусоидальный* характер (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*

$$P_{2k,2n} = u_{2k} u_{2n-2k}. \tag{12}$$

Доказательство. Выше было установлено, что $f_{2k} = u_{2(k-1)} - u_{2k}$. Покажем, что

$$u_{2k} = \sum_{r=1}^{k} f_{2r} u_{2(k-r)}. \tag{13}$$

Поскольку $\{S_{2k} = 0\} \subseteq \{\sigma_{2n} \leqslant 2k\}$, то

$$\{S_{2k} = 0\} = \{S_{2k} = 0\} \cap \{\sigma_2 \leqslant 2k\} = \sum_{1 \leqslant l \leqslant k} \{S_{2k} = 0\} \cap \{\sigma_{2n} = 2l\}.$$

Следовательно,

$$u_{2k} = \quad \{S_{2k} = 0\} = \sum_{1 \leqslant l \leqslant k} \quad \{S_{2k} = 0,\ \sigma_{2n} = 2l\} =$$

$$= \sum_{1 \leqslant l \leqslant k} \quad (S_{2k} = 0 \mid \sigma_{2k} = 2l) \quad \{\sigma_{2n} = 2l\}.$$

*) 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.

Но

$$
\begin{aligned}
(S_{2k}=0\,|\,\sigma_{2n}=2l) = \quad (S_{2k}=0\,|\,S_1\neq 0,\ \ldots,\ S_{2l-1}\neq 0,\ S_{2l}=0) = \\
= \quad (S_{2l}+(\xi_{2l+1}+\ldots+\xi_{2k})=0\,|\,S_1\neq 0,\ \ldots,\ S_{2l-1}\neq 0,\ S_{2l}=0) = \\
= \quad (S_{2l}+(\xi_{2l+1}+\ldots+\xi_{2k})=0\,|\,S_{2l}=0) = \\
= \quad \{\xi_{2l+1}+\ldots+\xi_{2k}=0\} = \quad \{S_{2(k-1)}=0\}.
\end{aligned}
$$

Поэтому

$$
u_{2k} = \sum_{1\leqslant l\leqslant k} \quad \{S_{2(k-l)}=0\} \quad \{\sigma_{2n}=2l\},
$$

что и доказывает формулу (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 $S_l\geqslant 0$, $l\leqslant 2r$ и $S_l\leqslant 0$, $l\leqslant 2r$.

Число путей, относящихся к первому случаю, равно, как нетрудно видеть,

$$
\left(\frac{1}{2}\cdot 2^{2r} f_{2r}\right) 2^{2(n-r)} P_{2(k-r),2(n-r)} = \frac{1}{2}\cdot 2^{2n} f_{2r} P_{2(k-r),2(n-r)}.
$$

Во втором случае соответствующее число путей равно

$$
\frac{1}{2}\cdot 2^{2n} f_{2r} P_{2k,2(n-r)}.
$$

Следовательно, для $1\leqslant k\leqslant n-1$

$$
P_{2k,2n} = \frac{1}{2}\sum_{r=1}^{k} f_{2r} P_{2(k-r),2(n-r)} + \frac{1}{2}\sum_{r=1}^{k} f_{2r} P_{2k,2(n-r)}. \tag{14}
$$

Предположим, что формула $P_{2k,2m}=u_{2m}u_{2m-2k}$ верна для $m=1,\ \ldots,\ n-1$. Тогда из (13) и (14) находим, что

$$
\begin{aligned}
P_{2k,2n} = \frac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{k} f_{2r} u_{2n-2r-2k} = \\
= \frac{1}{2} u_{2n-2k} u_{2k} + \frac{1}{2} u_{2k} u_{2n-2k} = u_{2k} u_{2n-2k}. \quad \square
\end{aligned}
$$

Пусть теперь $\gamma(2n)$ — число единиц времени в интервале $[0,\ 2n]$, которое частица проводит на *положительной* стороне. Тогда для $x<1$

$$
\left\{\frac{1}{2} < \frac{\gamma(2n)}{2n} \leqslant x\right\} = \sum_{\left\{k:\,\frac{1}{2}<\frac{2k}{2n}\leqslant x\right\}} P_{2k,2n}.
$$

Поскольку при $k \to \infty$

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}},$$

то

$$P_{2k,2n} = u_{2k} u_{2(n-k)} \sim \frac{1}{\pi\sqrt{k(n-k)}},$$

если $k \to \infty$, $n-k \to \infty$.

Поэтому

$$\sum_{\left\{k:\, \frac{1}{2} < \frac{2k}{2n} \leqslant x\right\}} P_{2k,2n} - \sum_{\left\{k:\, \frac{1}{2} < \frac{2k}{2n} \leqslant x\right\}} \frac{1}{\pi n}\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-1/2} \to 0, \quad n \to \infty,$$

откуда

$$\sum_{\left\{k:\, \frac{1}{2} < \frac{2k}{2n} \leqslant x\right\}} P_{2k,2n} - \frac{1}{\pi}\int_{1/2}^{x} \frac{dt}{\sqrt{t(1-t)}} \to 0, \quad n \to \infty.$$

Но из соображений симметрии

$$\sum_{\left\{k:\, \frac{k}{n} \leqslant \frac{1}{2}\right\}} P_{2k,2n} \to \frac{1}{2}$$

и

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

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 $u(t) = (t(1-t))^{-1/2}$ в интеграле

$$\frac{1}{\pi}\int_{0}^{x} \frac{dt}{\sqrt{t(1-t)}}$$

представляет U-образную кривую, уходящую в бесконечность в точках $t=0$ и $t=1$.

Отсюда следует, что при больших n

$$\left\{0 < \frac{\gamma(2n)}{2n} \leqslant \Delta\right\} > \quad \left\{\frac{1}{2} < \frac{\gamma(2n)}{2n} \leqslant \frac{1}{2} + \Delta\right\},$$

т. е., образно говоря, более вероятно, что доля времени, проводимого 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

$$\left\{\frac{\gamma(2n)}{2n} \leqslant 0{,}024\right\} \approx 0{,}1,$$
$$\left\{\frac{\gamma(2n)}{2n} \leqslant 0{,}1\right\} \approx 0{,}2,$$
$$\left\{\frac{\gamma(2n)}{2n} \leqslant 0{,}2\right\} \approx 0{,}3,$$
$$\left\{\frac{\gamma(2n)}{2n} \leqslant 0{,}65\right\} \approx 0{,}6.$$

Таким образом, если рассматривается временной интервал $[0,\ 1000]$, то вполне может случиться, что примерно в одном случае из десяти частица будет проводить всего 24 единицы времени на положительной стороне, но бо́льшую часть времени — 976 единиц — на отрицательной стороне.

3. Задачи.

1. С какой скоростью $\min(\sigma_{2n},\ 2n) \to \infty$ при $n \to \infty$?

2. Пусть $\tau_n = \min\{1 \leqslant k \leqslant n\colon S_k = 1\}$ и $\tau_n = \infty$, если $S_k < 1$ при всех $1 \leqslant k \leqslant n$. К чему стремится $\min(\tau_n,\ n)$ при $n \to \infty$ для симметричного $(p = q = 1/2)$ и несимметричного $(p \neq q)$ блужданий?

3. Основываясь на идеях и методах § 10, показать, что для симметричного $(p = q = 1/2)$ случайного блуждания Бернулли $\{S_k,\ k \leqslant n\}$ с $S_0 = 0$, $S_k = \xi_1 + \ldots + \xi_k$ имеют место следующие формулы (N — положительное целое число):

$$\left\{\max_{1 \leqslant k \leqslant n} S_k \geqslant N,\ S_n < N\right\} = \{S_n > N\},$$
$$\left\{\max_{1 \leqslant k \leqslant n} S_k \geqslant N\right\} = 2\ \{S_n \geqslant N\} - \{S_n = N\},$$
$$\left\{\max_{1 \leqslant k \leqslant n} S_k = N\right\} = \{S_n = N\} + \{S_n = N + 1\}.$$

§ 11. Мартингалы. Некоторые применения к случайному блужданию

1. Рассмотренные выше бернуллиевские величины $\xi_1, \ldots, \xi_n$ образовывали последовательность *независимых* случайных величин. В этом и следующем параграфах 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 некоторые 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 (5) § 10, полученного выше с применением принципа отражения.

2. Пусть $(\Omega, \mathscr{A}, \ )$ — конечное вероятностное пространство, $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \ldots \preccurlyeq \mathscr{D}_n$ — некоторая последовательность разбиений.

Определение 1. Последовательность случайных величин $\xi = (\xi_k)_{1 \leqslant k \leqslant n}$ называется *мартингалом* (относительно разбиений $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \ldots \preccurlyeq \mathscr{D}_n$), если:

1) ξ_k являются $\mathscr{D}_k$-измеримыми,

2) $\ (\xi_{k+1} \mid \mathscr{D}_k) = \xi_k$, $1 \leqslant k \leqslant n-1$.

Чтобы подчеркнуть, относительно какой *системы разбиений* случайные величины $\xi = (\xi_1, \ldots, \xi_n)$ образуют мартингал, будем использовать также запись

$$\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}, \tag{1}$$

часто опуская для простоты обозначений указание на то, что $1 \leqslant k \leqslant n$.

В том случае, когда разбиения $\mathscr{D}_k$ порождаются случайными 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 сама последовательность $\xi = (\xi_k)$ образует мартингал.

Остановимся на некоторых примерах мартингалов.

Пример 1. Пусть $\eta_1, \ldots, \eta_n$ — независимые бернуллиевские случайные величины с

$$\{\eta_k = 1\} = \ \{\eta_k = -1\} = 1/2,$$
$$S_k = \eta_1 + \ldots + \eta_k \quad \text{и} \quad \mathscr{D}_k = \mathscr{D}_{\eta_1, \ldots, \eta_k}.$$

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 $(S_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ образует мартингал.

Величины S_k являются $\mathscr{D}_k$-измеримыми и в силу (12), (18) и (24) § 8

$$(S_{k+1} \mid \mathscr{D}_k) = \quad (S_k + \eta_{k+1} \mid \mathscr{D}_k) = \quad (S_k \mid \mathscr{D}_k) + \quad (\eta_{k+1} \mid \mathscr{D}_k) = S_k + \quad \eta_{k+1} = S_k,$$

что и есть требуемое мартингальное свойство.

Если положить $S_0 = 0$ и взять $\mathscr{D}_0 = \{\Omega\}$ — 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 $\{\eta_i = 1\} = p$, $\{\eta_i = -1\} = q$. Если $p \neq q$, то каждая из последовательностей $\xi = (\xi_k)$ с

$$\xi_k = \left(\frac{q}{p}\right)^{S_k}, \qquad \xi_k = S_k - k(p - q), \text{ где } S_k = \eta_1 + \ldots + \eta_k,$$

образует мартингал.

Пример 3. Пусть η — некоторая случайная величина, $\mathscr{D}_1 \preccurlyeq \ldots \preccurlyeq \mathscr{D}_n$ и

$$\xi_k = \quad (\eta \mid \mathscr{D}_k). \tag{2}$$

Тогда последовательность $\xi = (\xi_k,\ \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ образует мартингал. В самом деле, $\mathscr{D}_k$-измеримость $\quad (\eta \mid \mathscr{D}_k)$ очевидна и, согласно (20) § 8,

$$(\xi_{k+1} \mid \mathscr{D}_k) = \quad [\quad (\eta \mid \mathscr{D}_{k+1}) \mid \mathscr{D}_k] = \quad (\eta \mid \mathscr{D}_k) = \xi_k.$$

В связи с этим заметим, что если $\xi = (\xi_k,\ \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ — произвольный мартингал, то в силу формулы (20) § 8

$$\xi_k = \quad (\xi_{k+1} \mid \mathscr{D}_k) = \quad [\quad (\xi_{k+2} \mid \mathscr{D}_{k+1}) \mid \mathscr{D}_k] = \quad (\xi_{k+2} \mid \mathscr{D}_k) = \ldots = \quad (\xi_n \mid \mathscr{D}_k). \tag{3}$$

Таким образом, множество всех мартингалов $\xi = (\xi_k,\ \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 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 6 в § 1 гл. VII.)

Пример 4. Пусть $\eta_1, \ldots, \eta_n$ — последовательность независимых одинаково распределенных случайных величин, $S_k = \eta_1 + \ldots + \eta_k$ и $\mathscr{D}_1 = \mathscr{D}_{S_n}$, $\mathscr{D}_2 = \mathscr{D}_{S_n, S_{n-1}}$, $\ldots$, $\mathscr{D}_n = \mathscr{D}_{S_n, \ldots, S_1}$. Покажем, что последовательность $\xi = (\xi_k,\ \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ с $\xi_1 = \dfrac{S_n}{n}$, $\xi_2 = \dfrac{S_{n-1}}{n-1}$, $\ldots$, $\xi_k = \dfrac{S_{n+1-k}}{n+1-k}$, $\ldots$, $\xi_n = S_1$ образует мартингал. Во-первых, ясно, что $\mathscr{D}_k \preccurlyeq \mathscr{D}_{k+1}$ и ξ_k — $\mathscr{D}_k$-измеримы. Далее, в силу симметрии для $j \leqslant n - k + 1$

$$(\eta_j \mid \mathscr{D}_k) = \quad (\eta_1 \mid \mathscr{D}_k) \tag{4}$$

(ср. с (26) § 8). Поэтому

$$(n - k + 1) \quad (\eta_1 \mid \mathscr{D}_k) = \sum_{j=1}^{n-k+1} \quad (\eta_j \mid \mathscr{D}_k) = \quad (S_{n-k+1} \mid \mathscr{D}_k) = S_{n-k+1}.$$

Значит,

$$\xi_k = \frac{S_{n-k+1}}{n-k+1} = \quad (\eta_1 \mid \mathscr{D}_k),$$

и мартингальность последовательности $\xi = (\xi_k,\ \mathscr{D}_k)$ следует из примера 3.

Замечание. Из установленного результата о мартингальном свойстве последовательности $\xi = (\xi_k,\ \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ понятно, почему иногда говорят, что последовательность $(S_k/k)_{1 \leqslant k \leqslant n}$ образует *обращенный мартингал.* (Ср. с задачей 5 в § 1 гл. VII.)

Пример 5. Пусть $\eta_1, \ldots, \eta_n$ — независимые бернуллиевские случайные величины с

$$\{\eta_i = +1\} = \quad \{\eta_i = -1\} = 1/2,$$

$S_k = \eta_1 + \ldots + \eta_k$. 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

$$\xi_k = (\cos\ \lambda)^{-k}\ \exp\Big\{ i\lambda\Big(S_k - \frac{B+A}{2}\Big)\Big\}, \qquad 1 \leqslant k \leqslant n \tag{5}$$

образует комплексный мартингал (т. е. действительная и комплексная части величин ξ_k, $1 \leqslant k \leqslant n$, являются мартингалами).

3. Из определения мартингала следует, что математическое ожидание ξ_k одно и то же для всех k:

$$\xi_k = \quad \xi_1.$$

Оказывается, что это свойство остается для мартингалов справедливым, если вместо (детерминированного) момента k брать так называемые *моменты остановки.*

Для соответствующей формулировки введем такое

Определение 2. Случайная величина $\tau = \tau(\omega)$, принимающая значения $1, 2, \ldots, n$, будет называться *моментом остановки* (относительно разбиений $(\mathscr{D}_k)_{1 \leqslant k \leqslant n}$, $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \ldots \preccurlyeq \mathscr{D}_n$), 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

$$(\xi_\tau \mid \mathscr{D}_1) = \xi_1, \tag{8}$$

где

$$\xi_\tau = \sum_{k=1}^{n} \xi_k I_{\{\tau = k\}} \tag{9}$$

и

$$\xi_\tau = \quad \xi_1. \tag{10}$$

Доказательство (ср. с доказательством формулы (29) § 9). Пусть $D \in \mathscr{D}_1$. Тогда, пользуясь свойствами условных математических ожиданий и принимая во внимание мартингальное свойство (3), находим, что

$$\begin{aligned} (\xi_\tau \mid D) &= \frac{(\xi_\tau I_D)}{(D)} = \frac{1}{(D)} \sum_{l=1}^{n} \quad (\xi_l I_{\{\tau = l\}} I_D) = \\ &= \frac{1}{(D)} \sum_{l=1}^{n} \quad [\quad (\xi_n \mid \mathscr{D}_l) I_{\{\tau = l\}} I_D] = \frac{1}{(D)} \sum_{l=1}^{n} \quad [\quad (\xi_n I_{\{\tau = l\}} I_D \mid \mathscr{D}_l)] = \\ &= \frac{1}{(D)} \sum_{l=1}^{n} \quad [\xi_n I_{\{\tau = l\}} I_D] = \frac{1}{(D)} \quad (\xi_n I_D) = \quad (\xi_n \mid D), \end{aligned}$$

а следовательно,

$$(\xi_\tau \mid \mathscr{D}_1) = \quad (\xi_n \mid \mathscr{D}_1) = \xi_1.$$

Равенство $\quad \xi_\tau = \quad \xi_1$ следует отсюда очевидным образом. □

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*

$$S_\tau = 0, \qquad S_\tau^2 = \quad \tau, \tag{11}$$

называемые тождествами Вальда (*ср. с* (29) *и* (30) § 9; *см. также задачу* 1 *и теорему* 3 *в* § 2 *гл.* VII).

Используем теорему 1 для доказательства следующего утверждения.

Теорема 2 (о баллотировке). *Пусть* $\eta_1, \ldots, \eta_n$ — *последовательность независимых одинаково распределенных случайных величин, принимающих конечное число значений из множества* $\{0, 1, \ldots\}$,

$$S_k = \eta_1 + \ldots + \eta_k, \quad 1 \leqslant k \leqslant n.$$

Тогда (*-п. н.*)

$$(S_k < k \text{ для всех } 1 \leqslant k \leqslant n \,|\, S_n) = \left(1 - \frac{S_n}{n}\right)^+, \tag{12}$$

где $a^+ = \max(a, 0)$.

Доказательство. На множестве $\{\omega\colon S_n \geqslant n\}$ формула очевидна. Будем поэтому доказывать (12) для тех элементарных исходов, для которых $S_n < n$.

Рассмотрим мартингал $\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$, введенный в примере 4, с $\xi_k = \dfrac{S_{n+1-k}}{n+1-k}$ и $\mathscr{D}_k = \mathscr{D}_{S_{n+1-k},\ldots,S_n}$.

Определим

$$\tau = \min\{1 \leqslant k \leqslant n\colon \xi_k \geqslant 1\},$$

полагая $\tau = n$ на множестве

$$\{\xi_k < 1 \text{ для всех } 1 \leqslant k \leqslant n\} = \left\{\max_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1\right\}.$$

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 $S_n < n$ и $\max\limits_{1\leqslant l\leqslant n} \dfrac{S_l}{l} \geqslant 1$. Обозначим $\sigma = n + 1 - \tau$. Нетрудно видеть, что

$$\sigma = \max\{1 \leqslant k \leqslant n\colon S_k \geqslant k\}$$

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 $\{S_n < n\}$

$$\Big(\max_{1\leqslant l\leqslant n} \frac{S_l}{l} \geqslant 1 \,|\, S_n\Big) = \quad (\xi_\tau = 1 \,|\, S_n) = \quad (\xi_\tau \,|\, S_n),$$

где последнее равенство следует из того, что ξ_τ принимает лишь два значения: 0 или 1.

Заметим теперь, что $\quad (\xi_\tau \,|\, S_n) = \quad (\xi_\tau \,|\, \mathscr{D}_1)$ и в силу теоремы 1 $\quad (\xi_\tau \,|\, \mathscr{D}_1) = \xi_1 = S_n/n$. Следовательно, на множестве $\{S_n < n\}$

$$(S_k < k \text{ для всех } 1 \leqslant k \leqslant n \,|\, S_n) = 1 - \frac{S_n}{n}. \qquad \square$$

Применим эту теорему для получения другого доказательства леммы 1 из § 10 и объясним ее название как *теоремы о баллотировке.*

Пусть $\xi_1, \ldots, \xi_n$ — последовательность независимых бернуллиевских случайных величин с

$$\{\xi_i = 1\} = \quad \{\xi_i = -1\} = 1/2,$$

$S_k = \xi_1 + \ldots + \xi_k$ и a, b — целые неотрицательные числа такие, что $a - b > 0$, $a + b = n$. Покажем, что тогда

$$(S_1 > 0, \ldots, S_n > 0 \,|\, S_n = a - b) = \frac{a-b}{a+b}. \tag{15}$$

В самом деле, в силу симметрии

$$\begin{aligned}
(S_1 > 0, \ldots, S_n > 0 \,|\, S_n = a - b) &= \quad (S_1 < 0, \ldots, S_n < 0 \,|\, S_n = -(a-b)) = \\
&= \quad (S_1 + 1 < 1, \ldots, S_n + n < n \,|\, S_n + n = n - (a-b))) = \\
&= \quad (\eta_1 < 1, \ldots, \eta_1 + \ldots + \eta_n < n \,|\, \eta_1 + \ldots + \eta_n = n - (a-b)) = \\
&= \left[1 - \frac{n-(a-b)}{n}\right]^+ = \frac{a-b}{n} = \frac{a-b}{a+b},
\end{aligned}$$

где мы положили $\eta_k = \xi_k + 1$ и воспользовались равенством (12).

Из (15) очевидным образом выводится формула (5) § 10, 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 $\quad (S_1 > 0, \ldots, S_n > 0 \,|\, S_n = a - b)$ 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 $\mathscr{D}_0 \preccurlyeq \mathscr{D}_1 \preccurlyeq \ldots \preccurlyeq \mathscr{D}_n$ — последовательность разбиений, $\mathscr{D}_0 = \{\Omega\}$; η_k — $\mathscr{D}_k$-измеримая величина, $1 \leqslant k \leqslant n$. Доказать, что последовательность

$\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ с

$$\xi_k = \sum_{l=1}^{k} [\eta_l - \quad (\eta_l \,|\, \mathscr{D}_{l-1})]$$

является мартингалом.

2. Пусть величины $\eta_1, \ldots, \eta_n$ таковы, что $\quad (\eta_k \,|\, \eta_1, \ldots, \eta_{k-1}) = 0$, $2 \leqslant k \leqslant n$. Доказать, что последовательность $\xi = (\xi_k)_{1 \leqslant k \leqslant n}$ с $\xi_1 = \eta_1$ и

$$\xi_{k+1} = \sum_{i=1}^{k} \eta_{i+1} f_i(\eta_1, \ldots, \eta_i), \qquad 1 \leqslant k \leqslant n-1,$$

где f_i — некоторые функции, образует мартингал.

3. Показать, что всякий мартингал $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ имеет *некоррелированные приращения*: если $a < b < c < d$, то

$$(\xi_d - \xi_c, \xi_b - \xi_a) = 0.$$

4. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — некоторая случайная последовательность такая, что ξ_k — $\mathscr{D}_k$-измеримы ($\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \ldots \preccurlyeq \mathscr{D}_n$). 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 $(\mathscr{D}_k)$) $\quad \xi_\tau = \quad \xi_1$. (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

$$[\xi_n I_{\{\tau = k\}}] = \quad [\xi_k I_{\{\tau = k\}}].$$

6. Пусть $\xi = (\xi_k, \mathscr{D}_k)$ и $\eta = (\eta_k, \mathscr{D}_k)$ — два мартингала, $\xi_1 = \eta_1 = 0$. Доказать, что

$$\xi_n \eta_n = \sum_{k=2}^{n} \quad (\xi_k - \xi_{k-1})(\eta_k - \eta_{k-1})$$

и, в частности,

$$\xi_n^2 = \sum_{k=2}^{n} \quad (\xi_k - \xi_{k-1})^2.$$

7. Пусть $\eta_1, \ldots, \eta_n$ — последовательность независимых одинаково распределенных случайных величин, $\quad \eta_i = 0$. Показать, что последовательно-

сти $\xi = (\xi_k)$ с

$$\xi_k = \left(\sum_{i=1}^{k} \eta_i\right)^2 - k \quad \eta_1^2,$$

$$\xi_k = \frac{\exp\{\lambda(\eta_1 + \ldots + \eta_k)\}}{(\quad \exp\{\lambda\eta_1\})^k}$$

являются мартингалами.

8. Пусть $\eta_1, \ldots, \eta_n$ — последовательность независимых одинаково распределенных случайных величин, принимающих значения в конечном множестве Y. Пусть $f_0(y) = \quad \{\eta_1 = y\} > 0$, $y \in Y$ и $f_1(y)$ — неотрицательная функция с $\sum\limits_{y \in Y} f_1(y) = 1$. Показать, что последовательность $\xi = (\xi_k, \mathscr{D}_k^\eta)$ с

$$\xi_k = \frac{f_1(\eta_1) \ldots f_1(\eta_k)}{f_0(\eta_1) \ldots f_0(\eta_k)}, \qquad \mathscr{D}_k^\eta = \mathscr{D}_{\eta_1, \ldots, \eta_k}$$

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 $\Omega = \{\omega \colon \omega = (x_1, \ldots, x_n),$ $x_i = 0, 1\}$ вероятность каждого исхода ω задавалась формулой $\quad (\{\omega\}) = p(\omega)$, где

$$p(\omega) = p(x_1) \ldots p(x_n) \tag{1}$$

с $p(x) = p^x q^{1-x}$. При этом условии случайные величины $\xi_1, \ldots, \xi_n$ с $\xi_i(\omega) = x_i$ оказывались *независимыми* и *одинаково* распределенными с

$$\{\xi_1 = x\} = \ldots = \quad \{\xi_n = x\} = p(x), \qquad x = 0, 1.$$

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*:

$$\{\xi_1 = x\} = p_1(x), \quad \ldots, \qquad \{\xi_n = x\} = p_n(x).$$

Рассмотрим 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 $p_0(x)$, $p_1(x, y)$, ..., $p_n(x, y)$ такие, что

$$\begin{gathered}\sum_{x\in X} p_0(x) = 1,\\ \sum_{y\in X} p_k(x, y) = 1, \quad k = 1, \ldots, n;\ \ x \in X.\end{gathered} \tag{2}$$

Для каждого исхода $\omega = (x_0, x_1, \ldots, x_n)$ положим $\quad(\{\omega\}) = p(\omega)$, где

$$p(\omega) = p_0(x_0)\,p_1(x_0, x_1)\ldots p_n(x_{n-1}, x_n). \tag{3}$$

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 $(\Omega, \mathscr{A}, \quad)$, которую принято называть *моделью испытаний, связанных в цепь Маркова*.

Введем в рассмотрение случайные величины $\xi_0, \xi_1, \ldots, \xi_n$ с $\xi_i(\omega) = x_i$ для $\omega = (x_1, \ldots, x_n)$. Простой подсчет показывает, что

$$\begin{gathered}\{\xi_0 = a\} = p_0(a),\\ \{\xi_0 = a_0, \ldots, \xi_k = a_k\} = p_0(a_0)\,p_1(a_0, a_1)\ldots p_k(a_{k-1}, a_k).\end{gathered} \tag{4}$$

Установим теперь справедливость для рассматриваемой вероятностной модели $(\Omega, \mathscr{A}, \quad)$ следующего важного свойства условных вероятностей:

$$(\xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k, \ldots, \xi_0 = a_0) = \quad(\xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k) \tag{5}$$

(в предположении $\quad\{\xi_k = a_k, \ldots, \xi_0 = a_0\} > 0$).

В силу (4) и определения условных вероятностей (§ 3)

$$\begin{aligned}(\xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k, \ldots, \xi_0 = a_0) &= \frac{\{\xi_{k+1} = a_{k+1}, \ldots, \xi_0 = a_0\}}{\{\xi_k = a_k, \ldots, \xi_0 = a_0\}} = \\ &= \frac{p_0(a_0)\,p_1(a_0, a_1)\ldots p_{k+1}(a_k, a_{k+1})}{p_0(a_0)\ldots p_k(a_{k-1}, a_k)} = p_{k+1}(a_k, a_{k+1}).\end{aligned}$$

Аналогичным образом проверяется равенство

$$(\xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k) = p_{k+1}(a_k, a_{k+1}), \tag{6}$$

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

$$(\xi_{k+1} = a_{k+1} \,|\, \mathscr{B}_k^{\xi}) = \quad (\xi_{k+1} = a_{k+1} \,|\, \xi_k), \tag{7}$$

или

$$(\xi_{k+1} = a_{k+1} \,|\, \xi_0, \ldots, \xi_k) = \quad (\xi_{k+1} = a_{k+1} \,|\, \xi_k).$$

Замечание. Прервем наше изложение, чтобы сделать важное для всего дальнейшего замечание в связи с формулами (5) и (7) и событиями *нулевой вероятности*.

При установлении формулы (5) предполагалось, что $\{\xi_k = a_k, \ldots, \xi_0 = a_0\} > 0$ (а значит, и $\{\xi_k = a_k\} > 0$). Нужно это было, в сущности, лишь потому, что (пока!) условные вероятности $(A \,|\, B)$ определялись только в предположении $(B) > 0$.

Но заметим, что если $B = \{\xi_k = a_k, \ldots, \xi_0 = a_0\}$ и $(B) = 0$ (а, значит, и $(C) = 0$ для $C = \{\xi_k \in a_k\}$), то «путь» $\{\xi_0 = a_0, \ldots, \xi_k = a_k\}$ должен рассматриваться как *нереализуемый* и тогда вопрос о том, что есть условная вероятность события $\{\xi_{k+1} = a_k\}$ при условии этого *нереализуемого* «пути», с практической точки зрения не представляет интереса.

В этой связи и для определенности мы будем далее *определять* условную вероятность $(A \,|\, B)$ формулой

$$(A \,|\, B) = \begin{cases} \dfrac{(AB)}{(B)}, & \text{если} \quad (B) > 0, \\ 0, & \text{если} \quad (B) = 0. \end{cases}$$

При таком определении формулы (5) и (7) становятся справедливыми и без всяких оговорок типа $\{\xi_k = a_k, \ldots, \xi_0 = a_0\} > 0$.

Подчеркнем, что отмеченная трудность, связанная с событиями нулевой вероятности, весьма типична для вероятностных рассмотрений. В § 7 гл. II будет приведено общее определение условных вероятностей (относительно *произвольных* разбиений, σ-алгебр, ...), которое и весьма естественно, и «работает» в ситуациях «нулевой вероятности».

Если воспользоваться очевидным равенством

$$(AB \,|\, C) = \quad (A \,|\, BC) \quad (B \,|\, C),$$

то из (7) получаем, что

$$(\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \,|\, \mathscr{B}_k^{\xi}) = \quad (\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \,|\, \xi_k), \tag{8}$$

или

$$(\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \,|\, \xi_0, \ldots, \xi_k) = \quad (\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \,|\, \xi_k). \tag{9}$$

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

$$
\begin{aligned}
&\text{Б} = \{\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1}\},\\
&\text{Н} = \{\xi_k = a_k\},\\
&\text{П} = \{\xi_{k-1} = a_{k-1}, \ldots, \xi_0 = a_0\}.
\end{aligned}
$$

Тогда из (9) следует, что

$$(\text{Б} \mid \text{НП}) = \quad (\text{Б} \mid \text{Н}),$$

откуда легко находим, что

$$(\text{БП} \mid \text{Н}) = \quad (\text{Б} \mid \text{Н}) \quad (\text{П} \mid \text{Н}). \tag{10}$$

Иначе говоря, из соотношения (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 «веса» $p(\omega)$ 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 $(\Omega, \mathscr{A}, \quad)$ — 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 $p_0(x) = \quad \{\xi_0 = x\}$ называют *начальным распределением*, а матрицу $\|p_k(x, y)\|$, $x, y \in X$, с $p_k(x, y) = \quad (\xi_k = y \mid \xi_{k-1} = x)$ — *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,

$$\sum_{y \in X} p(x, y) = 1, \qquad x \in X.$$

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

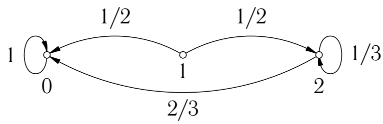
$$i \xrightarrow{p_{ij}} j$$

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

$$\| p_{ij} \| = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 2/3 & 0 & 1/3 \end{pmatrix}.$$

Этой матрице соответствует следующий граф:

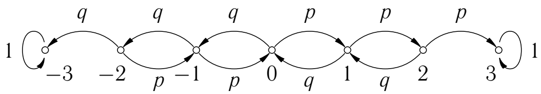


Отметим, что здесь состояние 0 является «поглощающим»: если частица в него попала, то она в нем и остается, поскольку $p_{00} = 1$. Из состояния 1 частица с равными вероятностями переходит в соседние состояния 0 и 2; состояние 2 таково, что частица остается в нем с вероятностью 1/3 и переходит в состояние 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 $\{\eta_i=+1\}=p$, $\{\eta_i=-1\}=q$ и $S_k=\eta_1+\ldots+\eta_k$ — величина выигрыша первого игрока у второго, то последовательность $S_0, S_1, \ldots, S_n$ с $S_0=0$ будет образовывать марковскую цепь с $p_0=1$ и матрицей переходных вероятностей (11), поскольку

$$\begin{aligned}(S_{k+1}=j \,|\, S_k=i_k,\, S_{k-1}=i_{k-1},\, \ldots,\, S_1=i_1) &= \\ = (S_k+\eta_{k+1}=j \,|\, S_k=i_k,\, S_{k-1}=i_{k-1},\, \ldots,\, S_1=i_1) &= \\ = (S_k+\eta_{k+1}=j \,|\, S_k=i_k) &= \{\eta_{k+1}=j-i_k\}.\end{aligned}$$

Марковская цепь $S_0, S_1, \ldots, S_n$ имеет весьма простую структуру:

$$S_{k+1}=S_k+\eta_{k+1}, \quad 0\leqslant k\leqslant n-1,$$

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 $\xi_0, \xi_1, \ldots, \xi_n$ с

$$\xi_{k+1}=f_k(\xi_k, \eta_{k+1}), \quad 0\leqslant k\leqslant n-1, \tag{12}$$

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,

$$\xi_{k+1} = (\xi_k - 1)^+ + \eta_{k+1}, \quad 0 \leqslant k \leqslant n-1,$$

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» превращается в j частиц с вероятностями p_j, $j = 0, 1, \ldots, M$. (Эта схема *процесса гибели-размножения* носит название «*модель Гальтона–Ватсона*»; [125, с. 127].)

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

$$(\xi_{k+1} = i_{k+1} \,|\, \xi_k = i_k,\ \xi_{k-1} = i_{k-1},\ \ldots) =$$
$$= \quad (\xi_{k+1} = i_{k+1} \,|\, \xi_k = i_k) = \quad \{\eta_1^{(k)} + \ldots + \eta_{i_k}^{(k)} = i_{k+1}\}.$$

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

$$p_{ij} = \quad (\xi_{k+1} = j \,|\, \xi_k = i)$$

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 $\xi = (\xi_k, \amalg, \mathbb{P})$ однородную марковскую цепь с вектором (строкой) начальных вероятностей $\amalg = \|p_i\|$ и матрицей переходных вероятностей $\mathbb{P} = \|p_{ij}\|$. Ясно, что

$$p_{ij} = \quad (\xi_1 = j \,|\, \xi_0 = i) = \ldots = \quad (\xi_n = j \,|\, \xi_{n-1} = i).$$

Обозначим

$$p_{ij}^{(k)} = \quad (\xi_k = j \,|\, \xi_0 = i) \quad (= \quad (\xi_{k+l} = j \,|\, \xi_l = i), \;\; l = 1, 2, \ldots)$$

— вероятность перехода за k шагов из состояния i в состояние j и

$$p_j^{(k)} = \quad \{\xi_k = j\}$$

— вероятность нахождения частицы в момент времени k в точке j. Пусть также

$$\amalg^{(k)} = \|p_i^{(k)}\|, \qquad \mathbb{P}^{(k)} = \|p_{ij}^{(k)}\|.$$

Покажем, что переходные вероятности $p_{ij}^{(k)}$ удовлетворяют «*уравнению Колмогорова—Чепмена*»

$$p_{ij}^{(k+l)} = \sum_{\alpha} p_{i\alpha}^{(k)} p_{\alpha j}^{(l)}, \tag{13}$$

или, в матричной форме,

$$\mathbb{P}^{(k+l)} = \mathbb{P}^{(k)} \cdot \mathbb{P}^{(l)}. \tag{14}$$

Доказательство соотношения (13) весьма просто и основано на формуле полной вероятности и марковском свойстве:

$$p_{ij}^{(k+l)} = \quad (\xi_{k+l} = j \,|\, \xi_0 = i) = \sum_{\alpha} \quad (\xi_{k+l} = j, \, \xi_k = \alpha \,|\, \xi_0 = i) =$$
$$= \sum_{\alpha} \quad (\xi_{k+l} = j \,|\, \xi_k = \alpha) \quad (\xi_k = \alpha \,|\, \xi_0 = i) = \sum_{\alpha} p_{\alpha j}^{(l)} p_{i\alpha}^{(k)}.$$

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 уравнения

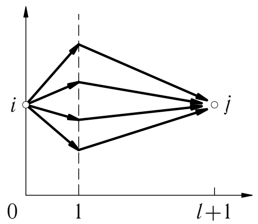


Рис. 22. К обратному уравнению

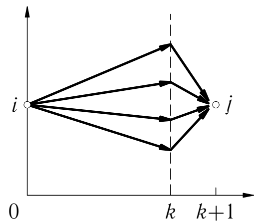


Рис. 23. К прямому уравнению

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,

$$\Pi^{(k+1)} = \Pi^{(k)} \cdot \mathbb{P} \quad (\textit{прямое уравнение})$$

и

$$\Pi^{(k+1)} = \Pi^{(1)} \cdot \mathbb{P}^{(k)} \quad (\textit{обратное уравнение}).$$

Поскольку $\mathbb{P}^{(1)} = \mathbb{P}$, $\Pi^{(0)} = \Pi$, то

$$\mathbb{P}^{(k)} = \mathbb{P}^k, \qquad \Pi^{(k)} = \Pi \cdot \mathbb{P}^k.$$

Тем самым для однородных марковских цепей вероятности перехода за k шагов $p_{ij}^{(k)}$ являются элементами k-х степеней матриц $\mathbb{P}$, 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, $i=0,\ 1$

$$\lim_n p_{i0}^{(n)}=\frac{1-p_{11}}{2-p_{00}-p_{11}},\quad \lim_n p_{i1}^{(n)}=\frac{1-p_{00}}{2-p_{00}-p_{11}},\quad i=0,\ 1.$$

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 $\pi_0>0$, $\pi_1>0$ (ср. далее с теоремой 1).

3. Следующая теорема описывает широкий класс марковских цепей, обладающих так называемым свойством *эргодичности*: пределы $\pi_j=\lim_n p_{ij}^{(n)}$ не только существуют, не зависят от i, образуют распределение вероятностей ($\pi_j\geqslant 0$, $\sum_j \pi_j=1$), но и таковы, что $\pi_j>0$ при всех j (такие распределения называются *эргодическими*; подробнее см. § 3 в гл. VIII).

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 для некоторого* n_0

$$\min_{i,j} \; p_{ij}^{(n_0)} > 0, \tag{21}$$

то найдутся числа $\pi_1, \ldots, \pi_N$ *со свойством*

$$\pi_j > 0, \qquad \sum_j \pi_j = 1 \tag{22}$$

такие, что для каждого $j \in X$ *и 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* $(\pi_1, \ldots, \pi_n)$ *из* a) *удовлетворяют системе уравнений*

$$\pi_j = \sum_\alpha \pi_\alpha p_{\alpha j}, \qquad j = 1, \ldots, N. \tag{24}$$

Доказательство. a) Обозначим

$$m_j^{(n)} = \min_i \; p_{ij}^{(n)}, \qquad M_j^{(n)} = \max_i \; p_{ij}^{(n)}.$$

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

$$M_j^{(n)} - m_j^{(n)} \to 0, \qquad n \to \infty, \qquad j = 1, \ldots, N.$$

Пусть $\varepsilon = \min_{i,j} \; p_{ij}^{(n_0)} > 0$. Тогда

$$p_{ij}^{(n_0+n)} = \sum_\alpha \; p_{i\alpha}^{(n_0)} p_{\alpha j}^{(n)} = \sum_\alpha \; [p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}] \, p_{\alpha j}^{(n)} + \varepsilon \sum_\alpha \; p_{j\alpha}^{(n)} p_{\alpha j}^{(n)} =$$

$$= \sum_\alpha \; [p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}] \, p_{\alpha j}^{(n)} + \varepsilon p_{jj}^{(2n)}.$$

Но $p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)} \geqslant 0$, поэтому

$$p_{ij}^{(n_0+n)} \geqslant m_j^{(n)} \sum_{\alpha} [p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}] + \varepsilon p_{jj}^{(2n)} = m_j^{(n)}(1-\varepsilon) + \varepsilon p_{jj}^{(2n)},$$

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

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \leqslant (M_j^{(n)} - m_j^{(n)})(1-\varepsilon)$$

и, следовательно,

$$M_j^{(kn_0+n)} - m_j^{(kn_0+n)} \leqslant (M_j^{(n)} - m_j^{(n)})(1-\varepsilon)^k \downarrow 0, \quad k \to \infty.$$

Итак, по некоторой подпоследовательности $\{n_\beta\}$ $M_j^{(n_\beta)} - m_j^{(n_\beta)} \to 0$ при $n_\beta \to \infty$. Но разность $M_j^{(n)} - m_j^{(n)}$ монотонна по n, а значит, $M_j^{(n)} - m_j^{(n)} \to 0$ при $n \to \infty$.

Если обозначить $\pi_j = \lim_n m_j^{(n)}$, то из полученных оценок вытекает, что для $n \geqslant n_0$

$$|p_{ij}^{(n)} - \pi_j| \leqslant M_j^{(n)} - m_j^{(n)} \leqslant (1-\varepsilon)^{[n/n_0]-1},$$

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). □

4. Система уравнений (ср. с (24))

$$x_j = \sum_{\alpha} x_\alpha p_{\alpha j}, \quad j = 1, \ldots, N \tag{24*}$$

играет большую роль в теории марковских цепей. Всякое ее *неотрицательное* решение $\mathbb{Q} = (q_1, \ldots, q_N)$, удовлетворяющее условию $\sum_{\alpha} q_\alpha = 1$, принято называть *стационарным* или *инвариантным* распределением вероятностей для марковской цепи с матрицей переходных вероятностей $\|p_{ij}\|$. Объяснение этого названия состоит в следующем.

Возьмем распределение $\mathbb{Q} = (q_1, \ldots, q_N)$ в качестве начального, т. е. пусть $p_j = q_j$, $j = 1, \ldots, N$. Тогда

$$p_j^{(1)} = \sum_\alpha q_\alpha p_{\alpha j} = q_j$$

и вообще $p_j^{(n)} = q_j$. Иначе говоря, если в качестве начального распределения взять $\mathbb{Q} = (q_1, \ldots, q_N)$, то это распределение *не будет изменяться* со временем, т. е. для любого k

$$\{\xi_k = j\} = \quad \{\xi_0 = j\}, \qquad j = 1, \ldots, N.$$

Более того, марковская цепь $\xi = (\xi, \mathbb{Q}, \mathbb{P})$ с таким начальным распределением $\mathbb{Q} = (q_1, \ldots, q_N)$ 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 $(\pi_1, \ldots, \pi_N)$ с $\pi_j > 0$. При этом распределение $(\pi_1, \ldots, \pi_N)$ оказывается также и *стационарным* распределением. Покажем сейчас, что этот 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

$$\tilde{\pi}_j = \sum_\alpha (\tilde{\pi}_\alpha \pi_j) = \pi_j.$$

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

$$\mathbb{P}^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbb{P}^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

и, следовательно, пределы $\lim\limits_n p_{ij}^{(n)}$ не существуют. В то же самое время система

$$q_j = \sum_\alpha q_\alpha p_{\alpha j}, \qquad j = 1, 2,$$

превращается в систему

$$\begin{aligned} q_1 &= q_2, \\ q_2 &= q_1, \end{aligned}$$

единственное решение (q_1, q_2) которой, удовлетворяющее условию $q_1 + q_2 = 1$, есть $(1/2, 1/2)$.

Отметим также, что для рассмотренного выше примера в п. 2 система (24*) с $x_j = q_j$ имеет вид

$$\begin{aligned} q_0 &= q_0 p_{00} + q_1 p_{10}, \\ q_1 &= q_0 p_{01} + q_1 p_{11}, \end{aligned}$$

откуда, учитывая условие $q_0 + q_1 = 1$, находим, что единственное стационарное распределение (q_0, q_1) совпадает с уже найденным:

$$q_0 = \frac{1 - p_{11}}{2 - p_{00} - p_{11}}, \qquad q_1 = \frac{1 - p_{00}}{2 - p_{00} - p_{11}}.$$

Рассмотрим теперь некоторые следствия, вытекающие из эргодической теоремы.

Пусть A — некоторая группа состояний, $A \subseteq X$, и

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

Введем величину

$$\nu_A(n) = \frac{I_A(\xi_0) + \ldots + I_A(\xi_n)}{n+1}$$

— *долю времени*, проводимого частицей в множестве A. Поскольку

$$[I_A(\xi_k) \,|\, \xi_0 = i] = \quad (\xi_k \in A \,|\, \xi_0 = i) = \sum_{j \in A} p_{ij}^{(k)} \quad (= p_i^{(k)}(A)),$$

то

$$[\nu_A(n) \,|\, \xi_0 = i] = \frac{1}{n+1} \sum_{k=0}^{n} p_i^{(k)}(A)$$

и, в частности,

$$[\nu_{\{j\}}(n) \,|\, \xi_0 = i] = \frac{1}{n+1} \sum_{k=0}^{n} p_{ij}^{(k)}.$$

Из анализа известно (см. также лемму 1 в § 3 гл. IV), что если последовательность $a_n \to a$, то $\dfrac{a_0 + \ldots + a_n}{n+1} \to a$, $n \to \infty$. Поэтому если $p_{ij}^{(k)} \to \pi_j$,

$k \to \infty$, то

$$\nu_{\{j\}}(n) \to \pi_j, \qquad \nu_A(n) \to \pi_A, \quad \text{где } \pi_A = \sum_{j \in A} \pi_j.$$

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* $\varepsilon > 0$, *любого множества* $A \subseteq X$ *и произвольного начального распределения*

$$\{|\nu_A(n) - \pi_A| > \varepsilon\} \to 0, \quad n \to \infty. \tag{26}$$

Прежде чем переходить к доказательству, заметим, что непосредственное применение результатов § 5 к бернуллиевским величинам $I_A(\xi_0), \ldots$ $\ldots, I_A(\xi_n), \ldots$ 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

$$|p_{ij}^{(n)} - \pi_j| \leqslant C\rho^n. \tag{27}$$

Рассмотрим состояния i и j (которые могут и совпадать) и покажем, что для $\varepsilon > 0$

$$(|\nu_{\{j\}}(n) - \pi_j| > \varepsilon \,|\, \xi_0 = i) \to 0, \quad n \to \infty. \tag{28}$$

В силу неравенства Чебышева

$$(|\nu_{\{j\}}(n) - \pi_j| > \varepsilon \,|\, \xi_0 = i) \leqslant \frac{\{|\nu_{\{j\}}(n) - \pi_j|^2 \,|\, \xi_0 = i\}}{\varepsilon^2}.$$

Поэтому надо лишь показать, что

$$\{|\nu_{\{j\}}(n) - \pi_j|^2 \,|\, \xi_0 = i\} \to 0, \quad n \to \infty.$$

Простой подсчет показывает, что

$$\{|\nu_{\{j\}}(n) - \pi_j|^2 \,|\, \xi_0 = i\} = \\ = \frac{1}{(n+1)^2} \quad \left\{ \left[\sum_{k=0}^{n} (I_{\{j\}}(\xi_k) - \pi_j) \right]^2 \,\Big|\, \xi_0 = i \right\} = \frac{1}{(n+1)^2} \sum_{k=0}^{n} \sum_{l=0}^{n} m_{ij}^{(k,l)},$$

где

$$m_{ij}^{(k,l)} = \quad \{[I_{\{j\}}(\xi_k) I_{\{j\}}(\xi_l)] \,|\, \xi_0 = i\} - \pi_j \quad [I_{\{j\}}(\xi_k) \,|\, \xi_0 = i] - \\ - \pi_j \quad [I_{\{j\}}(\xi_l) \,|\, \xi_0 = i] + \pi_j^2 = p_{ij}^{(s)} p_{jj}^{(t)} - \pi_j p_{ij}^{(k)} - \pi_j p_{ij}^{(l)} + \pi_j^2,$$

$s = \min(k, l)$ 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 $\xi = (\xi_0, \xi_1, \ldots)$ — марковская цепь с матрицей переходных вероятностей $\|p_{ij}\|$ и фазовым пространством $X = \{0, \pm 1, \ldots, \pm N\}$. 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 $A \leqslant x \leqslant B$

$$\beta_k(x) = \quad((\xi_0, \ldots, \xi_k) \in \mathscr{B}_{k+1} \mid \xi_0 = x).$$

С целью отыскания этих вероятностей (первого выхода марковской цепи из множества (A, B) через верхнюю границу) воспользуемся методом, примененным при выводе обратных уравнений.

Имеем

$$\beta_k(x) = \quad((\xi_0, \ldots, \xi_k) \in \mathscr{B}_{k+1} \mid \xi_0 = x) =$$
$$= \sum_y p_{xy} \quad((\xi_0, \ldots, \xi_k) \in \mathscr{B}_{k+1} \mid \xi_0 = x, \xi_1 = y),$$

где, как нетрудно убедиться, опираясь на марковское свойство и однородность цепи,

$$
\begin{aligned}
((\xi_0, \ldots, \xi_k) \in \mathscr{B}_{k+1} \,|\, \xi_0 = x,\ \xi_1 = y) = \\
= \quad ((x, y, \xi_2, \ldots, \xi_k) \in \mathscr{B}_{k+1} \,|\, \xi_0 = x,\ \xi_1 = y) = \\
= \quad ((y, \xi_2, \ldots, \xi_k) \in \mathscr{B}_k \,|\, \xi_1 = y) = \\
= \quad ((y, \xi_1, \ldots, \xi_{k-1}) \in \mathscr{B}_k \,|\, \xi_0 = y) = \beta_{k-1}(y).
\end{aligned}
$$

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 $m_k(x) = \ \{\tau_k \,|\, \xi_0 = x\}$, 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

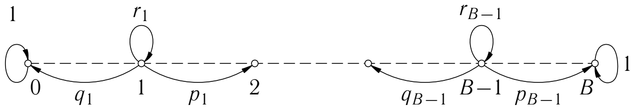
$$p_{00} = 1, \quad p_{BB} = 1$$

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 $\alpha(j+1) - 1 = \sum\limits_{i=0}^{j} (\alpha(i+1) - \alpha(i))$. Поэтому

$$\alpha(j+1) - 1 = (\alpha(1) - 1) \sum_{i=0}^{j} \rho_i.$$

Если $j = B-1$, то $\alpha(j+1) = \alpha(B) = 0$, и, значит,

$$\alpha(1) - 1 = -\frac{1}{\sum\limits_{i=0}^{B-1} \rho_i},$$

откуда

$$\alpha(1) = \frac{\sum\limits_{i=1}^{B-1} \rho_i}{\sum\limits_{i=0}^{B-1} \rho_i} \quad \text{и} \quad \alpha(j) = \frac{\sum\limits_{i=j}^{B-1} \rho_i}{\sum\limits_{i=0}^{B-1} \rho_i}, \qquad j = 1, \ldots, B.$$

(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

$$M(j) = m(j) - m(j-1), \qquad j = 1, \ldots, B.$$

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

$$\rho_j = \frac{q_1 \ldots q_j}{p_1 \ldots p_j}, \qquad R_j = \frac{1}{p_j}\left[1 + \frac{q_j}{p_{j-1}} + \ldots + \frac{q_j \ldots q_2}{p_{j-1} \ldots p_1}\right].$$

Поэтому

$$m(j) = m(j) - m(0) = \sum_{i=0}^{j-1} M(i+1) = \sum_{i=0}^{j-1} (\rho_i m(1) - R_i) = m(1) \sum_{i=0}^{j-1} \rho_i - \sum_{i=0}^{j-1} R_i.$$

Осталось лишь найти $m(1)$. Но $m(B) = 0$, значит,

$$m(1) = \frac{\sum\limits_{i=0}^{B-1} R_i}{\sum\limits_{i=0}^{B-1} \rho_i},$$

и для $1 < j \leqslant B$

$$m(j) = \sum_{i=0}^{j-1} \rho_i \cdot \frac{\sum\limits_{i=0}^{B-1} R_i}{\sum\limits_{i=0}^{B-1} \rho_i} - \sum_{i=0}^{j-1} R_i.$$

(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

$$\begin{aligned}(\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid B \cap \{\xi_k = a_k\}) &= \\ = \quad (\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid \xi_k = a_k) \quad & (29)\end{aligned}$$

(предполагается, что $(B \cap \{\xi_k = a_k\}) > 0$). Действительно, множество B можно представить в виде

$$B = \sum\nolimits^* \{\xi_0 = a_0^*, \ldots, \xi_k = a_k^*\},$$

где сумма $\sum^*$ берется по некоторым наборам $(a_0^*, \ldots, a_k^*)$. Поэтому

$$\begin{aligned}(\xi_n = a_n, \ldots, \xi_{k+1} = a_{k+1} \mid B \cap \{\xi_k = a_k\}) = \frac{(\{\xi_n = a_n, \ldots, \xi_k = a_k\} \cap B)}{(\{\xi_k = a_k\} \cap B)} = \\ = \frac{\sum^* \; (\{\xi_n = a_n, \ldots, \xi_k = a_k\} \cap \{\xi_0 = a_0^*, \ldots, \xi_k = a_k^*\})}{(\{\xi_k = a_k\} \cap B)}. \quad (30)\end{aligned}$$

Но в силу марковского свойства

$$
\begin{aligned}
&(\{\xi_n = a_n,\, \ldots,\, \xi_k = a_k\} \cap \{\xi_0 = a_0^*,\, \ldots,\, \xi_k = a_k^*\}) = \\
&\quad = \begin{cases} (\xi_n = a_n,\, \ldots,\, \xi_{k+1} = a_{k+1} \,|\, \xi_0 = a_0^*,\, \ldots,\, \xi_k = a_k^*) \times \\ \qquad\qquad \times \quad \{\xi_0 = a_0^*,\, \ldots,\, \xi_k = a_k^*\}, \quad \text{если } a_k = a_k^*, \\ 0, \quad \text{если } a_k \neq a_k^*, \end{cases} \\
&\quad = \begin{cases} (\xi_n = a_n,\, \ldots,\, \xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k) \times \\ \qquad\qquad \times \quad \{\xi_0 = a_0^*,\, \ldots,\, \xi_k = a_k^*\}, \quad \text{если } a_k = a_k^*, \\ 0, \quad \text{если } a_k \neq a_k^*, \end{cases} \\
&\qquad\qquad = \begin{cases} (\xi_n = a_n,\, \ldots,\, \xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k) \times \\ \qquad\qquad \times \quad (\{\xi_k = a_k\} \cap B), \quad \text{если } a_k = a_k^*, \\ 0, \quad \text{если } a_k \neq a_k^*. \end{cases}
\end{aligned}
$$

Тем самым сумма $\sum^*$ в (30) равна

$$(\xi_n = a_n,\, \ldots,\, \xi_{k+1} = a_{k+1} \,|\, \xi_k = a_k) \quad (\{\xi_k = a_k\} \cap B),$$

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 $\mathscr{B}_\tau^\xi$ 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 $\{A \cap B \cap (\xi_\tau = a_0)\} > 0$, то выполнены следующие строго марковские свойства:*

$$
\begin{aligned}
&(\xi_{\tau+l} = a_l,\, \ldots,\, \xi_{\tau+1} = a_1 \,|\, A \cap B \cap (\xi_\tau = a_0)) = \\
&\qquad\qquad = \quad (\xi_{\tau+l} = a_l,\, \ldots,\, \xi_{\tau+1} = a_1 \,|\, A \cap \{\xi_\tau = a_0)\}),
\end{aligned} \tag{32}
$$

и (в предположении $(A \cap \{\xi_\tau = a_0\}) > 0$)

$$(\xi_{\tau+l} = a_l,\, \ldots,\, \xi_{\tau+1} = a_1 \,|\, A \cap \{\xi_\tau = a_0\}) = p_{a_0 a_1} \ldots p_{a_{l-1} a_l}. \tag{33}$$

Доказательство проведем для простоты лишь в случае $l=1$. Поскольку $B\cap(\tau=k)\in\mathscr{B}_k^{\xi}$, то, согласно (29),

$$(\xi_{\tau+1}=a_1,\, A\cap B\cap\{\xi_\tau=a_0\})=\sum_{k\leqslant n-1}\quad\{\xi_{k+1}=a_1,\,\xi_k=a_0,\,\tau=k,\,B\}=$$

$$=\sum_{k\leqslant n-1}\quad(\xi_{k+1}=a_1\,|\,\xi_k=a_0,\,\tau=k,\,B)\quad\{\xi_k=a_0,\,\tau=k,\,B\}=$$

$$=\sum_{k\leqslant n-1}\quad(\xi_{k+1}=a_1\,|\,\xi_k=a_0)\quad\{\xi_k=a_0,\,\tau=k,\,B\}=$$

$$=p_{a_0a_1}\sum_{k\leqslant n-1}\quad\{\xi_k=a_0,\,\tau=k,\,B\}=p_{a_0a_1}\quad(A\cap B\cap\{\xi_\tau=a_0\}),$$

что доказывает и (32), и (33) (в случае (33) надо взять $B=\Omega$). □

Замечание 1. В случае $l=1$ строго марковское свойство (32), (33) эквивалентно, очевидно, тому, что для любого $C\subseteq X$

$$(\xi_{\tau+1}\in C\,|\,A\cap B\cap\{\xi_\tau=a_0\})=P_{a_0}(C), \tag{34}$$

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 $A=\{\tau\leqslant n-1\}$

$$(\xi_{\tau+1}\in C\,|\,\mathscr{B}_\tau^{\xi})=P_{\xi_\tau}(C), \tag{35}$$

что является одной из обычно используемых форм строго марковского свойства в общей теории однородных марковских процессов.

Замечание 2. Свойства (32) и (33) остаются справедливыми (если воспользоваться соглашениями, описанными в замечании 1) и без оговорок, что вероятности событий $A\cap B\cap\{\xi_\tau=a_0\}$ и $A\cap\{\xi_\tau=a_0\}$ должны быть положительными.

7. Пусть $\xi=(\xi_0,\,\ldots,\,\xi_n)$ — однородная марковская цепь с матрицей переходных вероятностей $\|p_{ij}\|$,

$$f_{ii}^{(k)}=\quad(\xi_k=i,\,\xi_l\neq i,\,1\leqslant l\leqslant k-1\,|\,\xi_0=i) \tag{36}$$

и для $i\neq j$

$$f_{ij}^{(k)}=\quad(\xi_k=j,\,\xi_l\neq j,\,1\leqslant l\leqslant k-1\,|\,\xi_0=i) \tag{37}$$

— вероятности *первого возвращения* в состояние i в момент времени k и *первого попадания* в состояние j в момент k соответственно.

Покажем, что

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, \quad \text{где } p_{jj}^{(0)} = 1. \tag{38}$$

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

$$\tau_j = \min\{1 \leqslant k \leqslant n \colon \xi_k = j\},$$

полагая $\tau_j = n+1$, если $\{\cdot\} = \varnothing$. Тогда $f_{ij}^{(k)} = \quad (\tau_j = k \,|\, \xi_0 = i)$ и

$$p_{ij}^{(n)} = \quad (\xi_n = j \,|\, \xi_0 = i) = \sum_{1 \leqslant k \leqslant n} \quad (\xi_n = j,\, \tau_j = k \,|\, \xi_0 = i) = \\ = \sum_{1 \leqslant k \leqslant n} \quad (\xi_{\tau_j + n - k} = j,\, \tau_j = k \,|\, \xi_0 = i), \tag{39}$$

где последнее равенство следует из того, что $\xi_{\tau_j+n-k} = \xi_n$ на множестве $\{\tau_j = k\}$. Далее, для всякого $1 \leqslant k \leqslant n$ множество $\{\tau_j = k\} = \{\tau_j = k,\, \xi_{\tau_j} = j\}$. Поэтому, если $\quad \{\xi_0 = i,\, \tau_j = k\} > 0$, то в силу теоремы 2

$$(\xi_{\tau_j+n-k} = j \,|\, \xi_0 = i,\, \tau_j = k) = \quad (\xi_{\tau_j+n-k} = j \,|\, \xi_0 = i,\, \tau_j = k,\, \xi_{\tau_j} = j) = \\ = \quad (\xi_{\tau_j+n-k} = j \,|\, \xi_{\tau_j} = j) = p_{jj}^{(n-k)}$$

и, согласно (37),

$$p_{ij}^{(n)} = \sum_{k=1}^{n} \quad (\xi_{\tau_j+n-k} = j \,|\, \xi_0 = i,\, \tau_j = k) \quad (\tau_j = k \,|\, \xi_0 = i) = \sum_{k=1}^{n} p_{jj}^{(n-k)} f_{ij}^{(k)},$$

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 $\lambda_1 = 1$ является собственным числом, а все остальные корни $\lambda_2, \ldots, \lambda_r$ по модулю не больше 1. Если все собственные

числа $\lambda_1, \ldots, \lambda_r$ различны, то $p_{ij}^{(k)}$ допускают представление

$$p_{ij}^{(k)} = \pi_j + a_{ij}(2)\lambda_2^k + \ldots + a_{ij}(r)\lambda_r^k,$$

где π_j, $a_{ij}(2), \ldots, a_{ij}(r)$ выражаются через элементы матрицы $\mathbb{P}$. (Из этого алгебраического подхода к анализу свойств марковских цепей, в частности, вытекает, что при $|\lambda_2| < 1, \ldots, |\lambda_r| < 1$ для каждого j существует предел $\lim_k p_{ij}^{(k)}$, не зависящий от i.)

3. Пусть $\xi = (\xi_0, \ldots, \xi_n)$ — однородная марковская цепь с множеством состояний X и матрицей переходных вероятностей $\mathbb{P} = \| p_{xy} \|$. Обозначим

$$T\varphi(x) = \quad [\varphi(\xi_1) \,|\, \xi_0 = x] \quad \Big(= \sum_y \varphi(y)\, p_{xy}\Big)$$

— *оператор перехода* за один шаг. Пусть неотрицательная функция $\varphi = \varphi(x)$ удовлетворяет уравнению

$$T\varphi(x) = \varphi(x), \qquad x \in X.$$

Доказать, что последовательность случайных величин

$$\zeta = (\zeta_k, \mathscr{D}_k^\xi) \quad \text{с} \quad \zeta_k = \varphi(\xi_k)$$

образует мартингал.

4. Пусть $\xi = (\xi_n, \mathbb{\Pi}, \mathbb{P})$ и $\tilde{\xi} = (\tilde{\xi}_n, \widetilde{\mathbb{\Pi}}, \mathbb{P})$ — две марковские цепи, отличающиеся начальными распределениями $\mathbb{\Pi} = (p_1, \ldots, p_r)$ и $\widetilde{\mathbb{\Pi}} = (\tilde{p}_1, \ldots, \tilde{p}_r)$. Пусть $\mathbb{\Pi}^{(n)} = (p_1^{(n)}, \ldots, p_r^{(n)})$, $\widetilde{\mathbb{\Pi}}^{(n)} = (\tilde{p}_1^{(n)}, \ldots, \tilde{p}_r^{(n)})$. Показать, что если

$$\min_{i,j} p_{ij} \geqslant \varepsilon > 0,$$

то

$$\sum_{i=1}^r |\tilde{p}_i^{(n)} - p_i^{(n)}| \leqslant 2(1 - r\varepsilon)^n.$$

5. Пусть P и Q — стохастические матрицы. Показать, что PQ и $\alpha P + (1-\alpha)Q$ с $0 \leqslant \alpha \leqslant 1$ также являются стохастическими матрицами.

6. Рассматривается однородная марковская цепь $(\xi_0, \ldots, \xi_n)$ со значениями в $X = \{0, 1\}$ и матрицей переходных вероятностей

$$\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

где $0 < p < 1$, $0 < q < 1$. Положим $S_n = \xi_0 + \ldots + \xi_n$. В обобщение теоремы

Муавра—Лапласа (§ 6) показать, что

$$\left\{ \frac{S_n - \dfrac{p}{p+q}n}{\sqrt{\dfrac{npq(2-p-q)}{(p+q)^3}}} \leqslant x \right\} \to \Phi(x), \qquad n \to \infty.$$

Убедиться в том, что в случае $p+q=1$ величины $\xi_0, \ldots, \xi_n$ независимы и сформулированное утверждение сводится к тому, что

$$\left\{ \frac{S_n - pn}{\sqrt{npq}} \leqslant x \right\} \to \Phi(x), \qquad n \to \infty.$$

§ 13. Производящие функции

1. В дискретной теории вероятностей, оперирующей с конечным или счетным множеством состояний, и, более обще, в дискретной математике метод производящих функций, восходящий к Л. Эйлеру (XVIII в.), является одним из мощных алгебраических средств решения задач комбинаторного характера, возникающих, в частности, и в теории вероятностей.

Прежде чем давать необходимые определения, связанные с производящими функциями, приведем формулировки двух вероятностных задач, для которых метод производящих функция дает быстрый способ получения решений.

2. Задача Галилея. Одновременно и независимым образом бросаются три правильные шестигранные кости с нанесенными на их гранях цифрами $1, 2, \ldots, 6$. Спрашивается, какова вероятность P того, что сумма выпавших очков равна 10. (Будет показано, что $P = \dfrac{1}{8}$.)

3. Задача о счастливых билетах. Будем считать, что случайно покупаемые билеты имеют шестизначные номера от 000 000 до 999 999 (общим числом 10^6). Спрашивается, какова вероятность P того, что у купленного билета сумма первых трех цифр совпадает с суммой последних трех цифр. (Случайность номера купленного билета означает, что его вероятность равна 10^{-6}.)

Далее мы увидим, что метод производящих функций полезен для решения не только вероятностных задач. Этот метод «работает» при отыскании формул для элементов последовательностей, удовлетворяющих *рекуррентным соотношениям*. Например, числа Фибоначчи F_n, $n \geqslant 0$, удовлетворяют соотношению $F_n = F_{n-1} + F_{n-2}$ при $n \geqslant 2$ с $F_0 = F_1 = 1$. Методом производящих функций далее (п. 6) будет найдено, что

$$F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right].$$

Ниже будет показано также, как этот метод используется для отыскания *целочисленных решений* соотношений

$$X_1 + \ldots + X_n = r$$

при тех или иных ограничениях на X_i, $i = 1, \ldots, n$, и заданном значении r из множества $\{0, 1, 2, \ldots\}$.

4. Перейдем к определениям.

Пусть $A = A(x)$ — действительнозначная функция, $x \in R$, допускающая при $|x| < \lambda$ с некоторым $\lambda > 0$ представление в виде ряда

$$A(x) = a_0 + a_1 x + a_2 x^2 + \ldots \tag{1}$$

с некоторыми коэффициентами

$$a = (a_0, a_1, a_2, \ldots).$$

Понятно, что по значениям этой функции $A(x)$, $|x| < \lambda$, можно однозначно восстановить коэффициенты $a = (a_0, a_1, a_2, \ldots)$, что и оправдывает название для $A = A(x)$ как *производящей функции последовательности* $a = (a_0, a_1, a_2, \ldots)$.

Наряду с функциями $A = A(x)$, задаваемыми в виде ряда (1), часто оказывается полезным обращение к *экспоненциальным производящим* функциям

$$E(x) = a_0 + a_1 \frac{x}{1!} + a_2 \frac{x^2}{2!} + \ldots, \tag{2}$$

называемым так по той очевидной причине, что последовательность $(a_0, a_1, a_2, \ldots) \equiv (1, 1, 1, \ldots)$ порождает экспоненциальную функцию $\exp(x)$ (см. пример 3 в § 14).

Во многих вопросах целесообразно также обращение к производящим функциям «бесконечных в обе стороны» последовательностей

$$a = (\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots)$$

(см. пример в п. 7).

Если $\xi = \xi(\omega)$ — случайная величина, принимающая значения $0, 1, 2, \ldots$ с вероятностями $p_0, p_1, p_2, \ldots$ (т. е. $p_i = \quad (\xi = i)$, $\sum\limits_{i=0}^{\infty} p_i = 1$), то функция

$$G(x) = \sum_{i=0}^{\infty} p_i x^i \tag{3}$$

заведомо определена для $|x| \leqslant 1$.

Эта функция, являющаяся производящей функцией последовательности $(p_0, p_1, p_2, \ldots)$, есть не что иное, как математическое ожидание $\quad x^{\xi}$, которое в случае обрывающихся последовательностей $(p_0, p_1, p_2, \ldots)$ было определено в § 4, а в общем случае определяется в § 6 главы II.

В теории вероятностей функция

$$G(x) = \quad x^{\xi} \ \left(= \sum_{i=0}^{\infty} p_i x^i \right), \tag{4}$$

по понятным причинам, называется *производящей функцией случайной величины* $\xi = \xi(\omega)$.

5. Отметим некоторые полезные свойства, устанавливаемые биективным (т. е. взаимно однозначным) соответствием

$$(a_n)_{n\geqslant 0} \leftrightarrow A(x), \tag{5}$$

где $A(x)$ задается рядом (1).

Если наряду с (5) также

$$(b_n)_{n\geqslant 0} \leftrightarrow B(x), \tag{6}$$

то для всяких констант c и d

$$(ca_n + db_n)_{n\geqslant 0} \leftrightarrow cA(x) + dB(x) \tag{7}$$

и выполнено также свойство *свертки*:

$$\left(\sum_{i=0}^{n} a_i b_{n-i} \right)_{n\geqslant 0} \ \leftrightarrow \ A(x)B(x). \tag{8}$$

Помимо операции свертки отметим также

– *операцию композиции* (или подстановки)

$$(A \circ B)(x) = A(B(x)), \tag{9}$$

означающую, что если A и B таковы, что выполнены (5) и (6), то

$$(A \circ B)(x) = \sum_{n\geqslant 0} a_n \left(\sum_{i\geqslant 0} b_i x^i \right)^n; \tag{10}$$

– *операцию* (формального) *дифференцирования* D, действующую на $A(x)$ по формуле

$$D(A(x)) = \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n. \tag{11}$$

Оператор D обладает следующими свойствами, хорошо известными для обычного дифференцирования:

$$D(AB) = D(A)\, B + A\, D(B), \qquad D(A \circ B) = (D(A) \circ B)\, D(B). \tag{12}$$

6. Начнем иллюстрацию собственно метода производящих функций с примера отыскания числа $N(r;\, n)$ решений $(X_1,\, \ldots,\, X_n)$ со значениями в $\{0,\, 1,\, 2,\, \ldots\}$ соотношения

$$X_1 + \ldots + X_n = r \tag{13}$$

при ограничениях

$$\begin{aligned} &X_1 \in \{k_1^{(1)},\, k_1^{(2)},\, \ldots\}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &X_n \in \{k_n^{(1)},\, k_n^{(2)},\, \ldots\}, \end{aligned} \tag{14}$$

где $0 \leqslant k_j^{(1)} < k_j^{(2)} < \ldots \leqslant r$, $j = 1,\, \ldots,\, n$, и r из множества $\{0,\, 1,\, 2,\, \ldots\}$.

Если, например, ищутся целочисленные решения X_1, X_2, X_3 соотношения

$$X_1 + X_2 + X_3 = 3$$

при ограничениях $0 \leqslant X_i \leqslant 3$, $i = 1,\, 2,\, 3$, то простым перебором можно найти все 10 решений: (1, 1, 1), (0, 1, 2), (0, 2, 1), (1, 0, 2), (1, 2, 0), (2, 0, 1), (2, 1, 0), (0, 0, 3), (0, 3, 0), (3, 0, 0). Но, разумеется, такой «переборный» способ становится нерациональным при больших n и r.

Введем следующие производящие функции:

$$\begin{aligned} &A_1(x) = x^{k_1^{(1)}} + x^{k_1^{(2)}} + \ldots, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &A_n(x) = x^{k_n^{(1)}} + x^{k_n^{(2)}} + \ldots, \end{aligned} \tag{15}$$

построенные в соответствии с условиями (14), и рассмотрим их произведение $A(x) = A_1(x) \ldots A_n(x)$:

$$A(x) = \left(x^{k_1^{(1)}} + x^{k_1^{(2)}} + \ldots\right) \ldots \left(x^{k_n^{(1)}} + x^{k_n^{(2)}} + \ldots\right). \tag{16}$$

При раскрытии скобок в правой части возникнут члены вида x^k с некоторыми коэффициентами. Возьмем $k = r$. Тогда видим, что x^r возникает от перемножения членов вида $x^{k_1^{(i_1)}},\, \ldots,\, x^{k_n^{(i_n)}}$ с

$$k_1^{(i_1)} + \ldots + k_n^{(i_n)} = r. \tag{17}$$

Число же разных возможностей типа (17) есть в точности число $N(r;\, n)$ решений системы (13)–(14). Тем самым установлена следующая простая, но важная лемма.

Лемма. *Число* $N(r;\, n)$ *целочисленных решений системы* (13)–(14) *есть в точности коэффициент при* x^r *в произведении* $A_1(x) \ldots A_n(x)$ *производящих функций из* (15).

7. Примеры. (a) Пусть все множества $\left\{k_i^{(1)}, k_i^{(2)}, \ldots\right\}$ имеют вид $\{0, 1, 2, \ldots\}$. Тогда

$$A_1(x)\ldots A_n(x) = (1 + x + x^2 + \ldots)^n$$

и коэффициент $N(r;\, n)$ при x^r в разложении $(1 + x + x^2 + \ldots)^n$ дает *число целочисленных решений соотношения* (13), подчиняющихся условию $X_i \geqslant 0$.

Поскольку

$$(1 + x + x^2 + \ldots)^n = \sum_{r \geqslant 0} C_{n+r-1}^r x^r, \tag{18}$$

в рассматриваемом случае число решений

$$N(r;\, n) = C_{n+r-1}^r \tag{19}$$

(задача 7).

(b) Обратимся к решению *задачи Галилея*, сформулированной в п. 1.

Если X_i — число очков на i-й кости ($X_i = 1, \ldots, 6$), $i = 1, 2, 3$, то число всех возможностей, дающих в сумме 10 очков,— это есть число $N(10;\, 3)$ целочисленных решений системы

$$X_1 + X_2 + X_2 = 10$$

с ограничениями $1 \leqslant X_i \leqslant 6$, $i = 1, 2, 3$.

Общее число возможностей (X_1, X_2, X_3) при бросании трех костей равно $6^3 = 216$. Поэтому интересующая нас вероятность

$$P = \frac{N(10;\, 3)}{216}.$$

Согласно лемме, число $N(10;\, 3)$ равно коэффициенту при x^{10} в разложении $(x + x^2 + \ldots + x^6)^3$ по степеням x. Для отыскания этого коэффициента заметим, что

$$x + x^2 + \ldots + x^6 = x(1 + x + \ldots + x^5) = x(1 - x^6)(1 + x + x^2 + \ldots + x^5 + \ldots).$$

Значит,

$$(x + x^2 + \ldots + x^6)^3 = x^3(1 - x^6)^3(1 + x + x^2 + \ldots)^3. \tag{20}$$

Мы уже видели в (18), что

$$(1 + x + x^2 + \ldots)^3 = C_2^0 + C_3^1 x + C_4^2 x^2 + \ldots \tag{21}$$

Из биномиальной формулы

$$(a + b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k} \tag{22}$$

находим, что

$$(1-x^6)^3 = C_3^0 - C_3^1 x^6 + C_3^2 x^{12} - C_3^3 x^{18}.$$

Тем самым, из (20)

$$(x + x^2 + \ldots + x^6)^3 = x^3(1 + 3x + 6x^2 + \ldots + 36x^7)(1 - 3x^6 + 3x^{12} - x^{18}).$$

Отсюда видно, что коэффициент при x^{10} равен $36-9=27$. Таким образом, интересующая нас вероятность

$$P = \frac{27}{216} = \frac{1}{8}.$$

(c) Задача о *счастливых билетах* решается сходным образом.

Действительно, пусть $(X_1, \ldots, X_6)$ — вектор, состоящий из независимых одинаково распределенных величин таких, что

$$p_k \equiv \quad (X_i = k) = \frac{1}{10}$$

для всех $k=0, 1, \ldots, 9$ и $i=1, \ldots, 6$.

Производящая функция случайной величины X_i

$$G_{X_i}(x) = \quad x^{X_i} = \sum_{k=0}^{9} p_k x^k = \frac{1}{10}(1 + x + \ldots + x^9) = \frac{1}{10}\,\frac{1-x^{10}}{1-x}.$$

В силу предполагаемой независимости величин $X_1, \ldots, X_6$

$$G_{X_1+X_2+X_3}(x) = G_{X_4+X_5+X_6}(x) = \frac{1}{10^3}\,\frac{(1-x^{10})^3}{(1-x)^3}.$$

и производящая функция $G_Y(x)$ величины $Y=(X_1+X_2+X_3)-(X_4+X_5+X_6)$ дается следующим выражением:

$$G_Y(x) = G_{X_1+X_2+X_3}(x)\, G_{X_4+X_5+X_6}\left(\frac{1}{x}\right) = \frac{1}{10^6}\,\frac{1}{x^{27}}\left(\frac{1-x^{10}}{1-x}\right)^6. \tag{23}$$

Записывая $G_Y(x)$ в виде

$$G_Y(x) = \sum_{k=0}^{\infty} q_k x^k, \tag{24}$$

видим, что требуемая вероятность

$$(X_1 + X_2 + X_3 = X_4 + X_5 + X_6),$$

т. е. вероятность $\quad(Y=0)$, равна коэффициенту q_0 (при x^0) в представлении $G_Y(x)$ формулой (24).

Имеем

$$\frac{1}{x^{27}}\left(\frac{1-x^{10}}{1-x}\right)^6=\frac{1}{x^{27}}\,(1-x^{10})^6(1+x+x^2+\ldots)^6.$$

Из представления (18) видим, что

$$(1+x+x^2+\ldots)^6=\sum_{r\geqslant 0}C_{r+5}^r x^r. \tag{25}$$

Биномиальная формула (22) дает

$$(1-x^{10})^6=\sum_{k=0}^{6}(-1)^k C_6^k x^{10k}. \tag{26}$$

Тем самым,

$$G_Y(x)=\frac{1}{10^6}\,\frac{1}{x^{27}}\left(\sum_{r\geqslant 0}C_{r+5}^r x^r\right)\left(\sum_{k=0}^{6}(-1)^k C_6^k x^{10k}\right)\quad\left(=\sum_{k=0}^{\infty}q_k x^k\right). \tag{27}$$

Произведя перемножение участвующих здесь сумм, (после простых, но все же несколько громоздких выкладок) можно найти, что требуемый коэффициент

$$q_0=\frac{55\,252}{10^6}=0,\,05525, \tag{28}$$

что и есть искомая вероятность $(Y=0)$ того, что $X_1+X_2+X_3=X_4+X_5+X_6$ (т. е. что сумма первых трех чисел равна сумме последних трех).

(d) Рассмотрим применение аппарата производящих функций для отыскания решений *разностных уравнений*, возникающих в комбинаторно-вероятностных задачах.

В п. 1 говорилось о числах Фибоначчи F_n, возникших (в связи с подсчетом числа кроликов в n-м поколении) в книге *Liber abacci*, опубликованной в 1220 г. Леонардо Пизанским, известным как Фибоначчи.

Теоретико-числовая интерпретация этих чисел такова: F_n — это число возможных представлений числа n как *упорядоченной суммы, составленной из единиц и двоек*. Ясно, что $F_1=1$, $F_2=2$ (поскольку $2=1+1=2$), $F_3=3$ (поскольку $3=1+1+1=1+2=2+1$).

Исходя из этой интерпретации, можно показать, что числа F_n удовлетворяют рекуррентным соотношениям

$$F_n=F_{n-1}+F_{n-2},\qquad n\geqslant 2 \tag{29}$$

(с $F_0=F_1=1$).

Пусть

$$G(x) = \sum_{n=0}^{\infty} F_n x^n \tag{30}$$

— производящая функция последовательности $\{F_n,\ n \geqslant 0\}$.

Ясно, что

$$G(x) = 1 + x + \sum_{n \geqslant 0} F_{n+2} x^{n+2}. \tag{31}$$

В силу (29)

$$F_{n+2} x^{n+2} = x(F_{n+1} x^{n+1}) + x^2 (F_n x^n),$$

и, значит,

$$\sum_{n \geqslant 0} F_{n+2} x^{n+2} = x \sum_{n \geqslant 0} F_{n+1} x^{n+1} + x^2 \sum_{n \geqslant 0} F_n x^n.$$

Отсюда и из (31)

$$G(x) - 1 - x = x[G(x) - 1] + x^2 G(x).$$

Следовательно,

$$G(x) = \frac{1}{1 - x - x^2}. \tag{32}$$

Заметим, что

$$1 - x - x^2 = -(x - a)(x - b), \tag{33}$$

где

$$a = \frac{1}{2}(-1 - \sqrt{5}), \qquad b = \frac{1}{2}(-1 + \sqrt{5}). \tag{34}$$

Из (32) и (33) получаем

$$G(x) = \frac{1}{a - b}\left[\frac{1}{a - x} - \frac{1}{b - x}\right] = \frac{1}{a - b}\left[\frac{1}{a} \sum_{n \geqslant 0} \left(\frac{x}{a}\right)^n - \frac{1}{b} \sum_{n \geqslant 0} \left(\frac{x}{b}\right)^n\right] =$$
$$= \sum_{n \geqslant 0} x^n \left[\frac{1}{a - b}\left(\frac{1}{a^{n+1}} - \frac{1}{b^{n+1}}\right)\right]. \tag{35}$$

Поскольку $G(x) = \sum_{n \geqslant 0} x^n F_n$, из (35) следует, что

$$F_n = \frac{1}{a - b}\left[\frac{1}{a^{n+1}} - \frac{1}{b^{n+1}}\right].$$

С учетом значений для a и b в (34) находим, что

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1 + \sqrt{5}}{2}\right)^{n+1} - \left(\frac{1 - \sqrt{5}}{2}\right)^{n+1}\right], \tag{36}$$

где

$$\frac{1+\sqrt{5}}{2} = 1,\ 6180\ldots, \qquad \frac{1-\sqrt{5}}{2} = -0,\ 6180\ldots$$

(e) Метод производящих функций весьма эффективен при отыскании разного рода вероятностей для *случайных блужданий*.

Проиллюстрируем это на двух моделях простого случайного блуждания, рассмотренных в § 2 и § 9.

Пусть $S_0 = 0$ и $S_k = \xi_1 + \ldots + \xi_k$ — сумма независимых одинаково распределенных случайных величин, $k \geqslant 1$, где ξ_i принимают два значения $+1$ и -1 с вероятностями p и q.

Пусть $P_n(i) = \quad (S_n = i)$ есть вероятность того, что случайное блуждание $\{S_k,\ 0 \leqslant k \leqslant n\}$ в момент времени n окажется в точке i. (Ясно, что $i \in \{0,\ \pm 1,\ \ldots,\ \pm n\}$.)

В § 2 (п. 1) с помощью комбинаторных рассмотрений было найдено, что (для целых $\frac{n+i}{2}$)

$$P_n(i) = C_n^{\frac{n+i}{2}}\, p^{\frac{n+i}{2}}\, q^{\frac{n-i}{2}}. \tag{37}$$

Покажем, как к этому ответу можно прийти, опираясь на метод производящих функций.

По формуле полной вероятности (см. § 3 и § 8)

$$\begin{aligned} (S_n = i) = {} & \quad (S_n = i \mid S_{n-1} = i-1) \quad (S_{n-1} = i-1) + \\ & + \quad (S_n = i \mid S_{n-1} = i+1) \quad (S_{n-1} = i+1) = \\ & = \quad (\xi_1 = 1) \quad (S_{n-1} = i-1) + \quad (\xi_1 = -1) \quad (S_{n-1} = i+1) = \\ & = p \quad (S_{n-1} = i-1) + q \quad (S_{n-1} = i+1). \end{aligned}$$

Следовательно, для $i = 0,\ \pm 1,\ \pm 2,\ \ldots$ получаем рекуррентные соотношения

$$P_n(i) = p\, P_{n-1}(i-1) + q\, P_{n-1}(i+1), \tag{38}$$

справедливые для любого $n \geqslant 1$ с $P_0(0) = 1$, $P_0(i) = 0$ при $i \neq 0$.

Введем производящие функции

$$G_k(x) = \sum_{i=-\infty}^{\infty} P_k(i) x^i \quad \left(= \sum_{i=-k}^{k} P_k(i) x^i \right). \tag{39}$$

В силу (38)

$$
\begin{aligned}
G_n(x) &= \sum_{i=-\infty}^{\infty} pP_{n-1}(i-1)x^i + \sum_{i=-\infty}^{\infty} qP_{n-1}(i+1)x^i = \\
&= px \sum_{i=-\infty}^{\infty} P_{n-1}(i-1)x^{i-1} + qx^{-1} \sum_{i=-\infty}^{\infty} P_{n-1}(i+1)x^{i+1} = \\
&= (px+qx^{-1})G_{n-1}(x) = \ldots = (px+qx^{-1})^n G_0(x) = \\
&= (px+qx^{-1})^n,
\end{aligned}
$$

поскольку $G_0(x)=1$.

Из леммы следует, что $P_n(0)$ есть коэффициент при x^i в разложении производящей функции $G_n(x)$ по степеням x.

Из биномиальной формулы (22) получаем

$$
\begin{aligned}
(px+qx^{-1})^n &= \sum_{k=0}^{n} C_n^k (px)^k (qx)^{n-k} = \\
&= \sum_{k=0}^{n} C_n^k p^k q^{n-k} x^{2k-n} = \sum_{i=-n}^{n} C_n^{\frac{n+i}{2}} p^{\frac{n+i}{2}} q^{\frac{n-i}{2}} x^i.
\end{aligned}
$$

Отсюда ясно, что вероятность $P_n(i)$, являющаяся коэффициентом при x^i, задается формулой (37).

Пусть теперь снова $S_0=0$, $S_k=\xi_1+\ldots+\xi_k$ — сумма независимых одинаково распределенных случайных величин ξ_i, принимающих на этот раз значения 1 и 0 с вероятностями p и q.

Согласно формуле (1) из § 2,

$$
P_n(i) = C_n^i p^i q^{n-i}. \tag{40}
$$

Методом производящих функций эта формула может быть получена следующим образом. Поскольку для $i \geqslant 1$

$$
P_n(i) = pP_{n-1}(i-1) + qP_{n-1}(i)
$$

и $P_n(0)=q^n$, то для производящих функций

$$
G_k(x) = \sum_{i=0}^{\infty} P_k(i)x^i
$$

находим, что

$$
G_n(x) = (px+q)G_{n-1}(x) = \ldots = (px+q)^n G_0(x) = (px+q)^n,
$$

поскольку $G_0(x)=1$.

Из полученной формулы $G_n(x) = (px+q)^n$ и биномиальной формулы (22) получаем, что коэффициент при x^i в разложении $(px+q)^n$ равен $C_n^i p^i q^{n-i}$, что в силу леммы дает формулу (40).

(f) Следующий пример показывает, как, опираясь на свойства производящих функций (в частности, на свойства свертки (9)), можно доказывать разного рода комбинаторные тождества.

Известно, например, следующее тождество:

$$(C_n^0)^2 + (C_n^1)^2 + \ldots + (C_n^n)^2 = C_{2n}^2. \tag{41}$$

Докажем его, опираясь на приведенную выше лемму.

Согласно биномиальной формуле (22),

$$(1+x)^{2n} = C_{2n}^0 + C_{2n}^1 x + C_{2n}^2 x^2 + \ldots + C_{2n}^n x^n. \tag{42}$$

Тем самым,

$$G_{2n}(x) = (1+x)^{2n} \tag{43}$$

есть производящая функция последовательности $\{C_{2n}^k,\ 0 \leqslant k \leqslant 2n\}$, причем C_{2n}^n есть коэффициент при x^n в разложении (42).

Запишем $(1+x)^{2n}$ в виде

$$(1+x)^{2n} = (1+x)^n (1+x)^n. \tag{44}$$

Тогда

$$G_{2n}(x) = G_n^{(a)}(x) G_n^{(b)}(x)$$

с

$$G_n^{(a)}(x) = (1+x)^n = \sum_{k=0}^{n} a_k x^k \quad \text{и} \quad G_n^{(b)}(x) = (1+x)^n = \sum_{k=0}^{n} b_k x^k,$$

где, очевидно, снова по (22)

$$a_k = b_k = C_n^k.$$

Воспользовавшись формулой свертки (9), находим, что коэффициент при x^n в произведении $G_n^{(a)}(x) G_n^{(b)}(x)$ равен

$$\begin{aligned} a_0 b_n + a_1 b_{n-1} + \ldots + a_n b_0 &= C_n^0 C_n^n + C_n^1 C_n^{n-1} + \ldots + C_n^n C_n^0 = \\ &= (C_n^0)^2 + (C_n^1)^2 + \ldots + (C_n^n)^2, \end{aligned} \tag{45}$$

поскольку $C_n^{n-k} = C_n^k$.

Так как в то же самое время коэффициент при x^n в разложении $G_{2n}(x)$ есть C_{2n}^n, а $G_{2n}(x) = G_n^{(a)}(x) G_n^{(b)}(x)$, то требуемая формула (41) вытекает из (45).

Из приведенного доказательства формулы (41), опирающегося на равенство (44), можно легко усмотреть, что если рассматривать вместо (44) равенство

$$(1+x)^n(1+x)^m = (1+x)^{n+m},$$

то, снова используя формулу свертки (9), получаем, что справедливо следующее тождество:

$$\sum_{j=1}^{n} C_n^j C_m^{k-j} = C_{n+m}^k, \tag{46}$$

известное как «свертка Вандермонда».

(g) В заключение остановимся на одном классическом примере применения метода производящих функций для отыскания вероятностей вырождения в ветвящихся процессах.

Продолжая рассмотрения примера 4 в § 12, будем сейчас предполагать, что временной параметр k не ограничен моментом n, а принимает любое из значений 0, 1, 2, ...

Пусть ξ_k, $k \geqslant 0$ — число индивидуумов («частиц») в момент времени k, $\xi_0 = 1$. В соответствии с указанным примером 4 § 12 будем предполагать, что

$$\xi_{k+1} = \eta_1^{(k)} + \ldots + \eta_{\xi_k}^{(k)}, \qquad k \geqslant 0 \tag{47}$$

(«модель Гальтона—Ватсона»; см. [125, с. 127] и задачу 18 к § 5 главы VIII в [137]), где $\{\eta_i^{(k)}, i \geqslant 1, k \geqslant 0\}$ — последовательность независимых одинаково распределенных случайных величин, каждая из которых распределена как величина η с вероятностями

$$p_k = \quad (\eta = k), \qquad k \geqslant 0, \qquad \sum_{k=0}^{\infty} p_k = 1.$$

Предполагается также, что при каждом k величины $\eta_i^{(k)}$ не зависят от $\xi_1, \ldots, \xi_k$. (Тем самым, в рассматриваемой модели Гальтона—Ватсона «процесс гибели-размножения» проходит так, что каждый индивидуум («частица») независимо от других индивидуумов и от «предыстории» превращается в j индивидуумов с вероятностями p_j, $j \geqslant 0$.)

Пусть $\tau = \inf\{k \geqslant 0 : \xi_k = 0\}$ — момент, когда происходит вырождение (семейства). Как принято, считаем, что $\tau = \infty$, если $\xi_k > 0$ при всех $k \geqslant 0$.

В теории ветвящихся процессов величина ξ_{k+1} интерпретируется как число «родителей» в $(k+1)$-м поколении, а $\eta_i^{(k)}$ — как число «детей», производимых i-м «родителем» k-го поколения.

Пусть

$$G(x) = \sum_{k=0}^{\infty} p_k x^k, \qquad |x| \leqslant 1,$$

есть производящая функция случайной величины η ($G(x) = \ x^{\eta}$) и $F_k(x) =$ $= \ x^{\xi_k}$ — производящая функция случайной величины ξ_k.

Из рекуррентной формулы (47) и свойства (16) (§ 8) условных математических ожиданий, находим, что

$$F_{k+1}(x) = \ x^{\xi_{k+1}} = \quad (x^{\xi_{k+1}} \,|\, \xi_k).$$

Здесь, в силу предполагаемой независимости,

$$(x^{\xi_{k+1}} \,|\, \xi_k = i) = \ \left(x^{\eta_1^{(k)} + \ldots + \eta_i^{(k)}}\right) = \left[G(x)\right]^i$$

и, значит,

$$F_{k+1}(x) = \ \left[G(x)\right]^{\xi_k} = F_k(G(x)).$$

В соответствии с (9) можно записать, что

$$F_{k+1}(x) = (F_k \circ G)(x),$$

т. е. производящая функция F_{k+1} является *композицией* производящих функций F_k и G.

Один из центральных вопросов теории ветвящихся процессов состоит в отыскании вероятности

$$q = \ (\tau < \infty),$$

т. е. *вероятности вырождения* за конечное время.

Заметим, что поскольку $\{\xi_k = 0\} \subseteq \{\xi_{k+1} = 0\}$, то

$$(\tau < \infty) = \ \left(\bigcup_{k=1}^{\infty} \{\xi_k = 0\}\right) = \lim_{N \to \infty} \ (\xi_N = 0).$$

Но $\ (\xi_N = 0) = F_N(0)$. Тем самым,

$$q = \lim_{N \to \infty} F_N(0).$$

Также заметим, что если $\xi_0 = 1$, то

$$F_0(x) = x, \quad F_1(x) = (F_0 \circ G)(x) = (G \circ F_0)(x),$$
$$F_2(x) = (F_1 \circ G)(x) = (G \circ G)(x) = (G \circ F_1)(x)$$

и, вообще,

$$F_N(x) = (F_{N-1} \circ G)(x) = \underbrace{[G \circ G \circ \ldots \circ G]}_{N \text{ раз}}(x) = G \circ \underbrace{[G \circ \ldots \circ G]}_{N-1 \text{ раз}}(x),$$

откуда

$$F_N(x) = (G \circ F_{N-1}(x))$$

и, значит, $q = \lim_{N\to\infty} F_N(0)$ является решением уравнения

$$q = G(q), \qquad 0 \leqslant q \leqslant 1.$$

Анализ решений этого уравнения показывает, что

(a) если $\eta > 1$, то вероятность вырождения $0 < q < 1$;

(b) если $\eta \leqslant 1$, то вероятность вырождения $q = 1$.

(См. детали и схему доказательств в [106, § 36 гл. 8] и [137, § 5 гл. VIII, задачи 18–21].)

8. Из приведенных выше задач видна важность умения по производящей функции $G(x)$ отыскивать ее коэффициенты a_i в разложении

$$G(x) = \sum_{i\geqslant 0} a_i x^i.$$

Приведем несколько стандартных производящих функций, для которых из математического анализа хорошо известны их полиномиальные представления:

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots$$

и, следовательно, функция $(1-x)^{-1}$ является производящей функцией последовательности $(1, 1, \ldots)$;

$$(1+x)^n = 1 + C_n^1 x^1 + C_n^2 x^2 + \ldots + C_n^n x^n;$$

$$\frac{1-x^{m+1}}{1-x} = 1 + x + x^2 + \ldots + x^m;$$

$$\frac{1}{(1-x)^n} = 1 + C_n^1 x^1 + C_{n+1}^2 x^2 + \ldots + C_{n+k-1}^k x^k + \ldots$$

Многие известные в математическом анализе числа (Бернулли, Эйлера и др.) определяются с помощью экспоненциальных производящих функций:

числа Бернулли $(b_0, b_1, \ldots)$:

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} b_n \frac{x^n}{n!}$$

$$\left(b_0 = 1,\quad b_1 = -\frac{1}{2},\quad b_2 = \frac{1}{6},\quad b_3 = 0,\quad b_4 = -\frac{1}{30},\quad b_5 = 0,\quad \ldots\right);$$

числа Эйлера $(e_0,\, e_1,\, \ldots)$:

$$\frac{2e^x}{e^{2x}+1} = \sum_{n=0}^{\infty} e_n\, \frac{x^n}{n!}$$

$$(e_0 = 1,\quad e_1 = 0,\quad e_2 = -1,\quad e_3 = 0,\quad e_4 = -5,\quad e_5 = 0,$$
$$e_6 = -61,\quad e_7 = 0,\quad e_8 = 1385,\quad \ldots);$$

9. Пусть ξ_1 и ξ_2 — две независимые случайные величины, имеющие пуассоновское распределение вероятностей с параметрами $\lambda_1 > 0$ и $\lambda_2 > 0$ (см. § 6 и таблицу 2 в § 2 главы II):

$$(\xi_i = k) = \frac{\lambda_i^k\, e^{-\lambda_i}}{k!}, \qquad k = 0,\, 1,\, 2,\, \ldots, \quad i = 1,\, 2. \tag{48}$$

Нетрудно подсчитать, что производящие функции $G_{\xi_i}(x)$ имеют вид $(|x| \leqslant 1)$

$$G_{\xi_i}(x) = \quad x^{\xi_i} = \sum_{i=0}^{\infty} \quad (\xi_i = k) x^k = e^{-\lambda_i (1-x)}. \tag{49}$$

Отсюда, в силу независимости величин ξ_1 и ξ_2, находим, что производящая функция $G_{\xi_1+\xi_2}(x)$ суммы $\xi_1 + \xi_2$ задается формулой

$$G_{\xi_1+\xi_2}(x) = \quad x^{\xi_1+\xi_2} = \quad x^{\xi_1} x^{\xi_2} = \quad x^{\xi_1} \cdot \quad x^{\xi_2} =$$
$$= G_{\xi_1}(x) \cdot G_{\xi_2}(x) = e^{-\lambda_1(1-x)} \cdot e^{-\lambda_2(1-x)} = e^{-(\lambda_1+\lambda_2)(1-x)}. \tag{50}$$

(Использованное здесь равенство $\quad x^{\xi_1} \cdot x^{\xi_2} = \quad x^{\xi_1} \cdot \quad x^{\xi_2}$ вытекает из независимости случайных величин x^{ξ_1} и x^{ξ_2} в силу свойств математического ожидания; см. свойство 5) в п. 5 § 4 и теорему 6 в § 6 главы II.)

Из (48)– (50) видим, что сумма независимых случайных величин ξ_1 и ξ_2, имеющих пуассоновское распределение с параметрами λ_1 и λ_2, также имеет пуассоновское распределение с параметром $\lambda_1 + \lambda_2$.

Сложнее дело обстоит при отыскании распределения разности $\xi_1 - \xi_2$ рассматриваемых величин ξ_1 и ξ_2.

Снова, пользуясь независимостью, находим, что

$$G_{\xi_1-\xi_2}(x) = G_{\xi_1}(x)\, G_{\xi_2}\!\left(\tfrac{1}{x}\right) = e^{-\lambda_1(1-x)}\, e^{-\lambda_2(1-1/x)} =$$
$$= e^{-(\lambda_1+\lambda_2)+\lambda_1 x+\lambda_2(1/x)} = e^{-(\lambda_1+\lambda_2)}\, e^{\sqrt{\lambda_1\lambda_2}(t+1/t)},$$

где $t = x\sqrt{\lambda_1/\lambda_2}$.

Из анализа известно, что для $\lambda \in R$

$$e^{\lambda(t+1/t)} = \sum_{k=-\infty}^{\infty} t^k I_k(2\lambda),$$

где $I_k(2\lambda)$ есть модифицированная функция Бесселя первого рода порядка k (см., например, [121, т. 5, с. 820–825]:

$$I_k(2\lambda) = \lambda^k \sum_{r=0}^{\infty} \frac{\lambda^{2r}}{r!\,\Gamma(k+r+1)}, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Таким образом,

$$(\xi_1 - \xi_2 = k) = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{k/2} I_k\bigl(2\sqrt{\lambda_1\lambda_2}\bigr)$$

для $k = 0, \pm 1, \pm 2, \ldots$

10. Задачи 1. Найти производящие функции последовательностей $\{a_r,\ r \geqslant 0\}$ с

$$\text{(a)}\quad a_r = r!, \qquad \text{(b)}\quad a_r = 2r^2, \qquad \text{(c)}\quad a_r = \frac{1}{r}.$$

2. Найти производящие функции для числа целочисленных решений систем

$$X_1 + X_2 + \ldots + X_n \leqslant r \quad \text{с} \ \ 1 \leqslant X_i \leqslant 4,$$
$$X_1 + 2X_2 + \ldots + nX_n = r \quad \text{с} \ \ X_i \geqslant 0.$$

3. Опираясь на аппарат производящих функций, вычислить

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

4. Пусть $a_1, \ldots, a_k$ — разные положительные числа. Показать, что число $N(k;\, n)$ разбиений числа n в сумму чисел из $\{a_1, \ldots, a_k\}$ есть коэффициент при x^n в разложении произведения

$$(1 + x^{a_1} + x^{2a_1} + \ldots)(1 + x^{a_2} + x^{2a_2} + \ldots)\ldots(1 + x^{a_k} + x^{2a_k} + \ldots).$$

(Допускаются и повторения, типа $8 = 2+2+2+2 = 3+3+2$, но порядок следования слагаемых несущественен, т. е. представления $3+2$ и $2+3$ числа 5 считаются за одно.)

5. Пусть a и b — два различных положительных числа. Показать, что число неотрицательных решений системы

$$aX_1 + bX_2 = n$$

есть коэффициент при x^n в разложении

$$(1 + x^a + x^{2a} + \ldots)(1 + x^b + x^{2b} + \ldots).$$

6. Показать, что
(a) число возможностей при распределении n неразличимых шаров по m разным ящиками равно C_{n+m-1}^{n};
(b) число векторов $(X_1, \ldots, X_m)$ с неотрицательными компонентами, удовлетворяющих соотношению $X_1+\ldots+X_m=n$, также равно C_{n+m-1}^{n};
(c) число возможностей выбора n шаров при осуществлении выбора с возвращением из урны с m различными шарами также равно C_{n+m-1}^{n}.

Указание. Провести доказательства, установив, что в каждом из случаев (a), (b) и (c) требуемое число есть коэффициент при x^n в разложении $(1+x+x^2+\ldots)^m$.

7. Доказать формулу (18).

8. Установить формулу (28) для q_0.

9. (*Задача Эйлера*.) Рассматриваются грузы, вес которых равен целому числу граммов. Спрашивается, какие грузы можно взвесить гирями в 1, 2^2, 2^3, …, 2^m, … граммов и сколькими способами это можно сделать?

10. Пусть $G_\xi(x)= x^\xi \left(=\sum\limits_{k\geqslant 0} x^k p_k\right)$ — производящая функция случайной величины ξ.

Показать, что

$$\begin{aligned}
\mathbb{E}_\xi &= G'_\xi(1),\\
\mathbb{E}_\xi^2 &= G''_\xi(1)+G'_\xi(1),\\
\mathbb{D}_\xi &= G''_\xi(1)+G'_\xi(1)-(G'_\xi(1))^2,
\end{aligned}$$

где $G'_\xi(x)$ и $G''_\xi(x)$ — первая и вторая производная (по x) функции $G_\xi(x)$.

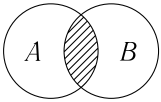
11. Пусть ξ — целочисленная случайная величина, принимающая значения 0, 1, … Положим $m_{(r)}= \xi(\xi-1)\ldots(\xi-r+1)$, $r=1, 2, \ldots$ Показать, что величины $m_{(r)}$, называемые факториальными моментами (порядка r), подсчитываются по производящей функции $G_\xi(x)$ по формулам

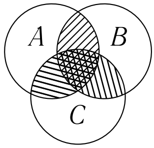
$$m_{(r)}=G_\xi^{(r)}(1),$$

где $G_\xi^{(r)}$ — r-я производная $G_\xi(x)$.

§ 14. Принцип включения-исключения

1. При оперировании с подмножествами A, B, C, … конечного множества Ω весьма полезны так называемые *диаграммы Венна* (Venn), дающие наглядный и удобный способ подсчета числа элементарных исходов $\omega\in\Omega$, входящих в такие комбинации из этих множеств, как, например, $A\cup B\cup C$, $A\cup B\cap\overline{C}$ и т. п.:





Если через $N(D)$ обозначать число элементов ω, входящих в D, то из приведенных диаграмм видим, что

$$N(A \cup B) = N(A) + N(B) - N(AB), \tag{1}$$

где член $N(AB)$ — число элементов, входящих в пересечение $AB = A \cap B$ — берется со знаком минус («исключение»), поскольку в сумме $N(A) + N(B)$ («включение») число элементов, входящих в пересечение множеств A и B, подсчитывалось дважды.

Аналогичным образом, для трех множеств A, B и C находим, что

$$\begin{aligned} N(A \cup B \cup C) = \big[N(A) + N(B) + N(C)\big] \\ - \big[N(AB) + N(AC) + N(BC)\big] + N(ABC). \end{aligned} \tag{2}$$

Возникающие здесь три члена соответствуют «включению», «исключению» и «включению».

При классическом способе задания вероятностей $\ (A)$ по формуле

$$\ (A) = \frac{N(A)}{N(\Omega)}$$

(см. формулу (10) в § I гл. 1) и в общем случае (см. § 2 гл. II) из свойств конечной аддитивности вероятности находим следующие формулы, аналогичные (1) и (2):

$$\ (A \cup B) = \ (A) + \ (B) - \ (AB), \tag{3}$$

$$\begin{aligned} \ (A \cup B \cup C) = \big[\ (A) + \ (B) + \ (C)\big] - \\ - \big[\ (AB) + \ (AC) + \ (BC)\big] + \\ + \ (ABC). \end{aligned} \tag{4}$$

Если воспользоваться (легко проверяемыми) *законами Моргана* (DeMorgan; см. задачу 1 в § 1 главы I)

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}, \tag{5}$$

устанавливающими связь между тремя основными теоретико-множественными операциями над множествами (объединение, пересечение и взятие дополнения), то из (1) находим, что

$$N(\overline{A}\overline{B}) = N(\Omega) - [N(A) + N(B)] + N(AB), \tag{6}$$

и, аналогично,

$$\begin{aligned} N(\overline{A}\overline{B}\overline{C}) = N(\Omega) - \big[N(A) + N(B) + N(C)\big] &+ \\ + \big[N(AB) + N(AC) + N(BC)\big] &- \\ - N(ABC). & \qquad (7) \end{aligned}$$

Событие $\overline{A}\,\overline{B}$ $(=\overline{A}\cap\overline{B})$ состоит из тех элементарных исходов ω, которые одновременно принадлежат и $\overline{A}$, и $\overline{B}$, т. е. тех ω, которые не принадлежат ни A, ни B. Приведем примеры, в которых эта интерпретация позволяет сводить вопрос о подсчете числа элементарных исходов к подсчету чисел типа $N(\overline{A}\,\overline{B})$, $N(\overline{A}\,\overline{B}\,\overline{C})$, для которых, в свою очередь, полезны формулы (6) и (7). (Обобщение этих формул на произвольное число множеств дается ниже в теореме.)

Пример 1. Пусть некоторая группа студентов состоит из 30 человек $(N(\Omega)=30)$. Из них 10 учат иностранный язык A $(N(A)=10)$ и 15 — иностранный язык B $(N(B)=15)$, при этом 5 из них учат оба языка $(N(AB)=5)$. Спрашивается, сколько студентов не учат ни одного языка. Ясно, что это число есть $N(\overline{A}\overline{B})$. По формуле (6)

$$N(\overline{A}\overline{B}) = 30 - [10 + 15 - 5] = 10.$$

Тем самым, никакого языка не учат 10 студентов.

Пример 2 (из теории чисел)**.** Каково число тех целых чисел между 1 и 300, которые

(A) не делятся на 3;

(B) не делятся на 3 и 5;

(C) не делятся на 3, 5 и 7?

Здесь $N(\Omega)=300$.

(A) Пусть $N(A)$ — число тех чисел, что делятся на 3. Ясно, что $N(A)=\frac{1}{3}\cdot 300=100$. Значит, число чисел, не делящихся на 3, есть $N(\overline{A}) = N(\Omega) - N(A) = 300 - 100 = 200$.

(B) Пусть $N(B)$ — число тех чисел, что делятся на 5. Тогда $N(B) = \frac{1}{5}\cdot 300 = 60$. Тогда число $N(AB)$ есть число чисел, делящихся и на 3, и на 5. Ясно, что $N(AB) = \frac{1}{15}\cdot 300 = 20$.

Требуемое число чисел, не делящихся ни на 3, ни на 5, есть

$$N(\overline{A}\,\overline{B}) = N(\Omega) - N(A) - N(B) + N(AB) = 300 - 100 - 60 + 20 = 160.$$

(C) Пусть $N(C)$ — число чисел, делящихся на 7. Тогда $N(C)=\left\lfloor \frac{300}{7}\right\rfloor = 42$ и $N(AC)=\left\lfloor \frac{300}{21}\right\rfloor = 14$, $N(BC)=\left\lfloor \frac{300}{35}\right\rfloor = 8$, $N(AB)=20$, $N(ABC)=2$. Следовательно, по формуле (7)

$$N(\overline{A}\,\overline{B}\,\overline{C}) = 300 - [100+60+42] + [20+35+21] - 2 = 172.$$

Итак, число чисел между 1 и 300, которые на делятся одновременно на 3, 5 и 7, равно 172.

2. Формулы (1), (2), (6), (7) допускают естественное обобщение на произвольное число множеств $A_1, \ldots, A_n$, $n \geqslant 2$, из Ω.

Теорема. *Справедливы следующие формулы включения-исключения:*

(a)

$$\begin{aligned} N(A_1 \cup \ldots \cup A_n) = \sum_{1 \leqslant i \leqslant n} N(A_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n} N(A_{i_1} \cap A_{i_1}) + \ldots + \\ + (-1)^{m+1} \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} N(A_{i_1} \cap \ldots \cap A_{i_m}) + \ldots + \\ + (-1)^{n+1} N(A_1 \cap \ldots \cap A_n), \end{aligned} \tag{8}$$

или, в более компактной форме,

$$N\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\varnothing \neq S \subseteq T} (-1)^{N(S)+1} N\left(\bigcap_{i \in S} A_i\right), \tag{9}$$

где $T = \{1, \ldots, n\}$;

(b)

$$\begin{aligned} N(A_1 \cap \ldots \cap A_n) = \sum_{1 \leqslant i \leqslant n} N(A_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n} N(A_{i_1} \cup A_{i_2}) + \ldots + \\ + (-1)^{m+1} \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} N(A_{i_1} \cup \ldots \cup A_{i_m}) + \ldots + \\ + (-1)^{n+1} N(A_1 \cup \ldots \cup A_n), \end{aligned} \tag{10}$$

или, в более компактной форме,

$$N\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\varnothing \neq S \subseteq T} (-1)^{N(S)+1} N\left(\bigcup_{i \in S} A_i\right); \tag{11}$$

(c)

$$N\left(\overline{A}_1 \cup \ldots \cup \overline{A}_n\right) = N(\Omega) - N(A_1 \cap \ldots \cap A_n), \tag{12}$$

$$N\left(\overline{A}_1 \cap \ldots \cap \overline{A}_n\right) = N(\Omega) - N(A_1 \cup \ldots \cup A_n), \tag{13}$$

или, в более компактной форме,

$$N\left(\bigcup_{i=1}^{n} \overline{A}_i\right) = \sum_{S\subseteq T}(-1)^{N(S)}N\left(\bigcup_{i\in S} A_i\right), \tag{14}$$

$$N\left(\bigcap_{i=1}^{n} \overline{A}_i\right) = \sum_{S\subseteq T}(-1)^{N(S)}N\left(\bigcap_{i\in S} A_i\right). \tag{15}$$

Доказательство. Достаточно доказать лишь формулу (8), поскольку все остальные могут быть выведены из нее переходом от событий к их дополнениям (по законам Моргана (5)).

Доказывать формулу (8) можно было бы по индукции (задача 1). Более элегантным является доказательство, основанное на свойствах индикаторов множеств.

Пусть

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

— индикатор (характеристическая функция) множества A. Если $B = A_1\cup\ldots$ $\ldots\cup A_n$, то согласно (5)

$$\overline{B} = \overline{A}_1\cap\ldots\cap\overline{A}_n$$

и

$$\begin{aligned} I_B &= 1 - I_{\overline{B}} = 1 - I_{\overline{A}_1}\ldots I_{\overline{A}_n} = 1 - (1 - I_{A_1})\ldots(1 - I_{A_n}) = \\ &= \sum_{i=1}^{n} I_{A_i} - \sum_{1\leqslant i_1<i_2\leqslant n} I_{A_{i_1}}I_{A_{i_2}} + \ldots + \\ &\quad + (-1)^{m+1}\sum_{1\leqslant i_1<\ldots<i_m\leqslant n} I_{A_{i_1}}\ldots I_{A_{i_m}} + \ldots + (-1)^n I_{A_1}\ldots I_{A_n}. \end{aligned} \tag{16}$$

Все участвующие здесь индикаторы являются функциями от ω: $I_B = I_B(\omega)$, $I_{A_i} = I_{A_i}(\omega)$, $I_{\overline{A}_i} = I_{\overline{A}_i}(\omega)$. Производя в (16) суммирование по всем $\omega\in\Omega$ и учитывая, что

$$\begin{aligned} &\sum_{\omega\in\Omega} I_B(\omega) = N(B), \qquad \sum_{\omega\in\Omega}\left(\sum_{i=1}^{n} I_{A_i}(\omega)\right) = \sum_{i=1}^{n} N(A_i), \\ &\sum_{\omega\in\Omega}\left(\sum_{1\leqslant i_1<i_2\leqslant n} I_{A_{i_1}}(\omega)I_{A_{i_1}}(\omega)\right) = \\ &\qquad = \sum_{\omega\in\Omega}\left(\sum_{1\leqslant i_1<i_2\leqslant n} I_{A_{i_1}\cap A_{i_1}}(\omega)\right) = \sum_{1\leqslant i_1<i_2\leqslant n} N(A_{i_1}\cap A_{i_1}) \end{aligned} \tag{17}$$

и т.д., получаем требуемую формулу (8). □

Замечание. Полезно отметить, что суммирование $\sum_{1\leqslant i_1<\ldots<i_m\leqslant n}$ в приведенных формулах производится по всевозможным неупорядоченным подмножествам размера m из множества $\{1, 2, \ldots, n\}$. Число таких подмножеств равно C_n^m. (См. таблицу 1 в § 1.)

3. Приведенные выше формулы выводились в предположении, что число элементарных исходов в Ω конечно ($N(\Omega)<\infty$). В этом предположении классическое определение вероятности (A) события A формулой

$$(A)=\frac{N(A)}{N(\Omega)}, \qquad A\subseteq\Omega, \tag{18}$$

сразу показывает, что утверждение теоремы остается в силе, если всюду $N(A_i)$ заменить на (A_i).

Так, например, формула (9) для вероятности события $\bigcup_{i=1}^n A_i$ примет следующий вид:

$$\left(\bigcup_{i=1}^n A_i\right)=\sum_{\varnothing\neq S\subseteq T}(-1)^{N(S)+1}\ \left(\bigcap_{i\in S} A_i\right), \tag{19}$$

где $T=\{1, \ldots, n\}$.

На самом деле все эти формулы для вероятностей событий $\bigcup_{i=1}^n A_i$, $\bigcap_{i=1}^n A_i$, $\bigcup_{i=1}^n \overline{A}_i$, $\bigcap_{i=1}^n \overline{A}_i$ остаются справедливыми не только для случая классического способа задания вероятностей посредством формулы (18).

Действительно, в случае $N(\Omega)<\infty$ доказательство формул в теореме было основано на соотношении (16) для индикаторов событий. Если взять от левой и правой частей этого соотношения математические ожидания, то в левой части получим $I_B=\ (B)$, а в правой части получим комбинации членов вида $I_{A_{i_1}}\ldots I_{A_{i_m}}$. Поскольку $I_{A_{i_1}}\ldots I_{A_{i_m}}=I_{A_{i_1}\cap\ldots\cap A_{i_m}}$, то

$$I_{A_{i_1}}\ldots I_{A_{i_m}}=\ I_{A_{i_1}\cap\ldots\cap A_{i_m}}=\ (A_{i_1}\cap\ldots\cap A_{i_m}).$$

Отсюда и из (16) получаем требуемую формулу включения-исключения для $(B)=\ (A_1\cup\ldots\cup A_n)$, из которой выводятся и другие формулы для событий $\bigcap_{i=1}^n A_i$, $\bigcup_{i=1}^n \overline{A}_i$, $\bigcap_{i=1}^n \overline{A}_i$.

Можно заметить, что в только что проведенном доказательстве на самом деле предположение $N(\Omega)<\infty$ вовсе и не использовалось и это доказательство сохраняет свою силу и для общих вероятностных пространств $(\Omega, \mathscr{F},\ )$, которые будут рассматриваться в следующей главе.

4. В дополнение к двум приведенным выше примерам на использование формул включения-исключения рассмотрим известную *задачу о беспорядках.*

Пример 3. Пусть

$$\Omega=\{1, 2, \ldots, n\}.$$

Рассмотрим всевозможные подстановки

$$\begin{pmatrix} 1, & 2, & \ldots, & n \\ a_1, & a_2, & \ldots, & a_n \end{pmatrix},$$

обладающие тем свойством, что у них отсутствуют номера i со свойством $a_i = i$.

Покажем, что число D_n, $n \geqslant 1$, таких подстановок («число беспорядков») задается формулой

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \sim \frac{n!}{e} \quad (n \to \infty). \tag{20}$$

(В случае $n = 0$ эта формула дает $D_0 = 1$.)

Обозначим A_i событие, состоящее в том, что $a_i = i$. Понятно, что

$$N(A_i) = (n-1)!.$$

Аналогично, для $i \neq j$

$$N(A_i A_j) = (n-2)!$$

и, вообще,

$$N(A_{i_1} A_{i_2} \ldots A_{i_k}) = (n-k)!. \tag{21}$$

Интересующее нас число

$$D_n = N\left(\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right)$$

по формулам (13) и (8) может быть представлено в виде

$$D_n = N(\Omega) - S_1 + S_2 + \ldots + (-1)^m S_m + \ldots + (-1)^n S_n, \tag{22}$$

где

$$S_m = \sum_{1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n} N(A_{i_1} A_{i_2} \ldots A_{i_m}). \tag{23}$$

Как уже было отмечено в замечании в конце п. 2, число членов в сумме (22) равно C_n^m. Поэтому с учетом (21) находим, что

$$S_m = (n-m)!\, C_n^m, \tag{24}$$

и, значит, из (22) получаем

$$\begin{aligned} D_n = n! - C_n^1 \cdot (n-1)! + C_n^2 \cdot (n-2)! + \ldots + \\ + (-1)^m \cdot C_n^m (n-m)! + \ldots + (-1)^n C_n^n \cdot 0! = \\ = \sum_{m=0}^{n} (-1)^m C_n^m (n-m)!. \end{aligned}$$

Здесь $C_n^m(n-m)! = \dfrac{n!}{m!}$. Следовательно,

$$D_n = n! \sum_{m=0}^{n} \frac{(-1)^m}{m!}, \tag{25}$$

и, значит, вероятность «полного беспорядка»

$$\left(\overline{A}_1 \ldots \overline{A}_n\right) = \frac{N(\overline{A}_1 \ldots \overline{A}_n)}{n!} = \sum_{m=0}^{n} \frac{(-1)^m}{m!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^n}{n!}.$$

Поскольку $\sum_{m=0}^{\infty} \dfrac{(-1)^m}{m!} = \dfrac{1}{e}$, то

$$\left(\overline{A}_1 \ldots \overline{A}_n\right) = \frac{1}{e} + O\left(\frac{1}{(n+1)!}\right)$$

и

$$D_n \sim \frac{n!}{e}.$$

Формулу (25) можно было бы также получить, обращаясь к методу (экспоненциальных) производящих функций.

Именно, положим (в соответствии с (20)) $D_0 = 1$ и заметим, что $D_1 = 0$ и для любого $n \geqslant 2$ для D_n справедливы рекуррентные соотношения

$$D_n = (n-1)\left[D_{n-1} + D_{n-2}\right] \tag{26}$$

(доказать это — задача 3), из которых вытекает, что

$$D_n - nD_{n-1} = -\left[D_{n-1} - (n-1)D_{n-2}\right].$$

Отсюда индукцией назад находим, что

$$D_n = nD_{n-1} + (-1)^n,$$

и, значит, для экспоненциальной производящей функции $E(x)$ последовательности $\{D_n,\ n \geqslant 0\}$ получаем

$$\begin{aligned} E(x) &= \sum_{n=0}^{\infty} D_n \frac{x^n}{n!} = 1 + \sum_{n \geqslant 2} D_n \frac{x^n}{n!} = 1 + \sum_{n \geqslant 2} \left\{nD_n + (-1)^n\right\} \frac{x^n}{n!} = \\ &= 1 + \sum_{n \geqslant 0} \frac{D_n}{n!} x^n + \left[e^{-x} - (1-x)\right] = xE(x) + \left[e^{-x} - (1-x)\right]. \end{aligned}$$

Тем самым,

$$E(x) = \frac{e^{-x}}{1-x} \tag{27}$$

и, следовательно,

$$E(x) = \left(1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \ldots\right)(1 + x + x^2 + \ldots). \tag{28}$$

Отсюда, представляя правую часть в виде $\sum_{n\geqslant 0} D_n \dfrac{x^n}{n!}$, можно, в принципе, найти, что коэффициенты D_n задаются формулой (25). Мы видим, однако, что метод включения-исключения приводит здесь к ответу быстрее, нежели метод производящих функций.

5. Задачи 1. Дать доказательство формул (8)–(9) и (10)–(11) по индукции.

2. Пусть B_1 — событие, состоящее в том, что произойдет в точности одно из событий $A_1, \ldots, A_n$. Показать, что

$$(B_1) = \sum_{i=1}^{n} \quad (A_i) - 2 \sum_{1\leqslant i_1<i_2\leqslant n} \quad (A_{i_1}A_{i_2}) + \ldots + (-1)^{n+1} n \quad (A_1\ldots A_n).$$

Указание. Воспользоваться тем, что

$$I_{B_1} = \sum_{i=1}^{n} I_{A_i} \prod_{j\neq i}(1 - I_{A_j}).$$

3. Доказать следующее обобщение формулы (15): для всякого множества I такого, что $\varnothing \subseteq I \subseteq T = \{1, \ldots, n\}$,

$$N\left(\left(\bigcap_{i\in T\setminus I} \overline{A}_i\right)\cap\left(\bigcap_{i\in I} A_i\right)\right) = \sum_{I\subseteq S\subseteq T} (-1)^{N(S\setminus I)} N\left(\bigcap_{i\in S} A_i\right).$$

4. Методом включения-исключения найти число всех целочисленных решений системы $X_1 + X_2 + X_3 + X_4 = 25$, удовлетворяющих условиям

$$-10 \leqslant X_i \leqslant 10.$$

Решить эту задачу и методом производящих функций. Какой метод быстрее в этой задаче приводит к цели?

5. Доказать формулу (26).

6. Используя формулу (8), найти число способов размещения n различных шаров по m различным ящикам с тем условием, что по меньшей мере один ящик остается пустым.

7. Сколькими способами n различных шаров можно разложить по m неразличимым ящикам так, чтобы не было пустых ящиков.

8. Пусть $A = A(n)$ и $B = B(m)$ — два множества, состоящие из n и k элементов соответственно.

Говорят, что отображение $\mathbb{F}\colon A \to B$ есть *функция*, если каждому $a \in A$ ставится в соответствие некоторое $b \in B$.

Говорят, что отображение $\mathbb{I}\colon A \to B$ есть *инъекция*, если разным элементам из A ставятся в соответствие разные элементы из B. (В этом случае $n \leqslant m$.)

Говорят, что отображение $\mathbb{S}\colon A \to B$ есть *сюръекция*, если для каждого $b \in B$ найдется $a \in A$ такое, что $S(a) = b$. (В этом случае $n \geqslant m$.)

Говорят, что отображение $\mathbb{F}\colon A \to B$ есть *биекция*, если это отображение является одновременно и инъекцией, и сюръекцией. (В этом случае $n = m$.)

Пользуясь принципом включения-исключения, показать, что числа $N(\mathbb{F})$, $N(\mathbb{I})$, $N(\mathbb{S})$, $N(\mathbb{B})$ (т. е. количество функций, инъекций, сюръекций и биекций) задаются следующими формулами:

$$
\begin{aligned}
&N(\mathbb{F}) = m^n,\\
&N(\mathbb{I}) = (m)_n \quad (= m(m-1)\ldots(m-n+1)),\\
&N(\mathbb{S}) = \sum_{i=0}^{m} (-1)^i C_m^i (m-i)^n,\\
&N(\mathbb{B}) = n! \quad \left(= \sum_{i=0}^{n} (-1)^i C_n^i (n-i)^n \right).
\end{aligned}
$$

Глава II

МАТЕМАТИЧЕСКИЕ ОСНОВАНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Теория вероятностей как математическая дисциплина может и должна быть аксиоматизирована совершенно в том же смысле, как геометрия и алгебра. Это означает, что, после того как даны названия изучаемым объектам и их основным отношениям, а также аксиомы, которым эти соотношения должны подчиняться, все дальнейшее изложение должно основываться исключительно лишь на этих аксиомах, не опираясь на обычное конкретное значение этих объектов и их отношений.

А. Н. Колмогоров. «Основные понятия теории вероятностей» [32]

§ 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 $(\Omega, \mathscr{A}, \quad)$ с $\Omega = \{\omega : \omega = (a_1, \ldots, a_n),\ a_i = 0, 1\}$, $\mathscr{A} = \{A : A \subseteq \Omega\}$ и $\quad(\{\omega\}) = p^{\sum a_i} q^{n - \sum a_i}$ — 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 можно вывести, 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 «вес» $p(\omega)$ каждого исхода $\omega\in[0, 1)$ непременно должен быть равен 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 «весам» $p(\omega)$ можно было найти вероятность (A) события A: $(A)=\sum\limits_{\omega\in A} p(\omega)$. 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 должна была бы быть равной $1/2$.

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) $A\in\mathscr{A} \Rightarrow \overline{A}\in\mathscr{A}$.

(Заметим, что в условии b) достаточно требовать лишь, чтобы либо $A\cup B\in\mathscr{A}$, либо $A\cap B\in\mathscr{A}$, поскольку, согласно законам Моргана (5) в § 14 гл. I, справедливы следующие соотношения: $A\cup B=\overline{\overline{A}\cap\overline{B}}$, $A\cap B=\overline{\overline{A}\cup\overline{B}}$.)

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) экспериментов с исходами («явлениями») из множества Ω.

Определение 3. Совокупность объектов

$$(\Omega, \mathscr{A}, \quad),$$

где

a) Ω — *множество* точек ω;

b) $\mathscr{A}$ — *алгебра* подмножеств Ω;

c) — *конечно-аддитивная вероятность* на $\mathscr{A}$,

называется *вероятностной моделью*, *вероятностной «теорией» (эксперимента) 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* (*сигма-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

$$\bigcup A_n \in \mathscr{F}, \quad \bigcap A_n \in \mathscr{F}$$

(при этом 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 $\sum\limits_{n=1}^{\infty} A_n \in \mathscr{A}$,

$$\mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

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 $\mu(\Omega_n) < \infty$, $n = 1, 2, \ldots$

Мера (подчеркнем — *счетно-аддитивная мера*) на алгебре $\mathscr{A}$, удовлетворяющая условию $(\Omega)=1$, будет называться *вероятностной мерой* или *вероятностью* (определенной на множествах алгебры $\mathscr{A}$).

Остановимся на некоторых свойствах вероятностных мер.

Если $\varnothing$ — пустое множество, то

$$(\varnothing)=0.$$

Если $A, B\in\mathscr{A}$, то

$$(A\cup B)= (A)+ (B)- (A\cap B).$$

Если $A, B\in\mathscr{A}$ и $B\subseteq A$, то

$$(B)\leqslant (A).$$

Если $A_n\in\mathscr{A}$, $n=1, 2, \ldots$, и $\bigcup A_n\in\mathscr{A}$, то

$$(A_1\cup A_2\cup\ldots)\leqslant (A_1)+ (A_2)+\ldots$$

Первые три свойства очевидны. Для доказательства последнего достаточно заметить, что $\bigcup\limits_{n=1}^{\infty} A_n=\sum\limits_{n=1}^{\infty} B_n$, 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,

$$\left(\bigcup_{n=1}^{\infty} A_n\right)= \left(\sum_{n=1}^{\infty} B_n\right)=\sum_{n=1}^{\infty} (B_n)\leqslant\sum_{n=1}^{\infty} (A_n).$$

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 — конечно-аддитивная функция множеств, заданная на алгебре $\mathscr{A}$, с $(\Omega)=1$. Тогда следующие четыре условия эквивалентны*:

1) *σ-аддитивна (— вероятность)*;

2) *непрерывна сверху, т. е. для любых множеств $A_1, A_2, \ldots\in\mathscr{A}$ таких, что $A_n\subseteq A_{n+1}$, $\bigcup\limits_{n=1}^{\infty} A_n\in\mathscr{A}$,*

$$\lim_n (A_n)= \left(\bigcup_{n=1}^{\infty} A_n\right);$$

3) *непрерывна снизу, т. е. для любых множеств $A_1, A_2, \ldots\in\mathscr{A}$ таких, что $A_n\supseteq A_{n+1}$, $\bigcap\limits_{n=1}^{\infty} A_n\in\mathscr{A}$,*

$$\lim_n (A_n)= \left(\bigcap_{n=1}^{\infty} A_n\right);$$

4) *непрерывна в «нуле», т. е. для любых множеств* $A_1, A_2, \ldots \in \mathscr{A}$ *таких, что* $A_{n+1} \subseteq A_n$, $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$,

$$\lim_n \quad (A_n) = 0.$$

Доказательство. 1) $\Rightarrow$ 2). Поскольку

$$\bigcup_{n=1}^{\infty} A_n = A_1 + (A_2 \setminus A_1) + (A_3 \setminus A_2) + \ldots,$$

то

$$\left(\bigcup_{n=1}^{\infty} A_n\right) = \quad (A_1) + \quad (A_2 \setminus A_1) + \quad (A_3 \setminus A_2) + \ldots =$$

$$= \quad (A_1) + \quad (A_2) - \quad (A_1) + \quad (A_3) - \quad (A_2) + \ldots = \lim_n \quad (A_n).$$

2) $\Rightarrow$ 3). Пусть $n \geqslant 1$, тогда

$$(A_n) = \quad (A_1 \setminus (A_1 \setminus A_n)) = \quad (A_1) - \quad (A_1 \setminus A_n).$$

Последовательность множеств $\{A_1 \setminus A_n\}_{n \geqslant 1}$ является неубывающей (см. таблицу на с. 197) и

$$\bigcup_{n=1}^{\infty} (A_1 \setminus A_n) = A_1 \setminus \bigcap_{n=1}^{\infty} A_n.$$

Тогда в силу 2)

$$\lim_n \quad (A_1 \setminus A_n) = \quad \left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right)$$

и, значит,

$$\lim_n \quad (A_n) = \quad (A_1) - \lim_n \quad (A_1 \setminus A_n) =$$

$$= \quad (A_1) - \quad \left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right) = \quad (A_1) - \quad \left(A_1 \setminus \bigcap_{n=1}^{\infty} A_n\right) =$$

$$= \quad (A_1) - \quad (A_1) + \quad \left(\bigcap_{n=1}^{\infty} A_n\right) = \quad \left(\bigcap_{n=1}^{\infty} A_n\right).$$

3) $\Rightarrow$ 4). Очевидно.

4) $\Rightarrow$ 1). Пусть множества $A_1, A_2, \ldots \in \mathscr{A}$ попарно не пересекаются и $\sum\limits_{n=1}^{\infty} A_n \in \mathscr{A}$. Тогда

$$\left(\sum_{i=1}^{\infty} A_i\right) = \left(\sum_{i=1}^{n} A_i\right) + \left(\sum_{i=n+1}^{\infty} A_i\right),$$

и поскольку $\sum\limits_{i=n+1}^{\infty} A_i \downarrow \varnothing$, $n \to \infty$, то

$$\begin{aligned}\sum_{i=1}^{\infty} (A_i) = \lim_n \sum_{i=1}^{n} (A_i) = \lim_n \left(\sum_{i=1}^{n} A_i\right) &= \\ = \lim_n \left[\left(\sum_{i=1}^{\infty} A_i\right) - \left(\sum_{i=n+1}^{\infty} A_i\right)\right] &= \\ = \left(\sum_{i=1}^{\infty} A_i\right) - \lim_n \left(\sum_{i=n+1}^{\infty} A_i\right) = \left(\sum_{i=1}^{\infty} A_i\right)&. \quad \square\end{aligned}$$

3. Теперь можно сформулировать ставшую общепринятой систему *аксиом Колмогорова*, лежащих в основе построения вероятностных моделей экспериментов с исходами (явлениями) из множества Ω.

Основное определение. Набор объектов

$$(\Omega, \mathscr{F}, \quad),$$

где

a) Ω — *множество* точек ω,

b) $\mathscr{F}$ — *σ-алгебра* подмножеств Ω,

c) — *вероятность* на $\mathscr{F}$,

называется *вероятностной моделью* (эксперимента) или *вероятностным пространством*. При этом пространство *исходов* Ω называется пространством *элементарных событий*, множества A из $\mathscr{F}$ — *событиями*, а (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 полезна таблица (см. с. 197—198), показывающая, 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.

Замечание. При построении моделей экспериментов, призванных описывать вероятностные связи «явлений с условиями», следует (согласно замечанию в § 1 гл. I) оговаривать, при каком «комплексе условий» эти

Таблица 1

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 исход $\omega \in A$, 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 по крайней мере одно из событий A или 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, ..., т. е. множество точек ω, входящих или в A_1, или в A_2, ...	событие, состоящее в наступлении по крайней мере одного из событий A_1, A_2, ...
$\sum\limits_{n=1}^{\infty} A_n$	сумма, т. е. объединение попарно непересекающихся множеств A_1, A_2, ...	событие, состоящее в наступлении одного из несовместных событий A_1, A_2, ...

Продолжение

$\bigcap\limits_{n=1}^{\infty} A_n$	пересечение множеств A_1, A_2, …, т. е. множество точек ω, входящих и в A_1, и в A_2, …	событие, состоящее в том, что одновременно произошли события A_1, A_2, …
$A_n \uparrow A$ (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 $\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$	множество исходов ω, которые бесконечное число раз (бесконечно часто) встречаются в последовательности A_1, A_2, …
$\underline{\lim\limits_n} A_n$ (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, … за исключением, быть может, только конечного числа

эксперименты рассматриваются. Как правило, мы этого не делаем, придерживаясь той точки зрения, что каждый раз должно быть ясно, в чем эти «условия» заключаются.

4. Задачи.

1. Пусть $\Omega=\{r\colon r\in[0,\,1]\}$ — множество рациональных точек на $[0,\,1]$, $\mathscr{A}$ — алгебра множеств, каждое из которых является конечной 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 $(A)=b-a$. Показать, что (A), $A\in\mathscr{A}$, является конечно-аддитивной, но не счетно-аддитивной функцией множеств.

2. Пусть Ω — некоторое счетное множество и $\mathscr{F}$ — совокупность всех его подмножеств. Положим $\mu(A)=0$, если A конечно, и $\mu(A)=\infty$, если A бесконечно. Показать, что функция множеств μ конечно-аддитивна, но не счетно-аддитивна.

3. Пусть μ — конечная мера на σ-алгебре $\mathscr{F}$, $A_n\in\mathscr{F}$, $n=1,\,2,\,\ldots$, и $A=\lim\limits_n A_n$, (т. е. $A=\overline{\lim\limits_n} A_n=\underline{\lim\limits_n} A_n$). Показать, что $\mu(A)=\lim\limits_n \mu(A_n)$.

4. Доказать, что $(A \triangle B) = \ (A) + \ (B) - 2 \ (A \cap B)$. (Ср. с задачей 4 § 1 гл. I.)

5. Показать, что «расстояния» $\rho_1(A, B)$ и $\rho_2(A, B)$, определенные по формулам

$$\rho_1(A, B) = \ (A \triangle B),$$

$$\rho_2(A, B) = \begin{cases} \dfrac{\ (A \triangle B)}{\ (A \cup B)}, & \text{если} \quad (A \cup B) \neq 0, \\ 0, & \text{если} \quad (A \cup B) = 0, \end{cases}$$

удовлетворяют «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 $A = \sum\limits_{i=1}^{\infty} A_i \in \mathscr{A}$. Тогда $\mu(A) \geqslant \geqslant \sum\limits_{i=1}^{\infty} \mu(A_i)$.

7. Доказать, что

$$\overline{\limsup A_n} = \liminf \bar{A}_n, \quad \overline{\liminf A_n} = \limsup \bar{A}_n,$$
$$\liminf A_n \subseteq \limsup A_n, \quad \limsup (A_n \cup B_n) = \limsup A_n \cup \limsup B_n,$$
$$\liminf (A_n \cap B_n) = \liminf A_n \cap \liminf B_n,$$
$$\limsup A_n \cap \liminf B_n \subseteq \limsup (A_n \cap B_n) \subseteq \limsup A_n \cap \limsup B_n.$$

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» $\varnothing$.

10. Проверить *неравенство Буля*: $\ (A \cap B) \geqslant 1 - \ (\bar{A}) - \ (\bar{B})$.

11. Пусть $A_1, \ldots, A_n$ — некоторые события из $\mathscr{F}$. Говорят, что эта система событий является *перестановочной* (*exchangeable* или *interchangeable*), если вероятности $\ (A_{i_1} \ldots A_{i_l})$ одни и те же ($= p_l$) для любого выбора индексов $1 \leqslant i_1 < \ldots < i_l \leqslant n$ и всех $1 \leqslant l \leqslant n$. Доказать, что для таких событий имеет место следующая формула:

$$\left(\bigcup_{i=1}^{n} A_i \right) = n p_1 - C_n^2 p_2 + C_n^3 p_3 - \ldots + (-1)^{n-1} p_n.$$

12. Пусть $(A_k)_{k\geqslant 1}$ — *бесконечная* последовательность перестановочных событий, т. е. для *любого* $n \geqslant 1$ и любого набора индексов $1 \leqslant i_1 < \ldots$ $\ldots < i_n$ вероятности $(A_{i_1} \ldots A_{i_n})$ имеют одно и то же значение $(= p_n)$. Доказать, что

$$\left(\varliminf_n A_n\right) = \left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{j\to\infty} p_j,$$

$$\left(\varlimsup_n A_n\right) = \left(\bigcup_{k=1}^{\infty} A_k\right) = 1 - \lim_{j\to\infty} (-1)^j \Delta^j(p_0),$$

где $p_0 = 1$, $\Delta^1(p_n) = p_{n+1} - p_n$, $\Delta^j(p_n) = \Delta^1(\Delta^{j-1}(p_n))$, $j \geqslant 2$.

13. Пусть $(A_n)_{n\geqslant 1}$ — некоторая последовательность множеств, $I(A_n)$ — индикатор множества A_n, $n \geqslant 1$. Показать, что

$$I\left(\varliminf_n A_n\right) = \varliminf_n I(A_n), \qquad I\left(\varlimsup_n A_n\right) = \varlimsup_n I(A_n),$$

$$I\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} I(A_n).$$

14. Показать, что

$$I\left(\bigcup_{n=1}^{\infty} A_n\right) = \max_{n\geqslant 1} I(A_n), \qquad I\left(\bigcap_{n=1}^{\infty} A_n\right) = \min_{n\geqslant 1} I(A_n).$$

15. Доказать, что

$$(\varlimsup A_n) \geqslant \varlimsup \quad (A_n), \qquad (\varliminf A_n) \leqslant \varliminf \quad (A_n).$$

16. Пусть $A^* = \varlimsup A_n$ и $A_* = \varliminf A_n$. Показать, что $(A_n - A_*) \to 0$, $(A^* - A_n) \to 0$.

17. Пусть множества $A_n \to A$ (в том смысле, что $A = A^* = A_*$). Показать, что $(A \triangle A_n) \to 0$.

18. Пусть множества A_n сходятся к множеству A в том смысле, что $(A \triangle A^*) = (A \triangle A_*) = 0$. Показать, что тогда также $(A \triangle A_n) \to 0$.

19. Доказать, что симметрическая разность $A \triangle B$ множеств A и B удовлетворяет следующему свойству:

$$I(A \triangle B) = I(A) + I(B) \pmod 2.$$

(Вывести отсюда, что $(A \triangle B) = (A) + (B) - 2 (A \cap B)$; ср. с задачей 4.) Проверить также следующие свойства симметрической разности:

$$(A \triangle B) \triangle C = A \triangle (B \triangle C), \qquad (A \triangle B) \triangle (B \triangle C) = A \triangle C,$$

$$A \triangle B = C \Leftrightarrow A = B \triangle C.$$

§ 2. Алгебры и σ-алгебры. Измеримые пространства

1. Алгебры и σ-алгебры являются составными элементами при построении вероятностных моделей (экспериментов). Приведем примеры и ряд результатов, относящихся к этим объектам.

Пусть Ω — некоторое пространство элементарных событий. Очевидным образом системы множеств

$$\mathscr{F}_* = \{\varnothing, \Omega\}, \qquad \mathscr{F}^* = \{A\colon A \subseteq \Omega\}$$

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

$$\mathscr{F}_A = \{A, \overline{A}, \varnothing, \Omega\}$$

является также примером алгебры (и σ-алгебры) и называется *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:

$$\Omega = D_1 + D_2 + \ldots; \qquad D_i \cap D_j = \varnothing, \qquad i \neq j.$$

Тогда система $\mathscr{A} = \alpha(\mathscr{D})$, образованная из множеств, являющихся объединением конечного или счетного числа элементов разбиения (с присоединенным пустым множеством), является алгеброй (и σ-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 $\mathscr{F}^*$ всех подмножеств пространства Ω есть σ-алгебра. Таким образом, по крайней мере одна алгебра и одна 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 1. Систему $\alpha(\mathscr{E})$ (соответственно $\sigma(\mathscr{E})$) часто называют (наименьшей) *алгеброй* (соответственно *σ-алгеброй*), *порожденной системой множеств* $\mathscr{E}$.

Как уже было отмечено, понятие σ-алгебры играет в теории вероятностей важную роль, входя в «основное определение» *вероятностного пространства* (п. 3 § 1). В этой связи понятно желание дать конструктивный способ получения σ-алгебры $\sigma(\mathscr{A})$, порожденной, скажем, некоторой алгеброй $\mathscr{A}$. (Лемма 1 устанавливает, что такая σ-алгебра существует, но не дает ее эффективного построения.)

Один из мыслимых и представляющихся естественными способов образования $\sigma(\mathscr{A})$ из $\mathscr{A}$ мог бы быть следующим.

Пусть $\mathscr{E}$ — некоторая система подмножеств Ω. Обозначим $\widehat{\mathscr{E}}$ систему подмножеств Ω, состоящую из множеств, входящих в $\mathscr{E}$, дополнений к ним и конечных или счетных объединений множеств из $\mathscr{E}$. Положим $\mathscr{A}_0=\mathscr{A}$, $\mathscr{A}_1=\widehat{\mathscr{A}_0}$, $\mathscr{A}_2=\widehat{\mathscr{A}_1}$ и т. д. Понятно, что при каждом n система $\mathscr{A}_n$ содержится в $\sigma(\mathscr{A})$, и можно было бы ожидать, что при некотором n $\mathscr{A}_n=\sigma(\mathscr{A})$ или, по крайней мере, $\bigcup \mathscr{A}_n=\sigma(\mathscr{A})$.

Однако это, вообще говоря, не так. Действительно, возьмем $\Omega=(0, 1]$ и в качестве алгебры $\mathscr{A}$ рассмотрим систему подмножеств Ω, порожденную пустым множеством $\varnothing$ и конечными суммами интервалов вида $(a, b]$ с рациональными концами a и b. Нетрудно убедиться, что в рассматриваемом случае система множеств $\bigcup\limits_{n=1}^{\infty} \mathscr{A}_n$ *строго* меньше σ-алгебры $\sigma(\mathscr{A})$.

В дальнейшем наш основной интерес будет связан не с тем, как, скажем, из алгебры $\mathscr{A}$ *сконструировать* наименьшую σ-алгебру $\sigma(\mathscr{A})$, а с вопросом о том, как установить, что та или иная *заданная* система множеств *является* σ-алгеброй.

Для получения ответа на такой вопрос нам понадобится важное понятие «монотонного класса».

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 $B_n = \bigcup\limits_{i=1}^{n} A_i \in \mathscr{A}$ и $B_n \subseteq B_{n+1}$. Следовательно, по определению монотонного класса $B_n \uparrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathscr{A}$. Аналогично устанавливается, что $\bigcap\limits_{i=1}^{\infty} A_i \in \mathscr{A}$. $\square$

Используя эту лемму, докажем справедливость следующего результата, проясняющего связь понятий «σ-алгебра» и «монотонный класс».

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

$$\widetilde{\mathscr{M}} = \{B\colon\ B \in \mathscr{M},\ \bar{B} \in \mathscr{M}\}$$

все те множества, которые обладают интересующим нас свойством. Ясно, что $\mathscr{A} \subseteq \widetilde{\mathscr{M}} \subseteq \mathscr{M}$. Установим, что $\widetilde{\mathscr{M}}$ — монотонный класс.

Пусть $B_n \in \widetilde{\mathscr{M}}$, 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, $\widetilde{\mathscr{M}}$ — монотонный класс. Но $\widetilde{\mathscr{M}} \subseteq \mathscr{M}$ и $\mathscr{M}$ — наименьший монотонный класс. Поэтому $\widetilde{\mathscr{M}} = \mathscr{M}$, 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

$$\lim\downarrow(A\cap B_n)=A\cap\lim\downarrow B_n,\qquad \lim\uparrow(A\cap B_n)=A\cap\lim\uparrow B_n$$

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, 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 $B\in\mathscr{M}$

$$(A\in\mathscr{M}_B)\ \Leftrightarrow\ (B\in\mathscr{M}_A=\mathscr{M}).$$

Поэтому, 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, $\mathscr{M}=\mu(\mathscr{A})$ — алгебра, что и завершает доказательство теоремы. □

Анализ проведенного доказательства показывает, что при рассмотрении систем множеств, образованных по *принципу подходящих множеств*, было важно то, что эти системы *замкнуты* относительно некоторых теоретико-множественных операций.

С этой точки зрения, во всей проблематике «монотонных классов» оказывается полезным выделение так называемых «π-*систем*» и «λ-*систем*» множеств, которые, в сущности, и были использованы в доказательстве теоремы 1. С помощью этих понятий можно дать еще ряд утверждений (теорема 2), которые относятся к рассматриваемой проблематике и часто оказываются более удобными, нежели непосредственное обращение к проверке того, что та или иная система множеств является «*монотонным классом*».

Определение 2 («π-λ-системы»). Пусть Ω — некоторое пространство. Система $\mathscr{P}$ подмножеств Ω называется *π-системой*, если она *замкнута* относительно взятия конечных пересечений: если $A_1, \ldots, A_n \in \mathscr{P}$, то $\bigcap\limits_{1\leqslant k\leqslant n} A_k \in \mathscr{P}$, $n \geqslant 1$.

Система $\mathscr{L}$ подмножеств Ω называется *λ-системой*, если

(λ_a) $\Omega \in \mathscr{L}$,

(λ_b) $(A, B \in \mathscr{L}$ и $A \subseteq B) \Rightarrow (B \setminus A \in \mathscr{L})$,

(λ_c) $(A_n \in \mathscr{L}$, $n \geqslant 1$, и $A_n \uparrow A) \Rightarrow (A \in \mathscr{L})$.

Система $\mathscr{D}$ подмножеств Ω, являющаяся *одновременно* π-системой и λ-системой, называется *π-λ-системой* или *d-системой* Дынкина.

Замечание 2. Полезно отметить, что группа условий (λ_a), (λ_b), (λ_c), определяющая λ-систему, равносильна (задача 9) группе условий (λ_a), (λ_b'), (λ_c'), где

(λ_b') если $A \in \mathscr{L}$, то $\overline{A} \in \mathscr{L}$,

(λ_c') если $A_n \in \mathscr{L}$, $n \geqslant 1$, $A_n \cap A_m = \varnothing$ для $m \neq n$, то $\bigcap A_n \in \mathscr{L}$.

Отметим также, что всякая алгебра, очевидно, является π-системой.

Если $\mathscr{E}$ — некоторая система множеств, то через $\pi(\mathscr{E})$, $\lambda(\mathscr{E})$ и $d(\mathscr{E})$ обозначаются соответственно наименьшие π-, λ- и d-системы, содержащие $\mathscr{E}$.

Роль π-λ-систем проясняется в приводимой ниже теореме. Чтобы полнее раскрыть смысл этой теоремы, заметим, что каждая *σ-алгебра является λ-системой*. Обратное же, вообще говоря, не верно. Так, если $\Omega = \{1, 2, 3, 4\}$, то система

$$\mathscr{L} = \{\varnothing, \Omega, (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$$

является λ-системой, но не σ-алгеброй.

Однако оказывается, что если *дополнительно* потребовать, чтобы λ-система была в то же самое время и π-системой, то тогда эта π-λ-система уже будет и σ-алгеброй.

Теорема 2 (о π-λ-системах). a) *Всякая π-λ-система $\mathscr{E}$ является σ-алгеброй.*

b) *Пусть $\mathscr{E}$ есть π-система множеств. Тогда $\lambda(\mathscr{E}) = d(\mathscr{E}) = \sigma(\mathscr{E})$.*

c) *Пусть $\mathscr{E}$ есть π-система множеств, $\mathscr{L}$ есть какая-то λ-система и $\mathscr{E} \subseteq \mathscr{L}$. Тогда $\sigma(\mathscr{E}) \subseteq \mathscr{L}$.*

Доказательство. a) Система $\mathscr{E}$ содержит Ω (в силу (λ_a)) и замкнута по отношению к взятию дополнений и конечных пересечений (согласно (λ_b') и предположению, что $\mathscr{E}$ является π-системой). Тем самым система множеств $\mathscr{E}$ является алгеброй (в соответствии с определением 1 из § 1). Чтобы теперь доказать, что $\mathscr{E}$ является также и σ-алгеброй, надо (в соответствии с определением 4 из § 1) убедиться в том, что если множества $B_1, B_2, \ldots$ принадлежат $\mathscr{E}$, то тогда и их объединение $\bigcup\limits_n B_n$ также принадлежит $\mathscr{E}$.

Положим $A_1 = B_1$ и $A_n = B_n \cap \overline{A}_1 \cap \ldots \cap \overline{A}_{n-1}$. Тогда, согласно (λ'_c), $\bigcap A_n \in \mathscr{E}$. Но $\bigcap B_n = \bigcap A_n$, следовательно, и $\bigcap B_n \in \mathscr{E}$.

Итак, всякая π-λ-система является σ-алгеброй.

b) Рассмотрим λ-систему $\lambda(\mathscr{E})$ и σ-алгебру $\sigma(\mathscr{E})$. Как уже отмечалось, всякая σ-алгебра является λ-системой. Тогда поскольку $\sigma(\mathscr{E}) \supseteq \mathscr{E}$, то $\sigma(\mathscr{E}) = \lambda(\sigma(\mathscr{E}) \supseteq \lambda(\mathscr{E})$. Тем самым $\lambda(\mathscr{E}) \subseteq \sigma(\mathscr{E})$.

Если теперь показать, что система $\lambda(\mathscr{E})$ является также и π-системой, то тогда, согласно утверждению a), получим, что $\lambda(\mathscr{E})$ есть σ-алгебра, содержащая $\mathscr{E}$. Но так как $\sigma(\mathscr{E})$ есть *минимальная* σ-алгебра, содержащая $\mathscr{E}$, и по доказанному $\lambda(\mathscr{E}) \subseteq \sigma(\mathscr{E})$, то $\lambda(\mathscr{E}) = \sigma(\mathscr{E})$.

Итак, обратимся к доказательству того, что $\lambda(\mathscr{E})$ является π-системой.

Как и при доказательстве теоремы 1, воспользуемся *принципом подходящих множеств*.

Пусть

$$\mathscr{E}_1 = \{B \in \lambda(\mathscr{E})\colon\ B \cap A \in \lambda(\mathscr{E}) \text{ для всех } A \in \mathscr{E}\}.$$

Если $B \in \mathscr{E}$, то $B \cap A \in \mathscr{E}$ (поскольку $\mathscr{E}$ есть π-система). Значит, $\mathscr{E} \subseteq \mathscr{E}_1$. Но система $\mathscr{E}_1$ есть λ-система (в силу самого определения $\mathscr{E}_1$). Поэтому $\lambda(\mathscr{E}) \subseteq \lambda(\mathscr{E}_1) = \mathscr{E}_1$. С другой стороны, по определению системы $\mathscr{E}_1$ имеет место включение $\mathscr{E}_1 \subseteq \lambda(\mathscr{E})$.

Таким образом, $\mathscr{E}_1 = \lambda(\mathscr{E})$.

Пусть теперь

$$\mathscr{E}_2 = \{B \in \lambda(\mathscr{E})\colon\ B \cap A \in \lambda(\mathscr{E}) \text{ для всех } A \in \lambda(\mathscr{E})\}.$$

Как и $\mathscr{E}_1$, система $\mathscr{E}_2$ является λ-системой.

Возьмем множество $B \in \mathscr{E}$. Тогда по определению системы $\mathscr{E}_1$ для всех $A \in \mathscr{E}_1 = \lambda(\mathscr{E})$ находим, что $B \cap A \in \lambda(\mathscr{E})$. Следовательно, из определения системы $\mathscr{E}_2$ видим, что $\mathscr{E} \subseteq \mathscr{E}_2$ и $\lambda(\mathscr{E}) \subseteq \lambda(\mathscr{E}_2) = \mathscr{E}_2$. Но $\lambda(\mathscr{E}) \supseteq \mathscr{E}_2$. Поэтому $\lambda(\mathscr{E}) = \mathscr{E}_2$, и, значит, для любых A и B из $\lambda(\mathscr{E})$ множество $A \cap B \in \lambda(\mathscr{E})$, т. е. система $\lambda(\mathscr{E})$ является π-системой. Итак, $\lambda(\mathscr{E})$ является π-λ-системой (а значит, $\lambda(\mathscr{E}) = d(\mathscr{E})$), и, как уже было отмечено выше, отсюда следует, что $\lambda(\mathscr{E}) = \sigma(\mathscr{E})$.

Тем самым утверждение b) установлено.

c) Из того, что $\mathscr{E} \subseteq \mathscr{L}$ и $\mathscr{L}$ есть λ-система, находим, что $\lambda(\mathscr{E}) \subseteq \lambda(\mathscr{L}) = \mathscr{L}$. Из b) следует, что $\lambda(\mathscr{E}) = \sigma(\mathscr{E})$. Поэтому $\sigma(\mathscr{E}) \subseteq \mathscr{L}$. $\square$

Замечание 3. Результаты теоремы 2 могут быть выведены непосредственно и из теоремы 1 (задача 10).

Сформулируем два утверждения, доказательство которых служит хорошей иллюстрацией применения *принципа подходящих множеств* и теоремы 2 о π-λ-системах.

Лемма 3. *Пусть   и   — две вероятностные меры, заданные на измеримом пространстве* $(\Omega, \mathscr{F})$. *Пусть* $\mathscr{E}$ *— некоторая* π*-система множеств из* $\mathscr{F}$ *и меры   и   совпадают на множествах из* $\mathscr{E}$. *Тогда эти меры совпадают и на* σ*-алгебре* $\sigma(\mathscr{E})$. *В частности, если* $\mathscr{A}$ *есть алгебра и меры   и   совпадают на ее множествах, то они совпадают и на множествах из* $\sigma(\mathscr{A})$.

Доказательство. Воспользуемся принципом *подходящих множеств*, беря в качестве таковых множества $\mathscr{L} = \{A \in \sigma(\mathscr{E}) \colon \quad (A) = \quad (A)\}$. Ясно, что $\Omega \in \mathscr{L}$. Если $A \in \mathscr{L}$, то, очевидно, и $\bar{A} \in \mathscr{L}$, поскольку $\quad (\bar{A}) = 1 - \quad (A) = 1 - \quad (A) = \quad (\bar{A})$. Если $A_1, A_2, \ldots$ есть система непересекающихся множеств в $\mathscr{L}$, тогда в силу *счетной аддитивности* мер   и

$$\left(\bigcup_n A_n\right) = \sum_n \quad (A_n) = \sum_n \quad (A_n) = \quad \left(\bigcup A_n\right).$$

Тем самым выполнены свойства (λ_a), (λ_b'), (λ_c'), и, значит, $\mathscr{L}$ является λ-системой.

По условию леммы $\mathscr{E} \subseteq \mathscr{L}$ и $\mathscr{E}$ является π-системой. Тогда из утверждения с) теоремы 2 следует, что $\sigma(\mathscr{E}) \subseteq \mathscr{L}$. В силу же самого определения подходящих множеств это свойство и означает, что меры   и   совпадают на σ-алгебре $\sigma(\mathscr{E})$. □

Лемма 4. *Пусть* $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$ *— независимые* (*относительно меры*) *алгебры событий. Тогда относительно этой меры будут независимы и* σ*-алгебры* $\sigma(\mathscr{A}_1), \sigma(\mathscr{A}_2), \ldots, \sigma(\mathscr{A}_n)$.

Доказательство. Отметим прежде всего, что в общих вероятностных моделях *независимость* множеств и систем множеств (алгебр, σ-алгебр, …) определяется точно так же, как и в элементарной теории вероятностей (см. определения 2—5 в § 3 гл. I).

Пусть $A_2, \ldots, A_n$ — множества из $\mathscr{A}_2, \ldots, \mathscr{A}_n$ соответственно и

$$\mathscr{L}_1 = \left\{A \in \sigma(\mathscr{A}_1) \colon \quad (A \cap A_2 \cap \ldots \cap A_n) = \quad (A) \prod_{k=2}^{n} \quad (A_k)\right\}. \tag{3}$$

Покажем, что $\mathscr{L}$ является λ-системой.

Ясно, что $\Omega \in \mathscr{L}$, т. е. выполнено свойство (λ_a). Пусть A и B принадлежат $\mathscr{L}_1$ и $A \subseteq B$. Тогда поскольку

$$(A \cap A_1 \cap \ldots \cap A_n) = \quad (A) \prod_{k=2}^{n} \quad (A_k)$$

и

$$(B\cap A_1\cap\ldots\cap A_n)=\ \ (B)\prod_{k=2}^{n}\ \ (A_k),$$

то, вычитая из второго соотношения первое, находим, что

$$((B\setminus A)\cap A_1\cap\ldots\cap A_n)=\ \ (B\setminus A)\prod_{k=2}^{n}\ \ (A_k).$$

Значит, выполнено свойство (λ_b). Наконец, если множества $B_k\in\sigma(\mathscr{A}_1)$, $k\geqslant 1$, и $B_k\uparrow B$, то $B_k\cap A_k\cap\ldots\cap A_n\uparrow B\cap A_2\cap\ldots\cap A_n$. Поэтому в силу *непрерывности сверху* вероятности (см. теорему в § 1) из $(B_k\cap A_2\cap\ldots\ldots\cap A_n)=\ \ (B_k)\prod_{i=2}^{n}\ \ (A_i)$ предельным переходом $(k\to\infty)$ находим, что

$$(B\cap A_2\cap\ldots\cap A_n)=\ \ (B)\prod_{i=2}^{n}\ \ (A_i),$$

т. е. выполнено свойство (λ_c).

Таким образом, система $\mathscr{L}_1$ является λ-системой и $\mathscr{L}_1\supseteq\mathscr{A}_1$. Применяя утверждение c) из теоремы 2, находим, что $\mathscr{L}_1\supseteq\sigma(\mathscr{A}_1)$.

Тем самым показано, что системы $\sigma(\mathscr{A}_1)$, $\mathscr{A}_2$, …, $\mathscr{A}_n$ являются *независимыми*.

Проводя аналогичные рассмотрения для системы $\mathscr{A}_2$, …, $\mathscr{A}_n$, $\sigma(\mathscr{A}_1)$ приходим к тому, что независимыми будут и системы $\sigma(\mathscr{A}_2)$, $\mathscr{A}_3$, …, $\mathscr{A}_n$, $\sigma(\mathscr{A}_2)$, или, равносильно, системы $\mathscr{A}_3$, …, $\mathscr{A}_n$, $\sigma(\mathscr{A}_1)$, $\sigma(\mathscr{A}_2)$.

Продолжая этот процесс, приходим к независимости системы, состоящей из σ-алгебр $\sigma(\mathscr{A}_1)$, $\sigma(\mathscr{A}_2)$, …, $\sigma(\mathscr{A}_n)$. ☐

Замечание 4. Проанализируем еще раз, что же нужно требовать от системы множеств, чтобы она образовывала σ-алгебру.

С этой целью будем говорить, что система множеств $\mathscr{E}$ является π^*-системой, если она замкнута относительно взятия *счетного* пересечения:

$$A_1,\,A_2,\,\ldots\in\mathscr{E}\ \Rightarrow\ \bigcap_{n=1}^{\infty}A_n\in\mathscr{E}.$$

Тогда из определения σ-алгебры следует, что если некоторая *алгебра $\mathscr{E}$ является в то же самое время и π^*-системой*, то она будет и σ-алгеброй.

Подход же, основанный на понятии «π-λ-система», несколько иной. Здесь мы отправляемся не от понятия «алгебра», а от понятия «λ-система». И, как следует из утверждения a) теоремы 2, если эта *λ-система является в то же самое время π-системой*, то она будет и σ-алгеброй.

Ясно, в чем состоит разница в этих подходах.

Когда мы проверяем, что некоторая система является σ-алгеброй, начиная при этом с проверки того, что эта система является *алгеброй*, это означает, что мы делаем эту проверку, привлекая к рассмотрению лишь *конечные* суммы (или пересечения) множеств. «*Счетность*» (а именно в ней-то «все дело») возникает тогда, когда мы осуществляем проверку того, что эта система множеств является также и π^*-системой.

При «λ-π-подходе» проверку свойства, что интересующая нас система множеств является σ-алгеброй, мы начинаем прежде всего с установления того, что эта система является λ-*системой*, свойства (λ_c) или (λ'_c) которой связаны уже со «*счетными*» *операциями*. Зато на втором этапе — при проверке того, что эта система является π-системой, — мы оперируем лишь с *конечными* пересечениями или суммами множеств.

Завершим изложение результатов о «монотонных классах» приведением одной из их *функциональных* версий. (Примером применения может служить доказательство леммы теоремы 1 в § 2 гл. VIII.)

Теорема 3. *Пусть $\mathscr{E}$ есть π-система множеств из $\mathscr{F}$ и $\mathscr{H}$ — совокупность тех действительнозначных $\mathscr{F}$-измеримых функций, для которых выполнены следующие свойства*:

(h_1) *если $A \in \mathscr{E}$, то функция $I_A \in \mathscr{H}$;*

(h_2) *если $f \in \mathscr{H}$, $h \in \mathscr{H}$, то $f + g \in \mathscr{H}$ и $cf \in \mathscr{H}$ для всякого действительного числа c;*

(h_3) *если функции $h_n \in \mathscr{H}$, $n \geqslant 1$, $0 \leqslant h_n \uparrow h$, то $h \in \mathscr{H}$.*

Тогда класс $\mathscr{H}$ содержит и все ограниченные функции, являющиеся измеримыми относительно σ-алгебры $\sigma(\mathscr{E})$.

Доказательство. Пусть $\mathscr{L} = \{A \in \mathscr{F} : I_A \in \mathscr{H}\}$. Из (h_1) следует, что $\mathscr{E} \subseteq \mathscr{L}$. Но система $\mathscr{L}$ в силу (h_2) и (h_3) является (задача 11) λ-системой. Поэтому, согласно утверждению с) теоремы 2, находим, что $\sigma(\mathscr{E}) \subseteq \mathscr{L}$. Таким образом, если $A \in \sigma(\mathscr{E})$, то функция $I_A \in \mathscr{H}$. По свойству (h_2) отсюда вытекает, что все простые функции (т. е. функции, являющиеся конечными линейными комбинациями функций вида I_{A_i}, где $A_i \in \sigma(\mathscr{E})$) тоже принадлежат классу $\mathscr{H}$. Наконец, в силу свойства (h_i) получаем, что и всякая ограниченная $\sigma(\mathscr{E})$-измеримая функция принадлежит классу $\mathscr{H}$. □

Замечание 5. Пусть $X_1, \ldots, X_n$ — случайные величины на $(\Omega, \mathscr{F})$, $\mathscr{F}^X = \sigma(X_1, \ldots, X_n)$ и $f = f(\omega)$ — $\mathscr{F}^X$-измеримая функция. Тогда найдется такая борелевская функция $F = F(x_1, \ldots, x_n)$, что $f(\omega) = F(X_1(\omega), \ldots, X_n(\omega))$.

Для доказательства этого утверждения достаточно воспользоваться теоремой 3, беря в качестве *подходящего множества функций $\mathscr{H}$* множество неотрицательных борелевских функций $F = F(x_1, \ldots, x_n)$ и в качестве множества $\mathscr{E}$ совокупность множеств

$$\mathscr{E} = \{\omega \colon X_1(\omega) \leqslant x_1, \ldots, X_n(\omega) \leqslant x_n;\ x_i \in R,\ i = 1, \ldots, n\}.$$

Применяя теорему 3, получаем, что всякая неотрицательная $\mathscr{F}^X$-измеримая функция $f = f(\omega)$ представлена в виде $f(\omega) = F(X_1(\omega), \ldots, X_n(\omega))$. Общий случай (не обязательно отрицательных) функций f сводится к рассмотренному предварительным переходом к представлению $f = f^+ - f^-$.

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, состоящую из *конечных* сумм непересекающихся интервалов вида $(a, b]$:

$$A \in \mathscr{A}, \quad \text{если} \quad A = \sum_{i=1}^{n} (a_i, b_i], \quad n < \infty.$$

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 $\bigcup\limits_n A_n = (0, 1) \notin \mathscr{A}$.

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

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right], \quad a < b, \qquad [a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right], \quad a < b,$$

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a\right].$$

Тем самым в борелевскую алгебру наряду с интервалами вида $(a, b]$ входят одноточечные множества $\{a\}$, а также любое из шести множеств

$$(a, b), \ [a, b], \ [a, b), \ (-\infty, b), \ (-\infty, b], \ (a, \infty). \tag{4}$$

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 множеств в R, состоящими из конечных сумм непересекающихся интервалов одного и того же типа из (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 системами множеств в $\bar{R}$, состоящими из конечных сумм непересекающихся интервалов вида

$$(a, b] = \{x \in \bar{R}\colon\ a < x \leqslant b\}, \qquad -\infty \leqslant a < b \leqslant \infty,$$

где под $(-\infty, b]$ понимается множество $\{x \in \bar{R}\colon\ -\infty \leqslant x \leqslant b\}$.

Замечание 1. Для измеримого пространства $(R, \mathscr{B}(R))$ часто используются также обозначения $(R, \mathscr{B})$, $(R^1, \mathscr{B}_1)$.

Замечание 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 конечными суммами непересекающихся открытых множеств вида $S_\rho(x^0) = \{x \in R : \rho_1(x, x^0) < \rho\}$, $\rho > 0$, $x^0 \in R$. 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 $\{x \in R^n : x_k \in I_k,\ k = 1, \ldots, n\}$, назовем *прямоугольником*, I_k — *сторонами* этого прямоугольника. Через $\mathscr{I}$ обозначим совокупность всех множеств, состоящих из конечных сумм непересекающихся прямоугольников 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 системами множеств, образованными из конечных сумм непересекающихся прямоугольников $I = I_1 \times \ldots \times I_n$, и системами множеств, образованными из конечных сумм непересекающихся прямоугольников $B = B_1 \times \ldots \times B_n$ с борелевскими сторонами, совпадают.

Доказательство существенно опирается на следующее предложение.

Лемма 5. *Пусть $\mathscr{E}$ — некоторый класс множеств из Ω, множество $B \subseteq \Omega$, и пусть по определению*

$$\mathscr{E} \cap B = \{A \cap B\colon\ A \in \mathscr{E}\} \tag{5}$$

и $\sigma(\mathscr{E} \cap B)$ — наименьшая σ-алгебра подмножеств B, порожденная системой $\mathscr{E} \cap B$. 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 в B, 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

$$\mathscr{C}_B = \{A \in \sigma(\mathscr{E})\colon\ A \cap B \in \sigma(\mathscr{E} \cap B)\}.$$

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, $\sigma(\mathscr{C}) \cap B \subseteq \sigma(\mathscr{E} \cap B)$. $\square$

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 $\mathscr{B}(R^2)\subseteq\mathscr{B}\otimes\mathscr{B}$, 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, $\widetilde{\mathscr{B}}_1=\mathscr{B}_1\times R_2$, $\widetilde{\mathscr{B}}_2=R_1\times\mathscr{B}_2$, 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 $\widetilde{\mathscr{I}}_1=\mathscr{I}_1\times R_2$, $\widetilde{\mathscr{I}}_2=R_1\times\mathscr{I}_2$. Тогда, если $\widetilde{B}_1=B_1\times R_2$, $\widetilde{B}_2=R_1\times B_2$, то в силу (6)

$$B_1\times B_2=\widetilde{B}_1\cap\widetilde{B}_2\in\widetilde{\mathscr{B}}_1\cap\widetilde{B}_2=\sigma(\widetilde{\mathscr{I}}_1)\cap\widetilde{B}_2=$$
$$=\sigma(\widetilde{\mathscr{I}}_1\cap\widetilde{B}_2)\subseteq\sigma(\widetilde{\mathscr{I}}_1\cap\widetilde{\mathscr{I}}_2)=\sigma(\mathscr{I}_1\times\mathscr{I}_2),$$

что и требовалось доказать.

Случай произвольного $n>2$ рассматривается аналогичным образом. $\square$

Замечание. Пусть $\mathscr{B}_0(R^n)$ — наименьшая σ-алгебра, порожденная системами множеств, образованными конечными суммами непересекающихся открытых «шаров»

$$S_\rho(x^0)=\{x\in R^n\colon\ \rho_n(x,\,x^0)<\rho\},\qquad x^0\in R^n,\qquad \rho>0,$$

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

$$\mathscr{I}(I_1\times\ldots\times I_n)=\{x\colon\ x=(x_1,\,x_2,\,\ldots),\,x_1\in I_1,\,\ldots,\,x_n\in I_n\},\tag{8}$$

$$\mathscr{I}(B_1\times\ldots\times B_n)=\{x\colon\ x=(x_1,\,x_2,\,\ldots),\,x_1\in B_1,\,\ldots,\,x_n\in B_n\},\tag{9}$$

$$\mathscr{I}(B^n)=\{x\colon\ x=(x_1,\,\ldots,\,x_n)\in B^n\},\tag{10}$$

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 $B^{n+1}=B^n\times R$.

Множества, составленные из конечных сумм непересекающихся цилиндров $\mathscr{I}(I_1\times\ldots\times I_n)$, образуют алгебру. Точно так же алгебру образуют множества, составленные из объединений непересекающихся цилиндров $\mathscr{I}(B_1\times\ldots\times B_n)$. Система цилиндров $\mathscr{I}(B^n)$ также образует алгебру. Обозначим $\mathscr{B}(R^\infty)$, $\mathscr{B}_1(R^\infty)$ 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 $n=1,\,2,\,\ldots$

$$\mathscr{C}_n=\{A\subseteq R^n\colon\ \{x\colon\ (x_1,\,\ldots,\,x_n)\in A\}\in\mathscr{B}(R^\infty)\}.$$

Пусть $B^n\in\mathscr{B}(R_n)$. Тогда

$$B^n\in\mathscr{C}_n.$$

Но $\mathscr{C}_n$ — σ-алгебра, а значит,

$$\mathscr{B}(R^n)\subseteq\sigma(\mathscr{C}_n)=\mathscr{C}_n.$$

Отсюда можно сделать заключение, что $\mathscr{B}_2(R^\infty)\subseteq\mathscr{B}(R^\infty)$.

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 системами множеств, образованными конечными суммами непересекающихся открытых «шаров»

$$S_\rho(x^0)=\{x\in R^\infty\colon\ \rho_\infty(x,\,x^0)<\rho\},\qquad x^0\in R^\infty,\qquad \rho>0,$$

в метрике

$$\rho_\infty(x,\, x^0) = \sum_{k=1}^{\infty} 2^{-k} \rho_1(x_k,\, x_k^0),$$

где $x = (x_1,\, x_2,\, \ldots)$ и $x^0 = (x_1^0,\, x_2^0,\, \ldots)$. Тогда $\mathscr{B}(R^\infty) = \mathscr{B}_0(R^\infty)$ (задача 7).

Приведем несколько примеров борелевских множеств в R^∞:

(a) $\{x \in R^\infty:\ \sup x_n > a\}$, $\{x \in R^\infty:\ \inf x_n < a\}$;

(b) $\{x \in R^\infty:\ \overline{\lim}\ x_n \leqslant a\}$, $\{x \in R^\infty:\ \underline{\lim}\ x_n > a\}$, где, как обычно,

$$\overline{\lim}\ x_n = \inf_n \sup_{m \geqslant n} x_m, \qquad \underline{\lim}\ x_n = \sup_n \inf_{m \geqslant n} x_m;$$

(c) $\{x \in R^\infty : x_n \to\}$ — множество тех $x \in R^\infty$, для которых $\lim x_n$ существует и конечен;

(d) $\{x \in R^\infty:\ \lim x_n > a\}$;

(e) $\left\{x \in R^\infty:\ \sum\limits_{n=1}^{\infty} |x_n| > a\right\}$;

(f) $\left\{x \in R^\infty:\ \sum\limits_{k=1}^{n} x_k = 0 \text{ по крайней мере для одного } n \geqslant 1\right\}$.

Чтобы убедиться, например, в том, что множества из (a) входят в систему $\mathscr{B}(R^\infty)$, достаточно заметить, что

$$\{x:\ \sup x_n > a\} = \bigcup_n \{x:\ x_n > a\} \in \mathscr{B}(R^\infty),$$

$$\{x:\ \inf x_n < a\} = \bigcup_n \{x:\ x_n < a\} \in \mathscr{B}(R^\infty).$$

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 $T = [0,\, \infty)$.

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.

---

*) В дальнейшем для функции из R^T используются также обозначения: $x = (x_t)_{t \in T}$, $x = (x_t)$, $t \in T$.

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).

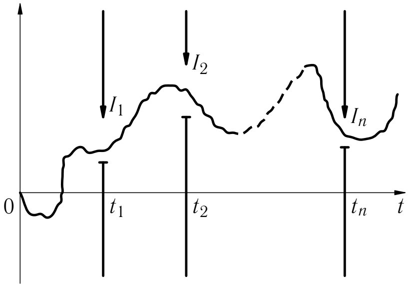


Рис. 24.

Обозначим через $\mathscr{B}(R^T)$, $\mathscr{B}_1(R^T)$ и $\mathscr{B}_2(R^T)$ наименьшие σ-алгебры, содержащие все цилиндрические множества (11), (12) и (13) соответственно. Ясно, что

$$\mathscr{B}(R^T) \subseteq \mathscr{B}_1(R^T) \subseteq \mathscr{B}_2(R^T). \qquad (14)$$

На самом же деле все три σ-алгебры совпадают между собой. Более того, исчерпывающим образом можно описать и структуру их множеств.

Теорема 4. *Пусть T — 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: найдутся не более чем счетное множество точек $t_1, t_2, \ldots$ из T и борелевское множество B из $\mathscr{B}(R^\infty)$ такие, что*

$$A = \{x\colon (x_{t_1}, x_{t_2}, \ldots) \in B\}. \qquad (15)$$

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 $T^{(1)} = (t_1^{(1)}, t_2^{(2)}, \ldots)$, $T^{(2)} = (t_1^{(2)}, t_2^{(2)}, \ldots)$, то множество $T^{(\infty)} = \bigcup\limits_k T^{(k)}$ можно взять в качестве единой системы такой, что все A_i будут представлены в виде

$$A_i = \{x\colon (x_{\tau_1}, x_{\tau_2}, \ldots) \in B_i\},$$

где B_i — 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 (13) и, поскольку $\mathscr{B}_2(R^T)$ есть *наименьшая* σ-алгебра, содержащая эти множества, то вместе с (14) это дает

$$\mathscr{B}(R^T) \subseteq \mathscr{B}_1(R^T) \subseteq \mathscr{B}_2(R^T) \subseteq \mathscr{E}. \qquad (16)$$

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. $\square$

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

$$A_1 = \{x\colon\ x_t < C \text{ для всех } t \in [0,\ 1]\},$$

$$A_2 = \{x\colon\ x_t = 0 \text{ по крайней мере для одного } t \in [0,\ 1]\},$$

$$A_3 = \{x\colon\ x_t \text{ непрерывна в фиксированной точке } t_0 \in [0,\ 1]\},$$

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

$$\{x\colon\ \sup_t x_t < C,\ t \in [0,\ 1]\} = \{x\colon\ (x_{t_1^0},\ x_{t_2^0},\ \ldots) \in B^0\}.$$

Ясно, что функция $y_t \equiv C - 1$ принадлежит A_1, и, следовательно, $(y_{t_1^0},\ y_{t_1^0},\ \ldots) \in B^0$. Образуем тогда функцию

$$z_t = \begin{cases} C-1, & t \in (t_1^0,\ t_2^0,\ \ldots), \\ C+1, & t \notin (t_1^0,\ t_2^0,\ \ldots). \end{cases}$$

Понятно, что

$$(y_{t_1^0},\ y_{t_2^0},\ \ldots) = (z_{t_1^0},\ z_{t_2^0},\ \ldots),$$

и, значит, функция $z = (z_t)$ принадлежит множеству

$$\{x\colon\ (x_{t_1^0},\ x_{t_2^0},\ \ldots) \in B^0\}.$$

Но ясно также, что она не принадлежит множеству $\{x\colon\ \sup x_t < C\}$. 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 пространстве 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 есть некоторая функция из C и $S_\rho(x^0) = \left\{x \in C : \sup\limits_{t \in T} |x_t - x_t^0| < \rho\right\}$ — открытая сфера с центром в x^0. В силу непрерывности функций из C

$$B_\rho = \{y \in C : y \in S_\rho(x^0)\} = \left\{y \in C : \max_t |y_t - x_t^0| < \rho\right\} = \\ = \bigcap_{t_k} \{y \in C : |y_{t_k} - x_{t_k}^0| < \rho\} \in \mathscr{B}(C), \quad (17)$$

где t_k — рациональные точки отрезка $[0, 1]$. Поэтому $\mathscr{B}_0(C) \subseteq \mathscr{B}(C)$.

Пространство $(C, \mathscr{B}_0(C), \rho)$ является польским, т. е. полным и сепарабельным; [5], [87].

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 задать метрику $d = d(x, y)$ так, что σ-алгебра $\mathscr{B}_0(D)$, порожденная *открытыми* множествами, будет совпадать с σ-алгеброй $\mathscr{B}(D)$, порожденной *цилиндрическими* множествами. При этом $(D, \mathscr{B}(D), d)$ станет сепарабельным пространством; [5], [87]. Эта метрика $d = d(x, y)$, введенная А. В. Скороходом, определяется следующим образом:

$$d(x, y) = \inf \left\{\varepsilon > 0 : \exists \lambda \in \Lambda : \sup_t |x_t - y_{\lambda(t)}| + \sup_t |t - \lambda(t)| \leqslant \varepsilon\right\}, \quad (18)$$

где Λ — множество строго возрастающих непрерывных на $[0, 1]$ функций $\lambda = \lambda(t)$ с $\lambda(0) = 0$, $\lambda(1) = 1$.

8. Измеримое пространство $\left(\prod\limits_{t \in T} \Omega_t, \boxtimes_{t \in T} \mathscr{F}_t\right)$**.** 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 $\left(\prod\limits_{t \in T} \Omega_t, \boxtimes_{t \in T} \mathscr{F}_t\right)$, 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 всех конечных объединений непересекающихся цилиндрических множеств

$$\mathscr{I}_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n) = \{\omega\colon\ \omega_{t_1} \in B_1,\ \ldots,\ \omega_{t_n} \in B_n\},$$

где $B_{t_i} \in \mathscr{F}_{t_i}$, образует, как нетрудно показать, алгебру. Наименьшую σ-алгебру, содержащую все цилиндрические множества, обозначают $\boxtimes_{t\in T} \mathscr{F}_t$, а измеримое пространство $(\prod \Omega_t,\ \boxtimes\, \mathscr{F}_t)$ называют *прямым произведением* измеримых пространств $(\Omega_t, \mathscr{F}_t)$, $t \in T$.

9. Задачи.

1. Пусть $\mathscr{B}_1$ и $\mathscr{B}_2$ — две σ-алгебры подмножеств пространства Ω. Будут ли σ-алгебрами системы множеств

$$\mathscr{B}_1 \cap \mathscr{B}_2 \equiv \{A\colon\ A \in \mathscr{B}_1 \text{ и } A \in \mathscr{B}_2\},$$
$$\mathscr{B}_1 \cup \mathscr{B}_2 \equiv \{A\colon\ A \in \mathscr{B}_1 \text{ или } A \in \mathscr{B}_2\}?$$

2. Пусть $\mathscr{D} = \{D_1, D_2, \ldots\}$ — некоторое счетное разбиение Ω и $\mathscr{B} = \sigma(\mathscr{D})$. Какова мощность σ-алгебры $\mathscr{B}$?

3. Показать, что

$$\mathscr{B}(R^n) \otimes \mathscr{B}(R) = \mathscr{B}(R^{n+1}).$$

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 (18) 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

$$\rho(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min\Big[\sup_{0\leqslant t\leqslant n} |x_t - y_t|,\ 1\Big], \qquad x, y \in C,$$

это пространство является (как и в случае $C = C[0, 1]$) польским, т. е. *полным сепарабельным метрическим пространством*, и σ-алгебра $\mathscr{B}_0(C)$, порожденная открытыми множествами, совпадает с σ-алгеброй $\mathscr{B}(C)$, порожденной цилиндрическими множествами.

9. Доказать равносильность групп условий (λ_a), (λ_b), (λ_c) и (λ_a), (λ'_b), (λ'_c) (см. с. 205).

10. Вывести теорему 2 из теоремы 1.

11. Доказать, что в теореме 3 система $\mathscr{L}$ является λ-системой.

12. Говорят, что σ-алгебра является *счетно-порожденной* или *сепарабельной*, если она порождается некоторым счетным классом подмножеств.

Показать, что σ-алгебра $\mathscr{B}$ борелевских подмножеств $\Omega = (0, 1]$ является счетно-порожденной.

Показать на примере, что возможна ситуация, когда две σ-алгебры $\mathscr{F}_1$ и $\mathscr{F}_2$ таковы, что $\mathscr{F}_1 \subset \mathscr{F}_2$, $\mathscr{F}_2$ — счетно порождена, но $\mathscr{F}_1$ такой не является.

13. Показать, что для того, чтобы σ-алгебра $\mathscr{G}$ была счетно-порожденной, необходимо и достаточно, чтобы $\mathscr{G} = \sigma(X)$ для некоторой случайной величины X (определение $\sigma(X)$ см. в п. 4 § 4).

14. Дать пример сепарабельной σ-алгебры, у которой есть несепарабельная под-σ-алгебра.

15. Показать, что $X_1, X_2, \ldots$ — независимая система случайных величин (§§ 4, 5), если $\sigma(X_n)$ и $\sigma(X_1, \ldots, X_{n-1})$ независимы при каждом $n \geqslant 1$.

16. Привести пример, показывающий, что соединение двух σ-алгебр не есть σ-алгебра.

17. Пусть $\mathscr{A}_1$ и $\mathscr{A}_2$ — две независимые системы множеств, каждая из которых является π-системой. Показать, что тогда и $\sigma(\mathscr{A}_1)$ и $\sigma(\mathscr{A}_2)$ также независимы. Привести пример двух независимых систем $\mathscr{A}_1$ и $\mathscr{A}_2$, *не* являющихся π-системами, для которых $\sigma(\mathscr{A}_1)$ и $\sigma(\mathscr{A}_2)$ уже зависимы.

18. Пусть $\mathscr{L}$ есть λ-система. Тогда $(A, B \in \mathscr{L}, A \cap B = \varnothing) \Rightarrow (A \cup B \in \in \mathscr{L})$.

19. Пусть $\mathscr{F}_1$ и $\mathscr{F}_2$ — две σ-алгебры подмножеств в Ω. Положим

$$d(\mathscr{F}_1, \mathscr{F}_2) = 4 \sup_{\substack{A_1 \in \mathscr{F}_1 \\ A_2 \in \mathscr{F}_2}} |\ \ (A_1 A_2) - \ \ (A_1)\ \ (A_2)|.$$

Показать, что эта величина, характеризующая степень зависимости между $\mathscr{F}_1$ и $\mathscr{F}_2$, имеет следующие свойства:

(a) $0 \leqslant d(\mathscr{F}_1, \mathscr{F}_2) \leqslant 1$;

(b) $d(\mathscr{F}_1, \mathscr{F}_2) = 0$, если и только если $\mathscr{F}_1$ и $\mathscr{F}_2$ независимы;

(c) $d(\mathscr{F}_1, \mathscr{F}_2) = 1$, если и только если пересечение $\mathscr{F}_1$ и $\mathscr{F}_2$ содержит множество, вероятность которого равна $1/2$.

20. Применяя метод доказательства леммы 1, доказать существование и единственность классов $\lambda(\mathscr{E})$ и $\pi(\mathscr{E})$, содержащих систему множеств $\mathscr{E}$.

21. Пусть $\mathscr{A}$ есть некоторая алгебра множеств, обладающая тем свойством, что любая последовательность $(A_n)_{n \geqslant 1}$ непересекающихся множеств $A_n \in \mathscr{A}$ такова, что $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{A}$. Доказать, что тогда $\mathscr{A}$ является σ-алгеброй.

22. Пусть $(\mathscr{F}_n)_{n\geqslant 1}$ есть возрастающая последовательность σ-алгебр: $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$, $n \geqslant 1$. Показать, что $\bigcup\limits_{n=1}^{\infty} \mathscr{F}_n$ есть (вообще говоря, только лишь) алгебра.

23. Пусть $\mathscr{F}$ есть алгебра (или σ-алгебра) и некоторое множество C не принадлежит $\mathscr{F}$. Рассмотрим наименьшую алгебру (соответственно σ-алгебру), порожденную множествами из $\mathscr{F} \cup \{C\}$. Показать, что все элементы этой алгебры (соответственно σ-алгебры) состоят из множеств вида $(A \cap C) \cup (B \cap \bar{C})$, где $A, B \in \mathscr{F}$.

24. Пусть $\bar{R} = R \cup \{-\infty\} \cup \{\infty\}$ — *расширенная* числовая прямая. Борелевская σ-алгебра $\mathscr{B}(\bar{R})$ может быть определена (ср. с определением в п. 2), как σ-алгебра, порожденная множествам $[-\infty, x]$, $x \in R$, где $[-\infty, x] = \{-\infty\} \cup (-\infty, x]$. Показать, что эта σ-алгебра $\mathscr{B}(\bar{R})$ совпадает также с любой из σ-алгебр, порожденных множествами

(a) $[-\infty, x)$, $x \in R$, или

(b) $(x, \infty]$, $x \in R$, или

(c) всеми конечными интервалами и $\{-\infty\}$ и $\{\infty\}$.

§ 3. Способы задания вероятностных мер на измеримых пространствах

1. Измеримое пространство $(R, \mathscr{B}(R))$**.** Пусть $\quad = \quad (A)$ — вероятностная мера, определенная на борелевских множествах A числовой прямой. Возьмем $A = (-\infty, x]$ и положим

$$F(x) = \quad (-\infty, x], \qquad x \in R. \tag{1}$$

Так определенная функция обладает следующими свойствами:

1) $F(x)$ — *неубывающая функция*;

2) $F(-\infty) = 0$, $F(+\infty) = 1$, *где*

$$F(-\infty) = \lim_{x \downarrow -\infty} F(x), \qquad F(+\infty) = \lim_{x \uparrow \infty} F(x);$$

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 $\quad$ 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   такая, что для любых $-\infty \leqslant a < b < \infty$*

$$(a,\, b] = F(b) - F(a). \tag{2}$$

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 ${}_0$, полагая

$${}_0(A) = \sum_{k=1}^{n} [F(b_k) - F(a_k)], \qquad A \in \mathscr{A}. \tag{3}$$

На алгебре $\mathscr{A}$ эта формула определяет по F, и, очевидно, однозначно, *конечно-аддитивную* функцию множеств. Поэтому, если показать, что на этой алгебре эта функция к тому же *счетно-аддитивна*, то существование и единственность требуемой меры   на $\mathscr{B}(R)$ будет непосредственно вытекать из следующего общего результата теории меры (приводимого здесь без доказательства, по поводу которого см., например, [42], [70]).

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 (как σ-аддитивная функция множеств) на $(\Omega, \mathscr{A})$. Тогда существует и притом единственная мера μ на $(\Omega, \mathscr{B})$, являющаяся продолжением μ_0, т. е. такая, что*

$$\mu(A) = \mu_0(A), \qquad A \in \mathscr{A}.$$

Итак, покажем, что функция ${}_0$ счетно-аддитивна (т. е. является вероятностной мерой) на алгебре $\mathscr{A}$. Согласно теореме из § 1, для этого достаточно проверить непрерывность ${}_0$ в $\varnothing$, т. е. проверить, что

$${}_0(A_n) \downarrow 0, \qquad A_n \downarrow \varnothing, \qquad A_n \in \mathscr{A}.$$

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 $F(x)$

$${}_0(a',\, b] = F(b) - F(a') \to F(b) - F(a) = {}_0(a,\, b]$$

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

$$ {}_0(A_n) - \quad {}_0(B_n) \leqslant \varepsilon 2^{-n}, $$

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 $\{[-N, N] \setminus [B_n]\}_{n \geqslant 1}$ образует *открытое покрытие* этого компакта. Тогда по *лемме Гейне—Бореля* (см., например, [1], [33]) существует *конечное* подпокрытие:

$$ \bigcup_{n=1}^{n_0} ([-N, N] \setminus [B_n]) = [-N, N], $$

а значит, $\bigcap\limits_{n=1}^{n_0} [B_n] = \varnothing$.)

Учитывая (4) и то, что $A_{n_0} \subseteq A_{n_0-1} \subseteq \ldots \subseteq A_1$, находим

$$ {}_0(A_{n_0}) = \quad {}_0\left(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\right) + \quad {}_0\left(\bigcap_{k=1}^{n_0} B_k\right) = \quad {}_0\left(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\right) \leqslant $$

$$ \leqslant \quad {}_0\left(\bigcup_{k=1}^{n_0} (A_k \setminus B_k)\right) \leqslant \sum_{k=1}^{n_0} \quad {}_0(A_k \setminus B_k) \leqslant \sum_{k=1}^{n_0} \varepsilon 2^{-k} \leqslant \varepsilon. $$

Поэтому $\quad {}_0(A_n) \downarrow 0$, $n \to \infty$.

Откажемся теперь от предположения, что все $A_n \subseteq [-N, N]$ для некоторого N. Зададим $\varepsilon > 0$ и выберем такое N, что $\quad {}_0[-N, N] > 1 - \varepsilon/2$. Тогда, поскольку

$$ A_n = A_n \cap [-N, N] + A_n \cap \overline{[-N, N]}, $$

то

$$ {}_0(A_n) = \quad {}_0(A_n \cap [-N, N]) + \quad {}_0(A_n \cap \overline{[-N, N]}) \leqslant \quad {}_0(A_n \cap [-N, N]) + \varepsilon/2, $$

и, применяя предшествующие рассуждения (с заменой A_n на $A_n \cap [-N, N]$), получаем, что для достаточно больших n $\quad {}_0(A_n \cap [-N, N]) \leqslant \varepsilon/2$. Тем самым снова $\quad {}_0(A_n) \downarrow 0$, $n \to \infty$. $\square$

Итак, между вероятностными мерами на $(R, \mathscr{B}(R))$ и функциями распределения F на числовой прямой R существует взаимно однозначное соответствие. Меру , 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 $\lambda(a, b] = b - a$, где $0 \leqslant a \leqslant b \leqslant 1$. 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

$$\mathscr{B}([0, 1]) = \{A \cap [0, 1] \colon A \in \mathscr{B}(R)\}$$

совокупность *борелевских* множеств отрезка $[0, 1]$. Наряду с этими множествами часто приходится рассматривать так называемые *лебеговские* множества отрезка $[0, 1]$. Мы говорим, 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 $\overline{\lambda}(\Lambda) = \lambda(A)$. Так определенная функция множеств $\overline{\lambda} = \overline{\lambda}(\Lambda)$, $\Lambda \in \overline{\mathscr{B}}([0, 1])$, является, как нетрудно проверить, вероятностной мерой на $([0, 1], \overline{\mathscr{B}}([0, 1]))$. Ее называют *лебеговской мерой* (на системе лебеговских множеств).

Замечание 1. Проведенная процедура пополнения (продолжения) меры применяется и оказывается полезной не только в рассмотренном случае. Например, пусть $(\Omega, \mathscr{F},\ )$ — некоторое вероятностное пространство. Обозначим через $\overline{\mathscr{F}}$ совокупность всех подмножеств Λ пространства Ω, для которых можно найти такие множества A и B из $\mathscr{F}$, что $A \subseteq \Lambda \subseteq B$ и $(B \setminus A) = 0$. Естественным образом (с помощью равенства $(\Lambda) =$ (B)) вероятностная мера определяется и для множеств $\Lambda \in \overline{\mathscr{F}}$. Полученное таким образом новое вероятностное пространство $(\Omega, \overline{\mathscr{F}},\ )$ называется *пополнением* пространства $(\Omega, \mathscr{F},\ )$ относительно меры .

Если вероятностная мера такова, что $\bar{\mathscr{F}}^{\,} = \mathscr{F}$, то она называется *полной*, а соответствующее пространство $(\Omega, \mathscr{F}, \ )$ — *полным вероятностным пространством.*

Замечание 2. Кратко остановимся на идее доказательства теоремы Каратеодори, считая, что $\mu_0(\Omega) = 1$.

Пусть A — множество из Ω и $A_1, A_2, \ldots$ — множества из $\mathscr{A}$, накрывающие множество A в том смысле, что $A \subseteq \bigcup\limits_{n=1}^{\infty} A_n$. Определим *внешнюю меру* $\mu^*(A)$ множества A, полагая

$$\mu^*(A) = \inf \sum_{n=1}^{\infty} \mu_0(A_n),$$

где inf берется по всем указанным покрытиям $(A_1, A_2, \ldots)$ множества A. *Внутренней мерой* $\mu_*(A)$ множества A будем называть величину

$$\mu_*(A) = 1 - \mu^*(\bar{A}).$$

Обозначим $\widehat{\mathscr{A}}$ совокупность тех множеств A из Ω, для которых $\mu^*(A) = \mu_*(A)$. Система $\widehat{\mathscr{A}}$ является, как нетрудно показать, *σ-алгеброй* (задача 12), и, следовательно, $\mathscr{A} \subseteq \sigma(\mathscr{A}) \subseteq \widehat{\mathscr{A}}$. Припишем множествам A из $\widehat{\mathscr{A}}$ «меру» $\mu(A)$, полагая ее равной $\mu^*(A)$ $(= \mu_*(A))$. Эта функция множеств $\mu(A)$, $A \in \widehat{\mathscr{A}}$, действительно является *мерой* (задача 13), т. е. является счетно-аддитивной функцией множеств (при этом вероятностной, поскольку $\mu(\Omega) = \mu_0(\Omega) = 1$).

Устанавливаемое равенством $(a, b] = F(b) - F(a)$ соответствие между вероятностными мерами и функциями распределения F дает возможность конструирования разных вероятностных мер с помощью задания соответствующих функций распределения.

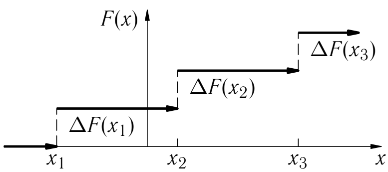


Рис. 25.

Дискретные меры. Так называют меры , для которых соответствующая функция $F = F(x)$ является функцией, меняющей свои значения в точках $x_1, x_2, \ldots$ $(\Delta F(x_i) > 0$, где $\Delta F(x) = F(x) - F(x-))$ (рис. 25). В этом случае мера сосредоточена в точках $x_1, x_2, \ldots$:

$$(\{x_k\}) = \Delta F(x_k) > 0, \qquad \sum_k \ (\{x_k\}) = 1.$$

Набор чисел $(p_1, p_2, \ldots)$, где $p_k = \quad (\{x_k\})$, называют *дискретным распределением вероятностей*, а соответствующую функцию распределения $F = F(x)$ — *дискретной*.

Приведем таблицу наиболее употребительных типов дискретных вероятностных распределений с соответствующими наименованиями.

Таблица 2

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 (распределение Паскаля)	$C_{k-1}^{r-1} p^r q^{k-r}$, $k = r, r+1, \ldots$	$0 < p \leqslant 1,\ \ q = 1 - p$, $r = 1, 2, \ldots$

Абсолютно непрерывные меры. Так называются меры, для которых соответствующая функция распределения $F = F(x)$ такова, что для некоторой неотрицательной борелевской функции $f = f(t)$, $t \in R$, имеет место равенство

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \tag{5}$$

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), а сама функция $F = F(x)$ — *абсолютно непрерывной*.

Понятно, что всякая неотрицательная функция $f = f(x)$, интегрируемая по Риману и такая, что $\int_{-\infty}^{\infty} f(x)\,dx = 1$, определяет формулой (5) функцию распределения некоторой вероятностной меры на $(R, \mathscr{B}(R))$. В таблице 3 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 не указанных в таблице значений x).

Сингулярные меры. Так называют меры, функции распределения $F(x)$ которых *непрерывны*, но *точки* их *роста* (x — точка роста $F(x)$, если

Таблица 3

Тип распределения	Плотность	Параметры
Равномерное на $[a, b]$	$\frac{1}{b-a},\ a \leqslant x \leqslant b$	$a, b \in R;\ a < b$
Нормальное, или гауссовское	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}},\ x \in R$	$m \in R,\ \sigma > 0$
Гамма	$\frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\,\beta^{\alpha}},\ x \geqslant 0$	$\alpha > 0,\ \beta > 0$
Бета	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)},\ 0 \leqslant x \leqslant 1$	$\alpha > 0,\ \beta > 0$
Экспоненциальное (гамма-распределение с $\alpha = 1,\ \beta = 1/\lambda$)	$\lambda e^{-\lambda x},\ x \geqslant 0$	$\lambda > 0$
Двустороннее экспоненциальное	$\frac{\lambda}{2} e^{-\lambda\lvert x-\alpha\rvert},\ x \in R$	$\lambda > 0,\ \alpha \in R$
Хи-квадрат, χ^2 (гамма-распределение с $\alpha = n/2,\ \beta = 2$) с n степенями свободы	$\frac{1}{2^{n/2}\,\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}},\ x \geqslant 0$	$n = 1, 2, \ldots$
Стьюдента, t-распределение с n степенями свободы	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}},\ x \in R$	$n = 1, 2, \ldots$
F-распределение	$\frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\mathrm{B}\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{\left(1+\frac{mx}{n}\right)^{\frac{m+n}{2}}},\ x \geqslant 0$	$m, n = 1, 2, \ldots$
Коши	$\frac{\theta}{\pi(x^2+\theta^2)},\ x \in R$	$\theta > 0$

$F(x+\varepsilon) - F(x-\varepsilon) > 0$ для любого $\varepsilon > 0$) образуют множество *нулевой меры Лебега*. Опуская подробности относительно структуры таких функций (см., например, [70]), приведем лишь один «классический» пример.

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.

Далее, каждый из интервалов $[0, 1/3]$ и $[2/3, 1]$ снова делим на три части и определяем функцию (рис. 27)

$$F_2(x) = \begin{cases} 1/2, & x \in (1/3,\, 2/3), \\ 1/4, & x \in (1/9,\, 2/9), \\ 3/4, & x \in (7/9,\, 8/9), \\ 0, & x = 0, \\ 1, & x = 1, \end{cases}$$

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

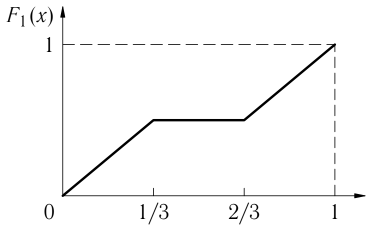


Рис. 26.

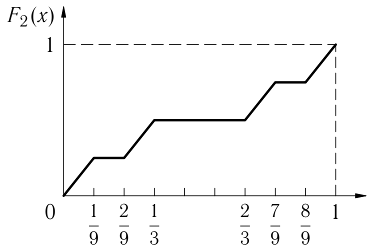


Рис. 27.

функции $F(x)$ (называемой канторовой), точки роста которой образуют 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 канторовой функции $F(x)$. Из (6) следует, что $\lambda(\mathscr{N}) = 0$. В то же самое время, если μ — мера, соответствующая канторовой 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 такие функции. Точнее, произвольная функция распределения может быть представлена в виде $\alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3$, где F_1 — дискретная, F_2 — абсолютно непрерывная, F_3 — сингулярная функции распределения, α_i — неотрицательные числа, $\alpha_1 + \alpha_2 + \alpha_3 = 1$ (задача 18).

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

$$\mu(a, b] = G(b) - G(a), \quad a < b,$$

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 на $(a, b]$ значение $\mu_0(a, b] = G(b) - G(a)$, 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 $\bar{\mathscr{B}}(R)$ ($\Lambda \in \bar{\mathscr{B}}(R)$, 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 $A \subseteq \Lambda \subseteq B$, $\Lambda \in \bar{\mathscr{B}}(R)$ и $\lambda(B \setminus A) = 0$).

3. Измеримое пространство $(R^n, \mathscr{B}(R^n))$**.** Как и в случае действительной прямой, предположим, что — некоторая вероятностная мера на $(R^n, \mathscr{B}(R^n))$.

Обозначим

$$F_n(x_1, \ldots, x_n) = \quad ((-\infty, x_1] \times \ldots \times (-\infty, x_n]),$$

или, в более компактной форме,

$$F_n(x) = \quad (-\infty, x],$$

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 $(a_i \leqslant b_i)$

$$\Delta_{a_i b_i} F_n(x_1, \ldots, x_n) = F_n(x_1, \ldots, x_{i-1}, b_i, x_{i+1}, \ldots, x_n) - \\ - F_n(x_1, \ldots, x_{i-1}, a_i, x_{i+1}, \ldots, x_n).$$

Простой подсчет показывает, что

$$\Delta_{a_1 b_1} \ldots \Delta_{a_n b_n} F_n(x_1, \ldots, x_n) = \quad (a, b], \tag{7}$$

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 $\quad (a, b]$, вообще говоря, *не равна* разности $F_n(b) - F_n(a)$.

Поскольку $(a, b] \geqslant 0$, 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 вытекает также, что $F_n(x_1, \ldots, x_n)$ 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 $F = F(x_1, \ldots, x_n)$ — некоторая функция распределения в R^n. Тогда на $(R^n, \mathscr{B}(R^n))$ существует и притом единственная вероятностная мера такая, что*

$$(a, b] = \Delta_{a_1 b_1} \ldots \Delta_{a_n b_n} F_n(x_1, \ldots, x_n). \tag{12}$$

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

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad x \in R,$$

с $\sigma > 0$ есть плотность (невырожденного) *гауссовского*, или *нормального*, *распределения*. Существуют естественные аналоги этой плотности и в случае $n > 1$.

Пусть $\mathbb{R} = \|r_{ij}\|$ — некоторая неотрицательно определенная симметрическая матрица порядка $n \times n$:

$$\sum_{i,j=1}^{n} r_{ij}\lambda_i\lambda_j \geqslant 0, \qquad \lambda_i \in R, \;\; i = 1, \ldots, n,$$

$$r_{ij} = r_{ji}.$$

В том случае, когда $\mathbb{R}$ — положительно определенная матрица, ее детерминант $|\mathbb{R}| \equiv \det \mathbb{R} > 0$, и, следовательно, определена обратная матрица $A = \|a_{ij}\|$. Тогда функция

$$f_n(x_1, \ldots, x_n) = \frac{|A|^{1/2}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum a_{ij}(x_i - m_i)(x_j - m_j)\right\}, \tag{13}$$

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

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \times \right.$$
$$\left. \times \left[\frac{(x_1-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\}, \quad (14)$$

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».

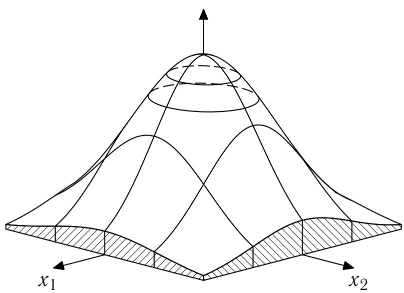


Рис. 28. Плотность двумерного нормального распределения

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

$$\mathscr{I}_n(B) = \{x \in R^\infty : (x_1, \ldots, x_n) \in \mathscr{B}\}, \quad B \in \mathscr{B}(R^n),$$

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 — некоторая вероятностная мера на $(R^\infty, \mathscr{B}(R^\infty))$. Обозначим для $n = 1, 2, \ldots$

$$P_n(B) = \ (\mathscr{I}_n(B)), \qquad B \in \mathscr{B}(R^n). \tag{15}$$

Последовательность вероятностных мер $P_1, P_2, \ldots$, определенных соответственно на $(R, \mathscr{B}(R))$, $(R^2, \mathscr{B}(R^2))$, …, 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 (Колмогорова о продолжении мер в $(R^\infty, \mathscr{B}(R^\infty))$). *Пусть* $P_1, P_2, \ldots$ — *последовательность вероятностных мер на* $(R, \mathscr{B}(R))$, $(R^2, \mathscr{B}(R^2))$, …, *обладающих свойством согласованности* (16). *Тогда существует и притом единственная вероятностная мера на* $(R^\infty, \mathscr{B}(R^\infty))$ *такая, что для каждого* $n = 1, 2, \ldots$

$$(\mathscr{I}_n(B)) = P_n(B), \qquad B \in \mathscr{B}(R^n). \tag{17}$$

Доказательство. Пусть $B^n \in \mathscr{B}(R^n)$ и $\mathscr{I}_n(B^n)$ — цилиндр с «основанием» B^n. Припишем этому цилиндру меру $(\mathscr{I}_n(B^n))$, полагая ее равной $P_n(B^n)$.

Покажем, что в силу условия согласованности такое определение является корректным, т. е. значение $(\mathscr{I}_n(B^n))$ не зависит от способа представления цилиндрического множества $\mathscr{I}_n(B^n)$. В самом деле, пусть один и тот же цилиндр представлен двумя способами:

$$\mathscr{I}_n(B^n) = \mathscr{I}_{n+k}(B^{n+k}).$$

Отсюда следует, что если $(x_1, \ldots, x_{n+k}) \in R^{n+k}$, то

$$(x_1, \ldots, x_n) \in B \ \Leftrightarrow \ (x_1, \ldots, x_{n+k}) \in B^{n+k}, \tag{18}$$

и, значит, в силу (16)

$$P_n(B^n) = P_{n+1}((x_1, \ldots, x_{n+1})\colon (x_1, \ldots, x_n) \in B^n) = \ldots$$
$$\ldots = P_{n+k}((x_1, \ldots, x_{n+k})\colon (x_1, \ldots, x_n) \in B) = P_{n+k}(B^{n+k}).$$

Обозначим $\mathscr{A}(R^\infty)$ совокупность всех цилиндрических множеств $\widehat{B}^n = \mathscr{I}_n(B^n)$, $B^n \in \mathscr{B}(R^n)$, $n = 1, 2\ldots$ Нетрудно видеть, что $\mathscr{A}(R^\infty)$ — алгебра.

Пусть теперь $\widehat{B}^1, \ldots, \widehat{B}^k$ — непересекающиеся множества из $\mathscr{A}(R^\infty)$. Без ограничения общности можно считать, что все они таковы, что для некоторого n $\widehat{B}_i = \mathscr{I}_n(B_i^n)$, $i = 1, \ldots, k$, где $B_1^n, \ldots, B_k^n$ — непересекающиеся множества из $\mathscr{B}(R^n)$. Тогда

$$\left(\sum_{i=1}^{k} \widehat{B}_i\right) = \quad \left(\sum_{i=1}^{k} \mathscr{I}_n(B_i^n)\right) = P_n\left(\sum_{i=1}^{k} B_i^n\right) = \sum_{i=1}^{k} P_n(B_i^n) = \sum_{i=1}^{n} \quad (\widehat{B}_i),$$

т. е. функция множеств — конечно-аддитивна на алгебре $\mathscr{A}(R^\infty)$.

Покажем, что непрерывна в «нуле» (а, значит, и σ-аддитивна на $\mathscr{A}(R^\infty)$; см. теорему в § 1), т. е. если последовательность множеств $\widehat{B}_n \downarrow \varnothing$, $n \to \infty$, то $(\widehat{B}_n) \to 0$, $n \to \infty$.

Предположим противное, т. е. пусть $\lim\limits_n \quad (\widehat{B}_n) = \delta > 0$. Без ограничения общности можно считать, что последовательность $\{\widehat{B}_n\}$ такова, что

$$\widehat{B}_n = \{x\colon (x_1, \ldots, x_n) \in B_n\}, \qquad B_n \in \mathscr{B}(R^n).$$

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

$$(\widehat{B}_n \setminus \hat{A}_n) = P_n(B_n \setminus A_n) \leqslant \delta/2^{n+1}.$$

Образуем множество $\widehat{C}_n = \bigcap\limits_{k=1}^{n} \hat{A}_k$, и пусть C_n таковы, что

$$\widehat{C}_n = \{x\colon (x_1, \ldots, x_n) \in C_n\}.$$

Тогда, учитывая, что множества $\widehat{B}_n$ убывают, находим

$$(\widehat{B}_n \setminus \widehat{C}_n) \leqslant \sum_{k=1}^{n} \quad (\widehat{B}_n \setminus \hat{A}_k) \leqslant \sum_{k=1}^{n} \quad (\widehat{B}_k \setminus \hat{A}_k) \leqslant \delta/2.$$

Но по предположению $\lim\limits_n \quad (\widehat{B}_n) = \delta > 0$, и, значит, $\lim\limits_n \quad (\widehat{C}_n) \geqslant \delta/2 > 0$. Покажем, что это противоречит тому, что $\widehat{C}_n \downarrow \varnothing$.

Действительно, выберем в множествах $\widehat{C}_n$ по точке $\hat{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots)$. Тогда для каждого $n \geqslant 1$ $(x_1^{(n)}, \ldots, x_n^{(n)}) \in C_n$.

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 $x_1^{(n_1)} \in C_1$, а C_1 — компакт.) Из последовательности (n_1) выберем подпоследовательность (n_2) такую, что $(x_1^{(n_2)}, x_2^{(n_2)}) \to (x_1^0, x_2^0) \in C_2$. Аналогичным образом пусть $(x_1^{(n_k)}, \ldots, x_k^{(n_k)}) \to$ $\to (x_1^0, \ldots, x_k^0) \in C_k$. 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 $(x_1^0, x_2^0, \ldots) \in \widehat{C}_n$ для любого $n = 1, 2, \ldots$, что, очевидно, противоречит предположению о том, что $\widehat{C}_n \downarrow \varnothing$, $n \to \infty$.

Итак, функция множеств на алгебре $\mathscr{A}(R^\infty)$ является σ-аддитивной и, значит, по теореме Каратеодори может быть продолжена до (вероятностной) меры на $(R^\infty, \mathscr{B}(R^\infty))$. □

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 $(\Omega_n, \mathscr{F}_n)$, если меры P_n, $n \geqslant 1$, 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 $n = 1, 2, \ldots$

$$\varphi_n(\omega) = \begin{cases} 1, & 0 < \omega < 1/n, \\ 0, & 1/n \leqslant \omega \leqslant 1, \end{cases}$$

$$\mathscr{C}_n = \{A \in \Omega\colon\ A = \{\omega\colon\ \varphi_n(\omega) \in B\},\ \ B \in \mathscr{B}(R)\}$$

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:

$$P_n\{\omega\colon\ (\varphi_1(\omega),\ \ldots,\ \varphi_n(\omega)) \in B^n\} = \begin{cases} 1, & \text{если } (1,\ \ldots,\ 1) \in B^n, \\ 0 & \text{в противном случае,} \end{cases}$$

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 такой, чтобы ее *сужение* $|\mathscr{F}_n$ (т. е. мера , рассматриваемая лишь на множествах из $\mathscr{F}_n$) совпадало с P_n, $n = 1,\ 2,\ \ldots$ В самом деле, допустим, что такая вероятностная мера существует. Тогда

$$\{\omega\colon\ \varphi_1(\omega) = \ldots = \varphi_n(\omega) = 1\} = P_n\{\omega\colon\ \varphi_1(\omega) = \ldots = \varphi_n(\omega) = 1\} = 1 \qquad (19)$$

для любого $n = 1,\ 2,\ \ldots$ Но

$$\{\omega\colon\ \varphi_1(\omega) = \ldots = \varphi_n(\omega) = 1\} = (0,\ 1/n) \downarrow \varnothing,$$

что противоречит (19) и предположению о счетной аддитивности (а значит, и непрерывности в «нуле» $\varnothing$) функции множеств .

Приведем теперь пример вероятностной меры в $(R^\infty,\ \mathscr{B}(R^\infty))$. Пусть $F_1(x),\ F_2(x),\ \ldots$ — последовательность одномерных функций распределения. Определим функции $G_1(x) = F_1(x)$, $G_2(x_1,\ x_2) = F_1(x_1)F_2(x_2),\ \ldots$ и соответствующие им вероятностные меры на $(R,\ \mathscr{B}(R))$, $(R^2,\ \mathscr{B}(R^2)),\ \ldots$ обозначим $P_1,\ P_2,\ \ldots$ Тогда из теоремы 3 следует, что в $(R^\infty,\ \mathscr{B}(R^\infty))$ существует такая мера , что

$$\{x \in R^\infty\colon\ (x_1,\ \ldots,\ x_n) \in B\} = P_n(B), \qquad B \in \mathscr{B}(R^n),$$

и, в частности,

$$\{x \in R^\infty\colon\ x_1 \leqslant a_1,\ \ldots,\ x_n \leqslant a_n\} = F_1(a_1)\ldots F_n(a_n).$$

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 $\mathscr{B}(R^\infty) \cap \Omega$ его борелевских подмножеств существует вероятностная мера такая, что для любого $n \geqslant 1$

$$\{x\colon\ x_1 = a_1,\ \ldots,\ x_n = a_n\} = p^{\sum a_i} q^{n - \sum a_i}.$$

Заметим, что именно этого результата нам не хватало в первой главе, чтобы сформулировать закон больших чисел в форме (8) § 5 гл. I.

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 $\sigma \subseteq \tau$,

$$P_\sigma\{(x_{s_1}, \ldots, x_{s_k}) \colon (x_{s_1}, \ldots, x_{s_k}) \in B\} = \\ = P_\tau\{(x_{t_1}, \ldots, x_{t_n}) \colon (x_{s_1}, \ldots, x_{s_k}) \in B\} \quad (20)$$

для любого $B \in \mathscr{B}(R^\sigma)$.

Теорема 4 (Колмогорова о продолжении мер 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 на $(R^T, \mathscr{B}(R^T))$ такая, что*

$$(\mathscr{I}_\tau(B)) = P_\tau(B)) \quad (21)$$

для всех неупорядоченных наборов $\tau = [t_1, \ldots, t_n]$ различных индексов $t_i \in T$, $B \in \mathscr{B}(R^\tau)$ и $\mathscr{I}_\tau(B) = \{x \in R^T : (x_{t_1}, \ldots, x_{t_n}) \in B\}$.

Доказательство. Пусть множество $\widehat{B} \in \mathscr{B}(R^T)$. Согласно теореме 3 из § 2, найдется не более чем счетное множество $S = \{s_1, s_2, \ldots\} \subseteq T$ такое, что $\widehat{B} = \{x \colon (x_{s_1}, x_{s_2}, \ldots) \in B\}$, где $B \in \mathscr{B}(R^S)$, $R^S = R_{s_1} \times R_{s_2} \times \ldots$ Иначе говоря, $\widehat{B} = \mathscr{I}_S(B)$ — цилиндрическое множество с «основанием» $B \in \mathscr{B}(R^S)$.

На таких цилиндрических множествах $\widehat{B} = \mathscr{I}_S(B)$ определим функцию множеств , полагая

$$(\mathscr{I}_S(B)) = P_S(B), \quad (22)$$

где P_S — та вероятностная мера, существование которой гарантируется теоремой 3.

Мы утверждаем, что — именно та мера, о существовании которой говорится в теореме. Чтобы установить это, надо, во-первых, проверить, что определение (22) корректно, т. е. приводит к одному и тому же значению $(\widehat{B})$ при разных способах представления $\widehat{B}$, и, во-вторых, что эта функция множеств счетно-аддитивна.

Итак, пусть $\widehat{B} = \mathscr{I}_{S_1}(B_1)$ и $\widehat{B} = \mathscr{I}_{S_2}(B_2)$. Ясно, что тогда $\widehat{B} = \mathscr{I}_{S_1 \cup S_2}(B_3)$ с некоторым $B_3 \in \mathscr{B}(R^{S_1 \cup S_2})$, и поэтому достаточно лишь убедиться в том, что если $S \subseteq S'$ и $B \in \mathscr{B}(R^S)$, то $P_{S'}(B') = P_S(B)$, где

$$B' = \{(x_{s'_1}, x_{s'_2}, \ldots)\colon (x_{s_1}, x_{s_2}, \ldots) \in B\}$$

с $S' = \{s'_1, s'_2, \ldots\}$, $S = \{s_1, s_2, \ldots\}$.

Но в силу условия согласованности (20) это равенство непосредственно вытекает из теоремы 3, что и доказывает независимость значений $(\widehat{B})$ от способа представления множества $\widehat{B}$.

Далее, для проверки свойства *счетной аддитивности* функции множеств предположим, что $\{\widehat{B}_n\}$ — некоторая последовательность попарно непересекающихся множеств из $\mathscr{B}(R^T)$. Тогда найдется такое не более чем счетное множество $S \subseteq T$, что для любого $n \geqslant 1$ $\widehat{B}_n = \mathscr{I}_S(B_n)$, где $B_n \in \mathscr{B}(R^S)$. Поскольку P_S — вероятностная мера, то

$$\Big(\sum \widehat{B}_n\Big) = \Big(\sum \mathscr{I}_S(B_n)\Big) = P_S\Big(\sum B_n\Big) =$$
$$= \sum P_S(B_n) = \sum (\mathscr{I}_S(B_n)) = \sum (\widehat{B}_n).$$

Наконец, свойство (21) непосредственно следует из самой конструкции меры . □

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. В этой связи важно подчеркнуть, что эти меры P_τ как функции от $\tau = [t_1, \ldots, t_n]$ являются, в сущности, функциями *множеств*, составленных из (разных) точек $\{t_1\}, \ldots, \{t_n\}$. (Скажем, неупорядоченные наборы $[a, b]$ и $[b, a]$ надо рассматривать как тождественные, поскольку они задают одно и то же множество, состоящее из точек $\{a\}$ и $\{b\}$.) Иногда же в качестве исходного берут семейство вероятностных мер $\{P_\tau\}$, где τ пробегает множество всех *упорядоченных* наборов $\tau = (t_1, \ldots, t_n)$ различных индексов. (Тогда наборы (a, b) и (b, a), составленные из одних и тех же точек, нужно рассматривать как разные, поскольку для упорядоченных наборов важен *порядок* следования их элементов.) В этом случае для справедливости теоремы 4 к условию (20) надо добавить еще одно *условие согласованности*:

$$P_{(t_1,\ldots,t_n)}(A_{t_1} \times \ldots \times A_{t_n}) = P_{(t_{i_1},\ldots,t_{i_n})}(A_{t_{i_1}} \times \ldots \times A_{t_{i_n}}), \tag{23}$$

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 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 $t_1<t_2<\ldots<t_n$. (Если, скажем, множество τ состоит из числовых точек $\{a_1\}, \{a_2\}, \ldots, \{a_n\}$, то τ всегда может быть представлено в виде $\tau=[t_1, t_2, \ldots, t_n]$, где $t_1<t_2<\ldots<t_n$, при этом $t_1=\min(a_1, \ldots, a_n)$, $\ldots$, $t_n=\max(a_1, \ldots, a_n)$.) Таким образом, все «конечномерные» вероятности в этом случае 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 $x=(x_t)_{t\geqslant 0}$. Важным примером вероятностной меры на $(R^{[0,\infty)}, \mathscr{B}(R^{[0,\infty)}))$ является так называемая *винеровская мера*, строящаяся следующим образом.

Рассмотрим семейство $\{\varphi_t(y\,|\,x)\}_{t\geqslant 0}$ гауссовских плотностей (по y при фиксированном x)

$$\varphi_t(y\,|\,x)=\frac{1}{\sqrt{2\pi t}}\,e^{-(y-x)^2/2t}, \qquad y\in R,$$

и определим для каждого набора $\tau=[t_1, \ldots, t_n]$, $t_1<t_2<\ldots<t_n$, и множеств $B=I_1\times\ldots\times I_n$, $I_k=(a_k, b_k)$, меру $P_\tau(B)$ по формуле

$$\begin{aligned}P_\tau(B)&=P_\tau(I_1\times\ldots\times I_n)=\\&=\int\limits_{I_1}\ldots\int\limits_{I_n}\varphi_{t_1}(a_1\,|\,0)\varphi_{t_2-t_1}(a_2\,|\,a_1)\ldots\varphi_{t_n-t_{n-1}}(a_n\,|\,a_{n-1})\,da_1\ldots da_n \qquad (24)\end{aligned}$$

(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 , полагая

$$(\mathscr{I}_{t_1,\ldots,t_n}(I_1\times\ldots\times I_n))=P_{[t_1,\ldots,t_n]}(I_1\times\ldots\times I_n).$$

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 $\varphi_{t_k-t_{k-1}}(a_k\,|\,a_{k-1})\,da_k$ как вероятность того, что частица, выходящая из точки a_{k-1}, за время t_k-t_{k-1} попадет в da_k-окрестность точки a_k. Тогда то, что в (24) рассматривается произведение плотностей, означает определенную независимость приращений смещений движущейся «частицы» на интервалах времени $[0, t_1]$, $[t_1, t_2]$, $\ldots$, $[t_{n-1}, t_n]$.

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 $F(x) = \ (-\infty, x]$. Показать справедливость следующих формул:

$$
\begin{aligned}
&(a, b] = F(b) - F(a), \qquad && (a, b) = F(b-) - F(a),\\
&[a, b] = F(b) - F(a-), \qquad && [a, b) = F(b-) - F(a-),\\
&(\{x\}) = F(x) - F(x-),
\end{aligned}
$$

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

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. Что можно сказать о соответствующем результате 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 некоторой 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 σ-алгебры лебеговских — 2^c.

8. Пусть $(\Omega, \mathscr{F}, \ )$ — 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

$$
(A \triangle B) \leqslant \varepsilon.
$$

9. Пусть — вероятностная мера на $(R^n, \mathscr{B}(R^n))$. Доказать, что для всякого $\varepsilon > 0$ и $B \in \mathscr{B}(R^n)$ можно найти *компактное множество* A_1 и открытое A_2 такие, что $A_1 \subseteq B \subseteq A_2$ и $\ (A_2 \setminus A_1) \leqslant \varepsilon$. (Этот результат используется в доказательстве теоремы 3.)

10. Проверить согласованность семейства мер $\{\ _\tau\}$, построенных по формуле $\ _\tau(B) = \ (\mathscr{I}_\tau(B))$, где — данная вероятностная мера (ср. с (21)).

11. Проверить, что приведенные в табл. 2 и 3 «распределения» действительно являются распределениями вероятностей.

12. Показать, что система $\widehat{\mathscr{A}}$ из замечания 2 в п. 1 является σ-алгеброй.

13. Показать, что функция множеств $\mu(A)$, $A \in \widehat{\mathscr{A}}$, введенная в замечании 2 п. 1, является *мерой*.

14. Привести пример, показывающий, что если мера μ_0 является на алгебре $\mathscr{A}$ *конечно-аддитивной* (но не счетно-аддитивной), то ее нельзя продолжить до счетно-аддитивной меры на $\sigma(\mathscr{A})$.

15. Показать, что всякая конечно-аддитивная вероятностная мера, заданная на алгебре $\mathscr{A}$ подмножеств Ω, может быть продолжена до конечно-аддитивной вероятности на *всех* подмножествах из Ω.

16. Пусть — вероятностная мера на σ-алгебре $\mathscr{F}$ подмножеств Ω. Пусть множество $C \subseteq \Omega$, но $C \notin \mathscr{F}$. Показать, что меру можно продолжить (с сохранением свойства счетной аддитивности) на σ-алгебру $\sigma(\mathscr{F} \cup \{C\})$.

17. Показать, что носитель *непрерывной* функции распределения F есть *совершенное* множество (т. е. носитель $\operatorname{supp} F$ есть замкнутое множество, обладающее тем свойством, что если $x \in \operatorname{supp} F$ и $\varepsilon > 0$, то найдется $y \in \operatorname{supp} F$ такое, что $0 < |x - y| < \varepsilon$). Показать, что носитель (произвольной) функции распределения является *замкнутым* множеством.

18. Доказать следующий фундаментальный результат (см. конец п. 1) о структуре функций распределения: каждая функция есть выпуклая комбинация

$$F = \alpha_1 F_{\mathrm{d}} + \alpha_2 F_{\mathrm{abc}} + \alpha_3 F_{\mathrm{sc}}$$

дискретной (F_{d}), абсолютно непрерывной (F_{abc}) и сингулярной непрерывной (F_{sc}) функций распределения; $\alpha_i \geqslant 0$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

19. Показать, что для канторовской функции распределения $F = F(x)$ для каждой точки из *канторова* множества $\mathscr{N}$ точек ее роста (совпадающего с носителем $\operatorname{supp} F$) имеет место представление $x = \sum\limits_{k=1}^{\infty} \dfrac{\alpha_k(x)}{3k}$, где $\alpha_k(x) = 0$ или 2, и что для таких точек $F(x) = \sum\limits_{k=1}^{\infty} \dfrac{\alpha_k(x)}{2^{k+1}}$.

20. Пусть C — некоторое замкнутое множество на R. Построить функцию распределения F, для которой носитель $\operatorname{supp} F = C$.

21. Показать, что в биномиальном случае (п. 1 § 2) функция распределения

$$B_n(m;\ p) = \ _n\{\nu \leqslant m\} = \sum_{k=0}^{m} C_n^k p^k q^{n-k}$$

может быть выражена через (неполную) бета-функцию:

$$B_n(m;\ p) = \frac{1}{\mathrm{B}(m+1,\, n-m)} \int\limits_p^1 x^m (1-x)^{n-m-1}\, dx.$$

22. Показать, что пуассоновская функция распределения $F(n;\, \lambda) = \sum\limits_{k=0}^{n} \frac{e^{-\lambda}\lambda^k}{k!}$ выражается следующим образом через (неполную) гамма-функцию:

$$F(n;\, \lambda) = \frac{1}{n!} \int\limits_\lambda^\infty x^n e^{-x}\, dx.$$

23. При описании формы плотностей распределений $f = f(x)$ помимо среднего значения и дисперсии стандартными характеристиками являются параметр «*скошенности*» (skewness)

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

и параметр «*пикообразности*» (peakedness, kurtosis)

$$\alpha_4 = \frac{\mu_4}{\sigma^4},$$

где $\mu_k = \int (x-\mu)^k f(x)\, dx$, $\mu = \int x f(x)\, dx$, $\sigma^2 = \mu_2$.

Рассмотреть вопрос о значениях параметров α_3 и α_4 для распределений, приведенных в таблице на с. 227.

24. Показать, что для случайной величины X с гамма-распределением (из таблицы на с. 227) с $\beta = 1$

$$X^k = \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)}.$$

В частности, $X = \alpha$, $X^2 = \alpha(\alpha+1)$ и, значит, $X = \alpha$.

Найти аналог этих формул, когда $\beta \neq 1$.

25. Показать, что для случайной величины X с бета-распределением (из таблицы на с. 227)

$$X^k = \frac{\mathrm{B}(r+k,\, s)}{\mathrm{B}(r,\, s)}.$$

26. При рассмотрении биномиального распределения фиксируется число испытаний n и рассматривается вероятность $_n\{\nu = r\}$ того, что число «успехов» в этих n испытаниях равно r. Эти вероятности $_n\{\nu = r\} = C_n^r p^r q^{n-r}$, $0 \leqslant r \leqslant n$, где p — вероятность единичного «успеха», образуют *биномиальное распределение* (n — задано). *Отрицательно-биномиальное* (или *обратно-биномиальное*) распределение возникает, когда

рассматривается вопрос о том, какова вероятность того, что r «успехов» впервые появятся на (случайном) шаге $\tau = k$ $(\geqslant r)$. Показать, что эта вероятность $\ ^r\{\tau = k\}$ задается формулой

$$\ ^r(\tau = k) = C_{k-1}^{r-1} p^r q^{k-r}, \quad k = r,\, r+1,\, \ldots,$$

с $r = 1, 2, \ldots$ (p — вероятность единичного «успеха»). Набор этих вероятностей $\ ^r\{\tau = k\}$ для $k = r,\, r+1, \ldots$ (r — задано) образует отрицательно-биномиальное распределение. Показать, что (для заданного r) $\ ^r\tau = rq/p$.

§ 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 $\xi = \xi(\omega)$, 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 , 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

$$P_\xi(B) = \ \{\omega\colon\ \xi(\omega) \in B\}, \quad B \in \mathscr{B}(R),$$

называется *распределением вероятностей случайной величины* ξ на $(R, \mathscr{B}(R))$.

Определение 3. Функция

$$F_\xi(x) = \quad \{\omega\colon \xi(\omega) \leqslant x\}, \qquad x \in R,$$

называется *функцией распределения случайной величины* ξ.

Для *дискретной* случайной величины ξ мера P_ξ сосредоточена не более чем в счетном числе точек и может быть представлена в виде

$$P_\xi(B) = \sum_{\{k:\, x_k \in B\}} p(x_k), \tag{3}$$

где $p(x_k) = \quad \{\xi = x_k\} = \Delta F_\xi(x_k)$.

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 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 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* (§ 2).

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:

$$\begin{gathered}\xi^{-1}\Big(\bigcup_{\alpha} B_\alpha\Big)=\bigcup_{\alpha}\xi^{-1}(B_\alpha),\\ \xi^{-1}\Big(\bigcap_{\alpha} B_\alpha\Big)=\bigcap_{\alpha}\xi^{-1}(B_\alpha),\\ \overline{\xi^{-1}(B_\alpha)}=\xi^{-1}(\bar B_\alpha).\end{gathered}\tag{6}$$

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 функция $\xi=\xi(\omega)$ 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

$$\begin{gathered}\mathscr{E}_1=\{x\colon\ x<c,\ c\in R\},\\ \mathscr{E}_2=\{x\colon\ x\leqslant c,\ c\in R\}\end{gathered}$$

порождает σ-алгебру $\mathscr{B}(R)$, т. е. $\sigma(\mathscr{E}_1)=\sigma(\mathscr{E}_2)=\mathscr{B}(R)$ (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 $\eta(\omega)=\varphi(\xi(\omega))$, также является случайной величиной.*

Доказательство следует из того, что для $B\in\mathscr{B}(R)$

$$\{\omega\colon\ \eta(\omega)\in B\}=\{\omega\colon\ \varphi(\xi(\omega))\in B\}=\{\omega\colon\ \xi(\omega)\in\varphi^{-1}(B)\}\in\mathscr{F},\tag{7}$$

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 3).

3. Отправляясь от заданной системы случайных величин $\{\xi_n\}$, можно строить новые функции, например, $\sum\limits_{n=1}^{\infty}|\xi_n|$, $\overline{\lim}\ \xi_n$, $\underline{\lim}\ \xi_n$ 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 $\bar{R} = [-\infty, \infty]$, называется *расширенной случайной величиной*, если для любого борелевского множества $B \in \mathscr{B}(\bar{R})$ (σ-алгебра $\mathscr{B}(\bar{R}) = \sigma(\mathscr{B}(R), \pm\infty)$) 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 $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$.*

b) *Если к тому же $\xi(\omega) \geqslant 0$, $\omega \in \Omega$, 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 $n = 1, 2, \ldots$

$$\xi_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\left\{\frac{k-1}{2^n} \leqslant \xi(\omega) < \frac{k}{2^n}\right\}}(\omega) + n I_{\{\xi(\omega) \geqslant n\}}(\omega).$$

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 $\xi = \xi^+ - \xi^-$. $\square$

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

$$\{\omega\colon\ \sup \xi_n > x\} = \bigcup_n \{\omega\colon \xi_n > x\} \in \mathscr{F},$$

$$\{\omega\colon\ \inf \xi_n < x\} = \bigcup_n \{\omega\colon \xi_n < x\} \in \mathscr{F}$$

и $\overline{\lim}\, \xi_n = \inf\limits_n \sup\limits_{m \geqslant n} \xi_n$, $\underline{\lim}\, \xi_n = \sup\limits_n \inf\limits_{m \geqslant n} \xi_m$.

Теорема 2. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность расширенных случайных величин и $\xi(\omega) = \lim \xi_n(\omega)$, $\omega \in \Omega$. Тогда $\xi = \xi(\omega)$ также является расширенной случайной величиной.*

Доказательство сразу следует из сделанного выше замечания и того, что

$$\begin{aligned}\{\omega\colon\ \xi(\omega) < x\} = \{\omega\colon\ & \lim \xi_n(\omega) < x\} = \\ = \{\omega\colon\ & \overline{\lim}\ \xi_n(\omega) = \underline{\lim}\ \xi_n(\omega)\} \cap \{\overline{\lim}\ \xi_n(\omega) < x\} = \\ & = \Omega \cap \{\overline{\lim}\ \xi_n(\omega) < x\} = \{\overline{\lim}\ \xi_n(\omega) < x\} \in \mathscr{F}.\end{aligned}$$

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 $\overline{R} = [-\infty, \infty]$ *).

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 $\infty - \infty$, $\dfrac{\infty}{\infty}$, $\dfrac{a}{0}\Big)$.

В самом деле, пусть $\{\xi_n\}$ и $\{\eta_n\}$ — последовательности случайных величин, сходящиеся к ξ и η (см. теорему 1). Тогда

$$\begin{gathered}\xi_n \pm \eta_n \to \xi \pm \eta, \\ \xi_n\eta_n \to \xi\eta, \\ \frac{\xi_n}{\eta_n + \dfrac{1}{n} I_{\{\eta_n = 0\}}(\omega)} \to \frac{\xi}{\eta}.\end{gathered}$$

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 $\xi = \xi(\omega)$ — случайная величина. Рассмотрим множества из $\mathscr{F}$ вида $\{\omega\colon \xi(\omega) \in B\}$, $B \in \mathscr{B}(R)$. Они образуют σ-алгебру, называемую *σ-алгеброй, порожденной случайной величиной* ξ. Будем ее обозначать $\mathscr{F}_\xi$ или $\sigma(\xi)$.

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 $\{\omega\colon \eta(\omega) \in B\} \in \mathscr{F}_\xi$, $B \in \mathscr{B}(R)$ (см. (7)). Оказывается, что справедлив и обратный результат.

---

*) В дальнейшем принимаются обычные соглашения относительно арифметических операций в $\overline{R}$: если $a \in R$, то $a \pm \infty = \pm\infty$, $\dfrac{a}{\pm\infty} = 0$; $a \cdot \infty = \infty$, если $a > 0$, и $a \cdot \infty = -\infty$, если $a < 0$; $0 \cdot (\pm\infty) = 0$, $\infty + \infty = \infty$, $-\infty - \infty = -\infty$

Теорема 3. *Пусть* $\eta = \eta(\omega)$ — $\mathscr{F}_\xi$*-измеримая случайная величина. Тогда найдется такая борелевская функция* φ, *что* $\eta = \varphi \circ \xi$, *т. е.* $\eta(\omega) = \varphi(\xi(\omega))$ *для каждого* $\omega \in \Omega$.

Доказательство. Пусть Φ — класс всех $\mathscr{F}_\xi$-измеримых функций $\eta = \eta(\omega)$, а $\widetilde{\Phi}_\xi$ — $\mathscr{F}_\xi$-измеримых функций, представимых в виде $\varphi \circ \xi$, где φ — некоторая борелевская функция. Ясно, что $\widetilde{\Phi}_\xi \subseteq \Phi_\xi$. Утверждение теоремы состоит в том, что на самом деле $\widetilde{\Phi}_\xi = \Phi_\xi$.

Пусть $A \in \mathscr{F}_\xi$ и $\eta(\omega) = I_A(\omega)$. Покажем, что $\eta \in \widetilde{\Phi}_\xi$. 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 $I_A(\omega) = \chi_B(\xi(\omega)) \in \widetilde{\Phi}_\xi$. Отсюда следует, что и любая простая $\mathscr{F}_\xi$-измеримая функция $\sum\limits_{i=1}^{n} c_i I_{A_i}(\omega)$, $A_i \in \mathscr{F}_\xi$, также принадлежит классу $\widetilde{\Phi}_\xi$.

Пусть теперь η — произвольная $\mathscr{F}_\xi$-измеримая функция. По теореме 1 найдется последовательность $\{\eta_n\}$ простых $\mathscr{F}_\xi$-измеримых функций $\eta_n = \eta_n(\omega)$ 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 6).

Но тогда, очевидно, $\eta(\omega) = \lim\limits_n \varphi_n(\xi(\omega)) = \varphi(\xi(\omega))$ для всех $\omega \in \Omega$. Следовательно, $\widetilde{\Phi}_\xi = \Phi_\xi$. □

6. Рассмотрим измеримое пространство $(\Omega, \mathscr{F})$ и некоторое конечное или счетное разбиение $\mathscr{D} = \{D_1, D_2, \ldots\}$ пространства Ω: $D_i \in \mathscr{F}$, $\sum\limits_i D_i = \Omega$. Образуем алгебру $\mathscr{A}$, содержащую пустое множество $\varnothing$ и множества вида $\sum\limits_\alpha D_\alpha$, где слагаемые берутся в конечном или счетном числе. Очевидно, что система $\mathscr{A}$ является монотонным классом, и поэтому, согласно лемме 2 § 2, алгебра $\mathscr{A}$ является в то же самое время и σ-алгеброй, обозначаемой $\sigma(\mathscr{D})$ и называемой *σ-алгеброй, порожденной разбиением* $\mathscr{D}$. Ясно, что $\sigma(\mathscr{D}) \subseteq \mathscr{F}$.

Лемма 3. *Пусть $\xi = \xi(\omega)$ является $\sigma(\mathscr{D})$-измеримой случайной величиной. Тогда ξ представима в виде*

$$\xi(\omega) = \sum_{k=1}^{\infty} x_k I_{D_k}(\omega), \tag{8}$$

где $x_k \in R$, $k \geqslant 1$, т. е. $\xi(\omega)$ принимает постоянные значения на элементах разбиения D_k, $k \leqslant 1$.

Доказательство. Возьмем некоторое множество D_k и покажем, что на нем $\sigma(\mathscr{D})$-измеримая функция ξ принимает постоянное значение.

С этой целью обозначим

$$x_k = \sup\,[c\colon\ D_k \cap \{\omega\colon\ \xi(\omega) < c\} = \varnothing].$$

Поскольку $\{\omega\colon \xi(\omega) < x_k\} = \bigcup \{\omega\colon \xi(\omega) < r\}$, где объединение берется по всем рациональным $r < x_k$, то $D_k \cap \{\omega\colon \xi(\omega) < x_k\} = \varnothing$.

Пусть теперь $c > x_k$. Тогда $D_k \cap \{\omega\colon \xi(\omega) < c\} \neq \varnothing$ и так как множество $\{\omega\colon \xi(\omega) < c\}$ имеет вид $\sum\limits_{\alpha} D_\alpha$, где сумма берется по конечному или счетному набору индексов, то

$$D_k \cap \{\omega\colon\ \xi(\omega) < c\} = D_k.$$

Отсюда вытекает, что для всех $c > x_k$

$$D_k \cap \{\omega\colon\ \xi(\omega) \geqslant c\} = \varnothing$$

и поскольку $\{\omega\colon \xi(\omega) > x_k\} = \bigcup \{\omega\colon \xi(\omega) \geqslant r\}$, где объединение берется по всем рациональным $r > x_k$, то

$$D_k \cap \{\omega\colon\ \xi(\omega) > x_k\} = \varnothing.$$

Таким образом, $D_k \cap \{\omega\colon \xi(\omega) \neq x_k\} = \varnothing$ и, значит,

$$D_k \subseteq \{\omega\colon\ \xi(\omega) = x_k\},$$

что и требовалось доказать. □

7. Задачи.

1. Показать, что случайная величина ξ непрерывна, если и только если $\{\xi = x\} = 0$ для всех $x \in R$.

2. Если $|\xi|$ является $\mathscr{F}$-измеримой, то верно ли, что ξ также $\mathscr{F}$-измерима?

3. Доказать, что функции x^n, $x^+ = \max(x, 0)$, $x^- = -\min(x, 0)$, $|x| = x^+ + x^-$ являются борелевскими.

4. Если ξ и η — $\mathscr{F}$-измеримы, то $\{\omega\colon \xi(\omega) = \eta(\omega)\} \in \mathscr{F}$.

5. Пусть ξ и η — две случайные величины на $(\Omega, \mathscr{F})$ и множество $A \in \mathscr{F}$. Тогда функция

$$\zeta(\omega) = \xi(\omega) I_A + \eta(\omega) I_{\bar{A}}$$

также является случайной величиной.

6. 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.

7. Пусть ξ и η — две случайные величины, принимающие значения $1, 2, \ldots, N$. Предположим, что $\mathscr{F}_\xi = \mathscr{F}_\eta$. Показать, что существует такая перестановка $(i_1, i_2, \ldots, i_N)$ чисел $(1, 2, \ldots, N)$, что для каждого $j = 1, 2, \ldots, N$ множества $\{\omega : \xi = j\}$ и $\{\omega : \eta = i + j\}$ совпадают.

8. Привести пример случайной величины ξ, функция распределения которой имеет плотность $f(x)$ такую, что $\lim\limits_{x \to \infty} f(x)$ не существует и, следовательно, $f(x)$ на бесконечности не аннулируется.

9. Пусть ξ и η — ограниченные случайные величины $(|\xi| \leqslant c_1,\ |\eta| \leqslant c_2)$. Доказать, что если для всех $m, n \geqslant 1$

$$\xi^m \eta^n = \quad \xi^m \cdot \quad \eta^n,$$

то ξ и η независимы.

10. Пусть ξ и η — случайные величины и их функции распредления F_ξ и F_η совпадают. Доказать, что если $x \in R$ и $\{\omega : \xi(\omega) = x\} \neq \varnothing$, то существует $y \in R$ такое, что $\{\omega : \xi(\omega) = x\} = \{\omega : \eta(\omega) = y\}$.

11. Пусть E — не более чем счетное подмножество R, ξ — отображение $\Omega \to E$. Доказать, что ξ является случайной величиной на $(\Omega, \mathscr{F})$ тогда и только тогда, когда $\{\omega : \xi(\omega) = x\} \in \mathscr{F}$ для каждого $x \in E$.

§ 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 $B \in \mathscr{B}(R)$

$$\begin{aligned}&\{\omega\colon \xi_k(\omega) \in B\} = \\ &\quad = \{\omega\colon \xi_1(\omega) \in R, \ldots, \xi_{k-1}(\omega) \in R, \xi_k(\omega) \in B, \xi_{k+1}(\omega) \in R, \ldots, \xi_n(\omega) \in R\} = \\ &\qquad\qquad = \{\omega\colon X(\omega) \in (R \times \ldots \times R \times B \times R \times \ldots \times R)\} \in \mathscr{F},\end{aligned}$$

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

$$\{\omega\colon X(\omega) \in (B_1 \times \ldots \times B_n)\} = \prod_{k=1}^{n} \{\omega\colon \xi_k(\omega) \in B_k\} \in \mathscr{F}.$$

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 § 4 сразу получаем, что для любого $B \in \mathscr{B}(R^n)$ множество $\{\omega\colon X(\omega) \in B\}$ принадлежит $\mathscr{F}$.

Пусть $(E, \mathscr{E}) = (\mathbf{Z}, \mathscr{B}(\mathbf{Z}))$, где $\mathbf{Z}$ — множество комплексных чисел $z = x + iy$, $x, y \in R$, а $\mathscr{B}(\mathbf{Z})$ — наименьшая σ-алгебра, содержащая множества вида $\{z\colon z = x + iy,\ a_1 < x \leqslant b_1,\ a_2 < y \leqslant b_2\}$. 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* (*стохастическим*) *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 $T = [0, 1]$, $(-\infty, \infty)$, $[0, \infty)$, ..., то $X = (\xi_t)_{t\in T}$ 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 (случайным элементом со значениями в 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

$$P_X(B) = \quad \{\omega\colon X(\omega) \in B\}, \qquad B \in \mathscr{B}(R^T),$$

называется *распределением вероятностей процесса* X. Вероятности

$$P_{t_1,\ldots,t_n}(B) \equiv \quad \{\omega\colon (\xi_{t_1}, \ldots, \xi_{t_n}) \in B\}, \qquad B \in \mathscr{B}(R^n),$$

с $t_1 < t_2 < \ldots < t_n$, $t_i \in T$, называются *конечномерными вероятностями* (или *распределениями вероятностей*). Функции

$$F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) \equiv \quad \{\omega\colon \xi_{t_1} \leqslant x_1, \ldots, \xi_{t_n} \leqslant x_n\}$$

с $t_1 < t_2 < \ldots < t_n$, $t_i \in T$, называются *конечномерными функциями распределения* процесса $X = (\xi_t)_{t\in T}$.

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 $A = \{x \in C : x_t < a\}$ 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 (17) § 2

$$\{x\in C:\ x\in S_\rho(x^0)\}=\bigcap_{t_k}\{x\in C:\ |x_{t_k}-x^0_{t_k}|<\rho\},$$

где t_k — рациональные точки отрезка $[0,\,1]$. Поэтому

$$\{\omega:\ X(\omega)\in S_\rho(X^0(\omega))\}=\bigcap_{t_k}\{\omega:\ |\xi_{t_k}(\omega)-\xi^0_{t_k}(\omega)|<\rho\}\in\mathscr{F},$$

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 $(D,\,\mathscr{B}_0(D))$, введенного в § 2 (п. 7), может рассматриваться как случайный процесс с траекториями из пространства функций без разрывов второго рода, и наоборот.

2. Пусть $(\Omega,\,\mathscr{F},\ \ )$ — 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.

$$\{X_{\alpha_1}\in B_{\alpha_1},\,\ldots,\,X_{\alpha_n}\in B_{\alpha_n}\}=\ \ \{X_{\alpha_1}\in B_{\alpha_1}\}\ldots\ \ \{X_{\alpha_n}\in B_{\alpha_n}\},\tag{3}$$

где $B_\alpha\in\mathscr{E}_\alpha$.

Пусть $\mathfrak{A}=\{1,\,2,\,\ldots,\,n\}$, ξ_α — случайные величины, $\alpha\in\mathfrak{A}$, и

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

— n-мерная функция распределения вектора $\xi=(\xi_1,\,\ldots,\,\xi_n)$. Пусть $F_{\xi_i}(x_i)$ есть функция распределения случайной величины ξ_i, $i=1,\,\ldots,\,n$.

Теорема 1. *Для того чтобы случайные величины $\xi_1,\,\ldots,\,\xi_n$ были независимы, необходимо и достаточно, чтобы для всех векторов $(x_1,\,\ldots,\,x_n)\in R^n$*

$$F_\xi(x_1,\,\ldots,\,x_n)=F_{\xi_1}(x_1)\ldots F_{\xi_n}(x_n).\tag{4}$$

Доказательство. Необходимость очевидна. Для доказательства достаточности положим $(a=(a_1,\,\ldots,\,a_n),\ b=(b_1,\,\ldots,\,b_n))$

$$P_\xi(a,\,b]=\ \ \{a_1<\xi_1\leqslant b_1,\,\ldots,\,a_n<\xi_n\leqslant b_n\},$$
$$P_{\xi_i}(a_i,\,b_i]=\ \ \{a_i<\xi_i\leqslant b_i\}.$$

Тогда в силу (7) § 3 и (4)

$$P_\xi(a,\, b] = \prod_{i=1}^{n} [F_{\xi_i}(b_i) - F_{\xi_i}(a_i)] = \prod_{i=1}^{n} P_{\xi_i}(a_i,\, b_i]$$

и, значит,

$$\{\xi_1 \in I_1,\, \ldots,\, \xi_n \in I_n\} = \prod_{i=1}^{n} \quad \{\xi_i \in I_i\}, \tag{5}$$

где $I_i = (a_i,\, b_i]$.

Зафиксируем $I_2,\, \ldots,\, I_n$ и покажем, что для любого $B_1 \in \mathscr{B}(R)$

$$\{\xi_1 \in B_1,\, \xi_2 \in I_2,\, \ldots,\, \xi_n \in I_n\} = \quad \{\xi_1 \in B_1\} \prod_{i=2}^{n} \quad \{\xi_i \in I_i\}. \tag{6}$$

Пусть $\mathscr{M}$ — совокупность множеств из $\mathscr{B}(R)$, для которых выполнено (6) («принцип подходящих множеств», § 2). 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)

$$\mu(\mathscr{M}) \subseteq \mathscr{M} \subseteq \mathscr{B}(R).$$

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

$$\{\xi_1 \in B_1,\, \ldots,\, \xi_n \in B_n\} = \quad \{\xi_1 \in B_1\} \ldots \quad \{\xi_n \in B_n\},$$

где $B_i \in \mathscr{B}(R)$. $\square$

3. Задачи.

1. Пусть $\xi_1,\, \ldots,\, \xi_n$ — дискретные случайные величины. Показать, что они независимы тогда и только тогда, когда для любых действительных чисел $x_1,\, \ldots,\, x_n$

$$\{\xi_1 = x_1,\, \ldots,\, \xi_n = x_n\} = \prod_{i=1}^{n} \quad \{\xi_i = x_i\}.$$

2. Провести доказательство того, что всякая случайная функция $X(\omega) = (\xi_t(\omega))_{t \in T}$ 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 $\mathscr{E}_1/\mathscr{E}_1'$, $\ldots$, $\mathscr{E}_n/\mathscr{E}_n'$-измеримыми функциями соответственно. Показать, что если $X_1, \ldots, X_n$ независимы, то независимы также и случайные элементы $g_1 \circ X_1, \ldots, g_n \circ X_n$.

4. Пусть $X_1, X_2, \ldots$ — бесконечная последовательность *перестановочных* случайных величин (т. е. таких, что совместное распределение каждой группы случайных величин, состоящей из k элементов с разными индексами, скажем, $X_{i_1}, \ldots, X_{i_k}$, зависит лишь от k и не зависит от конкретного выбора значений $i_1, \ldots, i_k$, где $i_1, \ldots, i_k$ попарно различны; ср. с определением в задаче 11 из § 1). Доказать, что если $X_n^2 < \infty$, $n \geqslant 1$, то ковариация $(X_1, X_2) \geqslant 0$.

5. Пусть ξ, η, ζ — независимые случайные величины. Доказать, что случайные величины $\xi + \eta$ и ζ^2 независимы.

6. Пусть $\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_n$ — случайные величины. Образуем случайные векторы $X = (\xi_1, \ldots, \xi_m)$ и $Y = (\eta_1, \ldots, \eta_n)$. Предположим, что выполнены следующие условия:

(i) случайные величины $\xi_1, \ldots, \xi_m$ независимы;

(ii) случайные величины $\eta_1, \ldots, \eta_n$ независимы;

(iii) случайные векторы X и Y, рассматриваемые как случайные элементы со значениями в R^m и R^n соответственно, независимы.

Доказать, что случайные величины $\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_n$ независимы.

7. Даны случайные векторы $X = (\xi_1, \ldots, \xi_m)$ и $Y = (\eta_1, \ldots, \eta_n)$. Известно, что случайные величины $\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_n$ независимы.

(i) Доказать, что случайные векторы X и Y, рассматриваемые как случайные элементы, независимы (ср. с задачей 6).

(ii) Пусть $f\colon R^m \to R$, $g\colon R^n \to R$ — борелевские функции. Доказать, что случайные величины $f(\xi_1, \ldots, \xi_m)$ и $g(\eta_1, \ldots, \eta_n)$ независимы.

§ 6. Интеграл Лебега. Математическое ожидание

1. В том случае, когда $(\Omega, \mathscr{F}, \ )$ — *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 ξ было определено в § 4 гл. I. Та же самая конструкция математического ожидания ξ от *простых* случайных величин ξ используется и в случае *произвольного* вероятностного пространства $(\Omega, \mathscr{F}, \ )$. А именно, *по определению* полагается

$$\xi = \sum_{k=1}^{n} x_k \ \ (A_k). \tag{2}$$

Это определение корректно (в том смысле, что значение $\ \xi$ 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 $\ \xi$ *произвольной* случайной величины. С точки зрения анализа математическое ожидание $\ \xi$ есть не что иное, как интеграл Лебега от $\mathscr{F}$-измеримой функции $\xi = \xi(\omega)$ по мере $\ $, для которого (наряду с $\ \xi$) используются также следующие обозначения: $\int_\Omega \xi(\omega) \ \ (d\omega)$ или $\int_\Omega \xi \, d \ $.

2. Пусть $\xi = \xi(\omega)$ — *неотрицательная* случайная величина. Построим последовательность простых неотрицательных случайных величин $\{\xi_n\}_{n \geqslant 1}$ таких, что $\xi_n(\omega) \uparrow \xi(\omega)$, $n \to \infty$, для каждого $\omega \in \Omega$ (см. теорему 1 в § 4).

Поскольку $\ \xi_n \leqslant \ \xi_{n+1}$ (ср. со свойством 3) из п. 5 § 4 гл. I), то существует $\lim_n \ \xi_n$, который может принимать и значение $+\infty$.

Определение 1. *Интегралом Лебега* от неотрицательной случайной величины $\xi = \xi(\omega)$, или ее *математическим ожиданием*, называется величина

$$\xi \equiv \lim_n \ \ \xi_n. \tag{3}$$

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 неотрицательных функций, то

$$\lim_n \ \ \xi_n = \lim_m \ \ \eta_m. \tag{4}$$

Лемма 1. *Пусть η и ξ_n — простые неотрицательные случайные величины, $n \geqslant 1$, причем*

$$\xi_n \uparrow \xi \geqslant \eta.$$

Тогда

$$\lim_n \ \ \xi_n \geqslant \ \ \eta. \tag{5}$$

Доказательство. Пусть $\varepsilon > 0$ и

$$A_n = \{\omega \colon \ \xi_n \geqslant \eta - \varepsilon\}.$$

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

$$\xi_n \geqslant \quad (\eta - \varepsilon) I_{A_n} = \quad \eta I_{A_n} - \varepsilon \quad (A_n) =$$
$$= \quad \eta - \quad \eta I_{\bar{A}_n} - \varepsilon \quad (A_n) \geqslant \quad \eta - C \quad (\bar{A}_n) - \varepsilon,$$

где $C = \max\limits_{\omega} \eta(\omega)$. Отсюда в силу произвольности $\varepsilon > 0$ вытекает требуемое неравенство (5). □

Из этой леммы следует, что $\lim\limits_n \quad \xi_n \geqslant \lim\limits_m \quad \eta_m$, и, по симметрии, $\lim\limits_m \quad \eta_m \geqslant \lim\limits_n \quad \xi_n$, что и доказывает (4).

Часто оказывается полезным следующее

Замечание 1. Для математического ожидания ξ от неотрицательной случайной величины ξ имеет место следующее представление:

$$\xi = \sup_{\{s \in S : s \leqslant \xi\}} \quad s, \tag{6}$$

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 ξ случайной величины ξ *существует*, или *определено*, если по крайней мере одна из величин ξ^+ или ξ^- конечна:

$$\min(\quad \xi^+, \quad \xi^-) < \infty.$$

В этом случае по *определению* полагают

$$\xi \equiv \quad \xi^+ - \quad \xi^-.$$

Математическое ожидание ξ называют иначе *интегралом Лебега* от функции ξ по вероятностной мере . (По поводу других подходов к определению интеграла Лебега см. замечание 2 п. 11.)

Определение 3. Говорят, что *математическое ожидание* случайной величины ξ *конечно* (или ξ — *интегрируема*), если $\xi^+ < \infty$ и $\xi^- < \infty$.

Поскольку $|\xi| = \xi^+ + \xi^-$, то конечность ξ эквивалентна тому, что $|\xi| < \infty$. (В этом смысле интегрируемость по Лебегу носит «абсолютный» характер.)

Замечание 2. Наряду с математическим ожиданием ξ важными числовыми характеристиками случайной величины ξ являются величины ξ^r (если они определены) и $|\xi|^r$, $r > 0$, называемые соответственно *моментом* r-го порядка (r-м моментом) и *абсолютным моментом* r-го порядка (r-м абсолютным моментом) случайной величины ξ.

Замечание 3. В данном выше определении интеграла Лебега $\int\limits_\Omega \xi\, d$ предполагалось, что мера является *вероятностной* ($(\Omega) = 1$), 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 $\bar{R} = [-\infty, \infty]$ (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 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 λ. В этом случае интеграл $\int\limits_R \xi(x)\, \lambda(dx)$ 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—S) $\int\limits_R \xi(x)\, G(dx)$, чтобы отличить его от соответствующего интеграла Римана—Стилтьеса (R—S) $\int\limits_R \xi(x)\, G(dx)$ (см. далее п. 11).

Из дальнейшего (свойство **D**) станет ясно, что если ξ определено, то определены также математические ожидания (ξI_A) для любого $A \in \mathscr{F}$. Для (ξI_A), или, что то же, $\int\limits_\Omega \xi I_A\, d$, часто используются обозначения $(\xi; A)$ и $\int\limits_A \xi\, d$. Интеграл $\int\limits_A \xi\, d$ принято называть *интегралом Лебега от* ξ *по мере* *на множестве* A.

Аналогично и в случае произвольной меры μ вместо $\int\limits_\Omega \xi I_A \, d\mu$ 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 $dx_1 \ldots dx_n$.

3. Свойства математического ожидания $\ \ \xi$ случайных величин ξ.

A. *Пусть c — постоянная и* $\ \ \xi$ *существует. Тогда* $\ \ (c\xi)$ *также существует и*

$$\ \ (c\xi) = c \ \ \xi.$$

B. *Пусть $\xi \leqslant \eta$, тогда*

$$\ \ \xi \leqslant \ \ \eta$$

в том смысле, что

$$\text{если } -\infty < \ \ \xi, \text{ то } -\infty < \ \ \eta \text{ и } \ \ \xi \leqslant \ \ \eta,$$

или

$$\text{если } \ \ \eta < \infty, \text{ то } \ \ \xi < \infty \text{ и } \ \ \xi \leqslant \ \ \eta.$$

C. *Если* $\ \ \xi$ *существует, то*

$$|\ \ \xi| \leqslant \ \ |\xi|.$$

D. *Если* $\ \ \xi$ *существует, то для каждого $A \in \mathscr{F}$* $\ \ (\xi I_A)$ *также существует; если* $\ \ \xi$ *конечно, то* $\ \ (\xi I_A)$ *также конечно.*

E. *Если ξ и η — неотрицательные случайные величины или такие, что* $\ \ |\xi| < \infty$, $\ \ |\eta| < \infty$, *то*

$$\ \ (\xi + \eta) = \ \ \xi + \ \ \eta.$$

(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,

$$\ \ (c\xi) = \lim_n \ \ (c\xi_n) = c \lim_n \ \ \xi_n = c \ \ \xi.$$

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 ξ и η определены и неравенство $\xi \leqslant \eta$ сразу следует из формулы (6). Пусть теперь $\xi > -\infty$, тогда $\xi^- < \infty$. Если $\xi \leqslant \eta$, то $\xi^+ \leqslant \eta^+$ и $\xi^- \geqslant \eta^-$. Поэтому $\eta^- \leqslant \xi^- < \infty$, следовательно, η определено и $\xi = \xi^+ - \xi^- \leqslant \eta^+ - \eta^- = \eta$. Аналогичным образом рассматривается случай, когда $\eta < \infty$.

C. Поскольку $-|\xi| \leqslant \xi \leqslant |\xi|$, то из свойств **A** и **B**

$$- |\xi| \leqslant \xi \leqslant |\xi|,$$

т. е. $|\xi| \leqslant |\xi|$.

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 $(\xi_n + \eta_n) = \xi_n + \eta_n$ и $(\xi_n + \eta_n) \uparrow (\xi + \eta)$, $\xi_n \uparrow \xi$, $\eta_n \uparrow \eta$ и, значит, $(\xi + \eta) = \xi + \eta$. Случай, когда $|\xi| < \infty$, $|\eta| < \infty$, 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 « -почти наверное». Будем говорить, что *некоторое свойство выполнено « -почти наверное», если существует множество $\mathscr{N} \in \mathscr{F}$ с $(\mathscr{N}) = 0$ такое, что это свойство выполнено для каждой точки $\omega \in \Omega \setminus \mathscr{N}$.* Вместо слов « -почти наверное» (-п. н.) часто говорят « -почти всюду» (-п. в.) или просто «почти наверное» (п. н.), «почти всюду» (п. в.).

F. *Если $\xi = 0$ (п. н.), то $\xi = 0$.*

В самом деле, если ξ — простая случайная величина, $\xi = \sum x_k I_{A_k}(\omega)$ и $x_k \neq 0$, то по условию $(A_k) = 0$, а значит, $\xi = 0$. Если же $\xi \geqslant 0$ и $0 \leqslant s \leqslant \xi$, где s — простая случайная величина, то $s = 0$ (п. н.), а следовательно, $s = 0$ и $\xi = \sup\limits_{\{s \in S:\, s \leqslant \xi\}} s = 0$. 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 $|\xi| < \infty$, то $|\eta| < \infty$ и $\xi = \eta$ (см. также задачу 3).*

В самом деле, пусть $\mathscr{N} = \{\omega : \xi \neq \eta\}$. Тогда $(\mathscr{N}) = 0$ и $\xi = \xi I_{\mathscr{N}} + \xi I_{\bar{\mathscr{N}}}$, $\eta = \eta I_{\mathscr{N}} + \xi I_{\bar{\mathscr{N}}}$. По свойствам **E** и **F** $\xi = \xi I_{\mathscr{N}} + \xi I_{\bar{\mathscr{N}}} = \xi I_{\bar{\mathscr{N}}} = \eta I_{\bar{\mathscr{N}}}$. Но $\eta I_{\mathscr{N}} = 0$, поэтому по свойству **E** $\xi = \eta I_{\bar{\mathscr{N}}} + \eta I_{\mathscr{N}} = \eta$.

H. *Пусть $\xi \geqslant 0$ и $\xi = 0$. 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 $0 \leqslant \xi I_{A_n} \leqslant \xi I_A$. Поэтому по свойству **B**

$$0 \leqslant \quad \xi I_{A_n} \leqslant \quad \xi = 0.$$

Следовательно,

$$0 = \quad \xi I_{A_n} \geqslant \frac{1}{n} \quad (A_n)$$

и, значит, $\quad (A_n) = 0$ для всех $n \geqslant 1$. Но $\quad (A) = \lim \quad (A_n)$ и, следовательно, $\quad (A) = 0$.

I. *Пусть* ξ *и* η *таковы, что* $\quad |\xi| < \infty$, $\quad |\eta| < \infty$ *и для всех* $A \in \mathscr{F}$ $\quad (\xi I_A) \leqslant \quad (\eta I_A)$. *Тогда* $\xi \leqslant \eta$ *(п. н.).*

В самом деле, пусть $B = \{\omega \colon \xi(\omega) > \eta(\omega)\}$. Тогда $\quad (\eta I_B) \leqslant \quad (\xi I_B) \leqslant \quad (\eta I_B)$ и, значит, $\quad (\xi I_B) = \quad (\eta I_B)$. В силу свойства **E** $\quad ((\xi - \eta) I_B) = 0$ и по свойству **H** $(\xi - \eta) I_B = 0$ (п. н.), откуда $\quad (B) = 0$.

J. *Пусть* ξ *— расширенная случайная величина и* $\quad |\xi| < \infty$. *Тогда* $|\xi| < \infty$ *(п. н.).*

Действительно, пусть $A = \{\omega \colon |\xi(\omega)| = \infty\}$ и $\quad (A) > 0$. Тогда $\quad |\xi| \geqslant \quad (|\xi| I_A) = \infty \cdot \quad (A) = \infty$, что противоречит предположению $\quad |\xi| < \infty$. (См. также задачу 4.)

4. 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* $n \geqslant 1$, $\quad \eta > -\infty$ *и* $\xi_n \uparrow \xi$, *то*

$$\xi_n \uparrow \quad \xi.$$

b) *Если* $\xi_n \leqslant \eta$ *для всех* $n \geqslant 1$, $\quad \eta < \infty$ *и* $\xi_n \downarrow \xi$, *то*

$$\xi_n \downarrow \quad \xi.$$

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

$$\xi = \quad \zeta = \lim \quad \zeta^{(n)} \leqslant \lim \quad \xi_n.$$

С другой стороны, очевидно, что поскольку $\xi_n \leqslant \xi_{n+1} \leqslant \xi$, то

$$\lim \quad \xi_n \leqslant \quad \xi.$$

Тем самым $\lim \quad \xi_n = \quad \xi$.

Пусть теперь η — произвольная случайная величина с $\quad \eta > -\infty$.

Если $\quad \eta = \infty$, то в силу **B** $\quad \xi_n = \quad \xi = \infty$ и утверждение доказано. Пусть $\quad \eta < \infty$. Тогда, учитывая сделанное предположение $\quad \eta > -\infty$, получаем, что $\quad |\eta| < \infty$. Ясно, что $0 \leqslant \xi_n - \eta \uparrow \xi - \eta$ для всех $\omega \in \Omega$. Поэтому, согласно доказанному, $\quad (\xi_n - \eta) \uparrow \quad (\xi - \eta)$ и, значит (по свойству **E** и задаче 2),

$$\xi_n - \quad \eta \uparrow \quad \xi - \quad \eta.$$

Но $\quad |\eta| < \infty$, поэтому $\quad \xi_n \uparrow \quad \xi$, $n \to \infty$.

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*

$$\sum_{n=1}^{\infty} \eta_n = \sum_{n=1}^{\infty} \quad \eta_n.$$

Доказательство следует из свойства **E** (см. также задачу 2), теоремы о монотонной сходимости и того замечания, что $\sum_{n=1}^{k} \eta_n \uparrow \sum_{n=1}^{\infty} \eta_n$, $k \to \infty$. □

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 $\quad \eta > -\infty$, то*

$$\underline{\lim} \ \xi_n \leqslant \underline{\lim} \quad \xi_n.$$

b) *Если $\xi_n \leqslant \eta$ для всех $n \geqslant 1$ и $\quad \eta < \infty$, то*

$$\overline{\lim} \quad \xi_n \leqslant \quad \overline{\lim} \ \xi_n.$$

c) *Если $|\xi_n| \leqslant \eta$ для всех $n \geqslant 1$ и $\quad \eta < \infty$, то*

$$\underline{\lim} \ \xi_n \leqslant \underline{\lim} \quad \xi_n \leqslant \overline{\lim} \quad \xi_n \leqslant \quad \overline{\lim} \ \xi_n. \tag{7}$$

Доказательство. a) Пусть $\zeta_n = \inf_{m \geqslant n} \xi_m$, 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

$$\underline{\lim}\, \xi_n = \lim_n \zeta_n = \lim_n \zeta_n = \underline{\lim}_n \zeta_n \leqslant \underline{\lim}\, \xi_n,$$

что и доказывает утверждение а). Второе утверждение следует из первого. Третье — есть следствие первых двух. □

Теорема 3 (теорема Лебега о мажорируемой сходимости)**.** *Пусть случайные величины таковы, что* $|\xi_n| \leqslant \eta$, $\eta < \infty$ *и* $\xi_n \to \xi$ *(п. н.). Тогда* $|\xi| < \infty$ *и*

$$\xi_n \to \xi, \quad n \to \infty, \tag{8}$$

и

$$|\xi_n - \xi| \to 0, \quad n \to \infty. \tag{9}$$

Доказательство. По предположению $\underline{\lim}\, \xi_n = \overline{\lim}\, \xi_n = \xi$ (п. н.). Поэтому в силу свойства **G** и леммы Фату (утверждение с))

$$\xi = \underline{\lim}\, \xi_n \leqslant \underline{\lim}\, \xi_n = \overline{\lim}\, \xi_n = \overline{\lim}\, \xi_n = \xi,$$

что и доказывает (8). Ясно также, что $|\xi| \leqslant \eta$. Поэтому $|\xi| < \infty$.

Утверждение (9) доказывается так же, если только заметить, что $|\xi_n - \xi| \leqslant 2\eta$. □

Следствие. *Пусть* $\eta, \xi, \xi_1, \xi_2, \ldots$ *— случайные величины такие, что* $|\xi_n| \leqslant \eta$, $\xi_n \to \xi$ *(п. н.) и* $\eta^p < \infty$ *для некоторого* $p > 0$. *Тогда* $|\xi|^p < \infty$ *и* $|\xi - \xi_n|^p \to 0$, $n \to \infty$.

Для доказательства достаточно заметить, что $|\xi| \leqslant \eta$ и $|\xi - \xi_n|^p \leqslant (|\xi| + |\xi_n|)^p \leqslant (2\eta)^p$.

Условие «$|\xi_n| \leqslant \eta$, $\eta < \infty$», входящее в лемму Фату и 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* (по мере), если

$$\sup_n \int_{\{|\xi_n| > c\}} |\xi_n| \,(d\omega) \to 0, \quad c \to \infty, \tag{10}$$

или (в других обозначениях)

$$\sup_n [|\xi_n| I_{\{|\xi_n| > c\}}] \to 0, \quad c \to \infty. \tag{11}$$

Ясно, что если случайные величины ξ_n, $n \geqslant 1$, таковы, что $|\xi_n| \leqslant \eta$, $\eta < \infty$, 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*

$$\underline{\lim}\,\xi_n \leqslant \underline{\lim}\quad \xi_n \leqslant \overline{\lim}\quad \xi_n \leqslant \quad \overline{\lim}\,\xi_n.$$

b) *Если к тому же* $\xi_n \to \xi$ (*п. н.*), *то случайная величина* ξ *интегрируема и*

$$\xi_n \to \quad \xi, \qquad n \to \infty,$$
$$|\xi_n - \xi| \to 0, \qquad n \to \infty.$$

Доказательство. a) Для всякого $c > 0$

$$\xi_n = \quad \xi_n I_{\{\xi_n < -c\}} + \quad \xi_n I_{\{\xi_n \geqslant -c\}}. \tag{12}$$

В силу равномерной интегрируемости для всякого $\varepsilon > 0$ можно выбрать c столь большим, что

$$\sup_n |\quad \xi_n I_{\{\xi_n < -c\}}| < \varepsilon. \tag{13}$$

В силу леммы Фату

$$\underline{\lim}\quad \xi_n I_{\{\xi_n \geqslant -c\}} \geqslant \quad \underline{\lim}\,\xi_n I_{\{\xi_n \geqslant -c\}}.$$

Но $\xi_n I_{\{\xi_n \geqslant -c\}} \geqslant \xi_n$, поэтому

$$\underline{\lim}\quad \xi_n I_{\{\xi_n \geqslant -c\}} \geqslant \quad \underline{\lim}\,\xi_n. \tag{14}$$

Из (12)—(14) находим, что

$$\underline{\lim}\quad \xi_n \geqslant \quad \underline{\lim}\,\xi_n - \varepsilon.$$

В силу произвольности $\varepsilon > 0$ отсюда следует, что $\underline{\lim}\quad \xi_n \geqslant \quad \underline{\lim}\,\xi_n$.

Аналогичным образом доказывается, что $\overline{\lim}\quad \xi_n \leqslant \quad \overline{\lim}\,\xi_n$.

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* $0 \leqslant \xi_n \to \xi$ (*-п. н.) и* $\xi_n < \infty$. *Тогда* $\xi_n \to \quad \xi < \infty$ *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 $A = \{a\colon \quad \{\xi = a\} > 0\}$. Тогда $\xi_n I_{\{\xi_n < a\}} \to \xi I_{\{\xi < a\}}$ для каждого $a \notin A$, причем семейство величин $\{\xi_n I_{\{\xi_n < a\}}\}_{n \geqslant 1}$ будет равномерно интегрируемым. Поэтому в силу «достаточности» $\quad \xi_n I_{\{\xi_n < a\}} \to \quad \xi I_{\{\xi < a\}}$, $a \notin A$, а значит,

$$\xi_n I_{\{\xi_n \geqslant a\}} \to \quad \xi I_{\{\xi \geqslant a\}}, \qquad a \notin A, \ \ n \to \infty. \tag{15}$$

Зафиксируем $\varepsilon > 0$ и выберем сначала $a_0 \notin A$ столь большим, что $\xi I_{\{\xi \geqslant a_0\}} < \varepsilon/2$, а затем N_0 таким, что для всех $n \geqslant N_0$

$$\xi_n I_{\{\xi_n \geqslant a_0\}} \leqslant \quad \xi I_{\{\xi \geqslant a_0\}} + \varepsilon/2,$$

и, значит, $\xi_n I_{\{\xi_n \geqslant a_0\}} \leqslant \varepsilon$. Выберем, наконец, $a_1 \geqslant a_0$ столь большим, что для всех $n \leqslant N_0$ $\quad \xi_n I_{\{\xi_n \geqslant a_1\}} \leqslant \varepsilon$. Тогда

$$\sup_n \quad \xi_n I_{\{\xi_n \geqslant a_1\}} \leqslant \varepsilon,$$

что и доказывает равномерную интегрируемость семейства случайных величин $\{\xi_n\}_{n \geqslant 1}$. □

5. Остановимся на некоторых критериях равномерной интегрируемости.

Прежде всего заметим, что если $\{\xi_n\}_{n \geqslant 1}$ — семейство равномерно интегрируемых случайных величин, то

$$\sup_n \quad |\xi_n| < \infty. \tag{16}$$

В самом деле, для фиксированного $\varepsilon > 0$ и достаточно больших $c > 0$

$$\sup_n \quad |\xi_n| = \sup_n \, [\quad |\xi_n| I_{\{|\xi_n| \geqslant c\}} + \quad |\xi_n| I_{\{|\xi_n| < c\}}] \leqslant$$
$$\leqslant \sup_n \quad |\xi_n| I_{\{|\xi_n| \geqslant c\}} + \sup_n \quad |\xi_n| I_{(|\xi_n| < c\}} \leqslant \varepsilon + c,$$

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* $|\xi_n|$, $n \geqslant 1$, *были равномерно ограничены* (*т. е. было выполнено условие* (16)) *и чтобы* $\{|\xi_n| I_A\}$, $n \geqslant 1$, *были равномерно непрерывны* (*т. е.* $\sup_n \quad \{|\xi_n| I_A\} \to 0$, *когда* $\quad (A) \to 0$).

Доказательство. Необходимость. Условие (16) было проверено выше. Далее,

$$\{|\xi_n| I_A\} = \quad \{|\xi_n| I_{A \cap \{|\xi_n| \geqslant c\}}\} + \quad \{|\xi_n| I_{A \cap \{|\xi_n| < c\}}\} \leqslant \quad \{|\xi_n| I_{\{|\xi_n| \geqslant c\}}\} + c \quad (A). \tag{17}$$

Выберем c столь большим, что $\sup_n \quad \{|\xi_n| I_{\{|\xi_n| \geqslant c\}}\} \leqslant \varepsilon/2$. Тогда если $\quad (A) \leqslant \varepsilon/2c$, то из (17)

$$\sup_n \quad \{|\xi_n| I_A\} \leqslant \varepsilon,$$

что и доказывает равномерную непрерывность.

Достаточность. Пусть $\varepsilon > 0$ и $\delta > 0$ таково, что из условия $\ (A) < \delta$ следует, что равномерно по n $\ (|\xi_n| I_A) \leqslant \varepsilon$. Поскольку для всякого $c > 0$

$$|\xi_n| \geqslant \ \ |\xi_n| I_{\{|\xi_n| \geqslant c\}} \geqslant c \ \ \{|\xi_n| \geqslant c\}$$

(ср. с неравенством Чебышева), то

$$\sup_n \ \ \{|\xi_n| \geqslant c\} \leqslant \frac{1}{c} \sup \ \ |\xi_n| \to 0, \quad c \to \infty,$$

а значит, для достаточно больших c в качестве множества A можно взять любое из множеств $\{|\xi_n| \geqslant c\}$, $n \geqslant 1$. Поэтому $\sup \ \ (|\xi_n| I_{\{|\xi_n| \geqslant c\}}) \leqslant \varepsilon$, что и доказывает равномерную интегрируемость. □

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*

$$\lim_{t \to \infty} \frac{G(t)}{t} = \infty, \tag{18}$$

$$\sup_n \ \ G(|\xi_n|) < \infty. \tag{19}$$

Тогда семейство случайных величин $\{\xi_n\}_{n \geqslant 1}$ *является равномерно интегрируемым.*

Доказательство. Пусть $\varepsilon > 0$, $M = \sup_n \ \ G(|\xi_n|)$, $a = \dfrac{M}{\varepsilon}$. Выберем c столь большим, что $\dfrac{G(t)}{t} \geqslant a$ для $t \geqslant c$. Тогда

$$[|\xi_n| I_{\{|\xi_n| \geqslant c\}}] \leqslant \frac{1}{a} \ \ [G(|\xi_n|) I_{\{|\xi_n| \geqslant c\}}] \leqslant \frac{M}{a} = \varepsilon$$

равномерно по всем $n \geqslant 1$. □

6. Если ξ и η — независимые *простые* случайные величины, то, как и в п. 5 § 4 гл. I, доказывается, что $\ \xi\eta = \ \ \xi \cdot \ \ \eta$. Установим теперь справедливость аналогичного утверждения в общем случае (см. также задачу 6).

Теорема 6. *Пусть* ξ *и* η — *независимые случайные величины с* $\ |\xi| < \infty$, $\ |\eta| < \infty$. *Тогда* $\ |\xi\eta| < \infty$ *и*

$$\xi\eta = \ \ \xi \cdot \ \ \eta. \tag{20}$$

Доказательство. Пусть сначала $\xi \geqslant 0$, $\eta \geqslant 0$. Положим

$$\xi_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{\left\{\frac{k}{n} \leqslant \xi < \frac{k+1}{n}\right\}}, \qquad \eta_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{\left\{\frac{k}{n} \leqslant \eta < \frac{k+1}{n}\right\}}.$$

Тогда $\xi_n \leqslant \xi$, $|\xi_n - \xi| \leqslant 1/n$ и $\eta_n \leqslant \eta$, $|\eta_n - \eta| \leqslant 1/n$. Поскольку $\xi < \infty$, $\eta < \infty$, то по теореме Лебега о мажорируемой сходимости

$$\lim \ \xi_n = \ \xi, \quad \lim \ \eta_n = \ \eta.$$

Далее, в силу независимости ξ и η

$$\xi_n\eta_n = \sum_{k,l \geqslant 0} \frac{kl}{n^2} \ I_{\left\{\frac{k}{n} \leqslant \xi < \frac{k+1}{n}\right\}} I_{\left\{\frac{l}{n} \leqslant \eta < \frac{l+1}{n}\right\}} =$$

$$= \sum_{k,l \geqslant 0} \frac{kl}{n^2} \ I_{\left\{\frac{k}{n} \leqslant \xi < \frac{k+1}{n}\right\}} \cdot \ I_{\left\{\frac{l}{n} \leqslant \eta < \frac{l+1}{n}\right\}} = \ \xi_n \cdot \ \eta_n.$$

Заметим теперь, что

$$| \ \xi\eta - \ \xi_n\eta_n| \leqslant \ |\xi\eta - \xi_n\eta_n| \leqslant \ [\xi|\eta - \eta_n|] + \ [\eta_n|\xi - \xi_n|] \leqslant$$

$$\leqslant \frac{1}{n} \ \xi + \frac{1}{n} \ \left(\eta + \frac{1}{n}\right) \to 0, \quad n \to \infty.$$

Поэтому $\xi\eta = \lim\limits_n \ \xi_n\eta_n = \lim \ \xi_n \cdot \lim \ \eta_n = \ \xi \cdot \ \eta$, причем $\xi\eta < \infty$.

Общий случай сводится к рассмотренному, если воспользоваться представлениями $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$, $\xi\eta = \xi^+\eta^+ - \xi^-\eta^+ - \xi^+\eta^- + \xi^-\eta^-$. □

7. Приводимые в этом пункте неравенства для математических ожиданий (многие уже рассматривались в элементарной теории вероятностей; §§ 4 и 5 в гл. I) систематически применяются и в теории вероятностей, и в математическом анализе.

Неравенство Чебышева. *Пусть ξ — неотрицательная случайная величина, тогда для всякого $\varepsilon > 0$*

$$\{\xi \geqslant \varepsilon\} \leqslant \frac{\ \xi}{\varepsilon}. \tag{21}$$

Доказательство сразу следует из того, что

$$\ \xi \geqslant \ [\xi I_{\{\xi \geqslant \varepsilon\}}] \geqslant \varepsilon \ I_{\{\xi \geqslant \varepsilon\}} = \varepsilon \ \{\xi \geqslant \varepsilon\}.$$

Из (21) получаем следующие разновидности неравенства Чебышева: *если ξ — произвольная случайная величина, то*

$$\{\xi \geqslant \varepsilon\} \leqslant \frac{\ \xi^2}{\varepsilon^2} \tag{22}$$

и

$$\{|\xi - \ \xi| \geqslant \varepsilon\} \leqslant \frac{\ \xi}{\varepsilon^2}, \tag{23}$$

где $\ \xi = \ (\xi - \ \xi)^2$ — *дисперсия случайной величины* ξ.

Неравенство Коши—Буняковского. *Пусть случайные величины* ξ *и* η *таковы, что* $\ \xi^2<\infty$, $\ \eta^2<\infty$. *Тогда* $\ |\xi\eta|<\infty$ *и*

$$(\ |\xi\eta|)^2 \leqslant \ \xi^2\cdot \ \eta^2. \tag{24}$$

Доказательство. Будем предполагать, что $\ \xi^2>0$, $\ \eta^2>0$. Тогда, обозначая $\tilde{\xi}=\dfrac{\xi}{\sqrt{\ \xi^2}}$, $\tilde{\eta}=\dfrac{\eta}{\sqrt{\ \eta^2}}$, находим, что поскольку $2|\tilde{\xi}\tilde{\eta}|\leqslant\tilde{\xi}^2+\tilde{\eta}^2$, то

$$2\ |\tilde{\xi}\tilde{\eta}| \leqslant \ \tilde{\xi}^2+\ \tilde{\eta}^2=2,$$

т. е. $\ |\tilde{\xi}\tilde{\eta}|\leqslant 1$, что и доказывает (24).

Если же, скажем, $\ \xi^2\equiv 0$, то тогда по свойству **I** $\xi=0$ (п. н.) и по свойству **F** $\ \xi\eta=0$, т. е. (24) также выполнено. □

Неравенство Иенсена. *Пусть* $g=g(x)$ — *выпуклая книзу борелевская функция, определенная на* R *и* ξ — *случайная величина такая, что* $\ |\xi|<\infty$. *Тогда*

$$g(\ \xi) \leqslant \ g(\xi). \tag{25}$$

Доказательство. Если функция $g=g(x)$ выпукла книзу, то для каждого $x_0\in R$ найдется число $\lambda(x_0)$ такое, что для всех $x\in R$

$$g(x)\geqslant g(x_0)+(x-x_0)\lambda(x_0). \tag{26}$$

Полагая $x=\xi$ и $x_0=\ \xi$, из (26) находим, что

$$g(\xi)\geqslant g(\ \xi)+(\xi-\ \xi)\lambda(\ \xi)$$

и, следовательно, $\ g(\xi)\geqslant g(\ \xi)$. □

Замечание. Неравенство Иенсена (25) справедливо и для векторных случайных величин $\xi=(\xi_1,\ldots,\xi_d)$ с $\ |\xi_i|<\infty$, $i=1,\ldots,d$ и выпуклых книзу функций $g=g(x)$, $x=R^d$ (т. е. функций $g\colon R^d\to R$ таких, что $g(px+(1-p)y)\leqslant p\,g(x)+(1-p)\,g(y)$, $x,\ y\in R^d$, $p\in[0,\,1]$).

Из неравенства Иенсена выводится целая серия полезных неравенств. Получим, к примеру,

Неравенство Ляпунова. *Если* $0<s<t$, *то*

$$(\ |\xi|^s)^{1/s} \leqslant (\ |\xi|^t)^{1/t}. \tag{27}$$

Для доказательства обозначим $r=t/s$. Тогда, полагая $\eta=|\xi|^s$ и применяя неравенство Иенсена к функции $g(x)=|x|^r$, находим, что $|\ \eta|^r\leqslant\ |\eta|^r$, т. е.

$$(\ |\xi|^s)^{t/s}\leqslant\ |\xi|^t,$$

что и доказывает (27).

Из неравенства Ляпунова вытекает следующая цепочка *неравенств между абсолютными моментами*:

$$|\xi| \leqslant (\ \ |\xi|^2)^{1/2} \leqslant \ldots \leqslant (\ \ |\xi|^n)^{1/n}. \tag{28}$$

Неравенство Гёльдера. *Пусть* $1<p<\infty$, $1<q<\infty$ *и* $\dfrac{1}{p}+\dfrac{1}{q}=1$. *Если* $\ \ |\xi|^p<\infty$, $\ \ |\eta|^q<\infty$, *то* $\ \ |\xi\eta|<\infty$ *и*

$$|\xi\eta| \leqslant (\ \ |\xi|^p)^{1/p}(\ \ |\eta|^q)^{1/q}. \tag{29}$$

Доказательство. Если $\ \ |\xi|^p=0$ или $\ \ |\eta|^q=0$, то (29) следует немедленно, так же как и в случае неравенства Коши—Буняковского (являющегося частным случаем неравенства Гёльдера при $p=q=2$).

Пусть теперь $\ \ |\xi|^p>0$, $\ \ |\eta|^q>0$. Положим

$$\tilde{\xi}=\frac{|\xi|}{(\ \ |\xi|^p)^{1/p}}, \qquad \tilde{\eta}=\frac{|\eta|}{(\ \ |\eta|^q)^{1/q}}.$$

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 $x=\tilde{\xi}^p$, $y=\tilde{\eta}^q$, $a=\dfrac{1}{p}$, $b=\dfrac{1}{q}$, находим, что

$$\tilde{\xi}\tilde{\eta} \leqslant \frac{1}{p}\,\tilde{\xi}^p+\frac{1}{q}\,\tilde{\eta}^q,$$

откуда

$$\tilde{\xi}\tilde{\eta} \leqslant \frac{1}{p}\quad \tilde{\xi}^p+\frac{1}{q}\quad \tilde{\eta}^q=\frac{1}{p}+\frac{1}{q}=1,$$

что и доказывает (29). □

Неравенство Минковского. *Если* $\ \ |\xi|^p<\infty$, $\ \ |\eta|^p<\infty$, $1\leqslant p<\infty$, *то* $\ \ |\xi+\eta|^p<\infty$ *и*

$$(\ \ |\xi+\eta|^p)^{1/p} \leqslant (\ \ |\xi|^p)^{1/p}+(\ \ |\eta|^p)^{1/p}. \tag{31}$$

Доказательство. 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 $F(x)=(a+x)^p-2^{p-1}(a^p+x^p)$. 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 $\ |\xi|^p < \infty$, $\ |\eta|^p < \infty$, то $\ |\xi + \eta|^p < \infty$.

Если $p = 1$, то неравенство (31) следует из (33).

Будем теперь предполагать, что $p > 1$. Возьмем $q > 1$ таким, что $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Тогда

$$|\xi + \eta|^p = |\xi + \eta| \cdot |\xi + \eta|^{p-1} \leqslant |\xi| \cdot |\xi + \eta|^{p-1} + |\eta| \cdot |\xi + \eta|^{p-1}. \tag{34}$$

Заметим, что $(p-1)q = p$. Поэтому

$$(|\xi + \eta|^{p-1})^q = \ |\xi + \eta|^p < \infty,$$

и, значит, в силу неравенства Гёльдера

$$(|\xi| \cdot |\xi + \eta|^{p-1}) \leqslant (\ |\xi|^p)^{1/p} (\ |\xi + \eta|^{(p-1)q})^{1/q} = (\ |\xi|^p)^{1/p} (\ |\xi + \eta|^p)^{1/q}.$$

Точно так же и

$$(|\eta| \cdot |\xi + \eta|^{p-1}) \leqslant (\ |\eta|^p)^{1/p} (\ |\xi + \eta|^p)^{1/q}.$$

Поэтому в силу (34)

$$|\xi + \eta|^p \leqslant (\ |\xi + \eta|^p)^{1/q} [(\ |\xi|^p)^{1/p} + (\ |\eta|^p)^{1/p}]. \tag{35}$$

Если $\ |\xi + \eta|^p = 0$, то требуемое неравенство (31) очевидно. Пусть теперь $\ |\xi + \eta|^p > 0$. Тогда из (35) находим

$$(\ |\xi + \eta|^p)^{1 - 1/q} \leqslant (\ |\xi|^p)^{1/p} + (\ |\eta|^p)^{1/p},$$

что и дает требуемое неравенство (31), поскольку $1 - \dfrac{1}{q} = \dfrac{1}{p}$. ☐

8. Пусть ξ — случайная величина, для которой определено математическое ожидание $\ \xi$. Тогда, согласно свойству **D**, определена *функция множеств*

$$(A) \equiv \int_A \xi \, d\ , \qquad A \in \mathscr{F}. \tag{36}$$

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

$$(A) = \quad (\xi I_A) = \quad (\xi I_{\sum A_n}) = \quad \Bigl(\sum \xi I_{A_n}\Bigr) = \sum \quad (\xi I_{A_n}) = \sum \quad (A_n).$$

Если же ξ — произвольная случайная величина, для которой ξ определено, то счетная аддитивность (A) следует из представления

$$(A) = \quad {}^{+}(A) - \quad {}^{-}(A), \tag{37}$$

где

$$ {}^{+}(A) = \int_A \xi^{+}\, d \quad , \qquad {}^{-}(A) = \int_A \xi^{-}\, d \quad ,$$

установленной счетной аддитивности для неотрицательных случайных величин и того факта, что $\min(\ {}^{+}(\Omega), \ {}^{-}(\Omega)) < \infty$.

Итак, если ξ определено, то функция множеств $\ = \ (A)$ является *мерой со знаком* — счетно-аддитивной функцией множеств, представимой в виде $\ = \ _1 - \ _2$, где по крайней мере одна из мер $_1$ или $_2$ конечна.

Покажем, что функция множеств $\ = \ (A)$ обладает следующим важным свойством *абсолютной непрерывности* относительно меры :

$$\textit{если} \quad (A) = 0, \ \textit{то} \quad (A) = 0 \quad (A \in \mathscr{F})$$

(это свойство кратко записывают в виде: $\ \ll \ $).

Для доказательства достаточно рассмотреть случай неотрицательных случайных величин. Если $\xi = \sum\limits_{k=1}^{n} x_k I_{A_k}$ — простая неотрицательная случайная величина и $(A) = 0$, то

$$(A) = \quad (\xi I_A) = \sum_{k=1}^{n} x_k \quad (A_k \cap A) = 0.$$

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

$$(A) = \quad (\xi I_A) = \lim \quad (\xi_n I_A) = 0,$$

поскольку $(\xi_n I_A) = 0$ для любого $n \geqslant 1$ и A такого, что $(A) = 0$.

Итак, интеграл Лебега $(A) = \int_A \xi\, d \quad$, рассматриваемый как функция множеств $A \in \mathscr{F}$, является *мерой со знаком*, абсолютно непрерывной относительно меры $\quad (\ \ll \ )$. 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* $\overline{R} = [-\infty, \infty]$, *такая, что*

$$\lambda(A) = \int_A f(\omega)\,\mu(d\omega), \qquad A \in \mathscr{F}. \tag{38}$$

С точностью до множеств μ-*меры нуль функция* $f(\omega)$ *единственна*: *если* $h = h(\omega)$ — *другая* $\mathscr{F}$-*измеримая функция такая, что* $\lambda(A) = \int_A h(\omega)\,\mu(d\omega)$, $A \in \mathscr{F}$, *то* $\mu\{\omega\colon f(\omega) \neq h(\omega)\} = 0$.

Если λ — *мера, то* $f = f(\omega)$ *принимает значения в* $\overline{R}_+ = [0, \infty]$.

Функция $f = f(\omega)$ в представлении (38) называется *производной Радона—Никодима* или *плотностью* меры λ относительно меры μ и обозначается $\dfrac{d\lambda}{d\mu}$ или $\dfrac{d\lambda}{d\mu}(\omega)$.

Ряд важных свойств этих производных изложен в лемме п. 8 следующего § 7. Особо отметим сейчас частный случай приводимой там формулы (35), часто используемый при пересчете математических ожиданий при замене меры.

Именно, пусть и $\widetilde{\ }$ — две вероятностные меры, и $\widetilde{\ }$ — соответствующие математические ожидания. Предположим, что мера $\widetilde{\ }$ абсолютно непрерывна относительно меры (обозначение: $\widetilde{\ } \ll$). Тогда для всякой неотрицательной случайной величины $\xi = \xi(\omega)$ справедлива следующая «*формула пересчета математических ожиданий*»:

$$\widetilde{\ }\xi = \left[\xi \frac{d\widetilde{\ }}{d}\right]. \tag{39}$$

Эта формула остается справедливой и без предположения о неотрицательности ξ в следующей формулировке: случайная величина ξ интегрируема по мере $\widetilde{\ }$ в том и только том случае, когда величина $\xi\dfrac{d\widetilde{\ }}{d}$ интегрируема по мере ; при этом справедливо равенство (39).

Доказательство формулы (39) весьма несложно: для простых функций ξ она непосредственно следует из определения производной $\dfrac{d\widetilde{\ }}{d}$, а для неотрицательных ξ надо воспользоваться теоремой 1 b) § 4, утверждающей существование простых функций $\xi_n \uparrow \xi$, $n \to \infty$, и затем теоремой 1 a) о монотонной сходимости. Если же ξ — произвольная случайная величина, то, согласно (39)), $\widetilde{\ }|\xi| = |\xi|\dfrac{d\widetilde{\ }}{d}$. Отсюда следует, что интегрируемость ξ

по мере $\tilde{}$ равносильна интегрируемости $\xi \frac{d\,\tilde{}}{d}$ по мере . Сама же формула (39) вытекает из рассмотрения представления $\xi = \xi^+ - \xi^-$.

Теорема Радона—Никодима, приводимая здесь без доказательства (по поводу ее доказательства см., например, [70]), будет играть ключевую роль в конструкции условных математических ожиданий (§ 7).

9. Если $\xi = \sum\limits_{i=1}^{n} x_i I_{A_i}$ — простая случайная величина, $A_i = \{\omega : \xi = x_i\}$, то

$$g(\xi) = \sum g(x_i) \quad (A_i) = \sum g(x_i) \Delta F_\xi(x_i).$$

Иначе говоря, для подсчета математического ожидания функции от (простой) случайной величины ξ нет надобности знать всю вероятностную меру , а достаточно знать лишь 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 в интеграле Лебега). *Пусть* $(\Omega, \mathscr{F})$ *и* $(E, \mathscr{E})$ — *два измеримых пространства*, $X = X(\omega)$ — $\mathscr{F}/\mathscr{E}$-*измеримая функция со значениями в* E. *Пусть* — *вероятностная мера на* $(\Omega, \mathscr{F})$ *и* P_X — *вероятностная мера на* $(E, \mathscr{E})$, *индуцируемая* $X = X(\omega)$:

$$P_X(A) = \quad \{\omega : X(\omega) \in A\}, \qquad A \in \mathscr{E}. \tag{40}$$

Тогда для всякой $\mathscr{E}$-*измеримой функции* $g = g(x)$, $x \in E$,

$$\int_A g(x)\, P_X(dx) = \int_{X^{-1}(A)} g(X(\omega)) \quad (d\omega), \qquad A \in \mathscr{E} \tag{41}$$

(*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

$$P_X(AB) = \quad (X^{-1}(A) \cap X^{-1}(B)), \tag{42}$$

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 $\int_{X^{-1}(A)} g^+(X(\omega)) \quad (d\omega) < \infty$, а значит, из существования $\int\limits_A g(x)\,P_X(dx)$ следует существование интеграла $\int\limits_{X^{-1}(A)} g(X(\omega))\quad(d\omega)$. $\square$

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* $\int\limits_A g(x)\,P_\xi(dx)$ *или* $\int\limits_{\xi^{-1}(A)} g(\xi(\omega))\quad(d\omega)$, *то*

$$\int\limits_A g(x)\,P_\xi(dx)=\int\limits_{\xi^{-1}(A)} g(\xi(\omega))\quad(d\omega).$$

В частности, при $A=R$ *получаем, что*

$$g(\xi(\omega))=\int\limits_\Omega g(\xi(\omega))\quad(d\omega)=\int\limits_R g(x)\,P_\xi(dx). \tag{43}$$

Мера P_ξ однозначно восстанавливается по функции распределения F_ξ (теорема 1 в § 3). Поэтому интегралы Лебега $\int\limits_R g(x)P_\xi(dx)$ часто обозначают $\int\limits_R g(x)\,F_\xi(dx)$ или $\int\limits_R g\,dF_\xi$ 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

$$F_\xi(x)=\int\limits_{-\infty}^{x} f_\xi(y)\,dy, \tag{44}$$

где $f_\xi=f_\xi(x)$ — неотрицательная борелевская функция, а интеграл понимается как интеграл Лебега по лебеговской мере на множестве $(-\infty,x]$ (см. замечание 3 в п. 2). В предположении (44) формула (43) принимает следующий вид:

$$g(\xi(\omega))=\int\limits_{-\infty}^{\infty} g(x)f_\xi(x)\,dx, \tag{45}$$

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

$$P_\xi(B)=\int\limits_B f_\xi(x)\,dx,\qquad B\in\mathscr{B}(R), \tag{46}$$

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.

10. 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

$$\mu_1 \times \mu_2(A \times B) = \mu_1(A)\mu_2(B), \quad A \in \mathscr{F}_1,\ B \in \mathscr{F}_2$$

(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 (Фубини). *Пусть $\xi = \xi(\omega_1, \omega_2)$ — $\mathscr{F}_1 \otimes \mathscr{F}_2$-измеримая функция, интегрируемая 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 μ_2-почти всех ω_2 и μ_1-почти всех ω_1;*

2) *являются $\mathscr{F}_2$- и $\mathscr{F}_1$-измеримыми функциями, соответственно,*

$$\begin{aligned} &\mu_2\left\{\omega_2 \colon \int_{\Omega_1} |\xi(\omega_1, \omega_2)|\,\mu_1(d\omega_1) = \infty\right\} = 0, \\ &\mu_1\left\{\omega_1 \colon \int_{\Omega_2} |\xi(\omega_1, \omega_2)|\mu_2(d\omega_2) = \infty\right\} = 0 \end{aligned} \tag{48}$$

и 3)

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} \xi(\omega_1, \omega_2)\,d(\mu_1 \times \mu_2) &= \int_{\Omega_1} \left[\int_{\Omega_2} \xi(\omega_1, \omega_2)\,\mu_2(d\omega_2)\right] \mu_1(d\omega_1) = \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} \xi(\omega_1, \omega_2)\,\mu_1(d\omega_1)\right] \mu_2(d\omega_2). \end{aligned} \tag{49}$$

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 через $F_{\omega_1} = \{\omega_2 \in \Omega_2 \colon (\omega_1, \omega_2) \in F\}$ — *сечение* множества F в точке ω_1, и пусть $\mathscr{C}_{\omega_1} = \{F \in \mathscr{F} : F_{\omega_1} \in \mathscr{F}_2\}$. 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 $\{F^n\}_{n \geqslant 1}$ — 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 $\xi(\omega_1, \omega_2) \geqslant 0$. Тогда, поскольку для каждого ω_1 функция $\xi_{\omega_1}(\omega_2) = \xi(\omega_1, \omega_2)$ является $\mathscr{F}_2$-измеримой, то определен интеграл $\int\limits_{\Omega_2} \xi(\omega_1, \omega_2)\, \mu_2(d\omega_2)$. Покажем, что этот интеграл является $\mathscr{F}_1$-измеримой функцией и

$$\int\limits_{\Omega_1} \left[\int\limits_{\Omega_2} \xi(\omega_1, \omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1) = \int\limits_{\Omega_1 \times \Omega_2} \xi(\omega_1, \omega_2)\, d(\mu_1 \times \mu_2). \tag{50}$$

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

$$\int\limits_{\Omega_2} I_{A \times B}(\omega_1, \omega_2)\, \mu_2(d\omega_2) = I_A(\omega_1) \int\limits_{\Omega_2} I_B(\omega_2)\, \mu_2(d\omega_2) \tag{51}$$

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 $f(\omega_1) = \int\limits_{\Omega_2} I_F(\omega_1, \omega_2)\, \mu_2(d\omega_2)$ является $\mathscr{F}$-измеримым. С этой целью обозначим $\mathscr{C} = \{F \in \mathscr{F} : f(\omega_1) — \mathscr{F}_1\text{-измерима}\}$. 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 $\int\limits_{\Omega_2} \xi(\omega_1, \omega_2)\, \mu_2(d\omega_2)$ 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 $\mu_1 \times \mu_2(A \times B) = \mu_1(A)\mu_2(B)$, $A \in \mathscr{F}_1$, $B \in \mathscr{F}_2$, действительно существует и единственна.

Положим для $F \in \mathscr{F}$

$$\mu(F) = \int_{\Omega_1} \left[\int_{\Omega_2} I_{F_{\omega_1}}(\omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1).$$

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

$$\mu\left(\sum_n F^n\right) = \int_{\Omega_1} \left[\int_{\Omega_2} I_{(\sum F^n)_{\omega_1}}(\omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1) =$$

$$= \int_{\Omega_1} \sum_n \left[\int_{\Omega_2} I_{F^n_{\omega_1}}(\omega_2) \mu_2(d\omega_2) \right] \mu_1(d\omega_1) =$$

$$= \sum_n \int_{\Omega_1} \left[\int_{\Omega_2} I_{F^n_{\omega_1}}(\omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1) = \sum_n \mu(F^n),$$

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

$$\int_{\Omega_1 \times \Omega_2} I_{A \times B}(\omega_1, \omega_2)\, d(\mu_1 \times \mu_2) = \mu_1 \times \mu_2(A \times B), \tag{52}$$

и так как $I_{A \times B}(\omega_1, \omega_2) = I_A(\omega_1) I_B(\omega_2)$, то

$$\int_{\Omega_1} \left[\int_{\Omega_2} I_{A \times B}(\omega_1, \omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1) =$$

$$= \int_{\Omega_1} \left[I_A(\omega_1) \int_{\Omega_2} I_B(\omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1) = \mu_1(A)\mu_2(B). \tag{53}$$

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

$$\lambda(F) = \int_{\Omega_1 \times \Omega_2} I_F(\omega_1, \omega_2)\, d(\mu_1 \times \mu_2), \qquad F \in \mathscr{F},$$

является, очевидно, σ-конечной мерой. Нетрудно проверить также, что таковой же является функция множеств

$$\nu(F) = \int_{\Omega_1} \left[\int_{\Omega_2} I_F(\omega_1, \omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1).$$

Как было установлено выше, λ и ν совпадают на множествах 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)

$$\int_{\Omega_1 \times \Omega_2} \xi^+(\omega_1, \omega_2)\, d(\mu_1 \times \mu_2) < \infty, \qquad \int_{\Omega_1 \times \Omega_2} \xi^-(\omega_1, \omega_2)\, d(\mu_1 \times \mu_2) < \infty.$$

Согласно доказанному, интеграл $\int_{\Omega_2} \xi^+(\omega_1, \omega_2)\, \mu_2(d\omega_2)$ является $\mathscr{F}_1$-измеримой функцией от ω_1 и

$$\int_{\Omega_1} \left[\int_{\Omega_2} \xi^+(\omega_1, \omega_2)\, \mu_2(d\omega_2) \right] \mu_1(d\omega_1) = \int_{\Omega_1 \times \Omega_2} \xi^+(\omega_1, \omega_2)\, d(\mu_1 \times \mu_2) < \infty.$$

Поэтому в силу задачи 4 (см. также свойство **J** в п. 3)

$$\int_{\Omega_2} \xi^+(\omega_1, \omega_2)\, \mu_2(d\omega_2) < \infty \quad (\mu_1\text{-п. н.}).$$

Точно так же и

$$\int_{\Omega_2} \xi^-(\omega_1, \omega_2)\, \mu_2(d\omega_2) < \infty \quad (\mu_1\text{-п. н.}),$$

а значит,

$$\int_{\Omega_2} |\xi(\omega_1, \omega_2)|\, \mu_2(d\omega_2) < \infty \quad (\mu_1\text{-п. н.}).$$

Ясно, что за исключением некоторого множества $\mathscr{N}$, имеющего μ_1-меру нуль,

$$\int_{\Omega_2} \xi(\omega_1, \omega_2)\, \mu_2(d\omega_2) = \int_{\Omega_2} \xi^+(\omega_1, \omega_2)\, \mu_2(d\omega_2) - \int_{\Omega_2} \xi^-(\omega_1, \omega_2)\, \mu_2(d\omega_2). \quad (54)$$

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

$$\int_{\Omega_1\times\Omega_2}\xi(\omega_1,\,\omega_2)\,d(\mu_1\times\mu_2)=\int_{\Omega_2}\left[\int_{\Omega_1}\xi(\omega_1,\,\omega_2)\,\mu_1(d\omega_1)\right]\mu_2(d\omega_2).\qquad\square$$

Следствие. *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.

Пример. Пусть $(\xi,\,\eta)$ — пара случайных величин, распределение которых имеет двумерную плотность $f_{\xi,\eta}(x,\,y)$, т. е.

$$\{(\xi,\,\eta)\in B\}=\int_B f_{\xi,\eta}(x,\,y)\,dx\,dy,\qquad B\in\mathscr{B}(R^2),$$

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

$$f_\xi(x)=\int_{-\infty}^{\infty}f_{\xi,\eta}(x,\,y)\,dy\quad\text{и}\quad f_\eta(y)=\int_{-\infty}^{\infty}f_{\xi,\eta}(x,\,y)\,dx.\tag{55}$$

В самом деле, если $A\in\mathscr{B}(R)$, то по теореме Фубини

$$\{\xi\in A\}=\quad\{(\xi,\,\eta)\in A\times R\}=\int_{A\times R}f_{\xi,\eta}(x,\,y)\,dx\,dy=\int_A\left[\int_R f_{\xi,\eta}(x,\,y)\,dy\right]dx,$$

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

$$F_{\xi,\eta}(x, y) = \int\limits_{(-\infty,x]\times(-\infty,y]} f_{\xi,\eta}(u, v)\,du\,dv = \int\limits_{(-\infty,x]\times(-\infty,y]} f_\xi(u) f_\eta(v)\,du\,dv =$$
$$= \int\limits_{(-\infty,x]} f_\xi(u)\,du \Bigg(\int\limits_{(-\infty,y]} f_\eta(v)\,dv \Bigg) = F_\xi(x) F_\eta(y)$$

и, следовательно, ξ и η независимы.

Обратно, если они независимы и имеют плотность $f_{\xi,\eta}(x, y)$, то опять-таки по теореме Фубини

$$\int\limits_{(-\infty,x]\times(-\infty,y]} f_{\xi,\eta}(u, v)\,du\,dv = \Bigg(\int\limits_{(-\infty,x]} f_\xi(u)\,du \Bigg) \Bigg(\int\limits_{(-\infty,y]} f_\eta(v)\,dv \Bigg) =$$
$$= \int\limits_{(-\infty,x]\times(-\infty,y]} f_\xi(u) f_\eta(v)\,du\,dv.$$

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).

11. В этом пункте будет рассмотрен вопрос о разных определениях интегралов Лебега и Римана и соотношениях между ними.

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 $G(x)$ — некоторая обобщенная функция распределения на R^1 (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

$$\overline{\sum_{\mathscr{P}}} = \sum_{i=1}^{n} \bar{g}_i [G(x_{i+1}) - G(x_i)], \qquad \underline{\sum_{\mathscr{P}}} = \sum_{i=1}^{n} \underline{g}_i [G(x_{i+1}) - G(x_i)],$$

где

$$\bar{g}_i = \sup_{x_{i-1} < y \leqslant x_i} g(y), \qquad \underline{g}_i = \inf_{x_{i-1} < y \leqslant x_i} g(y).$$

Определим простые функции $\bar{g}_{\mathscr{P}}(x)$ и $\underline{g}_{\mathscr{P}}(x)$, полагая на $x_{i-1} < x \leqslant x_i$

$$\bar{g}_{\mathscr{P}}(x) = \bar{g}_i, \qquad \underline{g}_{\mathscr{P}}(x) = \underline{g}_i$$

и определяя $\bar{g}_{\mathscr{P}}(a) = \underline{g}_{\mathscr{P}}(a) = g(a)$. Ясно, что тогда в соответствии с конструкцией интеграла *Лебега—Стилтьеса* (см. замечание 3 в п. 2)

$$\overline{\sum_{\mathscr{P}}} = (\mathrm{L{-}S}) \int_a^b \bar{g}_{\mathscr{P}}(x)\, G(dx)$$

и

$$\underline{\sum_{\mathscr{P}}} = (\mathrm{L{-}S}) \int_a^b \underline{g}_{\mathscr{P}}(x)\, G(dx).$$

Пусть теперь $\{\mathscr{P}_k\}$ — последовательность разбиений таких, что $\mathscr{P}_k \subseteq \subseteq \mathscr{P}_{k+1}$, причем $\mathscr{P}_k = \{x_0^{(k)}, \ldots, x_{n_k}^{(k)}\}$ таковы, что $\max\limits_{0 \leqslant i \leqslant n_k} |x_{i+1}^{(k)} - x_i^{(k)}| \to 0$, $k \to \infty$. Тогда $\bar{g}_{\mathscr{P}_1} \geqslant \bar{g}_{\mathscr{P}_2} \geqslant \ldots \geqslant g \geqslant \ldots \geqslant \underline{g}_{\mathscr{P}_2} \geqslant \underline{g}_{\mathscr{P}_1}$, и если $|g(x)| \leqslant C$, то по теореме о мажорируемой сходимости

$$\begin{aligned} \lim_{k\to\infty} \overline{\sum_{\mathscr{P}_k}} &= (\mathrm{L{-}S}) \int_a^b \bar{g}(x)\, G(dx), \\ \lim_{k\to\infty} \underline{\sum_{\mathscr{P}_k}} &= (\mathrm{L{-}S}) \int_a^b \underline{g}(x)\, G(dx), \end{aligned} \tag{57}$$

где $\bar{g}(x) = \lim\limits_k \bar{g}_{\mathscr{P}_k}(x)$, $\underline{g}(x) = \lim\limits_k \underline{g}_{\mathscr{P}_k}(x)$.

Если пределы $\lim\limits_{k} \overline{\sum\limits_{\mathscr{P}_k}}$ и $\lim\limits_{k} \underline{\sum\limits_{\mathscr{P}_k}}$ *конечны, совпадают и их общее значение не зависит от выбора последовательности разбиений* $\{\mathscr{P}_k\}$, то говорят, что функция $g=g(x)$ *интегрируема по Риману—Стилтьесу*, а соответствующее общее значение пределов обозначается

$$(\mathrm{R{-}S}) \int_a^b g(x)\, G(dx) \quad \text{или} \quad (\mathrm{R{-}S}) \int_a^b g(x)\, dG(x). \tag{58}$$

В том случае, когда $G(x)=x$, этот интеграл называется *интегралом Римана* и обозначается

$$(\mathrm{R}) \int_a^b g(x)\, dx.$$

Пусть теперь $(\mathrm{L{-}S}) \int\limits_a^b g(x)\, G(dx)$ — соответствующий интеграл *Лебега—Стилтьеса* (см. замечание 3 в п. 2).

Теорема 9. *Если функция* $g=g(x)$ *непрерывна на* $[a, b]$, *то она интегрируема по Риману—Стилтьесу и*

$$(\mathrm{R{-}S}) \int_a^b g(x)\, G(dx) = (\mathrm{L{-}S}) \int_a^b g(x)\, G(dx). \tag{59}$$

Доказательство. Так как функция $g(x)$ непрерывна, то $\overline{g}(x)=g(x)=$ $=\underline{g}(x)$. Поэтому в силу (57) $\lim\limits_{k\to\infty} \overline{\sum\limits_{\mathscr{P}_k}} = \lim\limits_{k\to\infty} \underline{\sum\limits_{\mathscr{P}_k}}$. 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. $\square$

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* $\overline{\lambda}$ *на* $\overline{\mathscr{B}}([a, b])$).

b) *Если* $g=g(x)$ *интегрируема по Риману, то она интегрируема по Лебегу и*

$$(\mathrm{R}) \int_a^b g(x)\, dx = (\mathrm{L}) \int_a^b g(x)\, \overline{\lambda}(dx). \tag{60}$$

Доказательство. a) Пусть функция $g = g(x)$ интегрируема по Риману. Тогда, согласно (57),

$$(\mathrm{L}) \int_a^b \overline{g}(x)\, \overline{\lambda}(dx) = (\mathrm{L}) \int_a^b \underline{g}(x)\, \overline{\lambda}(dx).$$

Но $\underline{g}(x) \leqslant g(x) \leqslant \overline{g}(x)$, поэтому в силу свойства **H**

$$\underline{g}(x) = g(x) = \overline{g}(x) \quad (\overline{\lambda}\text{-п. н.}), \tag{61}$$

откуда нетрудно вывести, что функция $g(x)$ непрерывна почти всюду (относительно меры $\overline{\lambda}$).

Обратно, пусть функция $g = g(x)$ непрерывна почти всюду (относительно меры $\overline{\lambda}$). В этом случае выполнено (61) и, следовательно, $g(x)$ отличается от измеримой (по Борелю) функции $\overline{g}(x)$ лишь на множестве $\mathscr{N}$ с $\overline{\lambda}(\mathscr{N}) = 0$. Но тогда

$$\begin{aligned} \{x\colon\ g(x) \leqslant c\} = \{x\colon\ g(x) \leqslant c\} \cap \bar{\mathscr{N}} + \{x\colon\ g(x) \leqslant c\} \cap \mathscr{N} = \\ = \{x\colon\ \overline{g}(x) \leqslant c\} \cap \bar{\mathscr{N}} + \{x\colon\ g(x) \leqslant c\} \cap \mathscr{N}. \end{aligned}$$

Ясно, что $\{x\colon \overline{g}(x) \leqslant c\} \cap \bar{\mathscr{N}} \in \bar{\mathscr{B}}([a,\, b])$, а множество $\{x\colon g(x) \leqslant c\} \cap \mathscr{N}$ является подмножеством множества $\mathscr{N}$, имеющего лебеговскую меру $\overline{\lambda}$, равную нулю, и, следовательно, также принадлежащего $\bar{\mathscr{B}}([a,\, b])$. Тем самым $g(x)$ $\bar{\mathscr{B}}([a,\, b])$-измерима и как ограниченная функция интегрируема по Лебегу. Поэтому по свойству **G**

$$(\mathrm{L}) \int_a^b \overline{g}(x)\, \overline{\lambda}(dx) = (\mathrm{L}) \int_a^b \underline{g}(x)\, \overline{\lambda}(dx) = (\mathrm{L}) \int_a^b g(x)\, \overline{\lambda}(dx),$$

что и завершает доказательство утверждения a).

b) Если функция $g = g(x)$ интегрируема по Риману, то, согласно a), она непрерывна ($\overline{\lambda}$-п. н.). Выше было показано, что тогда $g(x)$ интегрируема по Лебегу и ее интегралы Римана и Лебега совпадают. □

Замечание 1. Пусть μ некоторая мера Лебега—Стилтьеса на $\mathscr{B}([a,\, b])$. Обозначим $\bar{\mathscr{B}}_\mu([a,\, b])$ 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 $\overline{\mu}$ — *продолжение меры* μ на $\bar{\mathscr{B}}_\mu([a,\, b])$ ($\overline{\mu}(\Lambda) = \mu(A)$ 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 $\overline{\lambda}$ рассмотреть меру $\overline{\mu}$, а вместо интегралов Римана и Лебега рассмотреть соответствующие интегралы Римана—Стилтьеса и Лебега—Стилтьеса по мере $\overline{\mu}$.

Замечание 2. Определение интеграла Лебега (см. определения 1 и 2 и формулы (3) и (6) в п. 1) и концептуально, и «чисто внешне» отличается от определений интегралов Римана и Римана—Стилтьеса, требующих обращения к *верхним* и *нижним* суммам (см. (57)).

Остановимся более подробно на *сопоставлении* этих определений.

Пусть $(\Omega, \mathscr{F}, \mu)$ — некоторое измеримое пространство с мерой μ.

Для всякой $\mathscr{F}$-измеримой неотрицательной функции $f = f(\omega)$ определим два интеграла (*нижний* и *верхний*) $L_* f$ и $L^* f$ (обозначаемые также $\int_* f\,d\mu$ и $\int^* f\,d\mu$), полагая по определению

$$L_* f = \sup \sum_i \Big(\inf_{\omega \in A_i} f(\omega)\Big)\mu(A_i),$$

$$L^* f = \inf \sum_i \Big(\sup_{\omega \in A_i} f(\omega)\Big)\mu(A_i),$$

где sup и inf берутся по всем конечным разбиениям $(A_1, A_2, \ldots, A_n)$ пространства Ω на $\mathscr{F}$-измеримые множества $A_1, A_2, \ldots, A_n$ $\Big(\sum\limits_{i=1}^{n} A_i = \Omega\Big)$, $n \geqslant 1$.

Можно показать, что $L_* f \leqslant L^* f$ и если функция f ограничена, а мера μ конечна, то $L_* f = L^* f$ (задача 20).

Один из подходов (Дарбу—Юнг) к определению интеграла Lf от функции f по мере μ состоит в том, чтобы говорить, что *функция f является μ-интегрируемой, если $L_* f = L^* f$, и в этом случае полагать $Lf = L_* f$* $(= L^* f)$.

Если теперь обратиться к определению интеграла Лебега $\quad f$, данному в п. 1 (определение 1), то можно убедиться (задача 21), что

$$\quad f = L_* f.$$

Тем самым, можно сказать, что для *ограниченных неотрицательных* функций $f = f(\omega)$ подходы Лебега и Дарбу—Юнга приводят к одному и тому же результату ($\quad f = Lf = L^* f = L_* f$).

Отличия же в этих подходах к интегрированию проявляются тогда, когда рассматриваются *неограниченные* функции или когда мера μ может быть *бесконечной*.

Например, с точки зрения интегрирования в смысле Лебега интегралы $\int\limits_{(0,1]} \frac{dx}{x^{1/2}}$ и $\int\limits_{(1,\infty)} \frac{dx}{x^2}$ определены и совпадают с $L_* f$ для $f(x) = x^{-1/2} I(0, 1]$ и $f(x) = x^{-2} I(1, \infty)$ соответственно. Однако здесь $L^* f = \infty$.

Таким образом, $L_* f < L^* f$, и, значит, рассмотренные функции не интегрируемы в смысле Дарбу—Юнга, но интегрируемы в смысле Лебега.

В рамках изложенного подхода, оперирующего с нижним интегралом $L_* f$ и верхним интегралом $L^* f$, обратимся к интегрированию в смысле Римана.

Будем считать, что $\Omega=(0, 1]$, $\mathscr{F}=\mathscr{B}$ (борелевская σ-алгебра) и $\mu=\lambda$ — мера Лебега. Пусть $f=f(x)$, $x\in\Omega$, — некоторая ограниченная функция (условие ее измеримости пока не налагается).

По аналогии с $L_* f$ и $L^* f$ введем *нижние* и *верхние* римановские интегралы $R_* f$ и $R^* f$, полагая

$$R_* f=\sup\sum_i\Bigl(\inf_{\omega\in B_i} f(\omega)\Bigr)\lambda(B_i),$$
$$R^* f=\inf\sum_i\Bigl(\sup_{\omega\in B_i} f(\omega)\Bigr)\lambda(B_i),$$

где $(B_1, B_2, \ldots, B_n)$ образуют конечное разбиение $\Omega=(0, 1]$, причем B_i имеют вид $(a_i, b_i]$ (в отличие от множеств A_i в определении $L_* f$ и $L^* f$, которые были произвольными $\mathscr{F}$-измеримыми множествами).

Из приведенных определений очевидно, что

$$R_* f\leqslant L_* f\leqslant L^* f\leqslant R^* f.$$

Приведенные в теоремах 9 и 10 свойства интегрируемости по Риману могут быть переформулированы и дополнены с привлечением следующих условий:

(a) $R^* f=R_* F$;

(b) лебеговская мера множества D_f точек разрыва функции f равна нулю ($\lambda(D_f)=0$);

(c) существует константа $R(f)$ такая, что для всякого $\varepsilon>0$ найдется $\delta>0$ такое, что

$$\Bigl|R(f)-\sum_i f(\omega_i)\lambda((a_i, b_i])\Bigr|<\varepsilon$$

для всякой конечной системы непересекающихся интервалов $(a_i, b_i]$ с $\sum (a_i, b_i]=(0, 1]$ такой, что $\lambda((a_i, b_i])<\delta$, $\omega_i\in(a_i, b_i]$.

Воспользовавшись аргументами теорем 9 и 10, можно доказать (задача 22), что *если функция f ограничена, то*

(A) *условия* (a), (b), (c) *эквивалентны и*

(B) *при выполнении любого из условий* (a), (b), (c)

$$R(f)=R_* f=R^* f.$$

12. В этом пункте мы приведем полезную теорему об *интегрировании по частям* в интеграле Лебега—Стилтьеса.

Пусть на $(R, \mathscr{B}(R))$ заданы две обобщенные функции распределения $F=F(x)$ и $G=G(x)$.

Теорема 11. *Для любых действительных a и b, $a<b$, справедлива следующая формула интегрирования по частям*:

$$F(b)G(b) - F(a)G(a) = \int_a^b F(s-)\, dG(s) + \int_a^b G(s)\, dF(s), \tag{62}$$

или, что эквивалентно,

$$F(b)G(b) - F(a)G(a) = \\ = \int_a^b F(s-)\, dG(s) + \int_a^b G(s-)\, dF(s) + \sum_{a<s\leqslant b} \Delta F(s)\Delta G(s), \tag{63}$$

где

$$F(s-) = \lim_{t\uparrow s} F(t), \qquad \Delta F(s) = F(s) - F(s-).$$

Замечание 1. Символически формулу (62) можно записать в следующей «*дифференциальной*» форме:

$$d(FG) = F_-\, dG + G\, dF. \tag{64}$$

Замечание 2. Утверждение теоремы сохраняет свою силу для функций F и G ограниченной (на $[a, b]$) вариации. (Каждая такая непрерывная справа и имеющая пределы слева функция представима в виде разности двух монотонно неубывающих функций.)

Доказательство. Напомним прежде всего, что в соответствии с соглашениями п. 2 под интегралом $\int_a^b$ понимается интеграл $\int_{(a,b]}$. Поэтому (см. формулу (2) в § 3)

$$(F(b) - F(a))(G(b) - G(a)) = \int_a^b dF(s) \cdot \int_a^b dG(t).$$

Отсюда по теореме Фубини ($F \times G$ — обозначает прямое произведение мер, отвечающих F и G) находим, что

$$(F(b) - F(a))(G(b) - G(a)) = \int_{(a,b]\times(a,b]} d(F\times G)(s, t) =$$

$$= \int_{(a,b]\times(a,b]} I_{\{s\geqslant t\}}(s, t)\, d(F\times G)(s, t) + \int_{(a,b]\times(a,b]} I_{\{s<t\}}(s, t)\, d(F\times G)(s, t) =$$

$$= \int_{(a,b]} (G(s) - G(a))\, dF(s) + \int_{(a,b]} (F(t-) - F(a))\, dG(t) =$$

$$= \int_a^b G(s)\,dF(s) + \int_a^b F(s-)\,dG(s) - G(a)(F(b) - F(a)) - F(a)(G(b) - G(a)), \tag{65}$$

где I_A — индикатор множества A.

Из формулы (65) непосредственно следует (62). В свою очередь (63) вытекает из (62), если только заметить, что

$$\int_a^b (G(s) - G(s-))\,dF(s) = \sum_{a<s\leqslant b} \Delta G(s)\Delta F(s). \tag{66}$$

$\square$

Следствие 1. *Если $F(x)$ и $G(x)$ — функции распределения, то*

$$F(x)G(x) = \int_{-\infty}^{x} F(s-)\,dG(s) + \int_{-\infty}^{x} G(s)\,dF(s). \tag{67}$$

Если к тому же функция распределения

$$F(x) = \int_{-\infty}^{x} f(s)\,ds,$$

то

$$F(x)G(x) = \int_{-\infty}^{x} F(s)\,dG(s) + \int_{-\infty}^{x} G(s)f(s)\,ds. \tag{68}$$

Следствие 2. *Пусть ξ — случайная величина с функцией распределения $F = F(x)$ и $|\xi|^n < \infty$. Тогда*

$$\int_0^\infty x^n\,dF(x) = n\int_0^\infty x^{n-1}[1 - F(x)]\,dx, \tag{69}$$

$$\int_{-\infty}^{0} |x|^n\,dF(x) = n\int_0^\infty x^{n-1}F(-x)\,dx \tag{70}$$

и

$$|\xi|^n = \int_{-\infty}^{\infty} |x|^n\,dF(x) = n\int_0^\infty x^{n-1}[1 - F(x) + F(-x)]\,dx. \tag{71}$$

Для доказательства (69) заметим, что

$$\int_0^b x^n\,dF(x) = -\int_0^b x^n\,d(1 - F(x)) =$$

$$= -b^n(1 - F(b)) + n\int_0^b x^{n-1}(1 - F(x))\,dx. \tag{72}$$

Покажем, что в силу предположения $|\xi|^n < \infty$

$$b^n(1 - F(b) - F(-b)) \leqslant b^n \ \{|\xi| \geqslant b\} \to 0,\ b \to \infty. \tag{73}$$

Действительно,

$$|\xi|^n = \sum_{k=1}^{\infty} \int_{k-1}^{k} |x|^n \, dF(x) < \infty$$

и, значит,

$$\sum_{k \geqslant b+1} \int_{k-1}^{k} |x|^n \, dF(x) \to 0, \qquad b \to \infty.$$

Но

$$\sum_{k \geqslant b+1} \int_{k-1}^{k} |x|^n \, dF(x) \geqslant b^n \ \{|\xi| \geqslant b\},$$

что и доказывает (73).

Переходя в (72) к пределу при $b \to \infty$, получаем требуемую формулу (69). Формула (70) доказывается аналогично.

Формула же (71) следует из (69) и (70).

13. Пусть $A = A(t)$, $t \geqslant 0$, — непрерывная справа и имеющая пределы слева функция локально ограниченной вариации (т. е. имеющая ограниченную вариацию на каждом конечном интервале $[a, b]$). Рассмотрим уравнение

$$Z_t = 1 + \int_0^t Z_{s-} \, dA(s), \tag{74}$$

которое в дифференциальной форме записывают в виде

$$dZ_t = Z_{t-} \, dA(t), \qquad Z_0 = 1. \tag{75}$$

Доказанная выше формула интегрирования по частям позволяет (в классе локально ограниченных функций) найти явный вид решения уравнения (74).

Введем функцию (называемую *стохастической экспонентой*; [87])

$$\mathscr{E}_t(A) = e^{A(t) - A(0)} \prod_{0 < s \leqslant t} (1 + \Delta A(s)) e^{-\Delta A(s)}, \tag{76}$$

где $\Delta A(s) = A(s) - A(s-)$ при $s > 0$ и $\Delta A(0) = 0$.

Функция $A(s)$, $0 \leqslant s \leqslant t$, имеет ограниченную вариацию и, следовательно, допускает самое большее счетное число точек разрыва, а ряд

$\sum\limits_{0<s\leqslant t} |\Delta A(s)|$ сходится. Отсюда вытекает, что функция

$$\prod_{0<s\leqslant t} (1+\Delta A(s))e^{-\Delta A(s)}, \qquad t \geqslant 0,$$

является функцией локально ограниченной вариации.

Если обозначить $A^c(t) = A(t) - \sum\limits_{0<s\leqslant t} \Delta A(s)$ непрерывную составляющую функции $A(t)$, то (76) можно переписать в следующей форме:

$$\mathscr{E}_t(A) = e^{A^c(t)-A^c(0)} \prod_{0<s\leqslant t} (1+\Delta A(s)). \tag{77}$$

Обозначим

$$F(t) = e^{A^c(t)-A^c(0)}, \qquad G(t) = \prod_{0<s\leqslant t} (1+\Delta A(s)), \qquad G(0) = 1.$$

Тогда в силу (62)

$$\mathscr{E}_t(A) = F(t)G(t) = 1 + \int_0^t F(s)\,dG(s) + \int_0^t G(s-)\,dF(s) =$$
$$= 1 + \sum_{0<s\leqslant t} F(s)G(s-)\Delta A(s) + \int_0^t G(s-)F(s)\,dA^c(s) = 1 + \int_0^t \mathscr{E}_{s-}(A)\,dA(s).$$

Таким образом, $\mathscr{E}_t(A)$, $t \geqslant 0$, *является* (локально ограниченным) *решением уравнения* (74).

Покажем, что в классе локально ограниченных решений это решение единственное.

Предположим, что есть два локально ограниченных решения и $Y = Y(t)$, $t \geqslant 0$, — их разность. Тогда

$$Y(t) = \int_0^t Y(s-)\,dA(s).$$

Положим

$$T = \inf\{t \geqslant 0\colon\ Y(t) \neq 0\},$$

считая $T = \infty$, если $Y(t) = 0$ для всех $t \geqslant 0$.

Поскольку $A(t)$, $t \geqslant 0$, является функцией локально ограниченной вариации, то найдутся такие две обобщенные функции распределения $A_1(t)$ и $A_2(t)$, что $A(t) = A_1(t) - A_2(t)$. Если предположить, что $T < \infty$, то можно найти такое конечное $T' > T$, что

$$[A_1(T') + A_2(T')] - [A_1(T) + A_2(T)] \leqslant \frac{1}{2}.$$

Тогда из уравнения

$$Y(t) = \int_T^t Y(s-)\, dA(s), \qquad t \geqslant T,$$

следует, что

$$\sup_{t \leqslant T'} |Y(t)| \leqslant \frac{1}{2} \sup_{t \leqslant T'} |Y(t)|,$$

и поскольку $\sup\limits_{t \leqslant T'} |Y(t)| < \infty$, то $Y(t) = 0$ для $T < t \leqslant T'$, что противоречит предположению $T < \infty$.

Итак, доказана следующая

Теорема 12. *В классе локально ограниченных функций уравнение* (74) *имеет и притом единственное решение*, *задаваемое формулой* (76).

14. Задачи.

1. Доказать представление (6).

2. Показать, что справедливо следующее обобщение свойства **E**. Пусть ξ и η — случайные величины, для которых определены ξ и η и выражение $\xi + \eta$ имеет смысл (не имеет вида $\infty - \infty$ или $-\infty + \infty$). Тогда

$$(\xi + \eta) = \xi + \eta.$$

3. Обобщить свойство **G**, показав, что если $\xi = \eta$ (п. н.) и ξ существует, то η также существует и $\eta = \xi$.

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 $\int \xi\, d\mu$ и $\int \eta\, d\mu$ определены. Тогда, если для всех $A \in \mathscr{F}$ $\int_A \xi\, d\mu \leqslant \int_A \eta\, d\mu$, то $\xi \leqslant \eta$ (μ-п. н.). (Ср. со свойством **I**.)

6. Пусть ξ и η — независимые неотрицательные случайные величины. Показать, что тогда $\xi\eta = \xi \cdot \eta$.

7. Используя лемму Фату, показать, что

$$(\underline{\lim}\, A_n) \leqslant \underline{\lim}\ (A_n), \qquad (\overline{\lim}\, A_n) \leqslant \overline{\lim}\ (A_n).$$

8. Привести пример, показывающий, что в теореме о мажорируемой сходимости условие «$|\xi_n| \leqslant \eta$, $\eta < \infty$» не может быть, вообще говоря, ослаблено.

9. Привести пример, показывающий, что в лемме Фату условие «$\xi_n \leqslant \eta$, $\eta > -\infty$» не может быть, вообще говоря, отброшено.

10. Доказать справедливость следующего варианта леммы Фату. Пусть семейство случайных величин $\{\xi_n^+\}_{n\geqslant 1}$ равномерно интегрируемо. Тогда

$$\overline{\lim} \quad \xi_n \leqslant \quad \overline{\lim}\, \xi_n.$$

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 $(a_{ij};\ i,\ j \geqslant 1\})$ — последовательность действительных чисел таких, что $\sum\limits_{i,j} |a_{ij}| < \infty$. Вывести из теоремы Фубини, что

$$\sum_{(i,j)} a_{ij} = \sum_i \Bigl(\sum_j a_{ij}\Bigr) = \sum_j \Bigl(\sum_i a_{ij}\Bigr). \tag{78}$$

14. Привести пример последовательности $(a_{ij};\ i,\ j \geqslant 1)$, для которой $\sum\limits_{i,j} |a_{ij}| = \infty$ и равенства в (78) не справедливы.

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

$$\int\limits_{h(a)}^{h(b)} f(x)\, dx = \int\limits_a^b f(h(y))h'(y)\, dy.$$

16. Доказать формулу (70).

17. Пусть $\xi, \xi_1, \xi_2, \ldots$ — неотрицательные интегрируемые случайные величины такие, что $\xi_n \to \quad \xi$ и для всякого $\varepsilon > 0$ вероятность $(|\xi - \xi_n| > \varepsilon) \to 0$. Показать, что тогда $\quad |\xi_n - \xi| \to 0$, $n \to \infty$.

18. Пусть ξ — интегрируемая случайная величина ($\ |\xi| < \infty$). Доказать, что для всякого $\varepsilon > 0$ существует $\delta > 0$ такое, что для любого $A \in \mathscr{F}$ с $\ (A) < \delta$ выполнено свойство $\ I_A|\xi| < \varepsilon$ («свойство абсолютной непрерывности интеграла Лебега»).

19. Пусть ξ, η, ζ и ξ_n, η_n, ζ_n, $n \geqslant 1$, — случайные величины такие, что *)

$$\xi_n \to \xi, \ \ \eta_n \to \eta, \ \ \zeta_n \to \zeta, \quad \eta_n \leqslant \xi_n \leqslant \zeta_n, \ \ n \geqslant 1,$$
$$\zeta_n \to \quad \zeta, \qquad \eta_n \to \quad \eta,$$

и математические ожидания $\quad \xi, \quad \eta, \quad \zeta$ конечны. Показать, что тогда справедлива *лемма Пратта*: $\quad \xi_n \to \quad \xi$ и если к тому же $\eta_n \leqslant 0 \leqslant \zeta_n$, то $\quad |\xi_n - \xi| \to 0$.

Вывести отсюда, что если $\xi_n \to \xi$, $\quad |\xi_n| \to \quad |\xi|$ и $\quad |\xi| < \infty$, то $\quad |\xi_n - \xi| \to 0$.

Привести пример, показывающий, что в условиях леммы Пратта, вообще говоря, $\quad |\xi_n - \xi| \not\to 0$.

20. Доказать, что $L_* f \leqslant L^* f$ и если функция f ограничена и мера μ конечна, то $L_* f = L^* f$ (см. замечание 2 в п. 11).

21. Доказать, что для ограниченных функций f математическое ожидание $\quad f = L_* f$ (см. замечание 2 в п. 11).

22. Доказать заключительное утверждение в замечании 2 п. 11.

23. Пусть $F(x)$ — функция распределения случайной величины X. Показать, что

$$X^+ < \infty \ \Leftrightarrow \ \int_a^\infty \ln \frac{1}{F(x)}\, dx < \infty \text{ для некоторого } a.$$

24. Показать, что если $p > 0$ и $\lim\limits_{x \to \infty} x^p \quad \{|\xi| > x\} = 0$, то $\quad |\xi|^r < \infty$ для всех $r < p$. Привести пример, показывающий, что при $r = p$ может оказаться, что $\quad |\xi|^p = \infty$.

25. Дать пример плотности $f(x)$, не являющейся четной функцией, у которой, тем не менее, все нечетные моменты $\int\limits_{-\infty}^{\infty} x^k f(x)\, dx = 0$, $k = 1, 3, \ldots$

26. Привести пример случайных величин ξ_n, $n \geqslant 1$, таких, что

$$\sum_{n=1}^{\infty} \xi_n \neq \sum_{n=1}^{\infty} \quad \xi_n.$$

27. Пусть случайная величина X такова, что для любого $\alpha > 1$

$$\frac{\{|X| > \alpha n\}}{\{|X| > n\}} \to 0, \quad n \to \infty.$$

*) Сходимость $\xi_n \to \xi$, называемая сходимостью по вероятности, означает, что для всякого $\varepsilon > 0$ вероятность $\quad \{|\xi_n - \xi| > \varepsilon\} \to 0$, $n \to \infty$. Подробнее см. § 10.

Доказать, что тогда у X существуют все моменты. *Указание*: воспользоваться формулой

$$|X|^N = N \int_0^\infty x^{N-1} \quad (|X| > x)\, dx.$$

28. Пусть X — случайная величина, принимающая значения $k = 0,\ 1,\ 2,\ \ldots$ с вероятностями p_k. В соответствием с § 13 главы I функция $F(s) = \sum_{k=0}^{\infty} p_k s^k$, $|s| \leqslant 1$, называется *производящей функцией* случайной величины X. Установить следующие формулы:

(i) если X — пуассоновская случайная величина, т. е. $p_k = e^{-\lambda}\lambda^k/k!$, где $\lambda > 0$, $k = 0,\ 1,\ 2,\ \ldots$, то

$$F(s) = e^{-\lambda(1-s)}, \qquad |s| \leqslant 1;$$

(ii) если случайная величина X имеет геометрическое распределение, т. е. $p_k = pq^k$, где $0 < p < 1$, $q = 1 - p$, $k = 0,\ 1,\ 2,\ \ldots$, то

$$F(s) = \frac{p}{1 - sq}, \qquad |s| \leqslant 1.$$

29. Наряду с производящей функцией $F(s)$ полезно рассматривать *производящую функцию моментов*: $M(s) = \quad e^{sX}$ (в предположении, что s таковы, что $\quad e^{sX} < \infty$).

(a) Показать, что если производящая функция моментов $M(s)$ определена для всех s из некоторой окрестности нуля ($s \in [-a,\ a]$, $a > 0$), то существуют производные $M^{(k)}(s)$ при $s = 0$ для всех $k = 1,\ 2,\ \ldots$ и

$$M^{(k)}(0) = \quad X^k$$

(это свойство и оправдывает название для $M(s)$).

(b) Привести пример случайной величины, для которой $M(s) = \infty$ при всех $s > 0$.

(c) Показать, что для пуассоновской случайной величины X с $\lambda > 0$ функция $M(s) = e^{-\lambda(1-e^s)}$ для всех $s \in R$.

30. Пусть $0 < r < \infty$, $X_n \in L^r$, $X_n \to X$. Тогда следующие условия равносильны:

(i) семейство $\{|X_n|^r,\ n \geqslant 1\}$ равномерно интегрируемо;

(ii) $X_n \to X$ в L^r;

(iii) $\quad |X_n|^r \to \quad |X|^r < \infty$.

31. *Тождество Спицера*. Пусть $X_1,\ X_2,\ \ldots$ — независимые одинаково распределенные случайные величины с $\quad \{X_1 \leqslant 1\} = 1$, и пусть $S_n = X_1 + \ldots$

$\ldots + X_n$. Тогда для $|u|$, $|t| < 1$

$$\sum_{n=0}^{\infty} t^n \quad e^{uM_n} = \exp\left(\sum_{n=1}^{\infty} \frac{t^n}{n} \quad e^{uS_n^+}\right),$$

где $M_n = \max(0, X_1, X_2, \ldots, X_n)$, $S_n^+ = \max(0, S_n)$.

32. Пусть $S_0 = 0$, $S_n = X_1 + \ldots + X_n$, $n \geqslant 1$, — простое симметричное случайное блуждание и $\tau = \min\{n > 0\colon S_n \geqslant 0\}$. Показать, что

$$\min(\tau, 2m) = 2 \quad |S_{2m}| = 4m \quad \{S_{2m} = 0\}, \qquad m \geqslant 0.$$

33. Пусть ξ — стандартная гауссовская случайная величина ($\xi \sim \mathscr{N}(0, 1)$). Используя интегрирование по частям, показать, что $\quad \xi^k = (k-1) \quad \xi^{k-2}$. Вывести отсюда формулы

$$\xi^{2k-1} = 0 \quad \text{и} \quad \xi^{2k} = 1 \cdot 3 \cdot \ldots \cdot (2k-3)(2k-1) \quad (= (2k-1)!!).$$

34. Показать, что функция $x^{-1} \sin x$, $x \in R$, интегрируема по Риману, но не интегрируема по Лебегу (с лебеговой мерой на $(R, \mathscr{B}(R))$).

35. Показать, что функция

$$\xi(\omega_1, \omega_2) = e^{-\omega_1\omega_2} - 2e^{-2\omega_1\omega_2}, \quad \omega_1 \in \Omega_1 = [1, \infty), \ \ \omega_2 \in \Omega_2 = (0, 1],$$

такова, что (по мере Лебега)

(a) для каждого ω_2 она интегрируема по $\omega_1 \in \Omega_1$,

(b) для каждого ω_1 она интегрируема по $\omega_2 \in \Omega_2$, но теорема Фубини не имеет места.

36. Доказать *теорему Беппо Леви*: Пусть случайные величины $\xi_1, \xi_2, \ldots$ интегрируемы ($\quad |\xi_n| < \infty$ для всех $n \geqslant 1$), $\sup\limits_n \quad \xi_n < \infty$ и $\xi_n \uparrow \xi$; тогда случайная величина ξ интегрируема и $\quad \xi_n \uparrow \quad \xi$ (ср. с теоремой 1a).

37. Доказать следующую разновидность *леммы Фату*: если $0 \leqslant \xi_n \to \xi$ (-п. н.) и $\quad \xi_n \leqslant A < \infty$, $n \geqslant 1$, то ξ интегрируема и $\quad \xi \leqslant A$.

38. (*О связи интегрирования по Лебегу и по Риману.*) Пусть борелевская функция $f = f(x)$ интегрируема по мере Лебега: $\int\limits_R |f(x)|\, dx < \infty$. Доказать, что для всякого $\varepsilon > 0$ найдутся:

(a) ступенчатая функция $f_\varepsilon(x) = \sum\limits_{i=1}^{n} f_i I_{A_i}(x)$ с ограниченными интервалами A_i такая, что $\int\limits_R |f(x) - f_\varepsilon(x)|\, dx < \varepsilon$;

(b) интегрируемая непрерывная функция $g_\varepsilon(x)$ с ограниченным носителем такая, что $\int\limits_R |f(x) - g_\varepsilon(x)|\, dx < \varepsilon$.

39. Показать, что если ξ есть интегрируемая случайная величина, то

$$\xi = \int_0^\infty \{\xi > x\}\, dx - \int_{-\infty}^0 \{\xi < x\}\, dx.$$

40. Пусть ξ и η — интегрируемые случайные величины. Показать, что

$$\xi - \eta = \int_{-\infty}^\infty [\{\eta < x \leqslant \xi\} - \{\xi < x \leqslant \eta\}]\, dx.$$

41. Пусть ξ — неотрицательная случайная величина ($\xi \geqslant 0$) с преобразованием Лапласа $\varphi_\xi(\lambda) = e^{-\lambda\xi}$, $\lambda \geqslant 0$.

(a) Показать, что для всякого $0 < r < 1$

$$\xi^r = \frac{r}{\Gamma(1-r)} \int_0^\infty \frac{1 - \varphi_\xi(\lambda)}{\lambda^{r+1}}\, d\lambda.$$

Указание: воспользоваться тем, что для $s \geqslant 0$, $0 < r < 1$

$$\frac{1}{r}\,\Gamma(1-r)s^r = \int_0^\infty \frac{1 - e^{-s\lambda}}{\lambda^{r+1}}\, d\lambda.$$

(b) Показать, что если $\xi > 0$, то для всякого $r > 0$

$$\xi^{-r} = \frac{1}{r\,\Gamma(r)} \int_0^\infty \varphi_\xi(\lambda^{1/r})\, d\lambda.$$

Указание: воспользоваться тем, что для $s \geqslant 0$, $r > 0$

$$s = \frac{r}{\Gamma(1/r)} \int_0^\infty \exp\{-(\lambda/s)^r\}\, d\lambda.$$

§ 7. Условные вероятности и условные математические ожидания относительно σ-алгебр

1. Пусть $(\Omega, \mathscr{F}, \ )$ — вероятностное пространство и событие $A \in \mathscr{F}$ таково, что $(A) > 0$. Как и в случае конечных вероятностных пространств, *условной вероятностью* A (обозначение: $(B \mid A)$) будем называть величину $(BA)/(A)$, а *условной вероятностью события* B *относительно конечного или счетного разбиения* $\mathscr{D} = \{D_1, D_2, \ldots\}$ с $(D_i) > 0$, $i \geqslant 1$ (обозначение: $(B \mid \mathscr{D})$, $(B \mid \mathscr{D})(\omega)$) назовем *случайную величину*, равную $(B \mid D_i)$ для $\omega \in D_i$, $i \geqslant 1$:

$$(B \mid \mathscr{D})(\omega) = \sum_{i \geqslant 1} (B \mid D_i) I_{D_i}(\omega).$$

Аналогичным образом, если ξ — случайная величина, для которой определено $\ \xi$, то *условным математическим ожиданием ξ относительно события A* с $\ (A)>0$ (обозначение: $\ (\xi \,|\, A)$) будем называть величину $\dfrac{\ (\xi I_A)}{\ (A)}$ (ср. с (10) § 8 гл. I).

Случайная величина $\ (B \,|\, \mathscr{D})$ является, очевидно, измеримой относительно σ-алгебры $\mathscr{G}=\sigma(\mathscr{D})$, в связи с чем ее обозначают также $\ (B \,|\, \mathscr{G})$ (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 $\ (\nu = k \,|\, \xi = x)$»? Поскольку $\ \{\xi = x\} = 0$, то интересующая нас «условная вероятность $\ (\nu = k \,|\, \xi = x)$» пока не определена, хотя интуитивно понятно, что эта «вероятность должна была бы 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 $(\Omega, \mathscr{F}, \ )$ — вероятностное пространство, $\mathscr{G}$ — некоторая σ-алгебра, $\mathscr{G} \subseteq \mathscr{F}$ ($\mathscr{G}$ — *σ-подалгебра*, или *под-σ-алгебра* $\mathscr{F}$) и $\xi = \xi(\omega)$ — случайная величина. Напомним, что, согласно § 6, математическое ожидание $\ \xi$ определялось в два этапа: сначала для неотрицательных случайных величин ξ, а затем в общем случае с помощью равенства

$$\ \xi = \ \xi^+ - \ \xi^-$$

и только (чтобы избежать неопределенности вида $\infty - \infty$) в предположении, что

$$\min(\ \xi^-, \ \xi^+) < \infty.$$

Подобная двухэтапная конструкция применяется и при определении условных математических ожиданий $\ (\xi \,|\, \mathscr{G})$.

Определение 1. 1) *Условным математическим ожиданием неотрицательной* случайной величины ξ *относительно σ-алгебры $\mathscr{G}$* называется неотрицательная (расширенная) случайная величина, обозначаемая $\ (\xi \,|\, \mathscr{G})$ или $\ (\xi \,|\, \mathscr{G})(\omega)$, такая, что

a) $\ (\xi \,|\, \mathscr{G})$ является $\mathscr{G}$-измеримой;

b) для любого $A \in \mathscr{G}$

$$\int_A \xi\, d \quad = \int_A \quad (\xi\,|\,\mathscr{G})\, d \quad . \tag{1}$$

2) *Условное математическое ожидание* $\quad (\xi\,|\,\mathscr{G})$, или $\quad (\xi\,|\,\mathscr{G})(\omega)$, *произвольной* случайной величины ξ *относительно σ-алгебры* $\mathscr{G}$ считается определенным, если -п. н.

$$\min(\quad (\xi^+\,|\,\mathscr{G}), \quad (\xi^-\,|\,\mathscr{G})) < \infty,$$

и задается формулой

$$(\xi\,|\,\mathscr{G}) \equiv \quad (\xi^+\,|\,\mathscr{G}) - \quad (\xi^-\,|\,\mathscr{G}),$$

причем на множестве (нулевой вероятности) тех элементарных событий, для которых $\quad (\xi^+\,|\,\mathscr{G}) = \quad (\xi^-\,|\,\mathscr{G}) = \infty$, разность $\quad (\xi^+\,|\,\mathscr{G}) - \quad (\xi^-\,|\,\mathscr{G})$ определяется произвольно, например, полагается равной нулю.

Прежде всего покажем, что для неотрицательных случайных величин $\quad (\xi\,|\,\mathscr{G})$ действительно существует. Согласно п. 8 § 6, функция множеств

$$(A) = \int_A \xi\, d \quad , \qquad A \in \mathscr{G}, \tag{2}$$

является мерой на $(\Omega, \mathscr{G})$, которая абсолютно непрерывна относительно меры (рассматриваемой на $(\Omega, \mathscr{G})$, $\mathscr{G} \subseteq \mathscr{F}$). Поэтому (по теореме Радона—Никодима) существует такая неотрицательная $\mathscr{G}$-измеримая расширенная случайная величина $\quad (\xi\,|\,\mathscr{G})$, что

$$(A) = \int_A \quad (\xi\,|\,\mathscr{G})\, d \quad . \tag{3}$$

Из (2) и (3) следует соотношение (1).

Замечание 1. В соответствии с теоремой Радона—Никодима условное математическое ожидание $\quad (\xi\,|\,\mathscr{G})$ определяется однозначно лишь с точностью до множеств -меры нуль. Иначе говоря, в качестве $\quad (\xi\,|\,\mathscr{G})(\omega)$ можно взять любую $\mathscr{G}$-измеримую функцию $f(\omega)$, называемую *вариантом* условного математического ожидания, для которой $\quad (A) = \int_A f(\omega)\, d \quad$, $A \in \mathscr{G}$.

Отметим также, что, согласно замечанию к теореме Радона—Никодима,

$$(\xi\,|\,\mathscr{G}) \equiv \frac{d}{d}(\omega), \tag{4}$$

т. е. *условное математическое ожидание есть не что иное, как производная Радона—Никодима меры относительно меры* (рассматриваемых на $(\Omega, \mathscr{G})$).

Полезно заметить, что если неотрицательная случайная величина ξ такова, что $\xi < \infty$, то $(\xi \,|\, \mathscr{G}) < \infty$ (-п. н.), что непосредственно вытекает из (1). Аналогично, если $\xi \leqslant 0$ и $\xi > -\infty$, то $(\xi \,|\, \mathscr{G}) > -\infty$ (-п. н.).

Замечание 2. В связи с соотношением (1) отметим, что мы не можем, вообще говоря, положить $(\xi \,|\, \mathscr{G}) = \xi$, поскольку случайная величина ξ *не обязана быть* $\mathscr{G}$-измеримой.

Замечание 3. Предположим, что случайная величина ξ такова, что для нее существует ξ. Тогда $(\xi \,|\, \mathscr{G})$ можно было бы определить как такую $\mathscr{G}$-измеримую функцию, для которой справедливо (1). Обычно именно так и поступают. Приводимое нами определение $(\xi \,|\, \mathscr{G}) \equiv (\xi^+ \,|\, \mathscr{G}) - (\xi^- \,|\, \mathscr{G})$ обладает тем преимуществом, что в случае тривиальной σ-алгебры $\mathscr{G} = \{\varnothing, \Omega\}$ оно превращается в определение ξ и при этом оно не предполагает существования ξ. (Например, если ξ — случайная величина с $\xi^+ = \infty$, $\xi^- = \infty$, а $\mathscr{G} = \mathscr{F}$, то ξ не определено, но в смысле определения 1 $(\xi \,|\, \mathscr{G})$ существует и есть просто $\xi = \xi^+ - \xi^-$.)

Замечание 4. Пусть условное математическое ожидание $(\xi \,|\, \mathscr{G})$ определено. *Условной дисперсией* $(\xi \,|\, \mathscr{G})$ случайной величины ξ относительно σ-алгебры $\mathscr{G}$ называется случайная величина

$$(\xi \,|\, \mathscr{G}) = [(\xi - (\xi \,|\, \mathscr{G}))^2 \,|\, \mathscr{G}]$$

(ср. с определением $(\xi \,|\, \mathscr{D})$ относительно разбиения $\mathscr{D}$, данным в задаче 2 § 8 гл. I, и с определением *дисперсии* в § 8).

Определение 2. Пусть $B \in \mathscr{F}$. Условное математическое ожидание $(I_B \,|\, \mathscr{G})$ обозначается $P(B \,|\, \mathscr{G})$ или $(B \,|\, \mathscr{G})(\omega)$ и называется *условной вероятностью события B относительно σ-алгебры $\mathscr{G}$, $\mathscr{G} \subseteq \mathscr{F}$.*

Из определений 1 и 2 следует, что для каждого фиксированного $B \in \mathscr{F}$ условная вероятность $(B \,|\, \mathscr{G})$ есть такая *случайная величина*, что:

a) $(B \,|\, \mathscr{G})$ является $\mathscr{G}$-измеримой,

b) для любого $A \in \mathscr{G}$

$$(A \cap B) = \int_A (B \,|\, \mathscr{G}) \, d \quad . \tag{5}$$

Определение 3. Пусть ξ — случайная величина и $\mathscr{G}_\eta$ — σ-алгебра, порожденная некоторым случайным элементом η. Тогда $(\xi \,|\, \mathscr{G}_\eta)$, если оно определено, обозначается $(\xi \,|\, \eta)$ или $(\xi \,|\, \eta)(\omega)$ и называется *условным математическим ожиданием ξ относительно η.*

Условная вероятность $(B \,|\, \mathscr{G}_\eta)$ обозначается $(B \,|\, \eta)$ или $(B \,|\, \eta)(\omega)$, и называется *условной вероятностью события B относительно η.*

3. Покажем, что данное здесь определение $(\xi \,|\, \mathscr{G})$ согласуется с определением условного математического ожидания § 8 гл. 1.

Пусть $\mathscr{D}=\{D_1, D_2, \ldots\}$ — некоторое конечное или счетное разбиение с *атомами* D_i $\left(\sum\limits_i D_i=\Omega\right)$ такими, что $(D_i)>0$, $i\geqslant 1$.

Теорема 1. *Если $\mathscr{G}=\sigma(\mathscr{D})$ и ξ — случайная величина, для которой ξ определено, то*

$$(\xi\,|\,\mathscr{G})= \quad (\xi\,|\,D_i) \quad (\quad\text{-п. н. на } D_i) \tag{6}$$

или, что то же,

$$(\xi\,|\,\mathscr{G})=\frac{(\xi I_{D_i})}{(D_i)} \quad (\quad\text{-п. н. на } D_i).$$

(*Запись* «$\xi=\eta$ (-*п. н. на* A)» *или* «$\xi=\eta$ (A; -*п. н.*)» *означает, что* $(A\cap\{\xi\neq\eta\})=0$.)

Доказательство. Согласно лемме 3 из § 4, на D_i $(\xi\,|\,\mathscr{G})=K_i$, где K_i — постоянная. Но

$$\int\limits_{D_i}\xi\,d \quad =\int\limits_{D_i} \quad (\xi\,|\,\mathscr{G})\,d \quad =K_i \quad (D_i),$$

откуда

$$K_i=\frac{1}{(D_i)}\int\limits_{D_i}\xi\,d \quad =\frac{(\xi I_{D_i})}{(D_i)}= \quad (\xi\,|\,D_i). \qquad \square$$

Таким образом, введенное в гл. I понятие условного математического ожидания $(\xi\,|\,\mathscr{D})$ относительно конечного разбиения $\mathscr{D}=\{D_1, \ldots, D_n\}$ 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 ξ, η математические ожидания определены и σ-алгебра $\mathscr{G}\subseteq\mathscr{F}$.

A*. *Если C — постоянная и $\xi=C$ (п. н.), то* $(\xi\,|\,\mathscr{G})=C$ (*п. н.*).

B*. *Если $\xi\leqslant\eta$ (п. н.), то* $(\xi\,|\,\mathscr{G})\leqslant \quad (\eta\,|\,\mathscr{G})$ (*п. н.*).

C*. $| \quad (\xi\,|\,\mathscr{G})|\leqslant \quad (|\xi|\,|\,\mathscr{G})$ (*п. н.*).

D*. *Если a, b — постоянные и $a \quad \xi+b \quad \eta$ определено, то*

$$(a\xi+b\eta\,|\,\mathscr{G})=a \quad (\xi\,|\,\mathscr{G})+b \quad (\eta\,|\,\mathscr{G}) \quad (\text{п. н.}).$$

E*. *Пусть $\mathscr{F}_*=\{\varnothing, \Omega\}$ — тривиальная σ-алгебра. Тогда*

$$(\xi\,|\,\mathscr{F}_*)= \quad \xi \quad (\text{п. н.}).$$

F*. $(\xi\,|\,\mathscr{F})=\xi$ (*п. н.*).

G*. $(\quad (\xi\,|\,\mathscr{G}))= \quad \xi$.

H*. *Если $\mathscr{G}_1\subseteq\mathscr{G}_2$, то справедливо (первое) «телескопическое свойство»*:

$$[\quad (\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1]= \quad (\xi\,|\,\mathscr{G}_1) \quad (\text{п. н.}).$$

I*. *Если $\mathscr{G}_1 \supseteq \mathscr{G}_2$, то справедливо (второе) «телескопическое свойство»:*

$$[\ \ (\xi\,|\,\mathscr{G}_2)\,|\,\mathscr{G}_1] = \ \ (\xi\,|\,\mathscr{G}_2) \quad (\textit{п. н.}).$$

J*. *Пусть случайная величина ξ, для которой $\ \ \xi$ определено, не зависит от σ-алгебры $\mathscr{G}$ (т. е. не зависит от I_B, $B \in \mathscr{G}$). Тогда*

$$(\xi\,|\,\mathscr{G}) = \ \ \xi \quad (\textit{п. н.}).$$

K*. *Пусть η — $\mathscr{G}$-измеримая случайная величина, $\ \ |\xi| < \infty$ и $\ \ |\xi\eta| < \infty$. Тогда*

$$(\xi\eta\,|\,\mathscr{G}) = \eta \ \ (\xi\,|\,\mathscr{G}) \quad (\textit{п. н.}).$$

Приведем доказательства этих свойств.

A*. Функция, равная постоянной, измерима относительно $\mathscr{G}$. Поэтому остается лишь проверить равенство

$$\int_A \xi\, d \ \ = \int_A C\, d \ \ , \quad A \in \mathscr{G}.$$

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

$$\int_A \xi\, d \ \ \leqslant \int_A \eta\, d \ \ , \quad A \in \mathscr{G},$$

а значит,

$$\int_A \ \ (\xi\,|\,\mathscr{G})\, d \ \ \leqslant \int_A \ \ (\eta\,|\,\mathscr{G})\, d \ \ , \quad A \in \mathscr{G}.$$

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,

$$\int_A (a\xi + b\eta)\, d \ \ = \int_A a\xi\, d \ \ + \int_A b\eta\, d \ \ =$$
$$= \int_A a \ \ (\xi\,|\,\mathscr{G})\, d \ \ + \int_A b \ \ (\eta\,|\,\mathscr{G})\, d \ \ = \int_A [a \ \ (\xi\,|\,\mathscr{G}) + b \ \ (\eta\,|\,\mathscr{G})]\, d \ \ ,$$

что и доказывает свойство **D***.

E*. Это свойство следует из замечания, что $\ \ \xi$ является $\mathscr{F}_*$-измеримой функцией, и того факта, что если $A = \Omega$ или $A = \varnothing$, то очевидным образом

$$\int_A \xi\, d \ \ = \int_A \ \ \xi\, d \ \ .$$

F*. Поскольку ξ — $\mathscr{F}$-измерима и

$$\int_A \xi\, d \quad = \int_A \quad (\xi \mid \mathscr{F})\, d \quad , \qquad A \in \mathscr{F},$$

то $\quad (\xi \mid \mathscr{F}) = \xi$ (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

$$\int_A \quad (\xi \mid \mathscr{G}_1)\, d \quad = \int_A \xi\, d \quad .$$

Так как $\mathscr{G}_1 \subseteq \mathscr{G}_2$, то $A \in \mathscr{G}_2$ и, значит,

$$\int_A \quad [\quad (\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1]\, d \quad = \int_A \quad (\xi \mid \mathscr{G}_2)\, d \quad = \int_A \xi\, d \quad .$$

Следовательно, для $A \in \mathscr{G}_1$

$$\int_A \quad (\xi \mid \mathscr{G}_1)\, d \quad = \int_A \quad [\quad (\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1]\, d$$

и, рассуждая также, как и при доказательстве свойства **I** § 6 (см. также задачу 5 § 6), получаем, что

$$(\xi \mid \mathscr{G}_1) = \quad [\quad (\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1] \quad \text{(п. н.).}$$

I*. Если $A \in \mathscr{G}_1$, то по определению $\quad [\quad (\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1]$

$$\int_A \quad [\quad (\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1]\, d \quad = \int_A \quad (\xi \mid \mathscr{G}_2)\, d \quad .$$

Функция $\quad (\xi \mid \mathscr{G}_2)$ является $\mathscr{G}_2$-измеримой, и поскольку $\mathscr{G}_2 \subseteq \mathscr{G}_1$, то и $\mathscr{G}_1$-измеримой. Отсюда следует, что $\quad (\xi \mid \mathscr{G}_2)$ есть один из вариантов условного математического ожидания $\quad [\quad (\xi \mid \mathscr{G}_2) \mid \mathscr{G}_1]$, что и доказывает свойство **I***.

J*. Поскольку $\quad \xi$ является $\mathscr{G}$-измеримой функцией, то остается проверить, что для любого $B \in \mathscr{G}$

$$\int_B \xi\, d \quad = \int_B \quad \xi\, d \quad ,$$

т. е. что $\quad [\xi I_B] = \quad \xi \cdot \quad I_B$. Если $\quad |\xi| < \infty$, то это сразу следует из теоремы 6 § 6. В общем случае вместо теоремы 6 § 6 надо воспользоваться результатами задачи 6 из того же параграфа.

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.*

a) *Если* $|\xi_n| \leqslant \eta$, $\quad \eta < \infty$ *и* $\xi_n \to \xi$ *(п. н.), то*

$$(\xi_n \,|\, \mathscr{G}) \to \quad (\xi \,|\, \mathscr{G}) \quad (п.\,н.)$$

и

$$(|\xi_n - \xi| \,|\, \mathscr{G}) \to 0 \quad (п.\,н.).$$

b) *Если* $\xi_n \geqslant \eta$, $\quad \eta > -\infty$ *и* $\xi_n \uparrow \xi$ *(п. н.), то*

$$(\xi_n \,|\, \mathscr{G}) \uparrow \quad (\xi \,|\, \mathscr{G}) \quad (п.\,н.).$$

c) *Если* $\xi_n \leqslant \eta$, $\quad \eta < \infty$ *и* $\xi_n \downarrow \xi$ *(п. н.), то*

$$(\xi_n \,|\, \mathscr{G}) \downarrow \quad (\xi \,|\, \mathscr{G}) \quad (п.\,н.).$$

d) *Если* $\xi_n \geqslant \eta$, $\quad \eta > -\infty$, *то*

$$(\underline{\lim}\ \xi_n \,|\, \mathscr{G}) \leqslant \underline{\lim} \quad (\xi_n \,|\, \mathscr{G}) \quad (п.\,н.).$$

e) *Если* $\xi_n \leqslant \eta$, $\quad \eta < \infty$, *то*

$$\overline{\lim} \quad (\xi_n \,|\, \mathscr{G}) \leqslant \quad (\underline{\lim}\ \xi_n \,|\, \mathscr{G}) \quad (п.\,н.).$$

f) *Если* $\xi_n \geqslant 0$, *то*

$$\left(\sum \xi_n \,|\, \mathscr{G}\right) = \sum \quad (\xi_n \,|\, \mathscr{G}) \quad (п.\,н.).$$

Доказательство. a) Пусть $\zeta_n = \sup\limits_{m \geqslant n} |\xi_m - \xi_n|$. Поскольку $\xi_n \to \xi$ (п. н.), то $\zeta_n \downarrow 0$ (п. н.). Математические ожидания $\quad \xi_n$ и $\quad \xi$ конечны, поэтому в силу свойств **D*** и **C*** (п. н.)

$$| \quad (\xi_n \,|\, \mathscr{G}) - \quad (\xi \,|\, \mathscr{G})| = | \quad (\xi_n - \xi \,|\, \mathscr{G})| \leqslant \quad (|\xi_n - \xi| \,|\, \mathscr{G}) \leqslant \quad (\zeta_n \,|\, \mathscr{G}).$$

Поскольку $\quad (\zeta_{n+1} \,|\, \mathscr{G}) \leqslant \quad (\zeta_n \,|\, \mathscr{G})$ (п. н.), то (п. н.) существует предел $h = \lim\limits_n \quad (\zeta_n \,|\, \mathscr{G})$. Тогда

$$0 \leqslant \int\limits_\Omega h\, d \quad \leqslant \int\limits_\Omega \quad (\zeta_n \,|\, \mathscr{G})\, d \quad = \int\limits_\Omega \zeta_n\, d \quad \to 0, \quad n \to \infty,$$

где последнее утверждение следует из теоремы о мажорируемой сходимости, поскольку $0 \leqslant \zeta_n \leqslant 2\eta$, $\quad \eta < \infty$. Следовательно, $\int\limits_\Omega h\, d \quad = 0$ и по свойству **H** $h = 0$ (п. н.).

b) Пусть сначала $\eta \equiv 0$. Поскольку $\quad (\xi_n \,|\, \mathscr{G}) \leqslant \quad (\xi_{n+1} \,|\, \mathscr{G})$ (п. н.), то существует (п. н.) предел $\zeta(\omega) = \lim\limits_n \quad (\xi_n \,|\, \mathscr{G})$. Тогда из равенства

$$\int\limits_A \xi_n\, d \quad = \int\limits_A \quad (\xi_n \,|\, \mathscr{G})\, d \quad , \quad A \in \mathscr{G},$$

и теоремы о монотонной сходимости

$$\int_A \xi \, d \quad = \int_A \quad (\xi \,|\, \mathscr{G}) \, d \quad = \int_A \zeta \, d \quad , \qquad A \in \mathscr{G}.$$

Следовательно, по свойству, аналогичному свойству **I**, и задаче 5, $\xi = \zeta$ (п. н.).

Для доказательства в общем случае заметим, что $0 \leqslant \xi_n^+ \uparrow \xi^+$, и по доказанному

$$(\xi_n^+ \,|\, \mathscr{G}) \uparrow \quad (\xi^+ \,|\, \mathscr{G}) \quad (\text{п. н.}). \tag{7}$$

Но $0 \leqslant \xi_n^- \leqslant \xi^-$, $\quad \xi^- < \infty$, поэтому в силу a)

$$(\xi_n^- \,|\, \mathscr{G}) \to \quad (\xi^- \,|\, \mathscr{G}),$$

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 $\zeta = \underline{\lim}\, \xi_n$. Согласно b), $\quad (\zeta_n \,|\, \mathscr{G}) \uparrow \uparrow \quad (\zeta \,|\, \mathscr{G})$ (п. н.). Поэтому (п. н.)

$$(\underline{\lim}\, \xi_n \,|\, \mathscr{G}) = \quad (\zeta \,|\, \mathscr{G}) = \lim_n \quad (\zeta_n \,|\, \mathscr{G}) = \underline{\lim} \quad (\zeta_n \,|\, \mathscr{G}) \leqslant \underline{\lim} \quad (\xi_n \,|\, \mathscr{G}).$$

Утверждение e) вытекает из d).

f) Если $\xi_n \geqslant 0$, то по свойству **D***

$$\left(\sum_{k=1}^{n} \xi_k \,|\, \mathscr{G} \right) = \sum_{k=1}^{n} \quad (\xi_k \,|\, \mathscr{G}) \quad (\text{п. н.}),$$

что вместе с b) и доказывает требуемый результат. □

Приведем теперь доказательство свойства **K***. Пусть $\eta = I_B$, $B \in \mathscr{G}$. Тогда для всякого $A \in \mathscr{G}$

$$\int_A \xi\eta \, d \quad = \int_{A \cap B} \xi \, d \quad = \int_{A \cap B} \quad (\xi \,|\, \mathscr{G}) \, d \quad = \int_A I_B \quad (\xi \,|\, \mathscr{G}) \, d \quad = \int_A \eta \quad (\xi \,|\, \mathscr{G}) \, d \quad .$$

В силу аддитивности интеграла Лебега равенство

$$\int_A \xi\eta \, d \quad = \int_A \eta \quad (\xi \,|\, \mathscr{G}) \, d \quad , \qquad A \in \mathscr{G}, \tag{8}$$

останется справедливым и для простых случайных величин $\eta = \sum\limits_{k=1}^{n} y_k I_{B_k}$, $B_k \in \mathscr{G}$. Поэтому по свойству **I** (§ 6) для таких случайных величин

$$(\xi\eta \,|\, \mathscr{G}) = \eta \quad (\xi \,|\, \mathscr{G}) \quad (\text{п. н.}). \tag{9}$$

Пусть теперь η — произвольная $\mathscr{G}$-измеримая случайная величина 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)

$$(\xi\eta_n\,|\,\mathscr{G})=\eta_n \quad (\xi\,|\,\mathscr{G}) \quad (\text{п. н.}).$$

Ясно, что $|\xi\eta_n|\leqslant|\xi\eta|$, где $|\xi\eta|<\infty$. Поэтому по свойству a) теоремы 2

$$(\xi\eta_n\,|\,\mathscr{G})\to \quad (\xi\eta\,|\,\mathscr{G}) \quad (\text{п. н.}).$$

Далее, так как $|\xi|<\infty$, то $(\xi\,|\,\mathscr{G})$ конечно (п. н.) (см. свойство **C***, свойство **J** (§ 6)). Поэтому $\eta_n \quad (\xi\,|\,\mathscr{G})\to\eta \quad (\xi\,|\,\mathscr{G})$ (п. н.). (Предположение о конечности почти наверное $(\xi\,|\,\mathscr{G})$ существенно, поскольку, согласно сноске на с. 248, $0\cdot\infty=0$, но если $\eta_n=1/n$, $\eta\equiv 0$, то $\dfrac{1}{n}\cdot\infty\not\to 0\cdot\infty=0$.)

Замечание. Свойство **K*** сохраняет свою силу, если выполнены лишь следующие условия: *η является $\mathscr{G}$-измеримой и $(\xi\,|\,\mathscr{G})$ определено.*

5. Рассмотрим подробнее структуру условных математических ожиданий $(\xi\,|\,\mathscr{G}_\eta)$, обозначаемых, как было условлено выше, также через $(\xi\,|\,\eta)$.

Поскольку $(\xi\,|\,\eta)$ 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 $\omega\in\Omega$

$$m(\eta(\omega))= \quad (\xi\,|\,\eta)(\omega). \tag{10}$$

Эту функцию $m(y)$ будем обозначать через $(\xi\,|\,\eta=y)$ и называть *условным математическим ожиданием ξ относительно события $\{\eta=y\}$* или *условным математическим ожиданием ξ при условии, что $\eta=y$.*

В соответствии с определением

$$\int_A \xi\,d \quad =\int_A \quad (\xi\,|\,\eta)\,d \quad =\int_A m(\eta)\,d \quad, \qquad A\in\mathscr{G}_\eta. \tag{11}$$

Поэтому по теореме 7 § 6 (о замене переменных под знаком интеграла Лебега)

$$\int_{\{\omega:\,\eta\in B\}} m(\eta)\,d \quad =\int_B m(y)P_\eta(dy), \qquad B\in\mathscr{B}(\bar{R}), \tag{12}$$

где P_η — распределение вероятностей η. Следовательно, $m=m(y)$ есть борелевская функция такая, что для всякого $B\in\mathscr{B}(R)$

$$\int_{\{\omega:\,\eta\in B\}} \xi\,d \quad =\int_B m(y)\,P_\eta(dy). \tag{13}$$

Это замечание подсказывает, что к определению условного математического ожидания $(\xi\,|\,\eta=y)$ можно прийти и иначе.

Определение 4. Пусть ξ и η — случайные величины (быть может, и расширенные) и ξ 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

$$\int\limits_{\{\omega : \eta \in B\}} \xi \, d \quad = \int\limits_B m(y)\, P_\eta(dy), \qquad B \in \mathscr{B}(\bar{R}). \tag{14}$$

Тот факт, что такая функция существует, следует опять же из теоремы Радона—Никодима, если заметить, что функция множеств

$$(B) = \int\limits_{\{\omega : \eta \in B\}} \xi \, d$$

является мерой со знаком, которая абсолютно непрерывна относительно меры P_η.

Предположим теперь, что $m(y)$ есть условное математическое ожидание в смысле определения 4. Тогда, применяя снова теорему о замене переменных под знаком интеграла Лебега, находим, что

$$\int\limits_{\{\omega : \eta \in B\}} \xi \, d \quad = \int\limits_B m(y)\, P_\eta(dy) = \int\limits_{\{\omega : \eta \in B\}} m(\eta)\, d \quad , \qquad B \in \mathscr{B}(\bar{R}).$$

Функция $m(\eta)$ является $\mathscr{G}_\eta$-измеримой, и множествами $\{\omega \colon \eta \in B\}$, $B \in \mathscr{B}(\bar{R})$, исчерпываются все множества из $\mathscr{G}_\eta$.

Отсюда вытекает, что $m(\eta)$ есть математическое ожидание $(\xi \mid \eta)$. Тем самым, зная $(\xi \mid \eta = y)$, можно восстановить $(\xi \mid \eta)$, и, наоборот, по $(\xi \mid \eta)$ можно найти $(\xi \mid \eta = y)$.

С интуитивной точки зрения условное математическое ожидание $m(y) = (\xi \mid \eta = y)$ является более простым и понятным объектом, нежели $(\xi \mid \eta)$. Однако математическое ожидание $(\xi \mid \eta)$, рассматриваемое как $\mathscr{G}_\eta$-измеримая случайная величина, более удобно в работе.

Отметим, что приведенные выше свойства **A*—K*** и утверждения теоремы 2 легко переносятся на условные математические ожидания $(\xi \mid \eta = y)$ (с заменой «почти наверное» на «P_η-почти наверное»). Так, например, свойство **K*** переформулируется следующим образом: если $|\xi| < \infty$, $|\xi f(\eta)| < \infty$, где $f = f(y)$ — $\mathscr{B}(\bar{R})$-измеримая функция, то

$$(\xi f(\eta) \mid \eta = y) = f(y) \quad (\xi \mid \eta = y) \quad (P_\eta\text{-п. н.}). \tag{15}$$

Далее (ср. со свойством **J***), если ξ и η независимы, то

$$(\xi \mid \eta = y) = \quad \xi \quad (P_\eta\text{-п. н.}).$$

Отметим также, что если $B \in \mathscr{B}(R^2)$ и ξ и η независимы, то

$$[I_B(\xi, \eta) \mid \eta = y] = \quad I_B(\xi, y) \quad (P_\eta\text{-п. н.}), \tag{16}$$

и если $\varphi = \varphi(x,\, y)$ — $\mathscr{B}(R^2)$-измеримая функция такая, что $\quad |\varphi(\xi,\, \eta)| < \infty$, то

$$[\varphi(\xi,\, \eta)\,|\,\eta = y] = \quad [\varphi(\xi,\, y)] \quad (P_\eta\text{-п. н.}).$$

Для доказательства (16) заметим следующее. Если $B = B_1 \times B_2$, то для справедливости (16) надо лишь проверить, что

$$\int\limits_{\{\omega\,:\,\eta \in A\}} I_{B_1 \times B_2}(\xi,\, \eta) \quad (d\omega) = \int\limits_{(y \in A)} \quad I_{B_1 \times B_2}(\xi,\, y)\, P_\eta(dy).$$

Но левая часть есть $\quad \{\xi \in B_1,\, \eta \in A \cap B_2\}$, а правая — $\quad \{\xi \in B_1\} \quad \{\eta \in A \cap B_2\}$, и их равенство 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 5. *Условной вероятностью события $A \in \mathscr{F}$ при условии, что $\eta = y$* (обозначение: $\quad (A\,|\,\eta = y)$), будем называть $\quad (I_A\,|\,\eta = y)$.

Понятно, что $\quad (A\,|\,\eta = y)$ можно было бы определять как такую $\mathscr{B}(\bar{R})$-измеримую функцию, что

$$(A \cap \{\eta \in B\}) = \int\limits_B \quad (A\,|\,\eta = y)\, P_\eta(dy), \qquad B \in \mathscr{B}(\bar{R}). \tag{17}$$

6. Приведем некоторые примеры вычисления условных вероятностей и условных математических ожиданий.

Пример 1. Пусть η — дискретная случайная величина с $\quad \{\eta = y_k\} > 0$, $\sum\limits_{k=1}^{\infty} \quad \{\eta = y_k\} = 1$. Тогда

$$(A\,|\,\eta = y_k) = \frac{(A \cap \{\eta = y_k\})}{\{\eta = y_k\}}, \qquad k \geqslant 1.$$

Для $y \notin \{y_1,\, y_2,\, \ldots\}$ условную вероятность $\quad (A\,|\,\eta = y)$ можно определить произвольным образом, например, положить равной нулю.

Если ξ — случайная величина, для которой существует $\quad \xi$, то

$$(\xi\,|\,\eta = y_k) = \frac{1}{\{\eta = y_k\}} \int\limits_{\{\omega\,:\,\eta = y_k\}} \xi\, d \quad .$$

Условное математическое ожидание $\quad (\xi\,|\,\eta = y)$ для $y \notin \{y_1,\, y_2,\, \ldots\}$ определяется произвольно (например, полагается равным нулю).

Пример 2. Пусть $(\xi,\, \eta)$ — пара случайных величин, распределение которых обладает плотностью $f_{\xi,\eta}(x,\, y)$:

$$\{(\xi,\, \eta) \in B\} = \int\limits_B f_{\xi,\eta}(x,\, y)\, dx\, dy, \qquad B \in \mathscr{B}(R^2).$$

Пусть $f_\xi(x)$ и $f_\eta(y)$ — плотности распределения вероятностей случайных величин ξ и η (см. (46), (55), (56) § 6).

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

$$(\xi\in C\mid\eta=y)=\int\limits_C f_{\xi\mid\eta}(x\mid y)\,dx, \qquad C\in\mathscr{B}(R), \tag{19}$$

т. е. $f_{\xi\mid\eta}(x\mid y)$ 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 (43), (45) § 6 и теоремы Фубини

$$\begin{aligned}\int\limits_B\left[\int\limits_C f_{\xi\mid\eta}(x\mid y)\,dx\right]P_\eta(dy)=\int\limits_B\left[\int\limits_C f_{\xi\mid\eta}(x\mid y)\,dx\right]f_\eta(y)\,dy=\\ =\int\limits_{C\times B} f_{\xi\mid\eta}(x\mid y)f_\eta(y)\,dx\,dy=\int\limits_{C\times B} f_{\xi,\eta}(x,\,y)\,dx\,dy=\\ =\ \{(\xi,\,\eta)\in C\times B\}=\ (\{\xi\in C\}\cap\{\eta\in B\}),\end{aligned}$$

что и доказывает (17).

Аналогичным образом устанавливается следующий результат: если ξ существует, то

$$(\xi\mid\eta=y)=\int\limits_{-\infty}^{\infty} x f_{\xi\mid\eta}(x\mid y)\,dx. \tag{20}$$

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 $(\eta-a\mid\eta\geqslant a)$, т. е. среднее время, которое прибор *еще* проработает в предположении, что он уже проработал время a.

Пусть $\{\eta\geqslant a\}>0$. Тогда, согласно определению (см. п. 1) и (45) § 6,

$$(\eta-a\mid\eta\geqslant a)=\frac{[(\eta-a)I_{\{\eta\geqslant a\}}]}{\{\eta\geqslant a\}}=\frac{\int\limits_\Omega(\eta-a)I_{\{\eta\geqslant a\}}\quad(d\omega)}{(\eta\geqslant a)}=\frac{\int\limits_a^\infty(y-a)f_\eta(y)\,dy}{\int\limits_a^\infty f_\eta(y)\,dy}.$$

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 $\quad \eta = \quad (\eta \,|\, \eta \geqslant 0) = 1/\lambda$ и для любого $a > 0 \quad (\eta - a \,|\, \eta \geqslant a) = 1/\lambda$. 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 $\quad \eta$.

В предположении (21) найдем условное распределение $\quad (\eta - a \leqslant x \,|\, \eta \geqslant a)$.

Имеем

$$(\eta - a \leqslant x \,|\, \eta \geqslant a) = \frac{\{a \leqslant \eta \leqslant a + x\}}{\{\eta \geqslant a\}} = \frac{F_\eta(a+x) - F_\eta(a) + \quad \{\eta = a\}}{1 - F_\eta(a) + \quad \{\eta = a\}} =$$

$$= \frac{[1 - e^{-\lambda(a+x)}] - [1 - e^{-\lambda a}]}{1 - [1 - e^{-\lambda a}]} = \frac{e^{-\lambda a}[1 - e^{-\lambda x}]}{e^{-\lambda a}} = 1 - e^{-\lambda x}.$$

Таким образом, условное распределение $\quad (\eta - a \leqslant x \,|\, \eta \geqslant a)$ совпадает с безусловным распределением $\quad \{\eta \leqslant x\}$. Это замечательное свойство экспоненциального распределения является характеристическим: не существует других распределений с плотностями, обладающих свойством $\quad (\eta - a \leqslant x \,|\, \eta \geqslant a) = \quad \{\eta \leqslant x\}$, $a \geqslant 0$, $0 \leqslant x < \infty$.

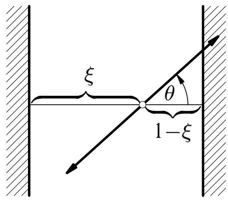


Рис. 29.

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 (на *геометрические вероятности*), 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 ($P_\theta(da) = da$) 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

$$B = \left\{(a, x)\colon\ |a| \leqslant \frac{\pi}{2},\ x \in \left[0,\ \frac{1}{2}\cos a\right] \cup \left[1 - \frac{1}{2}\cos a,\ 1\right]\right\},$$

то $A = \{\omega\colon (\theta,\, \xi) \in B\}$, и, значит, интересующая нас вероятность

$$(A) = \quad I_A(\omega) = \quad I_B(\theta(\omega),\, \xi(\omega)).$$

В силу свойства **G*** и формулы (16)

$$I_B(\theta(\omega),\, \xi(\omega)) = \quad (\quad [I_B(\theta(\omega),\, \xi(\omega))\,|\,\theta(\omega)]) =$$

$$= \int_\Omega \quad [I_B(\theta(\omega),\, \xi(\omega))\,|\,\theta(\omega)] \quad (d\omega) = \int_{-\pi/2}^{\pi/2} \quad [I_B(\theta(\omega),\, \xi(\omega))\,|\,\theta(\omega) = a]\, P_\theta(da) =$$

$$= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \quad I_B(a,\, \xi(\omega))\, da = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos a\, da = \frac{2}{\pi},$$

где мы воспользовались также тем, что

$$I_B(a,\, \xi(\omega)) = \quad \left\{\xi \in \left[0,\, \frac{1}{2}\cos a\right] \cup \left[1 - \frac{1}{2}\cos a\right]\right\} = \cos a.$$

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 η_i равным 1, если при i-м бросании игла пересекает коридор, и равным 0 в противном случае. Тогда в силу закона больших чисел (см., например, (6) § 5 гл. I) для всякого $\varepsilon > 0$

$$\left\{\left|\frac{\eta_1 + \ldots + \eta_N}{N} - \quad (A)\right| > \varepsilon\right\} \to 0, \qquad N \to \infty.$$

В этом смысле частота

$$\frac{\eta_1 + \ldots + \eta_N}{N} \approx \quad (A) = \frac{2}{\pi}$$

и, значит,

$$\frac{2N}{\eta_1 + \ldots + \eta_N} \approx \pi.$$

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 был одним из первых методов (известных теперь под названием «метода Монте-Карло») использования вероятностно-статистических закономерностей в численном анализе.

Замечание. Рассмотренный пример 4 (задача Бюффона) является типичным примером задач на *геометрические вероятности*. Весьма часто в таких задачах из простых геометрических соображений, типа «симметрии», видно, как задавать вероятности «элементарных событий».

(Ср. с пп. 3 и 4 в § 1 главы I и § 3 настоящей главы.) Формулируемые далее задачи 9—12 являются задачами на геометрические вероятности.

7. Если $\{\xi_n\}_{n\geqslant 1}$ — последовательность неотрицательных случайных величин, то, согласно утверждению f) теоремы 2,

$$\Big(\sum \xi_n \,|\, \mathscr{G}\Big) = \sum \quad (\xi_n \,|\, \mathscr{G}) \quad (\text{п. н.}).$$

В частности, если B_1, B_2, … — последовательность попарно непересекающихся множеств, то

$$\Big(\sum B_n \,|\, \mathscr{G}\Big) = \sum \quad (B_n \,|\, \mathscr{G}) \quad (\text{п. н.}). \tag{22}$$

Важно подчеркнуть, что это равенство выполнено лишь *почти наверное* и, следовательно, условную вероятность $(B \,|\, \mathscr{G})(\omega)$ нельзя рассматривать при *фиксированном* ω как *меру* по B. Можно было бы подумать, что, за исключением некоторого множества $\mathscr{N}$ меры нуль, $(\cdot \,|\, \mathscr{G})(\omega)$ является все же мерой для $\omega \in \overline{\mathscr{N}}$, где $\mathscr{N}$ — некоторое множество меры нуль. Однако это, вообще говоря, не так в силу следующего обстоятельства. Обозначим $\mathscr{N}(B_1, B_2, \ldots)$ то множество исходов ω, где для заданных B_1, B_2, … не выполнено свойство счетной аддитивности (22). Тогда исключительное множество $\mathscr{N}$ есть

$$\mathscr{N} = \bigcup \mathscr{N}(B_1, B_2, \ldots), \tag{23}$$

где объединение берется по всем непересекающимся B_1, B_2, … из $\mathscr{F}$. Хотя -мера каждого множества $\mathscr{N}(B_1, B_2, \ldots)$ равна нулю, -мера множества $\mathscr{N}$ может оказаться (в силу несчетности объединения в (23)) ненулевой. (Вспомним, что лебеговская мера отдельной точки равна нулю, а мера множества $\mathscr{N} = [0, 1)$, являющегося несчетной суммой одноточечных множеств $\{x\}$, $0 \leqslant x < 1$, равна единице.)

В то же время было бы удобно, чтобы условная вероятность $(\cdot \,|\, \mathscr{G})(\omega)$ являлась мерой для каждого $\omega \in \Omega$, поскольку тогда, например, подсчет условных математических ожиданий $(\xi \,|\, \mathscr{G})(\omega)$ можно было бы осуществлять (см. далее теорему 3) просто с помощью усреднения по мере $(\cdot \,|\, \mathscr{G})(\omega)$:

$$(\xi \,|\, \mathscr{G})(\omega) = \int\limits_{\Omega} \xi(\tilde{\omega}) \quad (d\tilde{\omega} \,|\, \mathscr{G})(\omega) \quad (\text{п. н.})$$

(ср. с (10) § 8 гл. I).

Введем такое

Определение 6. Функцию $P(\omega; B)$, определенную для всех $\omega \in \Omega$ и $B \in \mathscr{F}$, назовем *регулярной условной вероятностью* относительно σ-алгебры $\mathscr{G} \subseteq \mathscr{F}$, если:

a) для *каждого* $\omega \in \Omega$ $P(\omega;\, \cdot)$ есть *вероятностная мера* на $\mathscr{F}$;

b) для *каждого* $B \in \mathscr{F}$ $P(\omega;\, B)$ как функция от ω есть *один из вариантов условной вероятности* $(B\,|\,\mathscr{G})(\omega)$, т. е. $P(\omega;\, B) = (B\,|\,\mathscr{G})(\omega)$ (п. н.).

Теорема 3. *Пусть $P(\omega;\, B)$ — регулярная условная вероятность относительно $\mathscr{G}$ и ξ — интегрируемая случайная величина. Тогда*

$$(\xi\,|\,\mathscr{G})(\omega) = \int\limits_{\Omega} \xi(\tilde{\omega})\, P(\omega;\, d\tilde{\omega}) \quad (\textit{п. н.}). \tag{24}$$

Доказательство. Если $\xi = I_B$, $B \in \mathscr{F}$, то требуемая формула (24) превращается в равенство

$$(B\,|\,\mathscr{G})(\omega) = P(\omega;\, B) \quad (\text{п. н.}),$$

выполнимое в силу определения 6 b). Следовательно, (24) выполнено для простых функций.

Пусть теперь $\xi \geqslant 0$ и $\xi_n \uparrow \xi$, где ξ_n — простые функции. Тогда по свойству b) теоремы 2 $(\xi\,|\,\mathscr{G})(\omega) = \lim\limits_n (\xi_n\,|\,\mathscr{G})(\omega)$ (п. н.). Но поскольку для каждого $\omega \in \Omega$ $P(\omega;\, \cdot)$ есть мера, то по теореме о монотонной сходимости

$$\lim_n (\xi_n\,|\,\mathscr{G})(\omega) = \lim_n \int\limits_{\Omega} \xi_n(\tilde{\omega})\, P(\omega;\, d\tilde{\omega}) = \int\limits_{\Omega} \xi(\tilde{\omega})\, P(\omega;\, d\tilde{\omega}).$$

Общий случай сводится к рассмотренному с помощью представления $\xi = \xi^+ - \xi^-$. $\square$

Следствие. *Пусть $\mathscr{G} = \mathscr{G}_\eta$, где η — случайная величина, причем пара $(\xi,\, \eta)$ имеет плотность распределения вероятностей $f_{\xi,\eta}(x,\, y)$. Пусть $|g(\xi)| < \infty$. Тогда*

$$(g(\xi)|\eta = y) = \int\limits_{-\infty}^{\infty} g(x) f_{\xi\,|\,\eta}(x\,|\,y)\, dx,$$

где $f_{\xi\,|\,\eta}(x\,|\,y)$ — плотность условного распределения (*см.* (18)).

Чтобы сформулировать основной результат о *существовании* регулярных условных вероятностей, нам понадобятся следующие определения.

Определение 7. Пусть $(E,\, \mathscr{E})$ — измеримое пространство, $X = X(\omega)$ — 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 σ-алгебры $\mathscr{G}$*, 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* $(X \in B\,|\,\mathscr{G})(\omega)$, т. е.

$$Q(\omega;\, B) = (X \in B\,|\,\mathscr{G})(\omega) \quad (\text{п. н.}).$$

Определение 8. Пусть ξ — случайная величина. Функция $F = F(\omega;\, x)$, $\omega \in \Omega$, $x \in R$, называется *регулярной функцией распределения для* ξ *относительно* σ*-алгебры* $\mathscr{G}$, если:

a) для *каждого* $\omega \in \Omega$ $F(\omega;\, x)$ есть *функция распределения* на R;

b) для *каждого* $x \in R$ $F(\omega;\, x) = \quad (\xi \leqslant x \,|\, \mathscr{G})(\omega)$ (п. н.).

Теорема 4. *Всегда существуют регулярная функция распределения и регулярное условное распределение случайной величины* ξ *относительно* σ*-алгебры* $\mathscr{G} \subseteq \mathscr{F}$.

Доказательство. Для каждого рационального $r \in R$ положим $F_r(\omega) = \quad (\xi \leqslant r \,|\, \mathscr{G})(\omega)$, где $\quad (\xi \leqslant r \,|\, \mathscr{G})(\omega) = \quad (I_{\{\xi \leqslant r\}} \,|\, \mathscr{G})(\omega)$ — 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*** $\quad (\xi \leqslant r_i \,|\, \mathscr{G}) \leqslant \quad (\xi \leqslant r_j \,|\, \mathscr{G})$ (п. н.) и, значит, если $A_{ij} = \{\omega\colon F_{r_j}(\omega) < F_{r_i}(\omega)\}$, $A = \bigcup A_{ij}$, то $\quad (A) = 0$. Иначе говоря, множество тех ω, где у функций распределений $F_r(\omega)$, $r \in \{r_i\}$, нарушается монотонность, имеет меру нуль.

Пусть теперь $B_i = \{\omega\colon \lim\limits_{n\to\infty} F_{r_i+1/n}(\omega) \neq F_{r_i}(\omega)\}$, $B = \bigcup\limits_{i=1}^{\infty} B_i$. Ясно, что $I_{\{\xi \leqslant r_i+1/n\}} \downarrow I_{\{\xi \leqslant r_i\}}$, $n \to \infty$. Поэтому, согласно утверждению a) теоремы 2, $F_{r_i+1/n}(\omega) \to F_{r_i}(\omega)$ (п. н.) и, значит, множество B, где нарушается непрерывность справа (по рациональным числам), также имеет меру нуль, $\quad (B) = 0$.

Далее, пусть $C = \{\omega\colon \lim\limits_{n\to\infty} F_n(\omega) \neq 1\} \cup \{\omega\colon \lim\limits_{n\to-\infty} F_n(\omega) \neq 0\}$. Тогда, поскольку $\{\xi \leqslant n\} \uparrow \Omega$, $n \to \infty$, а $\{\xi \leqslant n\} \downarrow \varnothing$, $n \to -\infty$, то $\quad (C) = 0$.

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 8.

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 a) в определении 6.

Согласно конструкции, $\quad (\xi \leqslant r \,|\, \mathscr{G})(\omega) = F_r(\omega) = F(\omega;\, r)$. Если $r \downarrow x$, то $F(\omega;\, r) \downarrow F(\omega;\, x)$ для всех $\omega \in \Omega$ в силу установленной непрерывности справа. Но из утверждения a) теоремы 2 $\quad (\xi \leqslant r \,|\, \mathscr{G})(\omega) \to \quad (\xi \leqslant x \,|\, \mathscr{G})(\omega)$ (п. н.).

Поэтому $F(\omega;\, x) = \ \ (\xi \leqslant x \,|\, \mathscr{G})(\omega)$ (п. н.), что и доказывает свойство b) определения 8.

Обратимся теперь к доказательству существования регулярного условного распределения ξ относительно $\mathscr{G}$.

Пусть $F(\omega;\, x)$ — построенная выше функция. Положим

$$Q(\omega;\, B) = \int\limits_B F(\omega;\, dx),$$

где интеграл понимается как интеграл Лебега—Стилтьеса, из свойств которого (см. п. 8 § 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 $\ \ (\xi \in B \,|\, \mathscr{G})(\omega)$, воспользуемся *принципом подходящих множеств*.

Пусть $\mathscr{C}$ — совокупность множеств B из $\mathscr{B}(R)$, для которых $Q(\omega;\, B) = \ \ (\xi \in B \,|\, \mathscr{G})(\omega)$ (п. н.). Поскольку $F(\omega;\, x) = \ \ (\xi \leqslant x \,|\, \mathscr{G})(\omega)$ (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 $\mathscr{C} = \mathscr{B}(R)$. □

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 9. Измеримое пространство $(E, \mathscr{E})$ называется *борелевским пространством*, если оно борелевски эквивалентно некоторому борелевскому подмножеству числовой прямой, т. е. существует взаимно однозначное отображение $\varphi = \varphi(e)\colon (E, \mathscr{E}) \to (R, \mathscr{B}(R))$ такое, что:

1) $\varphi(E) \equiv \{\varphi(e)\colon e \in E\}$ 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 σ-алгебры $\mathscr{G} \subseteq \mathscr{F}$.*

Доказательство. Пусть $\varphi = \varphi(e)$ — функция из определения 9. 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 $\widetilde{Q}(\omega;\, B) = Q(\omega;\, \varphi(B))$, $B \in \mathscr{E}$. В силу 3) определения 9 $\varphi(B) \in \varphi(E) \cap \mathscr{B}(R)$ и, следовательно, $\widetilde{Q}(\omega;\, B)$ определено. Понятно, что при каждом ω $\widetilde{Q}(\omega;\, B)$ является мерой по $B \in \mathscr{E}$. Зафиксируем теперь $B \in \mathscr{E}$. В силу взаимной однозначности отображения $\varphi = \varphi(e)$

$$\widetilde{Q}(\omega;\, B) = Q(\omega;\, \varphi(B)) = \quad (\varphi(X) \in \varphi(B) \,|\, \mathscr{G})(\omega) = \quad (X \in B \,|\, \mathscr{G})(\omega) \quad \text{(п. н.)}.$$

Таким образом, $\widetilde{Q}(\omega;\, B)$ является регулярным условным распределением X относительно $\mathscr{G}$. □

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 $(E, \mathscr{E})$ 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 $\ (A_i) > 0$, то *теорема Байеса* (9) § 3 гл. I утверждает, что для всякого B с $\ (B) > 0$

$$(A_i \,|\, B) = \frac{(A_i) \quad (B \,|\, A_i)}{\sum\limits_{j=1}^{n} \ (A_j) \quad (B \,|\, A_j)}. \tag{25}$$

Поэтому, если $\theta = \sum\limits_{i=1}^{n} a_i I_{A_i}$ — дискретная случайная величина, то, согласно формуле (10) из § 8 гл. I,

$$[g(\theta) \,|\, B] = \frac{\sum\limits_{i=1}^{n} g(a_i) \quad (A_i), \quad (B \,|\, A_i)}{\sum\limits_{j=1}^{n} \ (A_j) \quad (B \,|\, A_j)}, \tag{26}$$

или

$$[g(\theta) \,|\, B] = \frac{\int\limits_{-\infty}^{\infty} g(a) \quad (B \,|\, \theta = a)\, P_\theta(da)}{\int\limits_{-\infty}^{\infty} \ (B \,|\, \theta = a)\, P_\theta(da)}, \tag{27}$$

где $P_\theta(A) = \quad \{\theta \in A\}$.

Основываясь на приведенном в этом параграфе определении $\ [g(\theta)|B]$, нетрудно установить, что формула (27) остается справедливой для любого события B с $\ (B)>0$, случайных величин $\theta=\theta(\omega)$ и функций $g=g(a)$ с $\ |g(\theta)|<\infty$.

Рассмотрим теперь аналог формулы (27) для условных математических ожиданий $\ [g(\theta)\,|\,\mathscr{G}]$ относительно некоторой σ-алгебры $\mathscr{G}$, $\mathscr{G}\subseteq\mathscr{F}$.

Пусть

$$(B)=\int_B g(\theta(\omega))\quad(d\omega),\qquad B\in\mathscr{G}. \tag{28}$$

Тогда в силу (4)

$$[g(\theta)\,|\,\mathscr{G}](\omega)=\frac{d}{d}(\omega). \tag{29}$$

Наряду с σ-алгеброй $\mathscr{G}$ рассмотрим σ-алгебру $\mathscr{G}_\theta$. Тогда, согласно (5),

$$(B)=\int_\Omega\quad(B\,|\,\mathscr{G}_\theta)\,d\quad, \tag{30}$$

или, по формуле замены переменных под знаком интеграла Лебега,

$$(B)=\int_{-\infty}^{\infty}\quad(B\,|\,\theta=a)\,P_\theta(da). \tag{31}$$

Поскольку

$$(B)=\quad[g(\theta)I_B]=\quad[g(\theta)\quad(I_B\,|\,\mathscr{G}_\theta)],$$

то

$$(B)=\int_{-\infty}^{\infty}g(a)\quad(B\,|\,\theta=a)\,P_\theta(da). \tag{32}$$

Предположим теперь, что условные вероятности $\ (B\,|\,\theta=a)$ являются регулярными и допускают представление

$$(B\,|\,\theta=a)=\int_B\rho(\omega;\,a)\,\lambda(d\omega), \tag{33}$$

где $\rho=\rho(\omega;\,a)$ — неотрицательная измеримая по паре переменных функция, а λ — некоторая σ-конечная мера на $(\Omega,\,\mathscr{G})$.

Пусть $\ |g(\theta)|<\infty$. Покажем, что (-п. н.)

$$[g(\theta)|\mathscr{G}](\omega)=\frac{\displaystyle\int_{-\infty}^{\infty}g(a)\,\rho(\omega;\,a)\,P_\theta(da)}{\displaystyle\int_{-\infty}^{\infty}\rho(\omega;\,a)\,P_\theta(da)} \tag{34}$$

(*обобщенная теорема Байеса*).

Для доказательства (34) нам понадобится следующая лемма.

Лемма. *Пусть* $(\Omega, \mathscr{F})$ — *некоторое измеримое пространство.*

a) *Пусть* μ *и* λ — σ*-конечные меры,* $\mu \ll \lambda$ *и* $f = f(\omega)$ — $\mathscr{F}$*-измеримая функция. Тогда*

$$\int_\Omega f\,d\mu = \int_\Omega f\,\frac{d\mu}{d\lambda}\,d\lambda \tag{35}$$

(*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*

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda} \quad (\lambda\text{-п. н.}) \tag{36}$$

и

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\lambda} \Big/ \frac{d\mu}{d\lambda} \quad (\mu\text{-п. н.}). \tag{37}$$

Доказательство. a) Поскольку

$$\mu(A) = \int_A \left(\frac{d\mu}{d\lambda}\right) d\lambda, \qquad A \in \mathscr{F},$$

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 (ср. с доказательством (39) в § 6).

b) Из утверждения a) с $f = \dfrac{d\nu}{d\mu}$ находим

$$\nu(A) = \int_A \left(\frac{d\nu}{d\mu}\right) d\mu = \int_A \left(\frac{d\nu}{d\mu}\right)\left(\frac{d\mu}{d\lambda}\right) d\lambda.$$

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 $\left\{\omega\colon \dfrac{d\mu}{d\lambda} = 0\right\}$ правую часть в (37) можно определить произвольно, например, положить равной нулю). $\square$

Для доказательства (34) заметим, что в силу теоремы Фубини и предположения (33)

$$(B) = \int_B \left[\int_{-\infty}^{\infty} g(a)\rho(\omega;\, a)\, P_\theta(da) \right] \lambda(d\omega), \tag{38}$$

$$(B) = \int_B \left[\int_{-\infty}^{\infty} \rho(\omega;\, a)\, P_\theta(da) \right] \lambda(d\omega). \tag{39}$$

Тогда в силу леммы

$$\frac{d\ \ }{d\ \ } = \frac{d\ \ /d\lambda}{d\ \ /d\lambda} \quad (\ \ \text{-п. н.}),$$

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, $\mathscr{G} = \sigma(\xi)$.

Предположим, что

$$(\xi \in A \,|\, \theta = a) = \int_A q(x;\, a)\, \lambda(dx), \qquad A \in \mathscr{B}(R), \tag{40}$$

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

$$[g(\theta) \,|\, \xi = x] = \frac{\int_{-\infty}^{\infty} g(a)\, q(x;\, a)\, P_\theta(da)}{\int_{-\infty}^{\infty} q(x;\, a)\, P_\theta(da)}. \tag{41}$$

Пусть, в частности, $(\theta,\, \xi)$ — 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

$$[g(\theta) \,|\, \xi = x_j] = \frac{\sum\limits_i g(a_i)\ \ (\xi = x_j \,|\, \theta = a_i)\ \ \{\theta = a_i\}}{\sum\limits_i\ \ (\xi = x_j \,|\, \theta = a_i)\ \ \{\theta = a_i\}}. \tag{42}$$

(Ср. с (26).)

Пусть теперь $(\theta,\ \xi)$ — пара абсолютно непрерывных величин с плотностью $f_{\theta,\xi}(a,\ x)$. Тогда в силу (19) представление (40) выполнено с $q(x;\ a) = f_{\xi\,|\,\theta}(x\,|\,a)$ и мерой Лебега λ. Поэтому

$$[\,g(\theta)\,|\,\xi = x] = \frac{\int\limits_{-\infty}^{\infty} g(a)\, f_{\xi\,|\,\theta}(x\,|\,a)\, f_\theta(a)\, da}{\int\limits_{-\infty}^{\infty} f_{\xi\,|\,\theta}(x\,|\,a)\, f_\theta(a)\, da}. \tag{43}$$

9. Приведем еще одну версию обобщенной теоремы Байеса (см. (34)), формулировка которой (даваемая ниже) особенно удобна в вопросах, связанных с заменой вероятностных мер.

Теорема 6. *Пусть и $\tilde{}$ — две вероятностные меры на измеримом пространстве* $(\Omega,\ \mathscr{F})$, *мера* $\tilde{}$ *абсолютно непрерывна по мере (обозначение:* $\tilde{} \ll$ *) и* $\dfrac{d\tilde{}}{d}$ *— производная Радона—Никодима меры* $\tilde{}$ *по мере . Пусть* $\mathscr{G}$ *— под-*σ*-алгебра* $\mathscr{F}$ $(\mathscr{G} \subseteq \mathscr{F})$ *и* $(\cdot\,|\,\mathscr{G})$ *и* $\tilde{}(\cdot\,|\,\mathscr{G})$ *— условные математические ожидания относительно* $\mathscr{G}$ *по мерам и* $\tilde{}$ *соответственно. Пусть* ξ *— неотрицательная* ($\mathscr{F}$*-измеримая) случайная величина. Тогда имеет место следующая «формула пересчета в условных математических ожиданиях»:*

$$\tilde{}(\xi\,|\,\mathscr{G}) = \frac{\left(\xi \dfrac{d\tilde{}}{d}\,\Big|\,\mathscr{G}\right)}{\left(\dfrac{d\tilde{}}{d}\,\Big|\,\mathscr{G}\right)} \quad (\tilde{}\text{-п. н.}). \tag{44}$$

Формула (44) *справедлива и для всякой случайной величины* ξ, *для которой условное математическое ожидание* $\tilde{}(\xi\,|\,\mathscr{G})$ *определено.*

Доказательство. Прежде всего, заметим, что $\tilde{}$-мера (а также и -мера) события $\left\{\omega\colon \ \left(\dfrac{d\tilde{}}{d}\,\Big|\,\mathscr{G}\right) = 0\right\}$ равна нулю. Действительно, если множество $A \in \mathscr{G}$, то

$$\int\limits_A \left(\frac{d\tilde{}}{d}\,\Big|\,\mathscr{G}\right) d \ = \int\limits_A \frac{d\tilde{}}{d}\, d \ = \int\limits_A d\tilde{} = \tilde{}(A),$$

и, значит, множество $A = \left\{\omega\colon \ \left(\dfrac{d\tilde{}}{d}\,\Big|\,\mathscr{G}\right) = 0\right\}$ имеет $\tilde{}$-меру нуль.

Пусть $\xi \geqslant 0$. По определению условного математического ожидания, $\tilde{}(\xi\,|\,\mathscr{G})$ — это есть такая $\mathscr{G}$-измеримая случайная величина, что для всякого $A \in \mathscr{G}$

$$\tilde{}[I_A \tilde{}(\xi\,|\,\mathscr{G})] = \tilde{}[I_A \xi]. \tag{45}$$

Поэтому для доказательства формулы (44) надо лишь убедиться в том, что ($\mathscr{G}$-измеримая) случайная величина, стоящая в правой части (44), удовлетворяет следующему равенству:

$$\tilde{}\left[I_A\cdot\frac{1}{\left(\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)}\cdot\quad\left(\xi\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)\right]=\tilde{}\,[I_A\xi]. \tag{46}$$

Пользуясь свойствами условных математических ожиданий и формулой (39) из § 6, находим:

$$\tilde{}\left[I_A\cdot\frac{1}{\left(\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)}\cdot\quad\left(\xi\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)\right]=\quad\left[I_A\cdot\frac{1}{\left(\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)}\cdot\quad\left(\xi\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)\cdot\frac{d\tilde{}}{d}\right]=$$

$$=\quad\left[I_A\cdot\frac{1}{\left(\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)}\cdot\quad\left(\xi\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)\cdot\quad\left(\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)\right]=$$

$$=\quad\left[I_A\quad\left(\xi\frac{d\tilde{}}{d}\middle|\mathscr{G}\right)\right]=\quad\left[I_A\xi\frac{d\tilde{}}{d}\right]=\tilde{}\,[I_A\xi],$$

что и доказывает формулу (45) для неотрицательных величин ξ. Переход к общему случаю аналогичен доказательству утверждения (39) в § 6 для произвольных интегрируемых случайных величин ξ. $\square$

10. Приведенная выше обобщенная теорема Байеса (формулы (34), (41) и (43)), являющаяся одним из основных средств при «байесовском подходе» в математической статистике, дает ответ на вопрос о том, как перераспределяется наше знание о распределении *случайной величины* θ в зависимости от результатов наблюдений над статистически с ней связанной случайной величиной ξ.

Ниже будет рассматриваться еще одно применение понятия условного математического ожидания в задачах оценивания *неизвестного параметра* θ по результатам наблюдений. (Подчеркнем, что, в отличие от рассмотренного выше случая, где θ — *случайная величина*, сейчас θ будет просто некоторым параметром из *a priori* данного параметрического множества Θ (ср. с § 7 гл. I).

В сущности, речь будет идти об одном важном понятии в математической статистике — понятии *достаточной под-σ-алгебры* (говорят также: *σ-подалгебры*).

Пусть $(\Omega,\mathscr{F})$ — некоторое измеримое пространство, на котором задано семейство $\mathscr{P}=\{\ {}_\theta,\ \theta\in\Theta\}$ вероятностных мер ${}_\theta$, зависящих от параметра θ, принадлежащего *параметрическому множеству* Θ. Часто говорят (ср. с § 7 гл. I), что набор $\mathscr{E}=(\Omega,\ \mathscr{F},\ \mathscr{P})$ задает *вероятностно-статистическую модель* (или *вероятностно-статистический эксперимент*).

Для пояснения приводимого ниже определения 10 предположим, что у нас имеется некотороая $\mathscr{F}$-измеримая функция $T=T(\omega)$ (статистика) от исходов ω и $\mathscr{G}=\sigma(T(\omega))$ — σ-алгебра, порожденная этой функцией. Ясно, что $\mathscr{G}\subseteq\mathscr{F}$ и, вообще говоря, в $\mathscr{F}$ могут входить события, и не принадлежащие $\mathscr{G}$ (т. е. $\mathscr{F}$ «богаче», нежели $\mathscr{G}$). Но вполне может оказаться, что с точки зрения определения того, какой «действует» параметр θ, никакой больше «информации», кроме знания $T=T(\omega)$, и не надо иметь. В этом смысле статистику T естественно было бы назвать *достаточной*.

Замечание 1. В случае $T(\omega)=\omega$, т. е. когда известными становятся сами *исходы* (а не функции от них), можно выделить следующие два крайних случая.

Один — когда все вероятности ${}_{\theta}$ *совпадают* при всех $\theta\in\Theta$. Понятно, что в этом случае никакой исход ω нам информации о значениях параметра θ не даст.

Другой случай — это когда *носители* всех мер ${}_{\theta}$, $\theta\in\Theta$, «сидят» на разных подмножествах $\mathscr{F}$ (т. е. для любых двух параметров θ_1 и θ_2 меры ${}_{\theta_1}$ и ${}_{\theta_2}$ являются *сингулярными*: существуют два множества (носителя) Ω_{θ_1} и Ω_{θ_2} такие, что ${}_{\theta_1}(\Omega\setminus\Omega_{\theta_1})=0$, ${}_{\theta_2}(\Omega\setminus\Omega_{\theta_2})=0$ и $\Omega_{\theta_1}\cap\Omega_{\theta_2}=\varnothing$). В этом случае по полученному исходу ω значение θ восстанавливается «однозначно».

Оба эти случая мало интересны. Интерес же представляют случаи, когда, скажем, все меры ${}_{\theta}$ являются *эквивалентными* (и тогда их носители не различаются) или же, когда для этих мер выполнено (более слабое, нежели эквивалентность) свойство *доминируемости*: существует σ-конечная мера λ такая, что меры ${}_{\theta}\ll\lambda$ при всех $\theta\in\Theta$. Роль этого свойства доминируемости, обычно предполагаемого выполнения в общей статистической теории (что позволяет исключить некоторые теоретико-множественные патологии), полно раскрывается в приводимой ниже факторизационной теореме 7.

Следующее определение можно рассматривать как один из способов формализации понятия достаточности «информации», содержащейся в некоторых под-σ-алгебрах $\mathscr{G}\subseteq\mathscr{F}$.

Определение 10. Пусть $(\Omega,\ \mathscr{F},\ \mathscr{P})$ — вероятностно-статистическая модель, $\mathscr{P}=\{\ {}_{\theta},\ \theta\in\Theta\}$ и $\mathscr{G}$ есть под-σ-алгебра $\mathscr{F}$ $(\mathscr{G}\subseteq\mathscr{F})$. Говорят, что под-$\sigma$-алгебра $\mathscr{G}$ является *достаточной* для семейства $\mathscr{P}$, если существуют варианты условных вероятностей ${}_{\theta}(\cdot\,|\,\mathscr{G})(\omega)$, $\theta\in\Theta$, $\omega\in\Omega$, которые не зависят от значений θ: найдется такая функция $P(A;\ \omega)$, $A\in\mathscr{F}$, $\omega\in\Omega$, что для всех $A \in \mathscr{F}$ и $\theta \in \Theta$

$$ {}_{\theta}(A \,|\, \mathscr{G})(\omega) = P(A;\, \omega) \quad ({}_{\theta}\text{-п. н.}), \tag{47}$$

иначе говоря, $P(A;\, \omega)$ есть версия ${}_{\theta}(A \,|\, \mathscr{G})(\omega)$ при всех $\theta \in \Theta$.

Если $\mathscr{G} = \sigma(T(\omega))$, то статистика $T = T(\omega)$ называется *достаточной статистикой* для семейства $\mathscr{P}$.

Замечание 2. Как уже было отмечено выше, интерес к отысканию в статистических исследованиях *достаточных* статистик обусловлен стремлением находить такие функции $T = T(\omega)$ от исходов ω, которые *редуцируют данные с сохранением информации* (о значениях параметра θ). Например, представим себе, что $\omega = (x_1, x_2, \ldots, x_n)$, где $x_i \in R$ и n очень велико. Тогда построение «хороших» оценок параметра θ (как, скажем, в § 7 гл. I) может оказаться весьма трудной задачей по причине высокой размерности имеющихся данных $x_1, x_2, \ldots, x_n$. Однако вполне может случиться (и это мы наблюдали в § 7 гл. I), что для построения «хороших» оценок вполне достаточно знания не *индивидуальных* значений $x_1, x_2, \ldots, x_n$, а лишь значения *суммарной* статистики $T(\omega) = x_1 + x_2 + \ldots + x_n$.

Понятно, что такая статистика действительно приводит к существенной редукции данных (и вычислительной работы), будучи в то же самое время достаточной для построения «хороших» оценок параметра θ.

Приводимая ниже *факторизационная теорема* дает условия, обеспечивающие достаточность σ-алгебры $\mathscr{G}$ для семейства $\mathscr{P}$.

Теорема 7. *Пусть $\mathscr{P} = \{ {}_{\theta},\, \theta \in \Theta\}$ — доминируемое семейство, т. е. существует такая σ-конечная мера λ на $(\Omega, \mathscr{F})$, что меры ${}_{\theta}$ абсолютно непрерывны относительно λ (${}_{\theta} \ll \lambda$) при всех $\theta \in \Theta$.*

Пусть $g_\theta^{(\lambda)}(\omega) = \dfrac{d\ {}_{\theta}}{d\lambda}(\omega)$ — производная Радона—Никодима меры ${}_{\theta}$ относительно меры λ.

Для того чтобы под-σ-алгебра $\mathscr{G}$ была достаточной для семейства $\mathscr{P}$, необходимо и достаточно, чтобы функции $g_\theta^{(\lambda)}(\omega)$ допускали факторизацию: нашлись такие неотрицательные функции $\hat{g}_\theta^{(\lambda)}(\omega)$ и $h(\omega)$, что $\hat{g}_\theta^{(\lambda)}(\omega)$ — $\mathscr{G}$-измеримы, $h(\omega)$ — $\mathscr{F}$-измерима и для всех $\theta \in \Theta$

$$ g_\theta^{(\lambda)}(\omega) = \hat{g}_\theta^{(\lambda)}(\omega)\, h(\omega) \quad (\lambda\text{-п. н.}). \tag{48}$$

Если в качестве меры λ может быть взята мера ${}_{\theta_0}$, где θ_0 — некоторый параметр из Θ, то для достаточности σ-алгебры $\mathscr{G}$ необходимо и достаточно, чтобы сама производная $g_\theta^{(\theta_0)}(\omega) = \dfrac{d\ {}_{\theta}}{d\ {}_{\theta_0}}$ была $\mathscr{G}$-измеримой.

Доказательство. Достаточность. По предположению доминирующая мера λ является σ-конечной. Это означает, что найдутся такие $\mathscr{F}$-измеримые непересекающиеся множества Ω_k, $k \geqslant 1$, что $\Omega = \sum\limits_{k\geqslant 1} \Omega_k$ и $0 < \lambda(\Omega_k) < \infty$, $k \geqslant 1$.

Образуем меру

$$\tilde{\lambda}(\cdot) = \sum_{k\geqslant 1} \frac{1}{2^k} \frac{\lambda(\Omega_k \cap \cdot)}{1+\lambda(\Omega_k)}.$$

Эта мера конечна, $\tilde{\lambda}(\Omega) < \infty$, и $\tilde{\lambda}(\Omega) > 0$. Без ограничения общности ее можно считать вероятностной, $\tilde{\lambda}(\Omega) = 1$.

Тогда из формулы (44) пересчета условных математических ожиданий находим, что для всякой $\mathscr{F}$-измеримой ограниченной случайной величины $X = X(\omega)$

$$ {}_{\theta}(X \mid \mathscr{G}) = \frac{{}_{\tilde{\lambda}}\Big(X \dfrac{d\ {}_{\theta}}{d\tilde{\lambda}} \Big| \mathscr{G}\Big)}{{}_{\tilde{\lambda}}\Big(\dfrac{d\ {}_{\theta}}{d\tilde{\lambda}} \Big| \mathscr{G}\Big)} \quad ({}_{\theta}\text{-п. н.}). \tag{49}$$

Согласно формуле (48),

$$g_{\theta}^{(\tilde{\lambda})} = \frac{d\ {}_{\theta}}{d\tilde{\lambda}} = \frac{d\ {}_{\theta}}{d\lambda} \cdot \frac{d\lambda}{d\tilde{\lambda}} = g_{\theta}^{(\lambda)} \frac{d\lambda}{d\tilde{\lambda}} = \hat{g}_{\theta}^{(\lambda)} h \frac{d\lambda}{d\tilde{\lambda}}. \tag{50}$$

Поэтому формула (49) принимает такой вид:

$$ {}_{\theta}(X \mid \mathscr{G}) = \frac{{}_{\tilde{\lambda}}\Big(X \hat{g}_{\theta}^{(\lambda)} h \dfrac{d\lambda}{d\tilde{\lambda}} \Big| \mathscr{G}\Big)}{{}_{\tilde{\lambda}}\Big(\hat{g}_{\theta}^{(\lambda)} h \dfrac{d\lambda}{d\tilde{\lambda}} \Big| \mathscr{G}\Big)} \quad ({}_{\theta}\text{-п. н.}). \tag{51}$$

Но $\hat{g}_{\theta}^{(\lambda)}$ — $\mathscr{G}$-измеримы и

$$ {}_{\theta}(X \mid \mathscr{G}) = \frac{{}_{\tilde{\lambda}}\Big(Xh \dfrac{d\lambda}{d\tilde{\lambda}} \Big| \mathscr{G}\Big)}{{}_{\tilde{\lambda}}\Big(h \dfrac{d\lambda}{d\tilde{\lambda}} \Big| \mathscr{G}\Big)} \quad ({}_{\theta}\text{-п. н.}). \tag{52}$$

Правая часть здесь не зависит от θ, и, следовательно, имеет место свойство (47). Тем самым, согласно определению 10, σ-алгебра $\mathscr{G}$ является достаточной.

Необходимость будет доказываться лишь при дополнительном предположении, что семейство $\mathscr{P} = \{\ {}_{\theta}, \theta \in \Theta\}$ не только доминируется некоторой σ-конечной мерой λ, но также существует некоторое θ_0 из Θ такое, что все меры ${}_{\theta} \ll {}_{\theta_0}$, т. е. при любом $\theta \in \Theta$ мера ${}_{\theta}$ абсолютно непрерывна относительно меры $\ _{\theta_0}$. (В общем случае доказательство становится более сложным; см. теорему 34.6 в [106].)

Итак, пусть $\mathscr{G}$ — достаточная σ-алгебра, т. е. выполнено свойство (47). Покажем, что в предположении $\ _\theta \ll \ _{\theta_0}$, $\theta \in \Theta$, производная $g_\theta^{(\theta_0)} = \dfrac{d\ \ _\theta}{d\ \ _{\theta_0}}$ является $\mathscr{G}$-измеримой функцией при каждом $\theta \in \Theta$.

Пусть $A \in \mathscr{F}$. Тогда для любого $\theta \in \Theta$, пользуясь свойствами условных математических ожиданий, находим, что $\left(\text{с } g_\theta^{(\theta_0)} = \dfrac{d\ \ _\theta}{d\ \ _{\theta_0}}\right)$

$$\begin{aligned} _\theta(A) = \ _\theta I_A = \ _\theta\ \ _\theta(I_A\,|\,\mathscr{G}) = \ _\theta\ \ _{\theta_0}(I_A\,|\,\mathscr{G}) = \ _{\theta_0}[g_\theta^{(\theta_0)}\ \ _{\theta_0}(I_A\,|\,\mathscr{G})] = \\ = \ _{\theta_0}\ \ _{\theta_0}(g_\theta^{(\theta_0)}\ \ _{\theta_0}(I_A\,|\,\mathscr{G})\,|\,\mathscr{G}) = \ _{\theta_0}([\ \ _{\theta_0}(g_\theta^{(\theta_0)}\,|\,\mathscr{G})]\cdot[\ \ _{\theta_0}(I_A\,|\,\mathscr{G})]) = \\ = \ _{\theta_0}\ \ _{\theta_0}(I_A\ \ _{\theta_0}(g_\theta^{(\theta_0)}\,|\,\mathscr{G})\,|\,\mathscr{G}) = \ _{\theta_0} I_A\ \ _{\theta_0}(g_\theta^{(\theta_0)}\,|\,\mathscr{G}) = \int_A \ _{\theta_0}(g_\theta^{(\theta_0)}\,|\,\mathscr{G})\,d\ \ _{\theta_0}. \end{aligned}$$

Следовательно, вариантом производной $g_\theta^{(\theta_0)} = \dfrac{d\ \ _\theta}{d\ \ _{\theta_0}}$ является $\mathscr{G}$-*измеримая* функция $\ _{\theta_0}(g_\theta^{(\theta_0)}\,|\,\mathscr{G})$.

Тем самым, в том случае, когда $\lambda = \ _{\theta_0}$, из достаточности σ-алгебры $\mathscr{G}$ вытекает факторизационное свойство (48) с $\hat{g}_\theta^{(\theta_0)} = g_\theta^{(\theta_0)}$ и $h \equiv 1$.

В общем же случае (опять же при дополнительном предположении $\ _\theta \ll \ _{\theta_0}$, $\theta \in \Theta$) находим, что

$$g_\theta^{(\lambda)} = \frac{d\ \ _\theta}{d\lambda} = \frac{d\ \ _\theta}{d\ \ _{\theta_0}} \cdot \frac{d\ \ _{\theta_0}}{d\lambda} = g_\theta^{(\theta_0)} \frac{d\ \ _{\theta_0}}{d\lambda}.$$

Обозначая

$$\hat{g}_\theta^{(\theta_0)} = g_\theta^{(\theta_0)}, \qquad h = \frac{d\ \ _{\theta_0}}{d\lambda},$$

получаем требуемое факторизационное представление (48). □

Замечание 3. Полезно подчеркнуть, что для всякого семейства $\mathscr{P} = \{\ _\theta,\ \theta \in \Theta\}$ заведомо существует (и без всяких предположений типа доминируемости) достаточная σ-алгебра. В качестве таковой можно взять самую «богатую» σ-алгебру $\mathscr{F}$.

Действительно, в этом случае $\ _\theta(X\,|\,\mathscr{F}) = X$ ($\ _\theta$-п. н.) для всякой интегрируемой случайной величины X и, следовательно, свойство (47) будет выполнено.

Понятно, что такая достаточная σ-алгебра не представляет большого интереса, поскольку в этом случае не происходит никакой «редукции данных». Реальный интерес представляет отыскание *минимальной* достаточной σ-алгебры $\mathscr{G}_{\min}$, т. е. σ-алгебры, являющейся пересечением всех достаточных под-σ-алгебр (ср. с доказательством леммы 1 в § 2, из которого следует, что минимальная σ-алгебра существует). Но, к сожалению, явные построения таких σ-алгебр, как правило, довольно-таки непросты (см., впрочем, §§ 13—15 гл. 2 в [107]).

Замечание 4. Предположим, что $\mathscr{P}=\{\ {}_{\theta},\, \theta\in\Theta\}$ — доминируемое семейство (${}_{\theta}\ll\lambda$, $\theta\in\Theta$, λ есть σ-конечная мера) и плотность $g_\theta^{(\lambda)}=\dfrac{d\ {}_{\theta}}{d\lambda}(\omega)$ представляется для всех $\theta\in\Theta$ в виде

$$g_\theta^{(\lambda)}(\omega)=G_\theta^{(\lambda)}(T(\omega))\,h(\omega)\quad(\lambda\text{-п. н.}),\tag{53}$$

где $T=T(\omega)$ — некоторая функция (случайный элемент; см. § 5), принимающая значения в множестве E (с выделенной на нем σ-алгеброй $\mathscr{E}$) и являющаяся $\mathscr{F}/\mathscr{E}$-измеримой. Функции $G_\theta^{(\lambda)}(t)$, $t\in E$, и $h(\omega)$, $\omega\in\Omega$, предполагаются неотрицательными и $\mathscr{E}$- и $\mathscr{F}$-измеримыми соответственно.

Из сопоставления (48) и (53) видим, что σ-алгебра $\mathscr{G}=\sigma(T(\omega))$ является достаточной, а сама функция $T=T(\omega)$ является достаточной статистикой (в смысле определения 10).

Заметим, что в доминируемых случаях обычно именно выполнение факторизационного представления (53) принимают как определение достаточности статистики $T=T(\omega)$, входящей в (53).

Пример 5 (экспоненциальное семейство)**.** Предположим, что $\Omega=R^n$, $\mathscr{F}=\mathscr{B}(R^n)$ и мера ${}_{\theta}$ образована следующим образом: если $\omega=(x_1,\,\ldots,\,x_n)$, то

$${}_{\theta}(d\omega)=\ {}_{\theta}(dx_1)\ldots\ {}_{\theta}(dx_n),\tag{54}$$

где мера ${}_{\theta}(dx)$, $x\in R$, имеет следующую структуру:

$${}_{\theta}(dx)=\alpha(\theta)\,e^{\beta(\theta)s(x)}\gamma(x)\,\lambda(dx).\tag{55}$$

Здесь $s=s(x)$ — некоторая $\mathscr{B}$-измеримая функция и $\alpha(\theta)$, $\beta(\theta)$, $\gamma(x)$, $\lambda(dx)$ имеют очевидный смысл. (Семейство мер ${}_{\theta}$, $\theta\in\Theta$, является простейшим примером так называемых *экспоненциальных семейств*.) Из (54) и (55) находим, что

$${}_{\theta}(d\omega)=\alpha^n(\theta)e^{\beta(\theta)[s(x_1)+\ldots+s(x_n)]}\gamma(x_1)\ldots\gamma(x_n)\,dx_1\ldots dx_n.\tag{56}$$

Сопоставляя (56) с (53), видим, что статистика $T(\omega)=s(x_1)+\ldots+s(x_n)$, $\omega=(x_1,\,\ldots,\,x_n)$, является (в рассматриваемом случае экспоненциального семейства) достаточной статистикой.

Если для $\omega=(x_1,\,\ldots,\,x_n)$ обозначить $X_1(\omega)=x_1,\,\ldots,\,X_n(\omega)=x_n$, то можно сказать, что структура мер ${}_{\theta}$ (образованных по принципу *прямого произведения мер*) такова, что относительно них $X_1,\,\ldots,\,X_n$ является последовательностью *независимых одинаково распределенных* случайных величин.

Таким образом, статистика $T(\omega)=s(X_1(\omega))+\ldots+s(X_n(\omega))$ будет достаточной статистикой, построенной по последовательности $X_1(\omega),\ \ldots,\ X_n(\omega)$. (В задаче 20 предлагается установить, является ли эта статистика *минимальной* достаточной статистикой.)

Пример 6. Пусть $\Omega=R^n$, $\mathscr{F}=\mathscr{B}(R^n)$, параметр $\theta>0$ и для $\omega=(x_1,\ \ldots,\ x_n)$ плотность (по мере Лебега λ)

$$\frac{d\mathsf{P}_\theta}{d\lambda}(\omega)=\begin{cases}\theta^{-n}, & \text{если } 0\leqslant x_i\leqslant\theta,\ i=1,\ \ldots,\ n,\\ 0 & \text{в других случаях.}\end{cases}$$

Если

$$T(\omega)=\max_{1\leqslant i\leqslant n} x_i,$$

$$h(\omega)=\begin{cases}1, & \text{если все } x_i\geqslant 0,\ i=1,\ \ldots,\ n,\\ 0 & \text{в других случаях,}\end{cases}$$

$$G_\theta^{(\lambda)}(t)=\begin{cases}\theta^{-n}, & \text{если } 0\leqslant t\leqslant\theta,\\ 0 & \text{в других случаях,}\end{cases}$$

то находим, что

$$\frac{d\mathsf{P}_\theta}{d\lambda}(\omega)=G_\theta^{(\lambda)}(T(\omega))h(\omega). \tag{57}$$

Тем самым, статистика $T(\omega)=\max\limits_{1\leqslant i\leqslant n} x_i$ является достаточной.

11. Предположим, что Θ есть некоторое подмножество на числовой прямой и $\mathscr{E}=(\Omega,\ \mathscr{F},\ \mathscr{P}=\{\mathsf{P}_\theta,\ \theta\in\Theta\})$ — вероятностно-статистическая модель. Рассматриваемый сейчас вопрос состоит в построении «хороших» оценок параметра θ.

Под оценкой будем понимать любую случайную величину $\hat\theta=\hat\theta(\omega)$ (ср. с § 7 в гл. I).

Приводимая ниже теорема показывает, как понятие достаточной σ-алгебры позволяет *улучшить* «качество» оценки, измеряемое ее среднеквадратическим отклонением от истинного значения параметра θ. Более точно, будем говорить, что оценка $\hat\theta$ параметра θ является *несмещенной*, если $\mathsf{E}_\theta|\hat\theta|<\infty$ и $\mathsf{E}_\theta\hat\theta=\theta$ для всех $\theta\in\Theta$ (ср. со свойством (2) в § 7 гл. I).

Теорема 8 (Рао и Блэкуэлл). *Пусть $\mathscr{G}$ является достаточной σ-алгеброй для семейства $\mathscr{P}$ и $\hat\theta=\hat\theta(\omega)$ — некоторая оценка.*

а) *Если $\hat\theta$ является несмещенной оценкой, то оценка*

$$T=\mathsf{E}_\theta(\hat\theta\,|\,\mathscr{G}) \tag{58}$$

также будет несмещенной.

b) *Оценка T «лучше» оценки $\hat{\theta}$ в том смысле, что*

$$ {}_{\theta}(T-\theta)^2 \leqslant \ {}_{\theta}(\hat{\theta}-\theta)^2, \quad \theta\in\Theta. \tag{59}$$

Доказательство. Свойство a) следует из того, что

$$ {}_{\theta}T = \ {}_{\theta}\ {}_{\theta}(\hat{\theta}\,|\,\mathscr{G}) = \ {}_{\theta}\hat{\theta} = \theta.$$

Для доказательства же b) надо лишь заметить, что, согласно неравенству Иенсена (см. задачу 5, в которой надо взять $g(x)=(x-\theta)^2$),

$$(\ {}_{\theta}(\hat{\theta}\,|\,\mathscr{G}) - \theta)^2 \leqslant \ {}_{\theta}[(\hat{\theta}-\theta)^2\,|\,\mathscr{G}].$$

Беря в обеих частях математическое ожидание $\ {}_{\theta}(\cdot)$, приходим к требуемому неравенству (59). □

12. Задачи.

1. Пусть ξ и η — независимые одинаково распределенные случайные величины и $\ \xi$ определено. Показать, что

$$(\xi\,|\,\xi+\eta) = \ (\eta\,|\,\xi+\eta) = \frac{\xi+\eta}{2} \quad (\text{п. н.}).$$

2. Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $\ |\xi_i| < \infty$. Показать, что

$$(\xi_1\,|\,S_n, S_{n+1}, \ldots) = \frac{S_n}{n} \quad (\text{п. н.}),$$

где $S_n = \xi_1 + \ldots + \xi_n$.

3. Предположим, что случайные элементы (X, Y) таковы, что существует регулярное распределение $P_x(B) = \ (Y\in B\,|\,X=x)$. Показать, что если $\ |g(X, Y)| < \infty$, то P_x-п. н.

$$[g(X, Y)\,|\,X=x] = \int g(x, y)P_x(dy).$$

4. Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$. Показать, что

$$(\xi\,|\,a<\xi\leqslant b) = \frac{\int_a^b x\,dF_\xi(x)}{F_\xi(b)-F_\xi(a)}$$

(предполагается, что $F_\xi(b)-F_\xi(a)>0$).

5. Пусть $g=g(x)$ — выпуклая книзу борелевская функция, $\ |g(\xi)|<\infty$. Показать, что для *условных математических ожиданий* справедливо (п. н.) *неравенство Иенсена*

$$g(\ (\xi\,|\,\mathscr{G})) \leqslant \ (g(\xi)\,|\,\mathscr{G}).$$

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 $(g(\xi)\,|\,\mathscr{G}) = \ g(\xi)$ для каждой борелевской функции $g(x)$ такой, что $\ |g(\xi)| < \infty$.

7. Пусть ξ — неотрицательная случайная величина и $\mathscr{G}$ — σ-алгебра, $\mathscr{G} \subseteq \mathscr{F}$. Показать, что $\ (\xi\,|\,\mathscr{G}) < \infty$ (п. н.) тогда и только тогда, когда мера , определенная на множествах $A \in \mathscr{G}$ равенством $\ (A) = \int_A \xi \, d\ $, является σ-конечной.

8. Показать, что условные вероятности $\ (A\,|\,B)$ «непрерывны» в том смысле, что если $\lim_n A_n = A$, $\lim_n B_n = B$, $\ (B_n) > 0$, $\ (B) > 0$, то $\lim_n \ (A_n\,|\,B_n) = \ (A\,|\,B)$.

9. Пусть $\Omega = (0,\,1)$, $\mathscr{F} = \mathscr{B}((0,\,1))$ и — мера Лебега. Пусть $X(\omega)$ и $Y(\omega)$ — две независимые случайные величины, равномерно распределенные на $(0,\,1)$. Рассмотрим третью величину $Z(\omega) = |X(\omega) - Y(\omega)|$ — расстояние между «точками» $X(\omega)$ и $Y(\omega)$. Доказать, что распределение $F_Z(z)$ имеет плотность $f_Z(z)$ и $f_Z(z) = 2(1-z)$, $0 \leqslant z \leqslant 1$. (Отсюда, конечно, следует, что $\ Z = 1/3$.)

10. На окружности радиуса R $(\{(x,\,y)\colon x^2 + y^2 \leqslant R^2\})$ «случайным образом» выбираются две точки A_1 и A_2, т. е. эти точки выбираются независимым образом с вероятностями для них (в полярных координатах, $A_i = (\rho_i,\,\theta_i)$, $i = 1,\,2$)

$$(\rho_i \in dr,\ \theta_i \in d\theta) = \frac{r\,dr\,d\theta}{\pi R^2}, \qquad i = 1,\,2.$$

Доказать, что расстояние ρ между точками A_1 и A_2 имеет плотность распределения $f_\rho(r)$ и

$$f_\rho(r) = \frac{2r}{\pi R^2}\left[2\arccos\left(\frac{r}{2R}\right) - \frac{r}{R}\sqrt{1 - \left(\frac{r}{2R}\right)^2}\,\right],$$

где $0 < r < 2R$.

11. На единичном квадрате (с вершинами $(0,\,0)$, $(0,\,1)$, $(1,\,1)$, $(1,\,0)$) «случайным образом» (поясните!) выбирается точка $P = (x,\,y)$. Найти вероятность того, что эта P точка будет ближе к точке $(1,\,1)$, нежели к точке $(1/2,\,1/2)$.

12. Два человека A и B договорились о встрече между 7 и 8 часами вечера. Но оба забыли точное время встречи и приходят между 7 и 8 часами «случайным образом» и ждут не более 10 минут. Показать, что вероятность их встречи равна $11/36$.

13. Пусть $X_1,\,X_2,\,\ldots$ — последовательность независимых случайных величин, $S_n = \sum_{i=1}^{n} X_i$. Показать, что S_1 и S_3 условно независимы относительно σ-алгебры $\sigma(S_2)$, порожденной величиной S_2.

14. Будем говорить, что σ-алгебры $\mathscr{G}_1$ и $\mathscr{G}_2$ условно независимы относительно σ-алгебры $\mathscr{G}_3$, если

$$(A_1 A_2 \,|\, \mathscr{G}_3) = \quad (A_1 \,|\, \mathscr{G}_3) \quad (A_2 \,|\, \mathscr{G}_3) \quad \text{для всех } A_i \in \mathscr{G}_i,\ i = 1,\ 2.$$

Показать, что условная независимость $\mathscr{G}_1$ и $\mathscr{G}_2$ относительно $\mathscr{G}_3$ равносильна выполнению (-п. н.) любого из следующих условий:

(a) $\quad (A_1 \,|\, \sigma(\mathscr{G}_2 \cup \mathscr{G}_3)) = \quad (A_1 \,|\, \mathscr{G}_3)$ для всех $A_1 \in \mathscr{G}_1$;

(b) $\quad (B \,|\, \sigma(\mathscr{G}_2 \cup \mathscr{G}_3)) = \quad (B \,|\, \mathscr{G}_3)$ для любого множества B из системы подмножеств $\mathscr{P}_1$, образующих π-систему такую, что $\mathscr{G}_1 = \sigma(\mathscr{P}_1)$;

(c) $\quad (B_1 B_2 \,|\, \sigma(\mathscr{G}_2 \cup \mathscr{G}_3)) = \quad (B_1 \,|\, \mathscr{G}_3) \quad (B_2 \,|\, \mathscr{G}_3)$ для любых множеств B_1 и B_2 из π-систем $\mathscr{P}_1$ и $\mathscr{P}_2$ соответственно таких, что $\mathscr{G}_1 = \sigma(\mathscr{P}_1)$ и $\mathscr{G}_2 = \sigma(\mathscr{P}_2)$;

(d) $\quad (X \,|\, \sigma(\mathscr{G}_2 \cup \mathscr{G}_3)) = \quad (X \,|\, \mathscr{G}_3)$ для любой $\sigma(\mathscr{G}_2 \cup \mathscr{G}_3)$-измеримой случайной величины X с определенным (см. определение 2 в § 6) математическим ожиданием $\quad X$.

15. Доказать следующую расширенную версию *леммы Фату* для *условных математических ожиданий* (ср. с (d) в теореме 2).

Пусть $(\Omega,\ \mathscr{F},\quad)$ — вероятностное пространство и $(\xi_n)_{n \geqslant 1}$ — последовательность случайных величин таких, что определены математические ожидания $\quad \xi_n$, $n \geqslant 1$, и $\quad \underline{\lim}\, \xi_n$ (которые могут принимать и значения $\pm\infty$; см. определение 2 в § 6).

Предположим, что $\mathscr{G}$ есть под-σ-алгебра событий из $\mathscr{F}$ и

$$\sup_{n \geqslant 1} \quad (\xi_n^- I(\xi_n \geqslant a) \,|\, \mathscr{G}) \to 0 \quad (\quad\text{-п. н.}), \qquad a \to \infty.$$

Тогда

$$(\underline{\lim}\, \xi_n \,|\, \mathscr{G}) \leqslant \underline{\lim} \quad (\xi_n \,|\, \mathscr{G}) \quad (\quad\text{-п. н.}).$$

16. Пусть, как и в предыдущей задаче, $(\xi_n)_{n \geqslant 1}$ — последовательность случайных величин, для которых определены математические ожидания $\quad \xi_n$, $n \geqslant 1$, и $\mathscr{G}$ — под-σ-алгебра событий из $\mathscr{F}$ такая, что

$$\sup_n \lim_{k \to \infty} \quad (|\xi_n| I(|\xi_n| \geqslant k) \,|\, \mathscr{G}) = 0 \quad (\quad\text{-п. н.}). \tag{60}$$

Тогда если $\xi_n \to \xi$ (-п. н.) и $\quad \xi$ определено, то

$$(\xi_n \,|\, \mathscr{G}) \to \quad (\xi \,|\, \mathscr{G}) \quad (\quad\text{-п. н.}).$$

17. Пусть в условиях предыдущей задачи вместо (60) выполнено условие $\sup\limits_n \quad (|\xi_n|^\alpha \,|\, \mathscr{G}) < \infty$ (-п. н.) для некоторого $\alpha > 1$. Тогда

$$(\xi_n \,|\, \mathscr{G}) \to \quad (\xi \,|\, \mathscr{G}) \quad (\quad\text{-п. н.}).$$

18. Пусть $\xi_n \xrightarrow{L^p} \xi$ для некоторого $p \geqslant 1$. Показать, что тогда $(\xi_n \,|\, \mathscr{G}) \xrightarrow{L^p}$ $\to$ $(\xi \,|\, \mathscr{G})$.

19. (a) Пусть $(X \,|\, Y) \equiv [(X - (X \,|\, Y))^2 \,|\, Y]$. Показать, что $X =$ $= (X \,|\, Y) + (X \,|\, Y)$.

(b) Показать, что $(X, Y) = (X, (Y \,|\, X))$.

20. Выясните, является ли в примере 5 достаточная статистика $T(\omega) =$ $= s(X_1(\omega)) + \ldots + s(X_n(\omega))$ минимальной.

21. Докажите справедливость факторизационного представления (57).

22. В примере 2 п. 10 покажите, что ${}_{\theta}(X_i \,|\, T) = \dfrac{n+1}{2n} T$, где $X_i(\omega) = x_i$ для $\omega = (x_1, \ldots, x_n)$, $i = 1, \ldots, n$.

§ 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 $(\Omega, \mathscr{F}, \ )$ — вероятностное пространство и $\xi = \xi(\omega)$ — случайная величина, для которой определено математическое ожидание ξ.

Дисперсией случайной величины ξ называется величина

$$\xi = (\xi - \xi)^2.$$

Величина $\sigma = +\sqrt{\ \xi}$ называется *стандартным отклонением* относительно случайной величины ξ от среднего значения ξ. (Ср. с определением 5 в § 4 гл. I.)

Если ξ — случайная величина с гауссовской (нормальной) плотностью

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}, \qquad \sigma > 0, \qquad -\infty < m < \infty, \tag{1}$$

то смысл параметров m и σ, входящих в (1), оказывается очень простым:

$$m = \xi, \qquad \sigma^2 = \xi.$$

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: $\xi \sim \mathscr{N}(m, \sigma^2)$.)

Пусть теперь (ξ, η) — пара случайных величин. Их *ковариацией* называется величина

$$(\xi, \eta) = (\xi - \xi)(\eta - \eta) \tag{2}$$

(предполагается, что математические ожидания определены).

Если $(\xi, \eta) = 0$, то говорят, что случайные величины ξ и η *не коррелированы*.

Если $0 < \quad \xi < \infty$, $0 < \quad \eta < \infty$, то величина

$$\rho(\xi, \eta) \equiv \frac{\quad(\xi, \eta)}{\sqrt{\quad \xi \cdot \quad \eta}} \tag{3}$$

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 ковариаций* (ковариационной матрицей) вектора ξ матрицу (порядка $n \times n$) $\mathbb{R} = \|R_{ij}\|$, где $R_{ij} = \quad (\xi_i, \xi_j)$. 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

$$\sum_{i,j}^{n} R_{ij} \lambda_i \lambda_j = \quad \left[\sum_{i=1}^{n} (\xi_i - \quad \xi_i) \lambda_i \right]^2 \geqslant 0.$$

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 $\mathscr{O}$ (т. е. $\mathscr{O}\mathscr{O}^* = E$ — единичная матрица), что

$$\mathscr{O}^* \mathbb{R} \mathscr{O} = D,$$

где

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

— диагональная матрица с неотрицательными элементами d_i, $i = 1, \ldots, n$.

Отсюда следует, что

$$\mathbb{R} = \mathscr{O} D \mathscr{O}^* = (\mathscr{O} B)(B^* \mathscr{O}^*),$$

где B — диагональная матрица с элементами $b_i = +\sqrt{d_i}$, $i = 1, \ldots, n$. Поэтому, если положить $A = \mathscr{O} B$, 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, $\mathscr{N}(0, 1)$. (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,

$$\xi\xi^* = \quad (A\eta)(A\eta)^* = A \cdot \quad \eta\eta^* \cdot A^* = AEA^* = AA^*.$$

(Если $\zeta = \|\zeta_{ij}\|$ — матрица, элементами которой являются случайные величины, то под $\quad \zeta$ понимается матрица $\| \quad \zeta_{ij}\|$.) □

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 (14) § 3), где $|m_1| < \infty$, $|m_2| < \infty$, $\sigma_1 > 0$, $\sigma_2 > 0$, $|\rho| < 1$. (См. рис. 28 в § 3.) Простой подсчет раскрывает смысл этих параметров:

$$m_1 = \quad \xi, \quad \sigma_1^2 = \quad \xi,$$
$$m_2 = \quad \eta, \quad \sigma_2^2 = \quad \eta,$$
$$\rho = \rho(\xi, \eta).$$

В § 4 гл. I было объяснено, что если величины ξ и η не корелированы ($\rho(\xi, \eta) = 0$), 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

$$f_{\xi,\eta}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{(x-m_1)^2}{2\sigma_1^2}\right\} \exp\left\{-\frac{(y-m_2)^2}{2\sigma_2^2}\right\}.$$

Но в силу (55) § 6 и (4)

$$f_\xi(x)=\int\limits_{-\infty}^{\infty} f_{\xi,\eta}(x,\,y)\,dy=\frac{1}{\sqrt{2\pi}\sigma_1}\,\exp\Big\{-\frac{(x-m_1)^2}{2\sigma_1^2}\Big\},$$

$$f_\eta(y)=\int\limits_{-\infty}^{\infty} f_{\xi,\eta}(x,\,y)\,dx=\frac{1}{\sqrt{2\pi}\sigma_2}\,\exp\Big\{-\frac{(y-m_2)^2}{2\sigma_2^2}\Big\}.$$

Поэтому

$$f_{\xi,\eta}(x,\,y)=f_\xi(x)f_\eta(y),$$

откуда следует, что величины ξ и η независимы (см. конец п. 9 § 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 $(\xi,\,\eta)$ — 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 $\quad [\eta-\varphi(\xi)]^2$ (*среднеквадратической*) *ошибкой* этой оценки. Оценку $\varphi^*(\xi)$ назовем *оптимальной* (в среднеквадратическом смысле), если

$$\Delta\equiv\quad[\eta-\varphi^*(\xi)]^2=\inf_\varphi\quad[\eta-\varphi(\xi)]^2, \tag{5}$$

где inf берется по классу всех борелевских функций $\varphi=\varphi(x)$.

Теорема 1. *Пусть* $\quad\eta^2<\infty$. *Тогда оптимальная оценка* $\varphi^*=\varphi^*(\xi)$ *существует и в качестве* $\varphi^*(x)$ *может быть взята функция*

$$\varphi^*(x)=\quad(\eta\,|\,\xi=x). \tag{6}$$

Доказательство. Без ограничения общности можно рассматривать только те оценки $\varphi(\xi)$, для которых $\quad\varphi^2(\xi)<\infty$. Тогда, если $\varphi(\xi)$ — такая оценка, а $\varphi^*(\xi)=\quad(\eta\,|\,\xi)$, то

$$\begin{aligned}[\eta-\varphi(\xi)]^2=\quad[(\eta-\varphi^*(\xi))+(\varphi^*(\xi)-\varphi(\xi))]^2=\quad[\eta-\varphi^*(\xi)]^2+\\ +\quad[\varphi^*(\xi)-\varphi(\xi)]^2+2\quad[(\eta-\varphi^*(\xi))(\varphi^*(\xi)-\varphi(\xi))]\geqslant\quad[\eta-\varphi^*(\xi)]^2,\end{aligned}$$

поскольку $\quad[\varphi^*(\xi)-\varphi(\xi)]^2\geqslant 0$ и по свойствам условных математических ожиданий

$$\begin{aligned}[(\eta-\varphi^*(\xi))(\varphi^*(\xi)-\varphi(\xi))]=\quad\{\quad[(\eta-\varphi^*(\xi))(\varphi^*(\xi)-\varphi(\xi))\,|\,\xi]\}=\\ =\quad\{(\varphi^*(\xi)-\varphi(\xi))\quad(\eta-\varphi^*(\xi)\,|\,\xi)\}=0.\quad\square\end{aligned}$$

Замечание 1. Из доказательства теоремы видно, что ее утверждение справедливо и в том случае, когда ξ не случайная величина, а 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 (18) § 7 находим, что плотность $f_{\eta|\xi}(y\,|\,x)$ условного распределения вероятностей задается формулой

$$f_{\eta|\xi}(y\,|\,x) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}}\ \exp\left\{-\frac{(y-m(x))^2}{2\sigma_2^2(1-\rho^2)}\right\}, \tag{7}$$

где

$$m(x) = m_2 + \frac{\sigma_2}{\sigma_1}\rho\cdot(x-m_1). \tag{8}$$

Тогда из следствия к теореме 3 § 7

$$(\eta\,|\,\xi = x) = \int_{-\infty}^{\infty} y f_{\eta|\xi}(y\,|\,x)\,dy = m(x) \tag{9}$$

и

$$\begin{aligned}(\eta\,|\,\xi = x) \equiv \ [(\eta - \ (\eta\,|\,\xi = x))^2\,|\,\xi = x] =& \\ = \int_{-\infty}^{\infty} (y-m(x))^2 f_{\eta\,|\,\xi}(y\,|\,x)\,dy = \sigma_2^2(1-\rho^2).& \qquad (10)\end{aligned}$$

Заметим, что условная дисперсия $\ (\eta\,|\,\xi = x)$ не зависит от x, и, значит,

$$\Delta = \ [\eta - \ (\eta\,|\,\xi = x)]^2 = \sigma_2^2(1-\rho^2). \tag{11}$$

Формулы (9), (11) получены в предположении $\ \xi > 0,\ \ \eta > 0$. Если же $\ \xi > 0$, а $\ \eta = 0$, то они выполняются очевидным образом.

Итак, справедлив следующий результат (ср. с (16), (17) § 4 гл. I).

Теорема 2 (теорема о нормальной корреляции)**.** *Пусть* (ξ, η) — *гауссовский вектор с* $\ \xi > 0$. *Оптимальная оценка* η *по* ξ *есть*

$$(\eta\,|\,\xi) = \ \eta + \frac{(\xi,\eta)}{\ \xi}(\xi - \ \xi), \tag{12}$$

а ее ошибка

$$\Delta \equiv \ [\eta - \ (\eta\,|\,\xi)]^2 = \ \eta - \frac{^2(\xi,\eta)}{\ \xi}. \tag{13}$$

Замечание 2. Кривая $y(x) = \ (\eta\,|\,\xi = x)$ называется *кривой регрессии* η *на* ξ *или* η *по отношению к* ξ. В гауссовском случае $\ (\eta\,|\,\xi = x) = a + bx$

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 (16) и (17) § 4 гл. I для оптимальной линейной оценки и ее ошибки.

Следствие. *Пусть* ε_1 *и* ε_2 *— независимые гауссовские случайные величины с нулевым средним и единичной дисперсией и*

$$\xi = a_1\varepsilon_1 + a_2\varepsilon_2, \quad \eta = b_1\varepsilon_1 + b_2\varepsilon_2.$$

Тогда $\ \xi = \ \eta = 0$, $\ \xi = a_1^2 + a_2^2$, $\ \eta = b_1^2 + b_2^2$, $\ (\xi, \eta) = a_1b_1 + a_2b_2$ *и если* $a_1^2 + a_2^2 > 0$, *то*

$$(\eta \,|\, \xi) = \frac{a_1b_1 + a_2b_2}{a_1^2 + a_2^2}\xi, \tag{14}$$

$$\Delta = \frac{(a_1b_2 - a_2b_1)^2}{a_1^2 + a_2^2}. \tag{15}$$

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

$$F_\eta(y) = \ \{\eta \leqslant y\} = \ \{\varphi(\xi) \in I_y\} = \ \{\xi \in \varphi^{-1}(I_y)\} = \int_{\varphi^{-1}(I_y)} F_\xi(dx), \tag{16}$$

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

$$F_\eta(y) = \ \left\{\xi \leqslant \frac{y-b}{a}\right\} = F_\xi\left(\frac{y-b}{a}\right). \tag{17}$$

Если $\eta = \xi^2$, то, очевидно, $F_\eta(y) = 0$ для $y < 0$, а для $y \geqslant 0$

$$\begin{aligned} F_\eta(y) = \ \{\xi^2 \leqslant y\} = \ \{-\sqrt{y} \leqslant \xi \leqslant \sqrt{y}\} = \\ = F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y}) + \ \{\xi = -\sqrt{y}\}. \end{aligned} \tag{18}$$

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 $y \in \{\varphi(x)\colon x \in I\}$

$$F_\eta(y) = \quad \{\eta \leqslant y\} = \quad \{\varphi(\xi) \leqslant y\} = \quad \{\xi \leqslant \varphi^{-1}(y)\} = \\ = \quad \{\xi \leqslant h(y)\} = \int\limits_{-\infty}^{h(y)} f_\xi(x)\,dx. \quad (19)$$

Согласно задаче 15 из § 6,

$$\int\limits_{-\infty}^{h(y)} f_\xi(x)\,dx = \int\limits_{-\infty}^{y} f_\xi(h(z))h'(z)\,dz \tag{20}$$

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 $h(y) = \dfrac{y-b}{a}$ и $f_\eta(y) = \dfrac{1}{|a|} f_\xi\Big(\dfrac{y-b}{a}\Big)$.

Если $\xi \sim \mathscr{N}(m, \sigma^2)$, а $\eta = e^\xi$, то из (22) находим, что

$$f_\eta(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma y}\ \exp\Big[-\dfrac{\ln\,(y/M)^2}{2\sigma^2}\Big], & y > 0, \\ 0, & y \leqslant 0, \end{cases} \tag{23}$$

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) для $\eta = \varphi(\xi)$ неприменима. Однако для многих приложений вполне достаточно следующее ее обобщение.

Пусть функция $\varphi = \varphi(x)$ определена на множестве $\sum\limits_{k=1}^{n} [a_k, b_k]$, причем на каждом открытом интервале $I_k = (a_k, b_k)$ является непрерывно дифференцируемой и 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):

$$f_\eta(y) = \sum_{k=1}^{n} f_\xi(h_k(y))|h_k'(y)| I_{D_k}(y), \tag{24}$$

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 $\xi \sim \mathscr{N}(0, 1)$, 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, $F_{\xi,\eta}(x, y) = F_\xi(x)F_\eta(y)$), то, применяя теорему Фубини, получим

$$F_\zeta(z) = \int\limits_{\{x,y:x+y \leqslant z\}} dF_\xi(x)\,dF_\eta(y) = \int\limits_{R^2} I_{\{x+y \leqslant z\}}(x, y)\,dF_\xi(x)\,dF_\eta(y) =$$

$$= \int\limits_{-\infty}^{\infty} dF_\xi(x) \left\{ \int\limits_{-\infty}^{\infty} I_{\{x+y \leqslant z\}}(x, y)\,dF_\eta(y) \right\} = \int\limits_{-\infty}^{\infty} F_\eta(z - x)\,dF_\xi(x) \tag{30}$$

и аналогично

$$F_\zeta(z) = \int\limits_{-\infty}^{\infty} F_\xi(z - y)\,dF_\eta(y). \tag{31}$$

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

$$\begin{aligned} F_\zeta(z) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-y} f_\xi(u)\,du\right] f_\eta(y)\,dy = \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z} f_\xi(u-y)\,du\right] f_\eta(y)\,dy = \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} f_\xi(u-y) f_\eta(y)\,dy\right] du, \end{aligned}$$

откуда

$$f_\zeta(z) = \int_{-\infty}^{\infty} f_\xi(z-y) f_\eta(y)\,dy, \tag{32}$$

и аналогично

$$f_\zeta(z) = \int_{\infty}^{\infty} f_\eta(z-x) f_\xi(x)\,dx. \tag{33}$$

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)

$$f_{\xi_1+\ldots+\xi_n}(x) = \begin{cases} \dfrac{1}{2^n(n-1)!} \sum\limits_{k=0}^{\left[\frac{n+x}{2}\right]} (-1)^k C_n^k (n+x-2k)^{n-1}, & |x| \leqslant n, \\ 0, & |x| > n. \end{cases}$$

Пусть теперь $\xi \sim \mathscr{N}(m_1, \sigma_1^2)$, $\eta \sim \mathscr{N}(m_2, \sigma_2^2)$. Если обозначить

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

то

$$f_\xi(x) = \frac{1}{\sigma_1}\,\varphi\Big(\frac{x-m_1}{\sigma_1}\Big), \qquad f_\eta(x) = \frac{1}{\sigma_2}\,\varphi\Big(\frac{x-m_2}{\sigma_2}\Big),$$

и из (32) получаем, что

$$f_{\xi+\eta}(x) = \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\,\varphi\Big(\frac{x-(m_1+m_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big).$$

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. 3 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. В случае $n=2$ это распределение называется распределением Рэлея.

Пусть снова ξ и η — независимые случайные величины с плотностями f_ξ и f_η. Тогда

$$F_{\xi\eta}(z) = \iint\limits_{\{x,y\,:\,xy\leqslant z\}} f_\xi(x) f_\eta(y)\,dx\,dy,$$

$$F_{\xi/\eta}(z) = \iint\limits_{\{x,y\,:\,x/y\leqslant z\}} f_\xi(x) f_\eta(y)\,dx\,dy.$$

Отсюда нетрудно получить, что

$$f_{\xi\eta}(z) = \int\limits_{-\infty}^{\infty} f_\xi\left(\frac{z}{y}\right) f_\eta(y)\,\frac{dy}{|y|} = \int\limits_{-\infty}^{\infty} f_\eta\left(\frac{z}{x}\right) f_\xi(x)\,\frac{dx}{|x|} \tag{36}$$

и

$$f_{\xi/\eta}(z) = \int\limits_{-\infty}^{\infty} f_\xi(zy) f_\eta(y)|y|\,dy. \tag{37}$$

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 $\xi_0 \Big/ \sqrt{\dfrac{1}{n}(\xi_1^2 + \ldots + \xi_n^2)}$ обычно обозначается через t, а ее распределение называется *t-распределением* или *распределением Стьюдента* с n степенями свободы (ср. с табл. 3 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 $\overline{\xi} = \max(\xi_1, \ldots, \xi_n)$, $\underline{\xi} = \min(\xi_1, \ldots, \xi_n)$, $\rho = \overline{\xi} - \underline{\xi}$. Показать, что

$$F_{\bar{\xi},\underline{\xi}}(y, x) = \begin{cases} (F(y))^n - (F(y) - F(x))^n, & y > x, \\ (F(y))^n, & y \leqslant x, \end{cases}$$

$$f_{\bar{\xi},\underline{\xi}}(y, x) = \begin{cases} n(n-1)[F(y) - F(x)]^{n-2} f(x) f(y), & y > x, \\ 0, & y \leqslant x, \end{cases}$$

$$F_\rho(x) = \begin{cases} n \int\limits_{-\infty}^{\infty} [F(y) - F(y-x)]^{n-1} f(y)\, dy, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

$$f_\rho(x) = \begin{cases} n(n-1) \int\limits_{-\infty}^{\infty} [F(y) - F(y-x)]^{n-2} f(y-x) f(y)\, dy, & x > 0, \\ 0, & x < 0. \end{cases}$$

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 случайные величины ξ и η независимы тогда и только тогда, когда $\rho^*(\xi, \eta) = 0$.

6. Пусть $\tau_1, \tau_2, \ldots, \tau_k$ — 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

$$\frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!}, \qquad t \geqslant 0,$$

и

$$(\tau_1 + \ldots + \tau_k > t) = \sum_{i=0}^{k-1} e^{-\lambda t} \frac{(\lambda t)^i}{i!}.$$

7. Пусть $\xi \sim \mathscr{N}(0, \sigma^2)$. Показать, что для всякого $p \geqslant 1$

$$|\xi|^p = C_p \sigma^p,$$

где

$$C_p = \frac{2^{p/2}}{\pi^{1/2}} \Gamma\Big(\frac{p+1}{2}\Big)$$

и $\Gamma(s)=\int\limits_0^\infty e^{-x}x^{s-1}\,dx$ — гамма-функция Эйлера. В частности, для любого целого $n\geqslant 1$

$$\xi^{2n}=(2n-1)!!\,\sigma^{2n}.$$

8. Пусть ξ и η — независимые случайные величины такие, что распределение $\xi+\eta$ совпадают с распределением ξ. Доказать, что $\eta=0$ п. н.

9. Пусть (X, Y) имеют равномерное распределение на единичном круге $\{(x, y)\colon x^2+y^2\leqslant 1\}$ и $W=X^2+Y^2$. Положим

$$U=X\sqrt{-\frac{2\ln W}{W}},\qquad V=Y\sqrt{-\frac{2\ln W}{W}}.$$

Показать, что X и Y являются независимыми $\mathscr{N}(0, 1)$-распределенными случайными величинами.

10. Пусть U и V — независимые равномерно распределенные на $(0, 1)$ случайные величины. Определим

$$X=\sqrt{-\ln V}\,\cos(2\pi U),\qquad Y=\sqrt{-\ln V}\,\sin(2\pi U).$$

Показать, что X и Y независимы и $\mathscr{N}(0, 1)$-распределены.

11. Привести пример гауссовских величин ξ и η, сумма которых $\xi+\eta$ имеет негауссовское распределение.

12. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины с плотностью распределения $f=f(x)$. Пусть $\mathscr{R}_n=\max(X_1, \ldots, X_n)-\min(X_1, \ldots, X_n)$ — «размах» выборки $(X_1, \ldots, X_n)$. Показать, что плотность $f_{\mathscr{R}_n}(x)$, $x>0$, величины $\mathscr{R}_n$ задается формулой

$$f_{\mathscr{R}_n}(x)=n(n-1)\int\limits_{-\infty}^{\infty}[F(y)-F(y-x)]^{n-2}f(y)f(y-x)\,dx,$$

где $F(y)=\int\limits_{-\infty}^{y}f(z)\,dz$. В частности, для случайных величин $X_1, \ldots, X_n$, имеющих равномерное распределение на $[0, 1]$,

$$f_{\mathscr{R}_n}(x)=\begin{cases} n(n-1)x^{n-2}(1-x), & 0\leqslant x\leqslant 1,\\ 0, & x<0 \text{ или } x>1.\end{cases}$$

13. Пусть $F(x)$ — функция распределения. Показать, что для всякого $a>0$ следующие функции также являются функциями распределения:

$$G_1(x)=\frac{1}{a}\int\limits_x^{x+a}F(u)\,du,\qquad G_2(x)=\frac{1}{2a}\int\limits_{x-a}^{x+a}F(u)\,du.$$

14. Пусть случайная величина X имеет экспоненциальное распределение с параметром $\lambda>0$ ($f_X(x)=\lambda e^{-\lambda x}$, $x\geqslant 0$). Найти плотность распределения (называемого *распределением Вейбулла*) случайной величины $Y=X^{1/\alpha}$, $\alpha>0$.

Пусть $\lambda=1$. Найти плотность распределения случайной величины $Y=\ln X$ (соответствующее распределение называется *двойным экспоненциальным*).

15. Пусть случайные величины X и Y имеют совместную плотность распределения $f(x, y)$ вида $f(x, y)=g(\sqrt{x^2+y^2})$.

Найти совместную плотность распределения случайных величин $\rho=\sqrt{X^2+Y^2}$ и $\theta=\operatorname{tg}^{-1}(Y/X)$. Показать, что ρ и θ независимы.

Пусть $U=(\cos\alpha)X+(\sin\alpha)Y$ и $V=(-\sin\alpha)X+(\cos\alpha)Y$. Показать, что совместная плотность распределения величин U и V совпадает с $f(x, y)$. (Это отражает факт инвариантности распределения величин (X, Y) относительно «вращения».)

16. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины с функцией распределения $F=F(x)$ и плотностью $f=f(x)$. Обозначим (ср. с задачей 12) $X^{(1)}=\min(X_1, \ldots, X_n)$ наименьшую из величин $X_1, \ldots, X_n$, $X^{(2)}$ — вторую наименьшую величину, и т. д., $X^{(n)}=\max(X_1, \ldots, X_n)$ — наибольшую из величин $X_1, \ldots, X_n$ (так определенные величины $X^{(1)}, \ldots, X^{(n)}$ называют *порядковыми статистиками* величин $X_1, \ldots, X_n$).

Показать, что: (a) плотность распределения вероятностей величины $X^{(k)}$ задается формулой

$$nf(x)C_{n-1}^{k-1}[F(x)]^{k-1}[1-F(x)]^{n-k};$$

(b) совместная плотность $f(x^1, \ldots, x^n)$ величин $X^{(1)}, \ldots, X^{(n)}$ задается формулой

$$f(x^1, \ldots, x^n)=\begin{cases} n!\, f(x^1)\ldots f(x^n), & \text{если } x^1<\ldots<x^n,\\ 0 & \text{в других случаях.}\end{cases}$$

17. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины, $\mathcal{N}(\mu, \sigma^2)$. Величина

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2, \quad n>1, \quad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

носит название *выборочной дисперсии*. Показать, что:

(a) $S^2=\sigma^2$;

(b) *выборочное среднее* $\overline{X}$ и выборочная дисперсия S^2 независимы;

(c) $\bar{X} \sim \mathscr{N}(\mu, \sigma^2/n)$, а $(n-1)S^2/\sigma^2$ имеет распределение χ^2 с $(n-1)$ степенью свободы.

18. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины, N — случайная величина, не зависящая от $X_1, \ldots, X_n$ $(N = 1, 2, \ldots)$, с $\ \ N < \infty$, $\ \ N < \infty$. Показать, что $(S_N = X_1 + \ldots + X_N)$

$$S_N = \ \ X_1 \ \ N + (\ \ X_1)^2 \ \ N, \qquad \frac{S_N}{S_N} = \frac{X_1}{X_1} + \ \ X_1 \frac{N}{N}.$$

19. Пусть $M(t) = \ \ e^{tX}$ — производящая функция случайной величины X. Показать, что $\ \ (X \geqslant 0) \leqslant M(t)$ для всякого $t > 0$.

20. Пусть $X, X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины, $S_n = \sum_{i=1}^{n} X_i$, $S_0 = 0$, $\bar{M}_n = \max_{0 \leqslant j \leqslant n} S_j$, $\bar{M} = \sup_{n \geqslant 0} S_n$. Показать, что («$\xi \overset{d}{=} \eta$» означает, что распределения ξ и η совпадают)

(a) $\bar{M}_n \overset{d}{=} (\bar{M}_{n-1} + X)^+$, $n \geqslant 1$;

(b) если $S_n \to \infty$ ($\ $-п. н.), то $\bar{M} \overset{d}{=} (\bar{M} + X)^+$;

(c) если $-\infty < \ \ X < 0$ и $\ \ X^2 < \infty$, то

$$\bar{M} = \frac{\ \ X - \ \ (S + X)^-}{-2 \ \ X}.$$

21. В предположениях предыдущей задачи пусть $\bar{M}(\varepsilon) = \sup_{n \geqslant 0}(S_n - n\varepsilon)$ для $\varepsilon > 0$. Показать, что $\lim_{\varepsilon \downarrow 0} \varepsilon \bar{M}(\varepsilon) = (\ \ X)/2$.

§ 9. Построение процесса с заданными конечномерными распределениями

1. Пусть $\xi = \xi(\omega)$ — случайная величина, заданная на вероятностном пространстве $(\Omega, \mathscr{F}, \ \ )$ и

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

— 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, оправдывающих 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 на $(\Omega, \mathscr{F})$, то правильная постановка вопроса о существовании случайной величины с заданной функцией распределения $F(x)$ такова:

Существуют ли вероятностное пространство $(\Omega, \mathscr{F}, \ )$ *и случайная величина* $\xi = \xi(\omega)$ *на нем такие, что*

$$\{\omega\colon \xi(\omega) \leqslant x\} = F(x)?$$

Покажем, что ответ на этот вопрос положительный и, в сущности, он содержится в теореме 1 § 3.

Действительно, положим

$$\Omega = R, \quad \mathscr{F} = \mathscr{B}(R).$$

Тогда из теоремы 1 § 3 следует, что на $(R, \mathscr{B}(R))$ существует (и притом единственная) вероятностная мера , для которой $(a, b] = F(b) - F(a)$, $a < b$.

Положим $\xi(\omega) \equiv \omega$. Тогда

$$\{\omega\colon \xi(\omega) \leqslant x\} = \ \{\omega\colon \omega \leqslant x\} = \ (-\infty, x] = F(x).$$

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 $(\Omega, \mathscr{F}, \ )$ 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*

$$F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \ \{\omega\colon \xi_{t_1} \leqslant x_1, \ldots, \xi_{t_n} \leqslant x_n\}, \tag{1}$$

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 $t_1 < t_2 < \ldots < t_n$ 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* (ср. с (20) из § 3):

$$F_{t_1,\ldots,t_k,\ldots,t_n}(x_1, \ldots, \infty, \ldots, x_n) = \\ = F_{t_1,\ldots,t_{k-1},t_{k+1},\ldots,t_n}(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n). \tag{2}$$

Естественно теперь поставить такой вопрос: при каких условиях 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 (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* $(\Omega, \mathscr{F}, \ )$ *и случайный процесс* $X = (\xi_t)_{t \in T}$ *такие, что*

$$\{\omega\colon \xi_{t_1} \leqslant x_1, \ldots, \xi_{t_n} \leqslant x_n\} = F_{t_1,\ldots,t_n}(x_1, \ldots, x_n). \tag{3}$$

Доказательство. Положим

$$\Omega = R^T, \qquad \mathscr{F} = \mathscr{B}(R^T),$$

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 притом единственную) вероятностную меру P_τ такую, что

$$P_\tau\{(\omega_{t_1}, \ldots, \omega_{t_n})\colon \omega_{t_1} \leqslant x_1, \ldots, \omega_{t_n} \leqslant x_n\} = F_{t_1,\ldots,t_n}(x_1, \ldots, x_n). \tag{4}$$

Из условий согласованности (2) вытекает, что семейство $\{P_\tau\}$ также является согласованным (см. (20) § 3). Согласно теореме 4 из § 3, на пространстве $(R^T, \mathscr{B}(R^T))$ существует вероятностная мера   такая, что

$$\{\omega\colon (\omega_{t_1}, \ldots, \omega_{t_n}) \in B\} = P_\tau(B)$$

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:

$$\xi_t(\omega) = \omega_t, \qquad t \in T. \tag{5}$$

$\square$

Замечание 1. Построенное вероятностное пространство $(R^T, \mathscr{B}(R^T), \ )$ 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 $(\Omega, \mathscr{F}, \ \ )$ и семейство $\mathscr{F}/\mathscr{E}_\alpha$-измеримых функций $(X_\alpha(\omega))_{\alpha\in\mathfrak{A}}$ такие, что

$$\{(X_{\alpha_1}, \ldots, X_{\alpha_n})\in B\}=P_\tau(B)$$

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 $\Omega=\prod\limits_\alpha E_\alpha$, $\mathscr{F}=\bigotimes\limits_\alpha \mathscr{E}_\alpha$ 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 $(\Omega, \mathscr{F}, \ \ )$ и последовательность независимых случайных величин ξ_1, ξ_2, ... такие, что*

$$\{\omega\colon \xi_i(\omega)\leqslant x\}=F_i(x). \tag{6}$$

В частности, существует вероятностное пространство $(\Omega, \mathscr{F}, \ \ )$, на котором определена бесконечная последовательность бернуллиевских случайных величин (*в этой связи см. п.* 2 § 5 *гл.* I). *В качестве Ω можно здесь взять пространство*

$$\Omega=\{\omega\colon \omega=(a_1, a_2, \ldots),\ \ a_i=0 \text{ или } 1\}$$

(*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—Чепмена*

$$P(s, x; \tau, B)=\int\limits_R P(s, x; t, dy)\,P(t, y; \tau, B). \tag{7}$$

Пусть, кроме того, $\pi=\pi(\cdot)$ — вероятностная мера на $(R, \mathscr{B}(R))$.

Тогда существуют вероятностное пространство $(\Omega, \mathscr{F}, \ )$ *и случайный процесс* $X=(\xi_t)_{t\geqslant 0}$ *на нем такие, что для* $0=t_0<t_1<\ldots$ $\ldots<t_n$

$$\{\xi_{t_0}\leqslant x_0,\ \xi_{t_1}\leqslant x_1,\ \ldots,\ \xi_{t_n}\leqslant x_n\}=$$
$$=\int_{-\infty}^{x_0}\pi(dy_0)\int_{-\infty}^{x_1}P(0,\,y_0;\,t_1,\,dy_1)\ldots\int_{-\infty}^{x_n}P(t_{n-1},\,y_{n-1};\,t_n,\,dy_n). \tag{8}$$

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,* $\pi=\pi(\,\cdot\,)$ — *вероятностная мера на* $(R,\,\mathscr{B}(R))$.

Тогда можно построить вероятностное пространство $(\Omega, \mathscr{F}, \ )$ *с семейством случайных величин* $X=\{\xi_0,\xi_1,\ldots\}$ *на нем таких, что*

$$\{\xi_0\leqslant x_0,\ \xi_1\leqslant x_1,\ \ldots,\ \xi_n\leqslant x_n\}=$$
$$=\int_{-\infty}^{x_0}\pi(dy_0)\int_{-\infty}^{x_1}P_1(y_0;\,dy_1)\ldots\int_{-\infty}^{x_n}P_n(y_{n-1};\,dy_n).$$

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 существуют вероятностное пространство $(\Omega, \mathscr{F}, \ )$ и последовательность *независимых* элементов $X_1, X_2, \ldots$ такие, что X_n — $\mathscr{F}/\mathscr{E}_n$-измеримы и $\ \ \{X_n\in B\}=P_n(B)$, $B\in\mathscr{E}_n$.

Оказывается, что этот результат остается справедливым и в том случае, когда пространства $(E_n, \mathscr{E}_n)$ являются *произвольными измеримыми пространствами*.

Теорема 2 (Ионеску Тулчи о продолжении меры и существовании случайной последовательности). *Пусть* $(\Omega_n, \mathscr{F}_n)$, $n=1,2,\ldots,$ — *произвольные измеримые пространства и* $\Omega=\prod\Omega_n$, $\mathscr{F}=\boxtimes\mathscr{F}_n$. *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 измеримыми* *функциями от* $(\omega_1, \ldots, \omega_n)$, *и пусть*

$$P_n(A_1 \times \ldots \times A_n) = \int_{A_1} P_1(d\omega_1) \int_{A_2} P(\omega_1; d\omega_2) \ldots \int_{A_n} P(\omega_1, \ldots, \omega_{n-1}; d\omega_n),$$

$$A_i \in \mathscr{F}_i, \quad n \geqslant 1. \tag{9}$$

Тогда на $(\Omega, \mathscr{F})$ *существуют единственная вероятностная мера такая, что для любого* $n \geqslant 1$

$$\{\omega\colon \omega_1 \in A_1, \ldots, \omega_n \in A_n\} = P_n(A_1 \times \ldots \times A_n), \tag{10}$$

и случайная последовательность $X = (X_1(\omega), X_2(\omega), \ldots)$ *такая, что*

$$\{\omega\colon X_1(\omega) \in A_1, \ldots, X_n(\omega) \in A_n\} = P_n(A_1 \times \ldots \times A_n), \tag{11}$$

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

$$P_n(B) = \int_{\Omega_1} P_1(d\omega_1) \int_{\Omega_2} P(\omega_1; d\omega_2) \ldots \int_{\Omega_{n-1}} P(\omega_1, \ldots, \omega_{n-2}; d\omega_{n-1}) \times$$

$$\times \int_{\Omega_n} I_B(\omega_1, \ldots, \omega_n)\, P(\omega_1, \ldots, \omega_{n-1}; d\omega_n). \tag{12}$$

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, показывается, что P_2 является мерой. Отсюда по индукции легко устанавливается, что P_n являются мерами 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 $J_n(B) = \{\omega \in \Omega\colon (\omega_1, \ldots, \omega_n) \in B\}$, $B \in \mathscr{F}_1 \otimes \ldots \otimes \mathscr{F}_n$, определим функцию множеств с помощью равенства

$$(J_n(B)) = P_n(B). \tag{13}$$

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 $(J_n(B))$ не зависит от способа представления цилиндрического множества.

Отсюда вытекает, что функция множеств , 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, $\{\widehat{B}_n\}_{n\geqslant 1}$ — последовательность цилиндрических множеств

$$\widehat{B}_n = \{\omega\colon\ (\omega_1, \ldots, \omega_n) \in B_n\},$$

убывающих к пустому множеству $\varnothing$, но

$$\lim_{n\to\infty} \quad (\widehat{B}_n) > 0. \tag{14}$$

Из (12) для $n > 1$

$$(\widehat{B}_n) = \int_{\Omega_1} f_n^{(1)}(\omega_1)\, P_1(d\omega_1),$$

где

$$f_n^{(1)}(\omega_1) = \int_{\Omega_2} P(\omega_1;\, d\omega_2) \ldots \int_{\Omega_n} I_{B_n}(\omega_1, \ldots, \omega_n)\, P(\omega_1, \ldots, \omega_{n-1};\, d\omega_n).$$

Поскольку $\widehat{B}_{n+1} \subseteq \widehat{B}_n$, 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

$$\lim_n \quad (\widehat{B}_n) = \lim_n \int_{\Omega_1} f_n^{(1)}(\omega_1)\, P_1(d\omega_1) = \int_{\Omega_1} f^{(1)}(\omega_1)\, P_1(d\omega_1).$$

По предположению $\lim_n \quad (\widehat{B}_n) > 0$. Отсюда следует, что найдется такое $\omega_1^0 \in B_1$, 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

$$f_n^{(2)}(\omega_2) = \int_{\Omega_3} P(\omega_1^0, \omega_2;\, d\omega_3) \ldots \int_{\Omega_n} I_{B_n}(\omega_1^0, \omega_2, \ldots, \omega_n)\, P(\omega_1^0, \omega_2, \ldots, \omega_{n-1};\, d\omega_n).$$

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 $(\omega_1^0, \ldots, \omega_n^0, \ldots) \in \bigcap \widehat{B}_n$, но в то же время, по предположению, $\bigcap \widehat{B}_n = \varnothing$. Полученное противоречие показывает, что $\lim_n \ (\widehat{B}_n) = 0$.

Итак, утверждение теоремы в части, касающейся существования вероятностной меры , доказано. Заключительная часть очевидным образом следует из предыдущей, если положить $X_n(\omega) = \omega_n$, $n \geqslant 1$. □

Следствие 4. *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 $(\Omega, \mathscr{F},\ )$ и семейство независимых случайных элементов $X_1, X_2, \ldots$ со значениями в измеримых пространствах $(E_1, \mathscr{E}_1)$, $(E_2, \mathscr{E}_2)$, … соответственно, такие, что*

$$\{\omega\colon X_n(\omega) \in B\} = P_n(B), \qquad B \in \mathscr{E}_n, \ \ n \geqslant 1.$$

Следствие 5. *Пусть $E = \{1, 2, \ldots\}$, $\{p_k(x, y)\}$, $k \geqslant 1$, $x, y \in E$, — семейство неотрицательных функций таких, что $\sum_{y\in E} p_k(x, y) = 1$, $x \in E$, $k \geqslant 1$. Пусть, кроме того, $\pi = \pi(\cdot)$ — распределение вероятностей на E ($\pi(x) \geqslant 0$, $\sum_{x\in E} \pi(x) = 1$).*

Тогда существуют вероятностное пространство $(\Omega, \mathscr{F},\ )$ и семейство случайных величин $X = \{\xi_0, \xi_1, \ldots\}$ на нем такие, что

$$\{\xi_0 = x_0, \xi_1 = x_1, \ldots, \xi_n = x_n\} = \pi(x_0)\, p_1(x_0, x_1) \ldots p_n(x_{n-1}, x_n) \tag{16}$$

(ср. с (4) § 12 гл. I) 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 $X = \{\xi_0, \xi_1, \ldots\}$ случайных величин, 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 и с определениями в § 1 гл. VIII.)

4. Теорема Колмогорова (теорема 1) утверждает существование процесса с заданной системой согласованных конечномерных функций распределения. При этом ее доказательство требует обращения к *каноническому* вероятностному пространству, а сами процессы строятся *координатным* образом, что само по себе говорит о сложности устройства их траекторий.

С этой точки зрения значительный интерес представляют случаи *конструктивного* построения случайных процессов с заданными свойствами, при этом с минимальным использованием «вероятностных структур».

Для демонстрации таких возможностей обратимся к так называемым процессам восстановления. (Их частный случай — процесс Пуассона; см. § 10 гл. II.)

Пусть $(\sigma_1, \sigma_2, \ldots)$ — последовательность независимых одинаково распределенных положительных случайных величин с функцией распределения $F = F(x)$. (Существование такой последовательности гарантируется следствием 1 к теореме 1.)

По последовательности $(\sigma_1, \sigma_2, \ldots)$ образуем новую последовательность $(T_0, T_1, \ldots)$ с $T_0 = 0$ и

$$T_n = \sigma_1 + \ldots + \sigma_n, \qquad n \geqslant 1.$$

Для наглядности будем интерпретировать момент T_n как момент появления n-го, скажем, вызова (например, телефонного). Тогда σ_n описывает длительность времени между $(n-1)$-м и n-м вызовами.

Процессом восстановления принято называть случайный процесс $N = (N_t)_{t \geqslant 0}$ с (конструктивно заданными) величинами

$$N_t = \sum_{n=1}^{\infty} I(T_n \leqslant t). \tag{17}$$

Понятно, что N_t можно было бы определить и так:

$$N_t = \max\{n\colon\ T_n \leqslant t\}, \tag{18}$$

т. е. N_t — это число вызовов, поступивших на интервале $(0, t]$; при этом очевидно, что

$$\{N_t \geqslant n\} = \{T_n \leqslant t\}. \tag{19}$$

Эта простая формула весьма полезна, поскольку она позволяет сводить рассмотрения о вероятностных свойствах процесса $N = (N_t)_{t \geqslant 0}$ к изучению свойств величин $T_n = \sigma_1 + \ldots + \sigma_n$, являющихся суммами независимых случайных величин $\sigma_1, \ldots, \sigma_n$, $n \geqslant 1$ (см. п. 4 § 3 гл. IV и п. 4 § 2 гл. VII).

Из формулы (17) сразу вытекает, что *функция восстановления* $m(t) =$ $=\ \ N_t$, $t \geqslant 0$, определяется по функциям распределения $F_n(t) = \ \ (T_n \leqslant t)$

следующим образом:

$$m(t) = \sum_{n=1}^{\infty} F_n(t). \tag{20}$$

5. Задачи.

1. Пусть $\Omega = [0, 1]$, $\mathscr{F}$ — класс борелевских множеств на $[0, 1]$, — мера Лебега на $[0, 1]$. Показать, что пространство $(\Omega, \mathscr{F},\ )$ является универсальным в том смысле, что для любой функции распределения $F(x)$ на $(\Omega, \mathscr{F},\ )$ можно так определить случайную величину $\xi = \xi(\omega)$, что ее функция распределения $F_\xi(x) = \ \{\xi \leqslant x\}$ совпадает с функцией $F(x)$. (*Указание*: $\xi(\omega) = F^{-1}(\omega)$, где $F^{-1}(\omega) = \sup\{x : F(x) < \omega\}$, $0 < \omega < 1$, 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.

4. Пусть F_n — функции распределения величин T_n, $n \geqslant 1$ (из п. 4). Показать, что $F_{n+1}(t) = \int\limits_0^t F_n(t-s)\, dF(s)$, $n \geqslant 1$, где $F_1 = F$.

5. Показать, что $\ \{N_t = n\} = F_n(t) - F_{n+1}(t)$ (см. (17)).

6. Показать, что введенная выше в п. 4 функция восстановления $m(t)$ удовлетворяет *уравнению восстановления*

$$m(t) = F(t) + \int\limits_0^t m(t-x)\, dF(x). \tag{21}$$

7. Показать, что в классе функций, ограниченных на конечных интервалах, единственным решением уравнения (21) является функция, определяемая формулой (20).

8. Пусть T — произвольное множество.

(i) Предположим, что для каждого $t \in T$ задано некоторое вероятностное пространство $(\Omega_t, \mathscr{F}_t,\ _t)$. Положим $\Omega = \prod\limits_{t \in T} \Omega_t$, $\mathscr{F} = \bigotimes\limits_{t \in T} \mathscr{F}_t$. Доказать, что на $(\Omega, \mathscr{F})$ существует и единственна вероятностная мера такая, что

$$\left(\prod_{t \in T} B_t \right) = \prod_{t \in T} \ (B_t),$$

где $B_t \in \mathscr{F}_t$, $t \in T$, и $B_t = \Omega_t$ для всех индексов t, за исключением конечного числа. (*Указание*. Задать на подходящей алгебре и воспользоваться методом доказательства в теореме Ионеску Тулчи.)

(ii) Пусть для каждого $t \in T$ заданы измеримое пространство $(E_t, \mathscr{E}_t)$ и вероятностная мера $\ _t$ на нем. Доказать, что существуют вероятностное пространство $(\Omega, \mathscr{F},\ )$ и независимые случайные элементы $(X_t)_{t \in T}$ такие, что X_t являются $\mathscr{F}/\mathscr{E}_t$-измеримыми и $\ \{X_t \in B\} = \ _t(B)$, $B \in \mathscr{E}_t$.

§ 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 ξ, ξ_1, ξ_2, … — случайные величины, заданные на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \quad)$.

Определение 1. Последовательность случайных величин $\xi_1, \xi_2, \ldots$ (обозначаемая также $\{\xi_n\}$, $\{\xi_n\}_{n\geqslant 1}$ или (ξ_n), $(\xi_n)_{n\geqslant 1}$) называется *сходящейся по вероятности* к случайной величине ξ (обозначение: $\xi_n \to \xi$), если

$$\{|\xi_n - \xi| > \varepsilon\} \to 0, \quad n \to \infty \tag{1}$$

для любого $\varepsilon > 0$.

С этим видом сходимости мы уже встречались в связи с законом больших чисел в схеме Бернулли, утверждающим, что

$$\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon\right\} \to 0, \quad n \to \infty$$

(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

$$\{\omega\colon\ \xi_n \not\to \xi\} = 0, \tag{2}$$

т. е. если множество исходов ω, для которых $\xi_n(\omega)$ *не сходятся* к $\xi(\omega)$, имеет нулевую вероятность.

Этот вид сходимости обозначают по-разному:

$$\xi_n \to \xi \ (\quad\text{-п. н.}),\ \text{ или } \xi_n \to \xi \ (\text{п. н.}),\ \text{ или } \xi_n \xrightarrow{\text{п. н.}} \xi,\ \text{ или } \xi_n \xrightarrow{\text{п.в.}} \xi.$$

Определение 3. Последовательность случайных величин $\xi_1, \xi_2, \ldots$ называется сходящейся *в среднем порядка* p, $0 < p < \infty$, к случайной величине ξ, если

$$|\xi_n - \xi|^p \to 0, \quad n \to \infty. \tag{3}$$

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 $\xi = \mathrm{l.i.m.}\ \xi_n$ (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: $\xi_n \xrightarrow{d} \xi$, $\xi_n \xrightarrow{law} \xi$; $Law(\xi_n) \to Law(\xi)$; d — от *distribution* — распределение, *law*, *Law* — закон), если для любой ограниченной непрерывной функции $f = f(x)$

$$f(\xi_n) \to \quad f(\xi), \qquad n \to \infty. \tag{4}$$

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* (*сходимость в основном*; обозначение: $F_{\xi_n} \Rightarrow F_\xi$).

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 $\{\xi_n\}_{n \geqslant 1}$ (или просто $\{\xi_n\}$) *фундаментальна по вероятности*, *с вероятностью единица* или *в среднем порядка* p, $0 < p < \infty$, если выполнены соответственно следующие условия: для любого $\varepsilon > 0$ $\quad \{|\xi_n - \xi_m| \geqslant \varepsilon\} \to 0$, $n, m \to \infty$; 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. $\quad |\xi_n - \xi_m|^p \to 0$, $n, m \to \infty$.

3. Теорема 1. a) *Для того чтобы* $\xi_n \to \xi$ (-*п. н.*), *необходимо и достаточно*, *чтобы для любого* $\varepsilon > 0$

$$\Big\{\sup_{k \geqslant n} |\xi_k - \xi| \geqslant \varepsilon\Big\} \to 0, \qquad n \to \infty. \tag{5}$$

b) *Последовательность* $\{\xi_n\}_{n \geqslant 1}$ *фундаментальна с вероятностью единица тогда и только тогда*, *когда для любого* $\varepsilon > 0$

$$\Big\{\sup_{k \geqslant n, l \geqslant n} |\xi_k - \xi_l| \geqslant \varepsilon\Big\} \to 0, \qquad n \to \infty, \tag{6}$$

или, что эквивалентно,

$$\left\{\sup_{k\geqslant 0} |\xi_{n+k} - \xi_n| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{7}$$

Доказательство. a) Пусть $A_n^\varepsilon = \{\omega\colon |\xi_n - \xi| \geqslant \varepsilon\}$, $A^\varepsilon = \overline{\lim} A_n^\varepsilon \equiv \equiv \bigcap_{n=1}^{\infty} \bigcup_{k\geqslant n} A_k^\varepsilon$. Тогда

$$\{\omega\colon \xi_n \not\to \xi\} = \bigcup_{\varepsilon>0} A^\varepsilon = \bigcup_{m=1}^{\infty} A^{1/m}.$$

Но

$$(A^\varepsilon) = \lim_n \left(\bigcup_{k\geqslant n} A_k^\varepsilon\right),$$

поэтому утверждение a) является результатом следующей цепочки импликаций:

$$\{\omega\colon \xi_n \not\to \xi\} = 0 \;\Leftrightarrow\; \left(\bigcup_{\varepsilon>0} A^\varepsilon\right) = 0 \;\Leftrightarrow\; \left(\bigcup_{m=1}^{\infty} A^{1/m}\right) = 0 \;\Leftrightarrow$$

$$\Leftrightarrow\; (A^{1/m}) = 0,\; m \geqslant 1 \;\Leftrightarrow\; (A^\varepsilon) = 0,\; \varepsilon > 0 \;\Leftrightarrow$$

$$\Leftrightarrow\; \left(\bigcup_{k\geqslant n} A_k^\varepsilon\right) \to 0,\; n\to\infty,\; \varepsilon>0 \;\Leftrightarrow\; \left\{\sup_{k\geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right\} \to 0,\; n\to\infty,\; \varepsilon>0.$$

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 $\{\omega\colon \{\xi_n(\omega)\}$ *не фундаментальна*$\} = \bigcup_{\varepsilon>0} B^\varepsilon$, и так же, как в а), показывается, что $\{\omega\colon \{\xi_n(\omega)\}$ *не фундаментальна*$\} = 0 \;\Leftrightarrow\;$ (6). Эквивалентность же утверждений (6) и (7) следует из очевидных неравенств

$$\sup_{k\geqslant 0} |\xi_{n+k} - \xi_n| \leqslant \sup_{k\geqslant n, l\geqslant n} |\xi_{n+k} - \xi_{n+l}| \leqslant 2 \sup_{k\geqslant 0} |\xi_{n+k} - \xi_n|. \qquad \square$$

Следствие. *Поскольку*

$$\left\{\sup_{k\geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right\} = \left(\bigcup_{k\geqslant n} \{|\xi_k - \xi| \geqslant \varepsilon\}\right) \leqslant \sum_{k\geqslant n} \{|\xi_k - \xi| \geqslant \varepsilon\},$$

то выполнение для каждого $\varepsilon > 0$ *условия*

$$\sum_{k=1}^{\infty} \{|\xi_k - \xi| \geqslant \varepsilon\} < \infty \tag{8}$$

достаточно для сходимости $\xi_n \to \xi$ (*-п. н.*).

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. 1 в § 1), что через $\{A_n \text{ б. ч.}\}$ обозначается событие $\overline{\lim}\, A_n$, состоящее в том, что произойдет *бесконечное число* событий из $A_1, A_2, \ldots$

Лемма 1 (Бореля—Кантелли). а) *Если* $\sum \quad (A_n) < \infty$, *то вероятность* $\quad \{A_n$ *б. ч.*$\} = 0$.

b) *Если* $\sum \quad (A_n) = \infty$ *и события* $A_1, A_2, \ldots$ *независимы, то вероятность* $\quad \{A_n$ *б. ч.*$\} = 1$.

Доказательство. а) По определению

$$\{A_n \text{ б. ч.}\} = \overline{\lim}\, A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k.$$

Поэтому

$$\{A_n \text{ б. ч.}\} = \quad \left(\bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k\right) = \lim \quad \left(\bigcup_{k \geqslant n} A_k\right) \leqslant \lim \sum_{k \geqslant n} \quad (A_k),$$

откуда и следует утверждение а).

b) Если события $A_1, A_2, \ldots$ *независимы*, то таковыми же будут и события $\overline{A}_1, \overline{A}_2, \ldots$ Тогда для любого $N \geqslant n$

$$\left(\bigcap_{k=n}^{N} \overline{A}_k\right) = \prod_{k=n}^{N} \quad (\overline{A}_k),$$

откуда нетрудно вывести, что

$$\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) = \prod_{k=n}^{\infty} \quad (\overline{A}_k). \tag{9}$$

В силу неравенства $\ln(1-x) \leqslant -x$, $0 \leqslant x < 1$,

$$\ln \prod_{k=n}^{\infty} [1 - \quad (A_k)] = \sum_{k=n}^{\infty} \ln[1 - \quad (A_k)] \leqslant -\sum_{k=n}^{\infty} \quad (A_k) = -\infty.$$

Следовательно, для любого n

$$\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) = 0$$

и, значит, $\quad \{A_n \text{ б. ч.}\} = 1$. □

Следствие 1. *Если* $A_n^\varepsilon = \{\omega\colon |\xi_n - \xi| \geqslant \varepsilon\}$, *то условие* (8) *означает, что* $\sum\limits_{n=1}^{\infty} \quad (A_n^\varepsilon) < \infty$, $\varepsilon > 0$, *и по лемме Бореля—Кантелли* $\quad (A^\varepsilon) = 0$, $\varepsilon > 0$, *где* $A^\varepsilon = \overline{\lim}\, A_n^\varepsilon$ $(= \{A_n^\varepsilon$ *б. ч.*$\})$. *Тем самым*

$$\sum_{k=1}^{\infty} \quad \{|\xi_k - \xi| \geqslant \varepsilon\} < \infty,\ \ \varepsilon > 0 \ \Rightarrow \quad (A^\varepsilon) = 0,\ \ \varepsilon > 0 \ \Leftrightarrow \quad \{\omega\colon \xi_n \not\to \xi\} = 0,$$

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 сходимость по вероятности* $\xi_n \to \xi$ *достаточно «быстрая», в том смысле, что*

$$\sum_{n=1}^{\infty} \quad \{|\xi_n - \xi| \geqslant \varepsilon_n\} < \infty, \tag{10}$$

то имеет место и сходимость почти наверное: $\xi_n \xrightarrow{\text{п. н.}} \xi$.

В самом деле, пусть $A_n = \{|\xi_n - \xi| \geqslant \varepsilon_n\}$. Тогда по лемме Бореля—Кантелли $\quad \{A_n$ 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*:

$$\xi_n \xrightarrow{\text{п. н.}} \xi \ \Rightarrow\ \xi_n \to \xi, \tag{11}$$

$$\xi_n \xrightarrow{L^p} \xi \ \Rightarrow\ \xi_n \to \xi, \qquad p > 0, \tag{12}$$

$$\xi_n \to \xi \ \Rightarrow\ \xi_n \xrightarrow{d} \xi. \tag{13}$$

Доказательство. Утверждение (11) следует из сравнения определения сходимости по вероятности с критерием (5), а импликация (12) — из неравенства Чебышева.

Докажем теперь импликацию (13). Пусть $f(x)$ — непрерывная функция, $|f(x)| \leqslant c$, $\varepsilon > 0$ и N таково, что $\quad \{|\xi| > N\} \leqslant \varepsilon/(4c)$. Выберем δ таким, чтобы для всех $|x| \leqslant N$ и $|x - y| \leqslant \delta$ было выполнено неравенство $|f(x) - f(y)| \leqslant \varepsilon/2$. Тогда (ср. с «вероятностным» доказательством теоремы Вейерштрасса в п. 5 § 5 гл. 1)

$$\begin{aligned}
|f(\xi_n) - f(\xi)| = {}&\quad (|f(\xi_n) - f(\xi)|;\ |\xi_n - \xi| \leqslant \delta,\ |\xi| \leqslant N) + {}\\
&+ \quad (|f(\xi_n) - f(\xi)|;\ |\xi_n - \xi| \leqslant \delta,\ |\xi| > N) + \quad (|f(\xi_n) - f(\xi)|;\ |\xi_n - \xi| > \delta) \leqslant \\
&\leqslant \varepsilon/2 + \varepsilon/2 + 2c \quad \{|\xi_n - \xi| > \delta\} = \varepsilon + 2c \quad \{|\xi_n - \xi| > \delta\}.
\end{aligned}$$

Но $\quad \{|\xi_n - \xi| > \delta\} \to 0$, поэтому для достаточно больших n $\quad |f(\xi_n) - f(\xi)| \leqslant \leqslant 2\varepsilon$, что в силу произвольности $\varepsilon > 0$ доказывает импликацию (13). □

Приведем ряд примеров, показывающих, в частности, что в (11), (12) обратные импликации, вообще говоря, несправедливы.

Пример 1 ($\xi_n \to \xi \;\not\Rightarrow\; \xi_n \xrightarrow{\text{п. н.}} \xi$; $\xi_n \xrightarrow{L^p} \xi \;\not\Rightarrow\; \xi_n \xrightarrow{\text{п. н.}} \xi$)**.** Пусть $\Omega = [0, 1]$, $\mathscr{F} = \mathscr{B}([0, 1])$, — мера Лебега. Положим

$$A_n^i = \left[\frac{i-1}{n}, \frac{i}{n}\right], \quad \xi_n^i = I_{A_n^i}(\omega), \quad i = 1, 2, \ldots, n; \quad n \geqslant 1.$$

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 ($\xi_n \to \xi \;\Leftarrow\; \xi_n \xrightarrow{\text{п. н.}} \xi \;\not\Rightarrow\; \xi_n \xrightarrow{L^p} \xi$, $p > 0$)**.** Снова пусть $\Omega = [0, 1]$, $\mathscr{F} = \mathscr{B}([0, 1])$, — 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 $p > 0$

$$|\xi_n|^p = \frac{e^{np}}{n} \to \infty, \quad n \to \infty.$$

Пример 3 ($\xi_n \xrightarrow{L^p} \xi \;\not\Rightarrow\; \xi_n \xrightarrow{\text{п. н.}} \xi$)**.** Пусть $\{\xi_n\}$ — последовательность независимых случайных величин с

$$\{\xi_n = 1\} = p_n, \qquad \{\xi_n = 0\} = 1 - p_n.$$

Тогда нетрудно установить, что

$$\xi_n \to 0 \;\Leftrightarrow\; p_n \to 0, \quad n \to \infty, \tag{14}$$

$$\xi_n \xrightarrow{L^p} 0 \;\Leftrightarrow\; p_n \to 0, \quad n \to \infty, \tag{15}$$

$$\xi_n \xrightarrow{\text{п. н.}} 0 \;\Leftrightarrow\; \sum_{n=1}^{\infty} p_n < \infty. \tag{16}$$

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 $\{\xi_n\}_{n \geqslant 1}$ — последовательность неотрицательных случайных величин таких, что $\xi_n \xrightarrow{\text{п. н.}} \xi$ и $\xi_n \to \xi < \infty$. Тогда*

$$|\xi_n - \xi| \to 0, \quad n \to \infty. \tag{17}$$

Доказательство. Для достаточно больших n $\xi_n < \infty$, поэтому для них

$$|\xi - \xi_n| = (\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} + (\xi_n - \xi) I_{\{\xi_n > \xi\}} = 2 (\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} + (\xi_n - \xi).$$

Но $0 \leqslant (\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} \leqslant \xi$. Тогда по теореме о мажорируемой сходимости $\lim\limits_n (\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} = 0$, что вместе с предположением $\xi_n \to \xi$ доказывает (17). $\square$

Замечание. Теорема о мажорируемой сходимости (теорема 3 § 6) 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 «$\xi_n \longrightarrow \xi$».

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* $\{\xi_n\}$ *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

$$\sup_{k \geqslant n, l \geqslant n} |\xi_k - \xi_l| \leqslant \sup_{k \geqslant n} |\xi_k - \xi| + \sup_{l \geqslant n} |\xi_l - \xi|,$$

откуда (см. теорему 1) вытекает необходимость условия теоремы.

Пусть теперь последовательность $\{\xi_n\}$ 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 $\{\xi_n(\omega)\}$ 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, $\xi_n \xrightarrow{\text{п. н.}} \xi$. $\square$

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 подпоследова-*

тельность $\{\xi_{n_k}\}$, *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 $s \geqslant n$, $t \geqslant n$

$$\{|\xi_t - \xi_s| > 2^{-k}\} < 2^{-k}.$$

Тогда

$$\sum_k \quad \{|\xi_{n_{k+1}} - \xi_{n_k}| > 2^{-k}\} < \sum 2^{-k} < \infty$$

и по лемме Бореля—Кантелли

$$\{|\xi_{n_{k+1}} - \xi_{n_k}| > 2^{-k} \text{ б. ч.}\} = 0.$$

Поэтому с вероятностью единица

$$\sum_{k=1}^{\infty} |\xi_{n_{k+1}} - \xi_{n_k}| < \infty.$$

Пусть $\mathscr{N} = \{\omega\colon \sum |\xi_{n_{k+1}} - \xi_{n_k}| = \infty\}$. Тогда, если положить

$$\xi(\omega) = \begin{cases} \xi_{n_1}(\omega) + \sum\limits_{k=1}^{\infty} (\xi_{n_{k+1}}(\omega) - \xi_{n_k}(\omega)), & \omega \in \Omega \setminus \mathscr{N}, \\ 0, & \omega \in \mathscr{N}, \end{cases}$$

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 6 (критерий Коши сходимости по вероятности). *Для того чтобы последовательность случайных величин* $\{\xi_n\}$ *была сходящейся по вероятности, необходимо и достаточно, чтобы она была фундаментальна по вероятности.*

Доказательство. Если $\xi_n \to \xi$, то

$$\{|\xi_n - \xi_m| \geqslant \varepsilon\} \leqslant \quad \{|\xi_n - \xi| \geqslant \varepsilon/2\} + \quad \{|\xi_m - \xi| \geqslant \varepsilon/2\} \tag{19}$$

и, значит, 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

$$\{|\xi_n - \xi| \geqslant \varepsilon\} \leqslant \quad \{|\xi_n - \xi_{n_k}| \geqslant \varepsilon/2\} + \quad \{|\xi_{n_k} - \xi| \geqslant \varepsilon/2\},$$

откуда ясно, что $\xi_n \longrightarrow \xi$. □

В связи со сходимостью в среднем порядка $p > 0$ сделаем прежде всего несколько замечаний о пространствах L^p.

Будем обозначать через $L^p = L^p(\Omega, \mathscr{F}, \quad)$ — пространство случайных величин $\xi = \xi(\omega)$ с $\quad |\xi|^p \equiv \int_\Omega |\xi|^p \, d \quad < \infty$. Предположим, что $p \geqslant 1$ и положим

$$\|\xi\|_p = (\quad |\xi|^p)^{1/p}.$$

Ясно, что

$$\|\xi\|_p \geqslant 0, \tag{20}$$

$$\|c\xi\|_p = |c| \, \|\xi\|_p, \quad c \text{ — постоянная}, \tag{21}$$

и в силу неравенства Минковского (31) § 6

$$\|\xi + \eta\|_p \leqslant \|\xi\|_p + \|\eta\|_p. \tag{22}$$

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

$$\|\xi\|_p = 0 \;\Rightarrow\; \xi = 0, \tag{23}$$

которое, конечно, вообще говоря, не выполнено, поскольку, согласно свойству **H** (§ 6), можно утверждать лишь, что $\xi = 0$ не тождественно, а только *почти наверное*.

Это обстоятельство приводит к несколько иному взгляду на пространство L^p. Именно, свяжем с каждой случайной величиной $\xi \in L^p$ класс $[\xi]$ эквивалентных ей случайных величин из L^p (ξ и η *эквивалентны*, если $\xi = \eta$ почти наверное). Нетрудно убедиться, что свойство эквивалентности рефлексивно, симметрично и транзитивно, а значит, линейное пространство L^p можно разбить на взаимно непересекающиеся *классы* эквивалентных между собой случайных величин. Если теперь под $[L^p]$ понимать совокупность всех таких классов $[\xi]$ эквивалентных между собой случайных величин $\xi \in L^p$ и положить по определению

$$[\xi] + [\eta] = [\xi + \eta],$$

$$a[\xi] = [a\xi], \quad a \text{ — константа},$$

$$\|[\xi]\|_p = \|\xi\|_p,$$

то $[L^p]$ становится линейным нормированным пространством.

В функциональном анализе об элементах пространства $[L^p]$ обычно принято говорить не как о классах эквивалентных функций, а просто как о функциях. В этом смысле мы не будем вводить нового обозначения $[L^p]$ и впредь под L^p будем понимать именно множество классов эквивалентных функций, по-прежнему называя их просто элементами, функциями, случайными величинами…

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 $\{\xi_n\}$ 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

$$A_k = \{\omega\colon\ |\xi_{n_{k+1}} - \xi_{n_k}| \geqslant 2^{-k}\}.$$

Тогда в силу неравенства Чебышева

$$(A_k) \leqslant \frac{|\xi_{n_{k+1}} - \xi_{n_k}|^p}{2^{-kp}} \leqslant \frac{2^{-2kp}}{2^{-kp}} = 2^{-kp} \leqslant 2^{-k}.$$

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 (§ 6)

$$|\xi_n - \xi|^p = \left\{\lim_{n_k \to \infty} |\xi_n - \xi_{n_k}|^p\right\} = \left\{\varliminf_{n_k \to \infty} |\xi_n - \xi_{n_k}|^p\right\} \leqslant$$
$$\leqslant \varliminf_{n_k \to \infty} |\xi_n - \xi_{n_k}|^p = \varliminf_{n_k \to \infty} \|\xi_n - \xi_{n_k}\|_p^p \leqslant \varepsilon.$$

Следовательно, $|\xi_n - \xi|^p \to 0$, $n \to \infty$. Ясно также, что поскольку $\xi = (\xi - \xi_n) + \xi_n$, то в силу неравенства Минковского $|\xi|^p < \infty$. $\square$

Замечание 1. В соответствии с терминологией функционального анализа полные нормированные линейные пространства называются *банаховыми* пространствами. Таким образом, пространства L^p, $p \geqslant 1$, являются банаховыми.

Замечание 2. Если $0 < p < 1$, то $\|\xi\|_p = (\ \ |\xi|^p)^{1/p}$ 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 $d(\xi, \eta) \equiv \ \ |\xi - \eta|^p$.

Замечание 3. Обозначим $L^\infty = L^\infty(\Omega, \mathscr{F}, \ \ )$ пространство (классов эквивалентности) случайных величин $\xi = \xi(\omega)$, для которых $\|\xi\|_\infty < \infty$, где величина $\|\xi\|_\infty$, называемая *существенным супремумом* ξ, определяется формулой

$$\|\xi\|_\infty \equiv \operatorname{ess\,sup} |\xi| \equiv \inf\{0 \leqslant c \leqslant \infty : \ \ (|\xi| > c) = 0\}.$$

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 $\xi_n \to \xi$ и в то же время $\xi_n \to \eta$, то ξ и η эквивалентны (в том смысле, что $\ \ \{\xi \neq \eta\} = 0$).

4. Пусть $\xi_n \to \xi$, $\eta_n \to \eta$ и случайные величины ξ и η эквивалентны. Показать, что для любого $\varepsilon > 0$

$$\{|\xi_n - \eta_n| \geqslant \varepsilon\} \to 0, \quad n \to \infty.$$

5. Пусть $\xi_n \to \xi$, $\eta_n \to \eta$. Показать, что если $\varphi = \varphi(x, y)$ — непрерывная функция, то $\varphi(\xi_n, \eta_n) \to \varphi(\xi, \eta)$ (*лемма Слуцкого*).

6. Пусть $(\xi_n - \xi)^2 \to 0$. Показать, что $\xi_n^2 \to \xi^2$.

7. Показать, что если $\xi_n \xrightarrow{d} C$, где C — постоянная, то имеет место и сходимость по вероятности:

$$\xi_n \xrightarrow{d} C \ \Rightarrow\ \xi_n \to C.$$

8. Пусть последовательность $\{\xi_n\}_{n\geqslant 1}$ такова, что для некоторого $p > 0$ $\sum\limits_{n=1}^{\infty} \ \ |\xi_n|^p < \infty$. Показать, что $\xi_n \to 0$ (-п. н.).

9. Пусть $\{\xi_n\}_{n\geqslant 1}$ — последовательность одинаково распределенных случайных величин.

Доказать, что

$$|\xi_1| < \infty \;\Leftrightarrow\; \sum_{n=1}^{\infty} \quad \{|\xi_1| > \varepsilon n\} < \infty, \;\; \varepsilon > 0 \;\Leftrightarrow$$

$$\Leftrightarrow \sum_{n=1}^{\infty} \quad \left\{\left|\frac{\xi_n}{n}\right| > \varepsilon\right\} < \infty, \;\; \varepsilon > 0 \;\Rightarrow\; \frac{\xi_n}{n} \to 0 \quad (\quad\text{-п. н.}).$$

10. Пусть $\{\xi_n\}_{n\geqslant 1}$ — некоторая последовательность случайных величин. Предположим, что существуют случайная величина ξ и подпоследовательность $\{n_k\}$ такие, что $\xi_{n_k} \to \xi$ (-п. н.) и $\max\limits_{n_{k-1} < l \leqslant n_k} |\xi_l - \xi_{n_{k-1}}| \to 0$ (-п. н.) при $k \to \infty$. Показать, что тогда $\xi_n \to \xi$ (-п. н.).

11. Определим «d-метрику» в множестве случайных величин, полагая

$$d(\xi, \eta) = \quad \frac{|\xi - \eta|}{1 + |\xi - \eta|}$$

и отождествляя случайные величины, совпадающие почти наверное. Показать, что $d = d(\xi, \eta)$ действительно задает метрику и сходимость по вероятности эквивалентна сходимости в этой метрике.

12. Показать, что не существует метрики в множестве случайных величин такой, что сходимость в ней эквивалентна сходимости почти наверное.

13. Пусть $X_1 \leqslant X_2 \leqslant \ldots$ и $X_n \longrightarrow X$. Показать, что $X_n \to X$ (-п. н.).

14. Пусть $X_n \to X$ (-п. н.). Тогда и $n^{-1} \sum\limits_{k=1}^{n} X_k \to X$ (-п. н.) (*суммирование по Чезаро*). Показать на примере, что сходимость -п. н. нельзя заменить на сходимость по вероятности.

15. Пусть $(\Omega, \mathscr{F},\quad)$ — вероятностное пространство и $X_n \longrightarrow X$. Показать, что если мера является атомической, то $X_n \to X$ также и с вероятностью единица. (Множество $A \in \mathscr{F}$ называется -*атомом*, если для всякого $B \in \mathscr{F}$ или $(B \cap A) = \quad (A)$, или $(B \cap A) = 0$. Мера называется *атомической*, если существует счетное семейство $\{A_n\}$ непересекающихся -атомов таких, что $\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = 1$.)

16. Согласно (первой) лемме Бореля—Кантелли, если $\sum\limits_{n=1}^{\infty} \quad (|\xi_n| > \varepsilon) < \infty$ для $\varepsilon > 0$, то последовательность $\xi_n \to 0$ (-п. н.). Дать пример, показывающий, что сходимость $\xi_n \to 0$ (-п. н.) может иметь место и при условии $\sum\limits_{n=1}^{\infty} \quad (|\xi_n| > \varepsilon) = \infty$, $\varepsilon > 0$.

17. (*Ко второй лемме Бореля—Кантелли.*) Пусть $\Omega = (0, 1)$, $\mathscr{B} = \mathscr{B}((0, 1))$, — лебегова мера. Рассмотрим события $A_n = (0, 1/n)$. Пока-

зать, что $\sum \quad (A_n) = \infty$, но каждое ω из (0, 1) может принадлежать только конечному числу множеств $A_1, \ldots, A_{[1/\omega]}$, т. е. $\quad \{A_n \text{ б. ч.}\} = 0$.

18. Привести пример последовательности случайных величин такой, что с вероятностью единица $\limsup \xi_n = \infty$, $\liminf \xi_n = -\infty$, но тем не менее существует случайная величина η такая, что $\xi_n \to \eta$.

19. Пусть Ω есть не более чем счетное множество. Доказать, что если $\xi_n \to \xi$, то $\xi_n \to \xi$ (-п. н.).

20. Пусть $A_1, A_2, \ldots$ — независимые события и $\sum\limits_{n=1}^{\infty} \quad (A_n) < \infty$. Доказать, что для $S_n = \sum\limits_{k=1}^{n} I(A_k)$ справедливо усиление «второй леммы Бореля—Кантелли»:

$$\lim_n \frac{S_n}{S_n} = 1 \quad (\text{ -п. н.}).$$

21. Пусть $(X_n)_{n \geqslant 1}$ и $(Y_n)_{n \geqslant 1}$ — две последовательности случайных величин, у которых совпадают все конечномерные распределения ($F_{X_1,\ldots,X_n} = F_{Y_1,\ldots,Y_n}$, $n \geqslant 1$). Пусть $X_n \to X$. Доказать, что тогда имеет место сходимость $Y_n \to Y$ к некоторой случайной величине Y, распределение которой совпадает с распределением X.

22. Пусть $(X_n)_{n \geqslant 1}$ — последовательность независимых случайных величин таких, что $X_n \to X$ для некоторой случайной величины X. Доказать, что X является вырожденной случайной величиной.

23. Показать, что для каждой последовательности случайных величин $\xi_1, \xi_2, \ldots$ можно найти такую последовательность *констант* $a_1, a_2, \ldots$, что $\xi_n / a_n \to 0$ (-п. н.).

24. Пусть $\xi_1, \xi_2, \ldots$ — последовательность случайных величин и $S_n = \xi_1 + \ldots + \xi_n$, $n \geqslant 1$. Показать, что множество $\{S_n \to\}$, т. е. множество тех $\omega \in \Omega$, где ряд $\sum\limits_{k \geqslant 1} \xi_k(\omega)$ сходится, может быть представлено в следующем виде:

$$\{S_n \to\} = \bigcap_{N \geqslant 1} \bigcup_{m \geqslant 1} \bigcap_{k \geqslant m} \left\{ \sup_{l \geqslant k} |S_l - S_k| \leqslant N^{-1} \right\}.$$

Соответственно, множество $\{S_n \not\to\}$, где ряд $\sum\limits_{k \geqslant 1} \xi_k(\omega)$ расходится, представимо в виде

$$\{S_n \not\to\} = \bigcup_{N \geqslant 1} \bigcap_{m \geqslant 1} \bigcup_{k \geqslant m} \left\{ \sup_{l \geqslant k} |S_l - S_k| > N^{-1} \right\}.$$

25. Доказать следующий вариант *второй леммы Бореля—Кантелли* (утверждение b) в лемме 1): пусть $A_1, A_2, \ldots$ — произвольные (не обязательно независимые) события такие, что

$$\sum_{n=1}^{\infty} \quad (A_n) = \infty \quad \text{и} \quad \liminf_n \frac{\sum\limits_{i,k \leqslant n} \quad (A_i \cap A_k)}{\left[\sum\limits_{1<k \leqslant n} \quad (A_k)\right]^2} \leqslant 1;$$

тогда $\quad (A_n \text{ б. ч.}) = 1$.

26. Показать, что во второй лемме Бореля—Кантелли вместо независимости событий $A_1, A_2, \ldots$ достаточно требовать лишь их *попарную* независимость.

27. Доказать следующий вариант *закона нуля или единицы* (ср. с законами нуля или единицы в § 1 гл. IV): если события $A_1, A_2, \ldots$ попарно независимы, то

$$\{A_n \text{ б. ч.}\} = \begin{cases} 0, & \text{если } \sum \quad (A_n) < \infty, \\ 1, & \text{если } \sum \quad (A_n) = \infty. \end{cases}$$

28. Пусть $A_1, A_2, \ldots$ — произвольная последовательность событий таких, что $\lim\limits_n \quad (A_n) = 0$ и $\sum\limits_n \quad (A_n \cap \bar{A}_{n+1}) < \infty$. Доказать, что тогда $\{A_n \text{ б. ч.}\} = 0$.

29. Доказать, что если $\sum\limits_n \quad \{|\xi_n| > n\} < \infty$, то $\limsup\limits_n (|\xi_n|/n) \leqslant 1$ (-п. н.).

30. Пусть $\xi_n \downarrow \xi$ (-п. н.), $\quad |\xi_n| < \infty$, $n \geqslant 1$, и $\inf\limits_n \quad \xi_n > -\infty$. Показать, что тогда $\xi_n \xrightarrow{L^1} \xi$, т. е. $\quad |\xi_n - \xi| \to 0$.

31. В связи с леммой Бореля—Кантелли показать, что $\quad \{A_n \text{ б. ч.}\} = 1$, если и только если $\sum\limits_n \quad (A \cap A_n) = \infty$ для каждого множества A с $\quad (A) > 0$.

32. Пусть события $A_1, A_2, \ldots$ независимы и $\quad (A_n) < 1$ для всех $n \geqslant 1$. Тогда $\quad \{A_n \text{ б. ч.}\} = 1$, если и только если $\quad (\bigcup A_n) = 1$.

33. Пусть $X_1, X_2, \ldots$ — независимые случайные величины с $\quad \{X_n = 0\} = 1/n$ и $\quad \{X_n = 1\} = 1 - 1/n$. Пусть $E_n = \{X_n = 0\}$. Показать, что $\sum\limits_{n=1}^{\infty} \quad (E_n) = \infty$, $\sum\limits_{n=1}^{\infty} \quad (\bar{E}_n) = \infty$. Заключить отсюда, что $\lim\limits_n X_n$ не существует (-п. н.).

34. Пусть $X_1, X_2, \ldots$ — последовательность случайных величин. Показать, что $X_n \to 0$ тогда и только тогда, когда

$$\frac{|X_n|^r}{1 + |X_n|^r} \to 0 \quad \text{для некоторого } r > 0.$$

В частности, если $S_n = X_1 + \ldots + X_n$, то

$$\frac{S_n - \quad S_n}{n} \to 0 \;\Leftrightarrow\; \quad \frac{(S_n - \quad S_n)^2}{n^2 + (S_n - \quad S_n)^2} \to 0.$$

Показать, что для произвольной последовательности $X_1, X_2, \ldots$

$$\max_{1 \leqslant k \leqslant n} |X_k| \to 0 \;\Rightarrow\; \frac{S_n}{n} \to 0.$$

35. Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные бернуллиевские величины: $\quad \{X_k = \pm 1\} = 1/2$. Пусть $U_n = \sum\limits_{k=1}^{n} \dfrac{X_k}{2^k}$, $n \geqslant 1$. Показать, что $U_n \to U$ (-п. н.), где U — случайная величина, равномерно распределенная на $(-1, +1)$.

§ 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 $L^2 = L^2(\Omega, \mathscr{F}, \quad)$ — пространство (классов эквивалентных) случайных величин с конечным вторым моментом.

Если $\xi, \eta \in L^2$, то положим

$$(\xi, \eta) \equiv \quad \xi\eta. \tag{1}$$

Ясно, что для $\xi, \eta, \zeta \in L^2$

$$(a\xi + b\eta, \zeta) = a(\xi, \zeta) + b(\eta, \zeta), \qquad a, b \in R,$$
$$(\xi, \xi) \geqslant 0$$

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 $(\xi, \eta) \equiv \quad \xi\eta = 0$. Согласно § 8, величины ξ и η называются некоррелированными, если $\quad (\xi, \eta) = 0$, т. е. если

$$\xi\eta = \quad \xi \cdot \quad \eta.$$

Отсюда следует, что для случайных величин с нулевыми средними понятия их *ортогональности* и *некоррелированности* совпадают.

Система $M \subseteq L^2$ будет называться *системой ортогональных* случайных величин, если $\xi \perp \eta$ для любых $\xi, \eta \in M$ $(\xi \neq \eta)$.

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

$$\left|\xi - \sum_{i=1}^{n} a_i \eta_i\right|^2 \equiv \left\|\xi - \sum_{i=1}^{n} a_i \eta_i\right\|^2 = \left(\xi - \sum_{i=1}^{n} a_i \eta_i,\ \xi - \sum_{i=1}^{n} a_i \eta_i\right) =$$
$$= \|\xi\|^2 - 2\sum_{i=1}^{n} a_i (\xi, \eta_i) + \left(\sum_{i=1}^{n} a_i \eta_i,\ \sum_{i=1}^{n} a_i \eta_i\right) =$$
$$= \|\xi\|^2 - 2\sum_{i=1}^{n} a_i (\xi, \eta_i) + \sum_{i=1}^{n} a_i^2 =$$
$$= \|\xi\|^2 - \sum_{i=1}^{n} |(\xi, \eta_i)|^2 + \sum_{i=1}^{n} |a_i - (\xi, \eta_i)|^2 \geqslant \|\xi\|^2 - \sum_{i=1}^{n} |(\xi, \eta_i)|^2, \quad (3)$$

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 инфимум $\left|\xi - \sum_{i=1}^{n} a_i \eta_i\right|^2$ 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 $\eta_2, \ldots, \eta_n$ является оценка

$$\hat{\xi} = \sum_{i=1}^{n} (\xi,\, \eta_i)\eta_i. \tag{4}$$

При этом

$$\Delta \equiv \inf \quad \left| \xi - \sum_{i=1}^{n} a_i \eta_i \right|^2 = \quad |\xi - \hat{\xi}|^2 = \|\xi\|^2 - \sum_{i=1}^{n} |(\xi,\, \eta_i)|^2 \tag{5}$$

(ср. с (17) § 4 гл. I и (13) § 8).

Из (3) вытекает также следующее *неравенство Бесселя*: если $M = \{\eta_1,\, \eta_2,\, \ldots\}$ — 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

$$\xi = \underset{n}{\text{l.i.m.}} \sum_{i=1}^{n} (\xi,\, \eta_i)\eta_i. \tag{7}$$

Оценку $\hat{\xi}$, являющуюся *оптимальной линейной оценкой*, часто обозначают $\widehat{\ }(\xi \,|\, \eta_1, \ldots, \eta_n)$ и называют *условным математическим ожиданием* (ξ относительно $\eta_1, \ldots, \eta_n$) *в широком смысле*.

Это название объясняется следующим. Если рассматривать *всевозможные* оценки $\varphi = \varphi(\eta_1, \ldots, \eta_n)$ случайной величины ξ по $\eta_1, \ldots, \eta_n$ (φ — борелевская функция), то оптимальной оценкой будет оценка $\varphi^* = (\xi \,|\, \eta_1, \ldots, \eta_n)$, т. е. условное математическое ожидание ξ относительно $\eta_1, \ldots, \eta_n$ (ср. с теоремой 1 § 8). Поэтому оптимальную *линейную* оценку по аналогии обозначают $\widehat{\ }(\xi \,|\, \eta_1, \ldots, \eta_n)$ и называют условным математическим ожиданием в широком смысле. В этой связи отметим, что если случайные величины $\eta_1, \ldots, \eta_n$ образуют гауссовскую систему (см. далее § 13), то $(\xi \,|\, \eta_1, \ldots, \eta_n)$ и $\widehat{\ }(\xi \,|\, \eta_1, \ldots, \eta_n)$ совпадают.

Остановимся на *геометрическом смысле* оценки $\hat{\xi} = \widehat{\ }(\xi \,|\, \eta_1, \ldots, \eta_n)$.

Обозначим через $\mathscr{L} = \mathscr{L}(\eta_1, \ldots, \eta_n)$ *линейное многообразие*, порожденное ортонормированной системой случайных величин $\eta_1, \ldots, \eta_n$ (т. е. совокупность случайных величин вида $\sum\limits_{i=1}^{n} a_i \eta_i$, $a_i \in R$).

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 $\eta_1, \ldots, \eta_n$ позволило просто найти оптимальную линейную оценку (проекцию) $\hat{\xi}$ для ξ по $\eta_1, \ldots, \eta_n$. Сложнее обстоит дело, если отказаться от предположения ортонормированности. Однако случай произвольных величин $\eta_1, \ldots, \eta_n$ 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

$$\sum_{i=1}^{n} a_i \eta_i = 0 \quad (\ \text{-п. н.})$$

выполнено лишь тогда, когда все a_i равны нулю.

Рассмотрим матрицу ковариаций

$$\mathbb{R} \equiv \ \ \eta\eta^*$$

вектора $\eta = (\eta_1, \ldots, \eta_n)$, рассматриваемого как вектор-столбец. Она является симметрической и неотрицательно определенной и, как отмечалось в § 8, найдется ортогональная матрица $\mathscr{O}$, приводящая ее к диагональному виду

$$\mathscr{O}^* \mathbb{R} \mathscr{O} = D,$$

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 $\det(\mathbb{R} - \lambda E) = 0$, где E — единичная матрица.

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

$$\beta = B^{-1} \mathscr{O}^* \eta. \tag{9}$$

Тогда матрица ковариаций вектора β

$$\beta\beta^* = B^{-1}\mathscr{O}^* \quad \eta\eta^*\mathscr{O}^* B^{-1} = B^{-1}\mathscr{O}^*\mathbb{R}\mathscr{O}B^{-1} = E,$$

и, следовательно, вектор $\beta = (\beta_1, \ldots, \beta_n)$ состоит из некоррелированных случайных величин. Ясно также, что

$$\eta = (\mathscr{O}B)\beta. \tag{10}$$

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 оказывается 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 оказывается 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 являются 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 $\varepsilon_1 = \eta_1/\|\eta_1\|$. Если $\varepsilon_1, \ldots, \varepsilon_{n-1}$ 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 $\|\varepsilon_n\| = 1$, $n \geqslant 1$, 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, для квадратичной формы

$$Q(a) = \sum_{i,j=1}^{n} r_{ij} a_i a_j, \quad a = (a_1, \ldots, a_n),$$

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

$$Q(a) = \left(\sum_{k=1}^{n} a_k \eta_k \right)^2.$$

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 $\bar{\mathscr{L}} = \bar{\mathscr{L}}(\eta_1, \eta_2, \ldots)$ 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) $\bar{\mathscr{L}}(\eta_1, \eta_2, \ldots) = L^2$.

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

$$\xi = \sum_{i=1}^{\infty} (\xi, \eta_i) \eta_i, \tag{13}$$

точнее,

$$\xi = \underset{n}{\mathrm{l.i.m.}} \sum_{i=1}^{n} (\xi, \eta_i) \eta_i.$$

Отсюда и из (3) тогда заключаем, что имеет место следующее *равенство Парсеваля*:

$$\|\xi\|^2 = \sum_{i=1}^{\infty} |(\xi, \eta_i)|^2, \quad \xi \in L^2. \tag{14}$$

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 — гауссовская мера,

$$(-\infty, a] = \int_{-\infty}^{a} \varphi(x)\, dx, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Обозначим $D = \dfrac{d}{dx}$ и введем функции

$$H_n(x) = \frac{(-1)^n D^n \varphi(x)}{\varphi(x)}, \quad n \geqslant 0. \tag{15}$$

Нетрудно найти, что

$$\begin{aligned} D\varphi(x) &= -x\varphi(x), \\ D^2\varphi(x) &= (x^2-1)\varphi(x), \\ D^3\varphi(x) &= (3x-x^3)\varphi(x), \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{16}$$

Отсюда следует, что $H_n(x)$ являются полиномами (называемыми *полиномами Эрмита*). Из (15), (16) находим, что

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2-1, \\ H_3(x) &= x^3-3x, \\ &\ldots\ldots\ldots\ldots \end{aligned}$$

Простой подсчет показывает, что

$$(H_m, H_n) = \int_{-\infty}^{\infty} H_m(x)\,H_n(x)\quad(dx) = \int_{-\infty}^{\infty} H_m(x)\,H_n(x)\,\varphi(x)\,dx = n!\,\delta_{nm},$$

где δ_{nm} — 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 (см., например, [33, гл. VII, § 3]), что если

$$\lim_{c\downarrow 0} \int_{-\infty}^{\infty} e^{c|x|}\quad(dx) < \infty, \tag{17}$$

то система функций $\{1, x, x^2, \ldots\}$ является полной в L^2, т. е. любая функция $\xi=\xi(x)$ из L^2 может быть представлена или в виде $\sum\limits_{i=1}^{n} a_i\eta_i(x)$, где $\eta_i(x)=x^i$, или в виде пределов (в среднеквадратическом смысле) этих сумм. Если применить процесс ортогонализации Грама—Шмидта к последовательности функций $\eta_1(x)$, $\eta_2(x)$, $\ldots$ с $\eta_i(x)=x^i$, то полученная ортонормированная система будет в точности совпадать с системой нормированный 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

$$\xi(x) = \underset{n}{\mathrm{l.i.m.}} \sum_{i=0}^{n} (\xi,\ h_i) h_i(x). \tag{18}$$

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*

$$\Pi_n(x) = \frac{(-1)^n \Delta^n P_x}{P_x}, \quad n \geqslant 1,\ \ \Pi_0 = 1. \tag{19}$$

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* $\{\pi_n(x)\}_{n\geqslant 0}$, $\pi_n(x) = \Pi_n(x)/\sqrt{c_n}$, есть ортонормированная система, 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 — 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

$$\frac{1}{2} = \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^2} + \ldots = \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots$$

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,

$$\begin{aligned}\{x\colon \xi_1=a_1,\ \ldots,\ \xi_n=a_n\}&= \\ &= \left\{x\colon \frac{a_1}{2}+\frac{a_2}{2^2}+\ldots+\frac{a_n}{2^n}\leqslant x<\frac{a_1}{2}+\frac{a_2}{2^2}+\ldots+\frac{a_n}{2^n}+\frac{1}{2^n}\right\}= \\ &= \left\{x\colon x\in\left[\frac{a_1}{2}+\ldots+\frac{a_n}{2^n},\ \frac{a_1}{2}+\ldots+\frac{a_n}{2^n}+\frac{1}{2^n}\right)\right\}=\frac{1}{2^n}.\end{aligned}$$

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 $R_n(x)=1-2\xi_n(x)$, $n\geqslant 1$, то нетрудно проверить, что система $\{R_n\}$ (*функций Радемахера*, рис. 31) является *ортонормированной*:

$$R_nR_m=\int_0^1 R_n(x)R_m(x)\,dx=\delta_{nm}.$$

Заметим, что $(1,\ R_n)\equiv\ \ R_n=0$. Отсюда следует, что эта система *не является* полной.

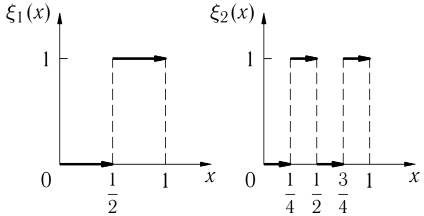


Рис. 30. Бернуллиевские величины

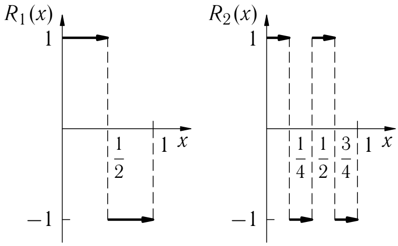


Рис. 31. Функции Радемахера

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*.

Снова пусть $\Omega=[0,\ 1)$, $\mathscr{F}=\mathscr{B}([0,\ 1))$. Положим

$$\begin{aligned}&H_1(x)=1,\\ &H_2(x)=R_1(x),\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &H_n(x)=\begin{cases}2^{j/2}R_{j+1}(x), & \text{если } \dfrac{k-1}{2^j}\leqslant x<\dfrac{k}{2^j},\ \ n=2^j+k,\\ & \qquad 1\leqslant k\leqslant 2^j,\ \ j\geqslant 1,\\ 0 & \text{в остальных случаях.}\end{cases}\end{aligned}$$

Нетрудно проверить, что $H_n(x)$, $n \geqslant 3$, можно записать и в таком виде:

$$H_{2^m+1}(x) = \begin{cases} 2^{m/2}, & 0 \leqslant x < 2^{-(m+1)}, \\ -2^{m/2}, & 2^{-(m+1)} \leqslant x < 2^{-m}, \\ 0 & \text{в других случаях,} \end{cases}$$

$$H_{2^m+j}(x) = H_{2^m+1}\left(x - \frac{j-1}{2^m}\right), \qquad j = 1, \ldots, 2^m, \qquad m = 1, 2, \ldots$$

На рис. 32 приведены графики первых восьми функций, дающих представление о структуре образования и поведении функций Хаара.

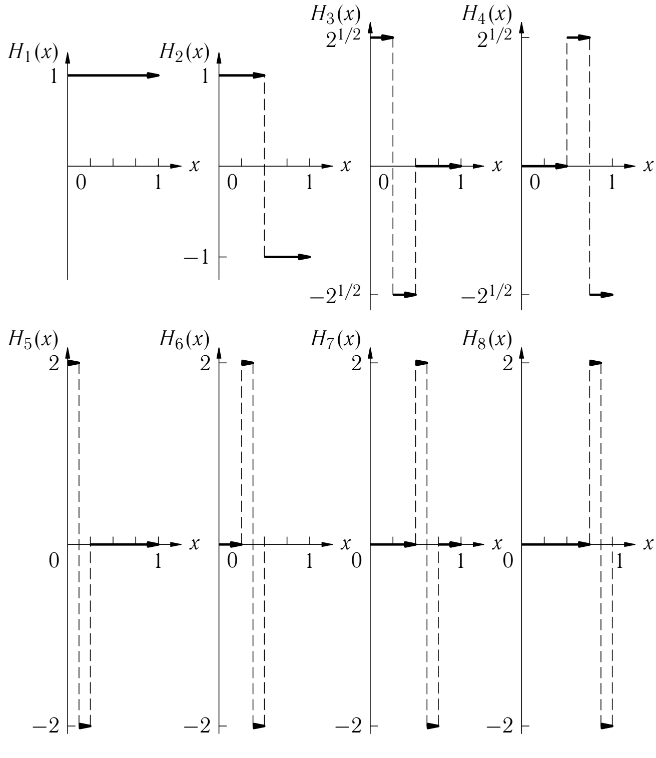


Рис. 32. Функции Хаара $H_1(x), \ldots, H_8(x)$

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

$$\int_0^1 \left| f(x) - \sum_{k=1}^{n} (f,\, H_k)\, H_k(x) \right|^p dx \to 0, \quad n \to \infty,$$

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 в теории функций.

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

$$\|\xi - \hat{\xi}\| = \inf\{\|\xi - \zeta\| \colon \zeta \in \mathscr{L}(\eta_1, \ldots, \eta_n)\}.$$

При этом $\hat{\xi} = \sum\limits_{i=1}^{n} (\xi,\, \eta_i)\eta_i$. 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*

$$\|\xi - \hat{\xi}\| = \inf\{\|\xi - \zeta\| \colon \zeta \in \overline{\mathscr{L}}\}. \tag{20}$$

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 $d = \inf\{\|\xi - \zeta\| \colon \zeta \in \overline{\mathscr{L}}\}$ 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

$$\|\zeta_n - \zeta_m\|^2 = 2\|\zeta_n - \xi\|^2 + 2\|\zeta_n - \xi\|^2 - 4\left\| \frac{1}{2}(\zeta_n + \zeta_m) - \xi \right\|^2.$$

Ясно, что $\frac{1}{2}(\zeta_n+\zeta_m)\in\bar{\mathscr{L}}$, поэтому $\left\|\frac{1}{2}(\zeta_n+\zeta_m)-\xi\right\|^2\geqslant d^2$ 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 $\bar{\mathscr{L}}$ замкнуто, поэтому $\hat{\xi}\in\bar{\mathscr{L}}$. Далее, $\|\zeta_n-\xi\|\to d$, следовательно, $\|\xi-\hat{\xi}\|=d$, что и доказывает существование требуемого элемента.

Покажем, что $\hat{\xi}$ — единственный элемент в $\bar{\mathscr{L}}$ с требуемым свойством. Пусть $\tilde{\xi}\in\bar{\mathscr{L}}$ 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

$$\|\hat{\xi}+\tilde{\xi}-2\xi\|^2=4\left\|\frac{1}{2}(\hat{\xi}+\tilde{\xi})-\xi\right\|^2\geqslant 4d^2.$$

Следовательно, $\|\hat{\xi}-\tilde{\xi}\|^2=0$, что и доказывает единственность «ближайшего» к ξ элемента из $\bar{\mathscr{L}}$.

Докажем теперь, что $\xi-\hat{\xi}\perp\zeta$, $\zeta\in\bar{\mathscr{L}}$. В силу (20) для любого $c\in R$

$$\|\xi-\hat{\xi}-c\zeta\|\geqslant\|\xi-\hat{\xi}\|.$$

Но

$$\|\xi-\hat{\xi}-c\zeta\|^2=\|\xi-\hat{\xi}\|^2+c^2\|\zeta\|^2-2(\xi-\hat{\xi},\ c\eta).$$

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 справедливо неравенство $\lambda^2\|\zeta\|^2-2\lambda<0$. Поэтому $(\xi-\hat{\xi},\ \zeta)=0$, $\zeta\in\bar{\mathscr{L}}$.

Осталось доказать представление (21).

Множество $\bar{\mathscr{L}}=\bar{\mathscr{L}}(\eta_1,\ \eta_2,\ \ldots)$ является замкнутым подпространством в L^2 и, значит, само является гильбертовым пространством (с тем же самым скалярным произведением). Для этого гильбертова пространства $\bar{\mathscr{L}}$ система η_1, η_2, … является базисом и, следовательно,

$$\hat{\xi}=\underset{n}{\mathrm{l.i.m.}}\sum_{k=1}^{n}(\hat{\xi},\ \eta_k)\eta_k. \tag{23}$$

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). $\square$

Замечание. Как и в конечномерном случае, $\hat{\xi}$ будем называть *проекцией* ξ на $\bar{\mathscr{L}} = \bar{\mathscr{L}}(\eta_1, \eta_2, \ldots)$, $\xi - \hat{\xi}$ — *перпендикуляром*, а представление

$$\xi = \hat{\xi} + (\xi - \hat{\xi})$$

— *ортогональным разложением.*

Величину $\hat{\xi}$ обозначают (ср. с $\widehat{\ }(\xi \mid \eta_1, \ldots, \eta_n)$ из п. 3) $\widehat{\ }(\xi \mid \eta_1, \eta_2, \ldots)$ 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

$$\Delta \equiv \ \ |\xi - \hat{\xi}|^2 \equiv \|\xi - \hat{\xi}\|^2 = \|\xi\|^2 - \sum_{i=1}^{\infty} |(\xi, \eta_i)|^2,$$

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 $\xi_1, \xi_2, \ldots$ — последовательность ортогональных случайных величин, $S_n = \xi_1 + \ldots + \xi_n$. Показать, что если $\sum\limits_{n=1}^{\infty} \ \ \xi_n^2 < \infty$, то найдется такая случайная величина S с $\ \ S^2 < \infty$, что $\text{l.i.m.}\ S_n = S$, т. е. $\|S_n - S\|^2 = \ \ |S_n - S|^2 \to 0$, $n \to \infty$.

6. Показать, что функции Радемахера R_n могут быть определены следующим образом:

$$R_n(x) = \operatorname{sign}(\sin\ 2^n \pi x), \quad 0 \leqslant x \leqslant 1, \ \ n = 1, 2, \ldots$$

7. Доказать, что

$$\|\xi\| \geqslant \| \ \ (\xi \mid \mathscr{G})\| \quad \text{для } \xi \in L^2(\mathscr{F}),$$

причем равенство имеет место тогда и только тогда, когда $\xi = \ \ (\xi \mid \mathscr{G})$ п. н.

8. Доказать, что если $\xi, \eta \in L^2(\mathscr{F})$, $\ \ (\xi \mid \eta) = \eta$, $\ \ (\eta \mid \xi) = \xi$, то $\xi = \eta$ п. н.

9. Даны три последовательности $(\mathscr{G}_n^{(1)})$, $(\mathscr{G}_n^{(2)})$ и $(\mathscr{G}_n^{(3)})$ под-σ-алгебр $\mathscr{F}$, ξ — ограниченная случайная величина. Известно, что для каждого n

$$\mathscr{G}_n^{(1)} \subseteq \mathscr{G}_n^{(2)} \subseteq \mathscr{G}_n^{(3)}, \qquad (\xi \,|\, \mathscr{G}_n^{(1)}) \to \eta, \qquad (\xi \,|\, \mathscr{G}_n^{(3)}) \to \eta.$$

Доказать, что $(\xi \,|\, \mathscr{G}_n^{(2)}) \to \eta$.

§ 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 ζ комплекснозначной случайной величины $\zeta = \xi + i\eta$ считается определенным, если определены математические ожидания ξ и η. В этом случае по определению полагаем $\zeta = \xi + i \eta$. Из определения 6 (§ 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 $|\zeta|^2 < \infty$, где $|\zeta|^2 = \xi^2 + \eta^2$, и скалярным произведением $(\zeta_1, \zeta_2) = \zeta_1 \overline{\zeta}_2$, где $\overline{\zeta}_2$ — 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

$$(\mathbb{R}\, a, a) = a^* \mathbb{R}\, a = \sum_{i,j=1}^{n} a_i r_{ij} a_j. \tag{1}$$

2. Определение 1. Пусть $F = F(x)$ — n-мерная функция распределения в $(R^n, \mathscr{B}(R^n))$, $x = (x_1, \ldots, x_n)$. Ее *характеристической функцией* называется функция

$$\varphi(t) = \int_{R^n} e^{i(t,x)}\, dF(x), \qquad t \in R^n. \tag{2}$$

Определение 2. Если $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор, определенный на вероятностном пространстве $(\Omega, \mathscr{F}, \;)$ и принимающей значения в R^n, то его *характеристической функцией* называется функция

$$\varphi_\xi(t) = \int_{R^n} e^{i(t,x)} dF_\xi(x), \qquad t \in R^n, \tag{3}$$

где $F_\xi = F_\xi(x)$ — функция распределения вектора $\xi = (\xi_1, \ldots, \xi_n)$, $x = (x_1, \ldots, x_n)$.

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 7 § 6 (о замене переменных под знаком интеграла Лебега) вытекает, что характеристическую функцию $\varphi_\xi(t)$ случайного вектора можно определить также равенством

$$\varphi_\xi(t) = \quad e^{i(t,\xi)}, \qquad t \in R^n. \tag{4}$$

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

$$\varphi_\xi(t) = \quad e^{it\xi}$$

— характеристическая функция.

Сразу отметим, что если $\eta = a\xi + b$, то

$$\varphi_\eta(t) = \quad e^{it\eta} = \quad e^{it(a\xi+b)} = e^{itb} \quad e^{ia\xi}.$$

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,

$$\varphi_{S_n}(t) = \quad e^{it(\xi_1+\ldots+\xi_n)} = \quad e^{it\xi_1}\ldots e^{it\xi_n} = \quad e^{it\xi_1}\ldots \quad e^{it\xi_n} = \prod_{j=1}^{n} \varphi_{\xi_j}(t),$$

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 $\{\xi = 1\} = p$, $\{\xi = 0\} = q$, $p + q = 1$, $1 > p > 0$, тогда

$$\varphi_\xi(t) = pe^{it} + q.$$

Если $\xi_1, \ldots, \xi_n$ — независимые одинаково распределенные (как ξ) случайные величины, то для $T_n = \dfrac{S_n - np}{\sqrt{npq}}$ находим, что

$$\varphi_{T_n}(t) = \varphi_{S_n - np/\sqrt{npq}}(t) = \quad e^{itS_n - np/\sqrt{npq}} = e^{-it\sqrt{np/q}}\left[pe^{it/\sqrt{npq}} + q\right]^n =$$
$$= \left[pe^{it\sqrt{q/np}} + qe^{-it\sqrt{p/nq}}\right]^n. \tag{7}$$

Заметим, что при $n \to \infty$ отсюда следует, что

$$\varphi_{T_n}(t) \to e^{-t^2/2}. \tag{8}$$

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 $\eta = \dfrac{\xi - m}{\sigma}$. 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

$$\varphi_\eta(t) = e^{-t^2/2}. \tag{10}$$

Следующая цепочка соотношений доказывает эту формулу (10):

$$\varphi_\eta(t) = \quad e^{it\eta} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2}\, dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{(itx)^n}{n!} e^{-x^2/2}\, dx = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n e^{-x^2/2}\, dx =$$

$$= \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!}(2n-1)!! = \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!} \frac{(2n)!}{2^n n!} = \sum_{n=0}^{\infty} \left(-\frac{t^2}{2}\right)^n \frac{1}{n!} = e^{-t^2/2},$$

где мы воспользовались тем (см. задачу 7 в § 8), что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2n} e^{-x^2/2}\, dx \equiv \quad \eta^{2n} = (2n-1)!!$$

Пример 3. Пусть ξ — пуассоновская случайная величина,

$$\{\xi = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

Тогда

$$e^{it\xi} = \sum_{k=0}^{\infty} e^{itk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = \exp\{\lambda(e^{it} - 1)\}. \tag{11}$$

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*

$$\varphi(t) = \quad e^{it\xi}$$

— 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* $\Bigl(\int\limits_{B} dF(x) = \int\limits_{-B} dF(x)$, $B \in \mathscr{B}(R)$, $-B = \{-x : x \in B\}\Bigr)$;

5) *если для некоторого* $n \geqslant 1$ $\quad |\xi|^n < \infty$, *то при всех* $r \leqslant n$ *существуют производные* $\varphi^{(r)}(t)$ *и*

$$\varphi^{(r)}(t) = \int\limits_{R} (ix)^r e^{itx}\, dF(x), \tag{12}$$

$$\xi^r = \frac{\varphi^{(r)}(0)}{i^r}, \tag{13}$$

$$\varphi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} \quad \xi^r + \frac{(it)^n}{n!} \varepsilon_n(t), \tag{14}$$

где $|\varepsilon_n(t)| \leqslant 3 \quad |\xi|^n$ *и* $\varepsilon_n(t) \to 0$, $t \to 0$;

6) *если существует и является конечной* $\varphi^{(2n)}(0)$, *то* $\quad \xi^{2n} < \infty$;

7) *если* $\quad |\xi|^n < \infty$ *для всех* $n \geqslant 1$ *и*

$$\overline{\lim_{n}} \frac{(\quad |\xi|^n)^{1/n}}{n} = \frac{1}{T} < \infty,$$

то при всех $|t| < T$

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \quad \xi^n. \tag{15}$$

Доказательство. Свойства 1) и 3) очевидны. Свойство 2) следует из оценки

$$|\varphi(t+h) - \varphi(t)| = | \quad e^{it\xi}(e^{ih\xi} - 1)| \leqslant \quad |e^{ih\xi} - 1|$$

и теоремы о мажорируемой сходимости, согласно которой $\quad |e^{ih\xi} - 1| \to 0$ при $h \to 0$.

Свойство 4). Пусть F симметрична. Тогда, если $g(x)$ — ограниченная борелевская нечетная функция, то $\int\limits_{R} g(x)\, dF(x) = 0$ (заметим, что для простых нечетных функций это следует сразу из определения симметричности F). Поэтому $\int\limits_R \sin\, tx\, dF(x) = 0$ и, значит,

$$\varphi(t) = \quad \cos\, t\xi.$$

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 1 § 3),

$$\{\xi \in B\} = \quad \{-\xi \in B\} = \quad \{\xi \in -B\}$$

для любого $B \in \mathscr{B}(R)$.

Свойство 5). Если $\quad |\xi|^n < \infty$, то в силу неравенств Ляпунова (28) § 6 $\; |\xi|^r < \infty$, $r \leqslant n$.

Рассмотрим отношение

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \quad e^{it\xi}\left(\frac{e^{ih\xi} - 1}{h}\right).$$

Поскольку

$$\left|\frac{e^{ihx} - 1}{h}\right| \leqslant |x|$$

и $\quad |\xi| < \infty$, то по теореме о мажорируемой сходимости существует

$$\lim_{h \to \infty} \quad e^{it\xi}\left(\frac{e^{ih\xi} - 1}{h}\right),$$

равный

$$e^{it\xi} \lim_{h \to 0}\left(\frac{e^{ih\xi} - 1}{h}\right) = i \quad (\xi e^{it\xi}) = i \int\limits_{-\infty}^{\infty} x e^{itx}\, dF(x). \tag{16}$$

Поэтому существует производная $\varphi'(t)$ и

$$\varphi'(t) = i \quad (\xi e^{it\xi}) = i \int\limits_{-\infty}^{\infty} x e^{itx}\, dF(x).$$

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

$$e^{it\xi}=\sum_{k=0}^{n-1}\frac{(it)^k}{k!}\quad \xi^k+\frac{(it)^n}{n!}[\quad \xi^n+\varepsilon_n(t)], \tag{18}$$

где

$$\varepsilon_n(t)=\quad [\xi^n(\cos\, \theta_1(\omega)t\xi+i\, \sin\, \theta_2(\omega)t\xi-1)].$$

Ясно, что $|\varepsilon_n(t)|\leqslant 3\quad |\xi^n|$, 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 $\quad \xi^2<\infty$. По правилу Лопиталя и лемме Фату

$$\begin{aligned}\varphi''(0)&=\lim_{h\to 0}\frac{1}{2}\left[\frac{\varphi'(2h)-\varphi'(0)}{2h}+\frac{\varphi'(0)-\varphi'(-2h)}{2h}\right]=\\&=\lim_{h\to 0}\frac{2\varphi'(2h)-2\varphi'(-2h)}{8h}=\lim_{h\to 0}\frac{1}{4h^2}[\varphi(2h)-2\varphi(0)+\varphi(-2h)]=\\&=\lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{e^{ihx}-e^{-ihx}}{2h}\right)^2 dF(x)=-\lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{\sin\, hx}{hx}\right)^2 x^2\, dF(x)\leqslant\\&\leqslant-\int_{-\infty}^{\infty}\lim_{h\to 0}\left(\frac{\sin\, hx}{hx}\right)^2 x^2\, dF(x)=-\int_{-\infty}^{\infty}x^2\, dF(x).\end{aligned}$$

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

$\int\limits_{-\infty}^{\infty} x^{2k} dF(x) > 0$. Тогда, согласно свойству 5),

$$\varphi^{(2k)}(t) = \int\limits_{-\infty}^{\infty} (ix)^{2k} e^{itx}\, dF(x)$$

и, значит,

$$(-1)^k \varphi^{(2k)}(t) = \int\limits_{-\infty}^{\infty} e^{itx}\, dG(x),$$

где $G(x) = \int\limits_{-\infty}^{x} u^{2k}\, dF(u)$.

Следовательно, функция $(-1)^k \varphi^{(2k)}(t) G^{-1}(\infty)$ является характеристической функцией вероятностного распределения $G(x) \cdot G^{-1}(\infty)$ и по доказанному

$$G^{-1}(\infty) \int\limits_{-\infty}^{\infty} x^2\, dG(x) < \infty.$$

Но $G^{-1}(\infty) > 0$, значит,

$$\int\limits_{-\infty}^{\infty} x^{2k+2}\, dF(x) = \int\limits_{-\infty}^{\infty} x^2\, dG(x) < \infty.$$

Свойство 7). Пусть $0 < t_0 < T$. Тогда, используя формулу Стирлинга (формула (6) в § 2 гл. I), находим, что

$$\varlimsup \frac{(\mathsf{E}|\xi|^n)^{1/n}}{n} < \frac{1}{t_0} \;\Rightarrow\; \varlimsup \frac{(\mathsf{E}|\xi|^n t_0^n)^{1/n}}{n} < 1 \;\Rightarrow\; \varlimsup \left(\frac{\mathsf{E}|\xi|^n t_0^n}{n!}\right)^{1/n} < 1.$$

Следовательно, по признаку Коши ряд $\sum \mathsf{E}|\xi|^n t_0^n / n!$ сходится, а значит, сходится и ряд $\sum\limits_{r=0}^{\infty} \frac{(it)^r}{r!} \mathsf{E}\xi^r$ для любого $|t| \leqslant t_0$. Но в силу (14)

$$\varphi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} \mathsf{E}\xi^r + R_n(t), \qquad n \geqslant 1,$$

где $|R_n(t)| \leqslant 3 \frac{|t|^n}{n!} \mathsf{E}|\xi|^n$. Поэтому для всех $|t| < T$

$$\varphi(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \mathsf{E}\xi^r. \qquad \square$$

Замечание 1. Аналогично доказательству (14) устанавливается, что если для некоторого $n \geqslant 1$ $\quad |\xi|^n < \infty$, 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 $|\varepsilon_n(t-s)| \leqslant 3 \quad |\xi^n|$ 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 о «единственности проблемы моментов».

4. Следующая теорема показывает, что характеристическая функция однозначно определяет функцию распределения.

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 $F(x) = G(x)$.

Доказательство. Зафиксируем $a,\ b \in R$, $\varepsilon > 0$ и рассмотрим функцию $f^\varepsilon = f^\varepsilon(x)$, изображенную на рис. 33. Покажем, что

$$\int_{-\infty}^{\infty} f^\varepsilon(x)\, dF(x) = \int_{-\infty}^{\infty} f^\varepsilon(x)\, dG(x). \tag{21}$$

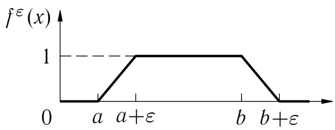


Рис. 33.

Пусть $n \geqslant 0$ таково, что $[a,\ b+\varepsilon] \subseteq [-n,\ n]$, 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

$$\begin{aligned}\left|\int_{-\infty}^{\infty} f^{\varepsilon}(x)\,dF(x)-\int_{-\infty}^{\infty} f^{\varepsilon}(x)\,dG(x)\right| = \left|\int_{-n}^{n} f^{\varepsilon}\,dF-\int_{-n}^{n} f^{\varepsilon}\,dG\right| \leqslant \\ \leqslant \left|\int_{-n}^{n} f_n^{\varepsilon}\,dF-\int_{-n}^{n} f_n^{\varepsilon}\,dG\right| + 2\delta_n \leqslant \\ \leqslant \left|\int_{-\infty}^{\infty} f_n^{\varepsilon}\,dF-\int_{-\infty}^{\infty} f_n^{\varepsilon}\,dG\right| + 2\delta_n + 2F(\overline{[-n,\,n]}) + 2G(\overline{[-n,\,n]}),\end{aligned} \tag{24}$$

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 $\varepsilon\to 0$ функции $f^{\varepsilon}(x)\to I_{(a,b]}(x)$. 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 $x\in R$. □

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,*

$$F(b)-F(a)=\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\,\varphi(t)\,dt. \tag{25}$$

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 правую 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

$$\Psi_c(x) = \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\,dt$$

и воспользовались теоремой Фубини, справедливость которой в данном случае следует из того, что

$$\left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| = \left| \frac{e^{-ita} - e^{-itb}}{it} \right| = \left| \int_a^b e^{-itx}\,dx \right| \leqslant b - a$$

и

$$\int_{-c}^{c}\int_{-\infty}^{\infty}(b-a)\,dF(x)\,dt \leqslant 2c(b-a)<\infty.$$

Далее,

$$\Psi_c(x)=\frac{1}{2\pi}\int_{-c}^{c}\frac{\sin t(x-a)-\sin t(x-b)}{t}\,dt=$$
$$=\frac{1}{2\pi}\int_{-c(x-a)}^{c(x-a)}\frac{\sin v}{v}\,dv-\frac{1}{2\pi}\int_{-c(x-b)}^{c(x-b)}\frac{\sin u}{u}\,du. \quad (30)$$

Функция

$$g(s,\,t)=\int_{s}^{t}\frac{\sin v}{v}\,dv$$

равномерно непрерывна по s и t и

$$g(s,\,t)\to\pi \quad (31)$$

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

$$\Psi_c(x)\to\Psi(x),\qquad c\to\infty,$$

где

$$\Psi(x)=\begin{cases}0, & x<a \text{ или } x>b,\\ 1/2, & x=a \text{ или } x=b,\\ 1, & a<x<b.\end{cases}$$

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, значит, она интегрируема на интервале $[a, b]$. Поэтому, снова применяя теорему Фубини, находим, что

$$\int\limits_a^b f(x)\,dx = \int\limits_a^b \frac{1}{2\pi}\left(\int\limits_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt\right)dx =$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \varphi(t)\left[\int\limits_a^b e^{-itx}\,dx\right]dt = \lim_{c\to\infty} \frac{1}{2\pi}\int\limits_{-c}^{c} \varphi(t)\left[\int\limits_a^b e^{-itx}\,dx\right]dt =$$

$$= \lim_{c\to\infty}\int\limits_{-c}^{c} \frac{1}{2\pi}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,dt = F(b) - F(a)$$

для всех точек a и b, являющихся точками непрерывности функции $F(x)$.

Отсюда вытекает, что

$$F(x) = \int\limits_{-\infty}^{x} f(y)\,dy, \qquad x \in R,$$

а так как $f(x)$ — непрерывная, а $F(x)$ — неубывающая функции, то $f(x)$ есть плотность $F(x)$. $\square$

Замечание. 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*:

$$e^{i(t_1\xi_1+\ldots+t_n\xi_n)} = \prod_{k=1}^{n} e^{it_k\xi_k}, \qquad (t_1, \ldots, t_n) \in R^n.$$

Доказательство. Необходимость следует из утверждения 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 $(t_1, \ldots, t_n) \in R^n$

$$\int_{R^n} e^{i(t_1x_1+\ldots+t_nx_n)} dG(x_1, \ldots, x_n) = \prod_{k=1}^{n} \int_{R} e^{it_kx_k} dF_k(x_k) =$$
$$= \prod_{k=1}^{n} \quad e^{it_k\xi_k} = \quad e^{i(t_1\xi_1+\ldots+t_n\xi_n)} = \int_{R^n} e^{i(t_1x_1+\ldots+t_nx_n)} dF(x_1, \ldots, x_n).$$

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 доказательств*) 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). (См. [69, т. 2, XIX.2].)

Теорема (Пойа). *Пусть непрерывная, четная и выпуклая книзу на $(-\infty, 0)$ (а значит, и на $(0, \infty)$) 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* ([69, т. 2, XV.2]).

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.

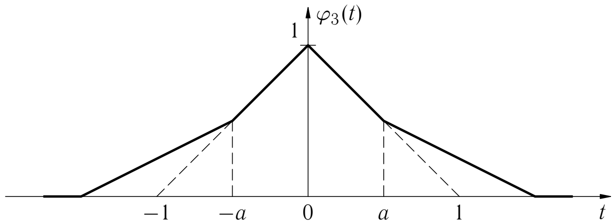


Рис. 34.

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* [135, 7.3].

Из этой теоремы вытекает, например, что функция $\exp(-t^4)$ 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.

а) *Если* $|\varphi_\xi(t_0)| = 1$ *для некоторого* $t_0 \neq 0$, *то случайная величина* ξ *является решетчатой с шагом* $h = \dfrac{2\pi}{|t_0|}$, *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*:

$$\{\xi = a\} = 1,$$

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

$$e^{it_0 a} = \int_{-\infty}^{\infty} e^{it_0 x}\, dF(x) \;\Rightarrow\; 1 = \int_{-\infty}^{\infty} e^{it_0(x-a)}\, dF(x) \;\Rightarrow$$

$$\Rightarrow\; 1 = \int_{-\infty}^{\infty} \cos\, t_0(x-a)\, dF(x) \;\Rightarrow\; \int_{-\infty}^{\infty} [1 - \cos\, t_0(x-a)]\, dF(x) = 0.$$

Поскольку $1 - \cos\, t_0(x-a) \geqslant 0$, то из свойства **H** (п. 2 § 6) следует, что (-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

$$\sum_{n=-\infty}^{\infty} \left\{\xi = a + \frac{2\pi}{t} n\right\} = \sum_{m=-\infty}^{\infty} \left\{\xi = b + \frac{2\pi}{\alpha t} m\right\} = 1.$$

Если ξ не является вырожденной, то тогда в 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 две совпадающие точки:

$$a + \frac{2\pi}{t} n_1 = b + \frac{2\pi}{\alpha t} m_1, \qquad a + \frac{2\pi}{t} n_2 = b + \frac{2\pi}{\alpha t} m_2,$$

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). □

8. Пусть $\xi = (\xi_1, \ldots, \xi_k)$ — случайный вектор,

$$\varphi_\xi(t) = \ e^{i(t,\xi)}, \qquad t = (t_1, \ldots, t_k),$$

— его характеристическая функция. Будем предполагать, что для некоторого $n \geqslant 1$ $\ |\xi_i|^n < \infty$, $i = 1, \ldots, k$. Из неравенства Гёльдера (29) § 6 и неравенства Ляпунова (27) § 6 следует, что существуют (смешанные) моменты $\ (\xi_1^{\nu_1} \ldots \xi_k^{\nu_k})$ 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

$$m_\xi^{(\nu_1,\ldots,\nu_k)}=\quad\xi_1^{\nu_1}\ldots\xi_k^{\nu_k}$$

— (*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 $z=re^{i\theta}$, $-\pi<\theta\leqslant\pi$, 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 «семиинвариант».)

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

$$\nu!=\nu_1!\ldots\nu_k!,\qquad|\nu|=\nu_1+\ldots+\nu_k,\qquad t^\nu=t_1^{\nu_1}\ldots t_k^{\nu_k}.$$

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* $\quad |\xi_i|^n < \infty$, $i = 1, \ldots, k$, $n \geqslant 1$. *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

$$\ln(1+z) = \sum_{q=1}^{n} \frac{(-1)^{q-1}}{q} z^q + o(z^n).$$

Применяя это разложение к (43) и сравнивая затем коэффициенты при t^{λ} с соответствующими коэффициентами в правой части (38), получим формулу (41). □

Следствие 1. *Справедливы следующие формулы, связывающие моменты и семиинварианты*:

$$m_{\xi}^{(\nu)} = \sum_{\{r_1\lambda^{(1)}+\ldots+r_x\lambda^{(x)}=\nu\}} \frac{1}{r_1!\ldots r_x!}\frac{\nu!}{(\lambda^{(1)}!)^{r_1}\ldots(\lambda^{(x)}!)^{r_x}} \prod_{j=1}^{x} \left[s_{\xi}^{(\lambda^{(j)})}\right]^{r_j}, \tag{44}$$

$$s_{\xi}^{(\nu)} = \sum_{\{r_1\lambda^{(1)}+\ldots+r_x\lambda^{(x)}=\nu\}} \frac{(-1)^{q-1}(q-1)!}{r_1!\ldots r_x!}\frac{\nu!}{(\lambda^{(1)}!)^{r_1}\ldots(\lambda^{(x)}!)^{r_x}} \prod_{j=1}^{x} \left[m_{\xi}^{(\lambda^{(j)})}\right]^{r_j}, \tag{45}$$

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 $\dfrac{q!}{r_1!\ldots r_x!}$ 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 частный случай, когда $\nu=(1, \ldots, 1)$. В этом случае моменты $m_{\xi}^{(\nu)} \equiv \xi_1\ldots\xi_k$ 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 $I_{\xi}=\{1, 2, \ldots, k\}$ — множество индексов компонент вектора $\xi = (\xi_1, \ldots, \xi_k)$. 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 взаимно однозначном соответствии с множествами

$I \subseteq I_\xi$. Поэтому обозначим

$$m_\xi(I) = m_\xi^{(\chi(I))}, \qquad s_\xi(I) = s_\xi^{(\chi(I))}.$$

Иначе говоря, $m_\xi(I)$ и $s_\xi(I)$ являются простыми моментами и семиинвариантами подвектора ξ_I вектора ξ.

Далее, в соответствии с определением, данным на с. 31, назовем разбиением множества I неупорядоченный набор непересекающихся непустых множеств I_p такой, что $\sum\limits_p I_p = I$.

С учетом этих обозначений имеют место формулы

$$m_\xi(I) = \sum_{\sum\limits_{p=1}^{q} I_p = I} \prod_{p=1}^{q} s_\xi(I_p), \tag{46}$$

$$s_\xi(I) = \sum_{\sum\limits_{p=1}^{q} I_p = I} (-1)^{q-1}(q-1)! \prod_{p=1}^{q} m_\xi(I_p), \tag{47}$$

где $\sum\limits_{\sum\limits_{p=1}^{q} I_p = I}$ *означает суммирование по всевозможным разбиениям множества* I, $1 \leqslant q \leqslant N(I)$, $N(I)$ — *число элементов множества* I.

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 $\{\chi(I_1), \ldots, \chi(I_q)\}$ взаимно однозначно соответствует разбиение $I = \sum\limits_{p=1}^{q} I_p$. Следовательно, формула (46) вытекает из (44).

Аналогичным образом из (45) выводится справедливость представления (47).

Пример 4. Пусть ξ — случайная величина ($k = 1$) и $m_n = m_\xi^{(n)} = \quad \xi^n$, $s_n = s_\xi^{(n)}$. Тогда из (40) и (41) получаем следующие формулы:

$$\begin{aligned} m_1 &= s_1, \\ m_2 &= s_2 + s_1^2, \\ m_3 &= s_3 + 3s_1 s_2 + s_1^3, \\ m_4 &= s_4 + 3s_2^2 + 4s_1 s_3 + 6s_1^2 s_2 + s_1^4, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{48}$$

и

$$\begin{aligned} s_1 &= m_1 = \quad \xi, \\ s_2 &= m_2 - m_1^2 = \quad \xi, \\ s_3 &= m_3 - 3m_1 m_2 + 2m_1, \\ s_4 &= m_4 - 3m_2^2 - 4m_1 m_3 + 12m_1^2 m_2 - 6m_1^4, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{49}$$

Пример 5. Пусть $\xi \sim \mathscr{N}(m, \sigma^2)$. 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 $n \geqslant 3$, обращаются в нуль.

Пример 6. 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 7. Пусть $\xi = (\xi_1, \ldots, \xi_k)$ — случайный вектор. Тогда

$$\begin{aligned} m_\xi(1) &= s_\xi(1), \\ m_\xi(1, 2) &= s_\xi(1, 2) + s_\xi(1)s_\xi(2), \\ m_\xi(1, 2, 3) &= s_\xi(1, 2, 3) + s_\xi(1, 2)s_\xi(3) + s_\xi(1, 3)s_\xi(2) + \\ &\quad + s_\xi(2, 3)s_\xi(1) + s_\xi(1)s_\xi(2)s_\xi(3), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{51}$$

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 (51) получатся, конечно, формулы (48).

Из (51) становится понятным «групповое» происхождение коэффициентов в формулах (48). Из (51) следует также, что

$$s_\xi(1, 2) = m_\xi(1, 2) - m_\xi(1)m_\xi(2) = \quad \xi_1\xi_2 - \quad \xi_1 \quad \xi_2, \tag{52}$$

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 $F(x)$ и характеристической функцией $\varphi(t)$. Предположим, что существуют все моменты $m_n = \quad \xi^n$, $n \geqslant 1$.

Из теоремы 2 следует, что характеристическая функция *однозначно* определяет распределение вероятностей. Поставим сейчас следующий вопрос (единственность проблемы моментов): *однозначно* ли определяют моменты $\{m_n\}_{n\geqslant 1}$ *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

$$g(x) = ke^{-\alpha x^\lambda}[1 + \varepsilon \sin(\beta x^\lambda)], \quad |\varepsilon| < 1, \ \ x > 0.$$

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 $p=(n+1)/\lambda$, $q=\alpha+i\beta$, $t=x^{\lambda}$. Тогда

$$\int_0^\infty x^{\lambda\left(\frac{n+1}{\lambda}-1\right)} e^{-(\alpha+i\beta)x^{\lambda}} \lambda x^{\lambda-1}\,dx = \lambda \int_0^\infty x^n e^{-(\alpha+i\beta)x^{\lambda}}\,dx =$$

$$= \lambda \int_0^\infty x^n e^{-\alpha x^{\lambda}} \cos \beta x^{\alpha} dx - i\lambda \int_0^\infty x^n e^{-\alpha x^{\lambda}} \sin \beta x^{\lambda}\,dx =$$

$$= \frac{\Gamma\left(\dfrac{n+1}{\lambda}\right)}{\alpha^{\frac{n+1}{\lambda}}(1+i\,\operatorname{tg}\lambda\pi)^{\frac{n+1}{\lambda}}}. \quad (55)$$

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 для $n\geqslant 1$*

$$\mu_n = \int_{-\infty}^{\infty} |x|^n\,dF(x).$$

Если

$$\varlimsup_{n\to\infty} \frac{\mu_n^{1/n}}{n} < \infty, \quad (56)$$

то моменты $\{m_n\}_{n\geqslant 1}$, где $m_n = \int_{-\infty}^{\infty} x^n\,dF(x)$, 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), выводится, что для всех $|t-s|\leqslant t_0$

$$\varphi(t)=\sum_{k=0}^{\infty}\frac{(t-s)^k}{k!}\varphi^{(k)}(s),$$

где

$$\varphi^{(k)}(s)=i^k\int\limits_{-\infty}^{\infty}x^k e^{isx}\,dF(x)$$

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 $\{m_n\}_{n\geqslant 1}$ определяют однозначно $\varphi(t)$ при всех t, а значит, и функцию распределения $F(x)$. □

Следствие 1. *Моменты однозначно определяют распределение вероятностей, сосредоточенное на конечном интервале.*

Следствие 2. *Для единственности проблемы моментов достаточно, чтобы*

$$\varlimsup_{n\to\infty}\frac{(m_{2n})^{1/2n}}{2n}<\infty. \tag{57}$$

Для доказательства нужно лишь заметить, что нечетные моменты оцениваются по четным, и затем воспользоваться условием (56).

Пример. Пусть $F(x)$ — функция нормального распределения,

$$F(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\int\limits_{-\infty}^{x}e^{-\frac{t^2}{2\sigma^2}}\,dt.$$

Тогда $m_{2n+1}=0$, $m_{2n}=\dfrac{(2n)!}{2^n n!}\sigma^{2n}$ и из (57) следует, что эти моменты являются моментами только нормального распределения.

Приведем в заключение (без доказательства)

Критерий Карлемана (единственности проблемы моментов); [69, т. 2, VII.3].

а) *Пусть* $\{m_n\}_{n\geqslant 1}$ — *моменты некоторого распределения вероятностей, причем*

$$\sum_{n=0}^{\infty}\frac{1}{(m_{2n})^{1/2n}}=\infty.$$

Тогда они определяют распределение вероятностей однозначно.

b) *Если* $\{m_n\}_{n\geqslant 1}$ — *моменты распределения, сосредоточенного на* $[0, \infty)$, *то для однозначности достаточно потребовать, чтобы*

$$\sum_{n=0}^{\infty} \frac{1}{(m_n)^{1/2n}} = \infty.$$

10. Пусть $F = F(x)$ и $G = G(x)$ — функции распределения с характеристическими функциями $f = f(t)$ и $g = g(t)$ соответственно. Следующая теорема, которую мы приводим без доказательства, показывает, как можно оценить близость F и G (в равномерной метрике) в терминах близости f и g. (По поводу применений этой теоремы см. § 11 гл. III.)

Теорема (неравенство Эссеена)**.** *Пусть* $G(x)$ *имеет производную* $G'(x)$ *с* $\sup\limits_x |G'(x)| \leqslant C$. *Тогда для каждого* $T > 0$

$$\sup_x |F(x) - G(x)| \leqslant \frac{2}{\pi} \int_0^T \left| \frac{f(t) - g(t)}{t} \right| dt + \frac{24}{\pi T} \sup_x |G'(x)|.$$

11. На следующей странице приведены две таблицы характеристических функций $\varphi(t)$ некоторых часто встречающихся распределений вероятностей; см. таблицы распределений (вместе с их параметрами) на с. 226 и 227.

12. Задачи.

1. Пусть ξ и η — независимые случайные величины, $f(x) = f_1(x) + i f_2(x)$, $g(x) = g_1(x) + i g_2(x)$, где $f_k(x)$, $g_k(x)$ — борелевские функции, $k = 1, 2$. Показать, что если $\quad |f(\xi)| < \infty$, $\quad |g(\xi)| < \infty$, то

$$|f(\xi)\, g(\eta)| < \infty$$

и

$$f(\xi)\, g(\eta) = \quad f(\xi) \cdot \quad g(\eta).$$

2. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ и $\quad \|\xi\|^n < \infty$, где $\|\xi\| = \sqrt{\sum \xi_i^2}$. Показать, что

$$\varphi_\xi(t) = \sum_{k=0}^{n} \frac{i^k}{k!} \quad (t, \xi)^k + \varepsilon_n(t) \|t\|^n,$$

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 (12) § 3, установить справедливость *формулы обращения*

$$(a, b] = \lim_{c \to \infty} \frac{1}{(2\pi)^n} \int_{-c}^{c} \ldots \int_{-c}^{c} \prod_{k=1}^{n} \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} \varphi(t_1, \ldots, t_n)\, dt_1 \ldots dt_n.$$

Таблица 4

Дискретные распределения	Характеристические функции
Дискретное равномерное	$\frac{1}{N}\frac{e^{it}}{1-e^{it}}(1-e^{itN})$
Бернуллиевское	$q+pe^{it}$
Биномиальное	$[q+pe^{it}]^n$
Пуассоновское	$\exp\{\lambda(e^{it}-1)\}$
Геометрическое	$\frac{p}{1-qe^{it}}$
Отрицательно-биномиальное	$\left[\frac{p}{1-qe^{it}}\right]^r$

Таблица 5

Распределения, имеющие плотность	Характеристические функции
Равномерное на $[a, b]$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
Нормальное, или гауссовское	$\exp\left\{itm-\frac{\sigma^2t^2}{2}\right\}$
Гамма	$(1-it\beta)^{-\alpha}$
Бета	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{k=0}^{\infty}\frac{(it)^k\Gamma(\alpha+k)}{k!\,\Gamma(\alpha+\beta+k)\,\Gamma(1+k)}$
Экспоненциальное	$\frac{\lambda}{\lambda-it}$
Двустороннее экспоненциальное	$\frac{\lambda^2e^{it\alpha}}{t^2+\lambda^2}$
Хи-квадрат	$(1-2it)^{-n/2}$
Стьюдента, или t-распределение	$\frac{\sqrt{\pi}\Gamma((n+1)/2)}{\Gamma(n/2)}\,\frac{\exp\{-\sqrt{n}\}\lvert t\rvert}{2^{2(m-1)}(m-1)!}\sum_{k=0}^{m-1}(2k)!\,C_{n-1+k}^{2k}(2\sqrt{n}\lvert t\rvert)^{m-1-k}$, если $m=\frac{n+1}{2}$ — целое число
Коши	$e^{-\theta\lvert t\rvert}$

(В приведенной выше *формуле обращения* $(a, b]$ является интервалом непрерывности функции $(a, b]$, 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. Доказать справедливость формул, приведенных для характеристических функций в табл. 4 и 5.

9. Пусть ξ — целочисленная случайная величина и $\varphi_\xi(t)$ — ее характеристическая функция. Показать, что

$$\{\xi = k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi_\xi(t)\, dt, \qquad k = 0, \pm 1, \pm 2, \ldots$$

10. Показать, что в пространстве $L^2 = L^2\big([-\pi, \pi], \mathscr{B}([-\pi, \pi])\big)$ с мерой Лебега μ система функций $\left\{\dfrac{1}{\sqrt{2\pi}} e^{i\lambda n},\ n = 0, \pm 1, \ldots\right\}$ образует ортонормированный базис.

11. В теореме Бохнера—Хинчина предполагается, что рассматриваемая функция $\varphi(t)$ является *непрерывной*. Доказать следующий результат (Рисс), показывающий, в какой степени можно отказаться от предположения непрерывности.

Пусть $\varphi = \varphi(t)$ — комплекснозначная измеримая по Лебегу функция с $\varphi(0) = 1$. Тогда функция $\varphi = \varphi(t)$ является положительно определенной в том и только том случае, когда она совпадает (почти всюду относительно лебеговой меры на числовой прямой) с некоторой характеристической функцией.

12. Какие из функций

$$\varphi(t) = e^{-|t|^k}, \quad 0 \leqslant k \leqslant 2, \qquad \varphi(t) = e^{-|t|^k}, \quad k > 2,$$
$$\varphi(t) = (1 + |t|)^{-1}, \qquad \varphi(t) = (1 + t^4)^{-1},$$
$$\varphi(t) = \begin{cases} 1 - |t|^3, & |t| \leqslant 1, \\ 0, & |t| > 1, \end{cases} \qquad \varphi(t) = \begin{cases} 1 - |t|, & |t| \leqslant 1/2, \\ 1/(4|t|), & |t| > 1/2, \end{cases}$$

являются характеристическими?

13. Пусть $\varphi(t)$ — характеристическая функция распределения $F = F(x)$. Пусть $\{x_n\}$ — множество точек разрыва функции F $(\Delta F(x_n) \equiv F(x_n) - F(x_n-) > 0)$. Показать, что

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} |\varphi(t)|^2\, dt = \sum_{n \geqslant 1} (\Delta F(x_n))^2.$$

14. *Функцией концентрации* случайной величины X называется функция

$$Q(X;\, l) = \sup_{x\in R} \quad \{x \leqslant X \leqslant x+l\}.$$

Показать, что:

(a) если X и Y — независимые случайные величины, то

$$Q(X+Y;\, l) \leqslant \min(Q(x;\, l),\ Q(Y;\, l)) \quad \text{для всех } l \geqslant 0;$$

(b) найдется такое x_l^*, что $Q(X;\, l) = \ \{x_l^* \leqslant X \leqslant x_l^* + l\}$, и функция распределения величины X непрерывна тогда и только тогда, когда $Q(X;0)=0$.

15. Пусть $(m_n)_{n\geqslant 1}$ — последовательность моментов случайной величины X с функцией распределения $F=F(x)$ $\left(m_k = \int\limits_{-\infty}^{\infty} x^k\, dF(x)\right)$. Показать, что если ряд $\sum\limits_{k=1}^{\infty} \frac{m_k}{k!} s^k$ сходится абсолютно для некоторого $s>0$, то $(m_n)_{n\geqslant 1}$ однозначно определяет функцию распределения $F=F(x)$.

16. Пусть $\varphi(t) = \int\limits_{-\infty}^{\infty} e^{itx}\, dF(x)$ — характеристическая функция распределения $F=F(x)$. Показать, что:

$$\lim_{c\to\infty} \frac{1}{2c} \int\limits_{-c}^{c} e^{-itx} \varphi(t)\, dt = F(x) - F(x-),$$

$$\lim_{c\to\infty} \frac{1}{2c} \int\limits_{-c}^{c} |\varphi(t)|^2\, dt = \sum_{x\in R} [F(x) - F(x-)]^2.$$

17. Показать, что каждая характеристическая функция $\varphi(t)$ удовлетворяет неравенству $1 - \mathrm{Re}\ \varphi(2t) \leqslant 4[1 - \mathrm{Re}\ \varphi(t)]$.

18. Пусть характеристическая функция $\varphi(t)$ такова, что $\varphi(t) = 1 + f(t) + o(t^2)$, $t\to 0$, где $f(t) = -f(-t)$. Показать, что тогда $\varphi(t) \equiv 1$.

19. Показать, что для каждого $n \geqslant 1$ функции

$$\varphi_n(t) = \frac{e^{it} - \sum\limits_{k=0}^{n-1} (it)^k/k!}{(it)^n/n!}$$

являются характеристическими.

20. Доказать, что

$$\frac{2}{\pi} \int\limits_{-\infty}^{\infty} \frac{1 - \mathrm{Re}\ \varphi(t)}{t^2}\, dt = \int\limits_{-\infty}^{\infty} |x|\, dF(x).$$

21. Пусть характеристическая функция $\varphi(t) = 1 + O(|t|^{\alpha})$, $t \to 0$, где $\alpha \in (0, 2]$. Показать, что случайная величина ξ, имеющая своей характеристической функцией функцию $\varphi(t)$, обладает следующим свойством:

$$\{|\xi| > x\} = O(x^{-\alpha}), \qquad x \to 0.$$

22. Если $\varphi(t)$ — характеристическая функция, то функция $|\varphi(t)|^2$ — также характеристическая.

23. Пусть X и Y — независимые одинаково распределенные величины с нулевым средним и единичной дисперсией. Доказать, основываясь на рассмотрении характеристических функций, что если распределение величины $(X+Y)/\sqrt{2}$ совпадает с распределением F величин X и Y, то это F является нормальным распределением.

24. Если φ есть характеристическая функция, то таковой же является функция $e^{\lambda(\varphi-1)}$ для каждого $\lambda \geqslant 0$.

25. *Преобразованием Лапласа* неотрицательной случайной величины X с функцией распределения F называется функция $\widehat{F} = \widehat{F}(\lambda)$, определенная для $\lambda \geqslant 0$ формулой

$$\widehat{F}(\lambda) = \quad e^{-\lambda X} = \int_{[0,\infty)} e^{-\lambda x}\, dF(x).$$

Доказать следующий критерий (С. Н. Бернштейн): функция $f = f(\lambda)$ на $(0, \infty)$ есть преобразование Лапласа функции распределения $F = F(x)$ на $[0, \infty)$ в том и только том случае, когда эта функция является полностью (вполне) монотонной (т. е. существуют производные $f^{(n)}(\lambda)$ всех порядков $n \geqslant 0$ и $(-1)^n f^{(n)}(\lambda) \geqslant 0$).

26. Пусть $\varphi(t)$ — характеристическая функция. Показать, что следующие функции также являются характеристическими:

$$\int_0^1 \varphi(ut)\, du, \qquad \int_0^\infty e^{-u}\varphi(ut)\, du.$$

§ 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 является 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 $\sigma=+\sqrt{\sigma^2}$. (Величина σ носит название *стандартное отклонение* значений случайной величины ξ от ее среднего значения ξ; ср. с определением 5 в § 4 гл. I.)

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 $\{\xi=m\}=1$.

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 $\varphi_\xi(t)\equiv e^{it\xi}$, $t\in R$.

Если $\{\xi=m\}=1$, то очевидно, что

$$\varphi_\xi(t)=e^{itm}, \tag{2}$$

а если $\xi\sim\mathscr{N}(m,\sigma^2)$, $\sigma^2>0$, то, согласно (9) § 12,

$$\varphi_\xi(t)=e^{itm-\frac{t^2\sigma^2}{2}}. \tag{3}$$

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

$$\varphi_\xi(t) = \quad e^{i(t,\xi)}, \qquad t = (t_1, \ldots, t_n) \in R^n, \tag{4}$$

— 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

$$x - m = \mathcal{O}u, \qquad t = \mathcal{O}v,$$

где $\mathcal{O}$ — ортогональная матрица такая, что

$$\mathcal{O}^*\mathbb{R}\mathcal{O} = D,$$

и

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

— 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)

$$\begin{aligned} i(t, x-m) - \frac{1}{2}(A(x-m), x-m) = \\ = i(\mathscr{O}v, \mathscr{O}u) - \frac{1}{2}(A\mathscr{O}u, \mathscr{O}u) = i(\mathscr{O}v)^* \mathscr{O}u - \frac{1}{2}(\mathscr{O}u)^* A(\mathscr{O}u) = \\ = iv^* u - \frac{1}{2}u^* \mathscr{O}^* A \mathscr{O} u = iv^* u - \frac{1}{2} u^* D^{-1} u. \end{aligned}$$

Вместе с (9) § 12 и (8) это дает

$$\begin{aligned} I_n = \frac{1}{(2\pi)^{n/2}(d_1 \ldots d_n)^{1/2}} \int\limits_{R^n} e^{iv^* u - \frac{1}{2} u^* D^{-1} u}\, du = \\ = \prod_{k=1}^{n} \frac{1}{(2\pi d_k)^{1/2}} \int\limits_{-\infty}^{\infty} e^{iv_k u_k - \frac{u_k^2}{2d_k}}\, du_k = \prod_{k=1}^{n} e^{-\frac{v_k^2 d_k}{2}} = \\ = e^{-\frac{1}{2} v^* D v} = e^{-\frac{1}{2} v^* \mathscr{O}^* \mathbb{R} \mathscr{O} v} = e^{-\frac{1}{2} t^* \mathbb{R} t} = e^{-\frac{1}{2}(\mathbb{R}t, t)}. \end{aligned}$$

Из (6) следует также, что

$$\int\limits_{R^n} f(x)\, dx = 1. \tag{9}$$

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 $\mathbb{R}^\varepsilon \equiv \mathbb{R} + \varepsilon E$, где E — единичная матрица. Тогда по доказанному функция

$$\varphi^\varepsilon(t) = \exp\left\{ i(t, m) - \frac{1}{2}(\mathbb{R}^\varepsilon t, t) \right\}$$

является характеристической:

$$\varphi^\varepsilon(t) = \int\limits_{R^n} e^{i(t,x)} dF_\varepsilon(x),$$

где $F_\varepsilon(x) = F_\varepsilon(x_1, \ldots, x_n)$ — n-мерная функция распределения.

При $\varepsilon \to 0$

$$\varphi^\varepsilon(t) \to \varphi(t) = \exp\left\{ i(t, m) - \frac{1}{2}(\mathbb{R}\, t, t) \right\}.$$

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 (35) § 12 и формул связи моментов и семиинвариантов находим, что

$$m_1 = s_\xi^{(1,0,\ldots,0)} = \quad \xi_1, \quad \ldots, \quad m_n = s_\xi^{(0,\ldots,0,1)} = \quad \xi_n.$$

Аналогично

$$r_{11} = s_\xi^{(2,0,\ldots,0)} = \quad \xi_1, \qquad r_{12} = s_\xi^{(1,1,0,\ldots,0)} = \qquad (\xi_1, \xi_2),$$

и вообще

$$r_{kl} = \qquad (\xi_k, \xi_l).$$

Таким образом, m есть *вектор средних значений* ξ, а $\mathbb{R}$ — *матрица ковариаций*.

Если матрица $\mathbb{R}$ невырожденная, то к этому результату можно было прийти и иначе. Именно, в этом случае вектор ξ имеет плотность $f(x)$, задаваемую формулой (6). Тогда прямой подсчет показывает, что

$$\begin{gathered} \xi_k \equiv \int x_k f(x)\, dx = m_k, \\ (\xi_k, \xi_l) = \int (x_k - m_k)(x_l - m_l) f(x)\, dx = r_{kl}. \end{gathered} \tag{11}$$

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 $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\lambda_k \in R$, случайная величина $(\xi, \lambda) = \lambda_1 \xi_1 + \ldots + \lambda_n \xi_n$ 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 (см. (5))

$$\exp\{it(\xi_1\lambda_1+\ldots+\xi_n\lambda_n)\}=\exp\left\{it\left(\sum\lambda_k m_k\right)-\frac{t^2}{2}\left(\sum r_{kl}\lambda_k\lambda_l\right)\right\},\quad t\in R,$$

и, поэтому,

$$(\xi,\lambda)\sim\mathscr{N}\left(\sum\lambda_k m_k,\sum r_{kl}\lambda_k\lambda_l\right).$$

Обратно, гауссовость случайной величины $(\xi,\lambda)=\xi_1\lambda_1+\ldots+\xi_n\lambda_n$ означает, в частности, что

$$e^{i(\xi,\lambda)}=e^{i\quad(\xi,\lambda)-\frac{(\xi,\lambda)}{2}}=e^{i\sum\lambda_k\quad\xi_k-\frac{1}{2}\sum\lambda_k\lambda_l\qquad(\xi_k,\xi_l)}.$$

В силу произвольности $\lambda_1,\ldots,\lambda_n$ отсюда следует, что $\xi=(\xi_1,\ldots,\xi_n)$ — гауссовский вектор (см. определение 1). ☐

Замечание. Пусть (θ,ξ) — гауссовский вектор с $\theta=(\theta_1,\ldots,\theta_k)$, $\xi=(\xi_1,\ldots,\xi_l)$. Если векторы θ и ξ некоррелированы, т. е. $(\theta_i,\xi_j)=0$, $i=1,\ldots,k$, $j=1,\ldots,l$, то они и независимы.

Доказательство — то же, что и для утверждения a) 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

$$\mathscr{O}^*\mathbb{R}\mathscr{O}=D.$$

Как уже отмечалось в п. 3, 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

$$e^{i(t,\beta)}=\quad e^{i\beta^*t}=e^{-\frac{1}{2}(Et,t)},$$

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, $\xi_k = 0$, $k=1, \ldots, n$. 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

$$\hat{\xi}_k= \quad (\xi_k \,|\, \xi_{k-1}, \ldots, \xi_1). \tag{17}$$

Отсюда в силу (16) и (14) следует, что в гауссовском случае условное математическое ожидание $\quad(\xi_k \,|\, \xi_{k-1}, \ldots, \xi_1)$ является *линейной* функцией от $\xi_1, \ldots, \xi_{k-1}$:

$$(\xi_k \,|\, \xi_{k-1}, \ldots, \xi_1)=\sum_{i=1}^{k-1} a_i\xi_i. \tag{18}$$

(В случае $k=2$ этот результат был установлен в § 8.)

Поскольку, согласно замечанию к теореме 1 § 8, $\quad(\xi_k \,|\, \xi_{k-1}, \ldots, \xi_1)$ 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

$$m_\theta = \quad \theta, \;\; m_\xi = \quad \xi$$

— вектор-столбцы средних значений и

$$\begin{aligned} {}_{\theta\theta} &\equiv \quad (\theta,\, \theta) \equiv \| \quad (\theta_i,\, \theta_j)\|, \quad 1 \leqslant i,\; j \leqslant k, \\ {}_{\theta\xi} &\equiv \quad (\theta,\, \xi) \equiv \| \quad (\theta_i,\, \xi_j)\|, \quad 1 \leqslant i \leqslant k,\; 1 \leqslant j \leqslant l, \\ {}_{\xi\xi} &\equiv \quad (\xi,\, \xi) \equiv \| \quad (\xi_i,\, \xi_j)\|, \quad 1 \leqslant i,\; j \leqslant l, \end{aligned}$$

— матрицы ковариаций. Предположим, что матрица ${}_{\xi\xi}$ имеет обратную матрицу. Тогда (ср. с теоремой 2 в § 8) справедлива следующая

Теорема 2 (теорема о нормальной корреляции, векторный случай). *Для гауссовского вектора $(\theta,\, \xi)$ оптимальная оценка $(\theta\,|\,\xi)$ вектора θ по ξ и ее матрица ошибок*

$$\Delta = \quad [\theta - \quad (\theta\,|\,\xi)]\,[\theta - \quad (\theta\,|\,\xi)]^*$$

задаются следующими формулами:

$$(\theta\,|\,\xi) = m_\theta + \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1}(\xi - m_\xi), \tag{19}$$

$$\Delta = \quad {}_{\theta\theta} - \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1} \quad {}_{\theta\xi}^{*}. \tag{20}$$

Доказательство. Образуем вектор

$$\eta = (\theta - m_\theta) - \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1}(\xi - m_\xi). \tag{21}$$

Тогда непосредственно проверяется, что $\eta(\xi - m_\xi)^* = 0$, т. е. вектор η не коррелирован с вектором $\xi - m_\xi$. Но в силу гауссовости $(\theta,\, \xi)$ вектор $(\eta,\, \xi)$ также будет гауссовским. Отсюда в силу замечания к теореме 1 векторы η и $\xi - m_\xi$ независимы. Значит, независимы η и ξ и, следовательно, $(\eta\,|\,\xi) = \quad \eta = 0$. Поэтому

$$[\theta - m_\theta|\xi] - \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1}(\xi - m_\xi) = 0,$$

что и доказывает представление (19).

Для доказательства (20) рассмотрим условную ковариацию

$$(\theta,\, \theta\,|\,\xi) \equiv \quad [(\theta - \quad (\theta\,|\,\xi))(\theta - \quad (\theta\,|\,\xi))^*|\xi]. \tag{22}$$

Поскольку $\theta - \quad (\theta\,|\,\xi) = \eta$, то в силу независимости η и ξ находим, что

$$\begin{aligned} &(\theta,\, \theta\,|\,\xi) = \quad (\eta\eta^*\,|\,\xi) = \quad \eta\eta^* = \\ &= \quad {}_{\theta\theta} + \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1} \quad {}_{\xi\xi} \quad {}_{\xi\xi}^{-1} \quad {}_{\theta\xi}^{*} - 2 \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1} \quad {}_{\xi\xi} \quad {}_{\xi\xi}^{-1} \quad {}_{\theta\xi}^{*} = \quad {}_{\theta\theta} - \quad {}_{\theta\xi} \quad {}_{\xi\xi}^{-1} \quad {}_{\theta\xi}^{*}. \end{aligned}$$

Поскольку $(\theta,\, \theta\,|\,\xi)$ не зависит от «случая», то

$$\Delta = \qquad (\theta,\, \theta\,|\,\xi) = \qquad (\theta,\, \theta\,|\,\xi),$$

что и доказывает представление (20). □

Следствие. *Пусть* $(\theta, \xi_1, \ldots, \xi_n)$ *есть* $(n+1)$*-мерный гауссовский вектор, причем* $\xi_1, \ldots, \xi_n$ *независимы. Тогда*

$$(\theta\,|\,\xi_1, \ldots, \xi_n) = \quad \theta + \sum_{i=1}^{n} \frac{(\theta, \xi_i)}{\xi_i}(\xi_i - \quad \xi_i),$$

$$\Delta = \quad \theta - \sum_{i=1}^{n} \frac{^2(\theta, \xi_i)}{\xi_i}$$

(*ср. с формулами* (12), (13) § 8).

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 $m_n = \quad \xi_n$, $\sigma_n^2 = \quad \xi_n$. Тогда по теореме Лебега о мажорируемой сходимости

$$\lim_{n\to\infty} e^{itm_n - \frac{1}{2}\sigma_n^2 t^2} = \lim_{n\to\infty} \quad e^{it\xi_n} = \quad e^{it\xi}.$$

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,

$$e^{it\xi} = e^{itm - \frac{1}{2}\sigma^2 t^2},$$

т. е. $\xi \sim \mathscr{N}(m, \sigma^2)$.

Отсюда, в частности, вытекает, что замкнутое линейное многообразие $\bar{\mathscr{L}}(\xi_1, \xi_2, \ldots)$, 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.

а) Если $\xi = (\xi_\alpha)$, $\alpha \in \mathfrak{A}$, — гауссовская система, то всякая ее подсистема $\xi' = (\xi'_{\alpha'})$, $\alpha' \in \mathfrak{A}' \subset \mathfrak{A}$, 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 $m_\alpha = \ \xi_\alpha$. «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 $\mathbb{R}^{\mathfrak{A}}$, 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 $(\Omega, \mathscr{F}, \ )$ и на нем гауссовская система случайных величин $\xi = (\xi_\alpha)_{\alpha\in\mathfrak{A}}$ такие, что

$$\xi_\alpha = m_\alpha,$$
$$(\xi_\alpha, \xi_\beta) = r_{\alpha\beta}, \qquad \alpha, \beta \in \mathfrak{A}?$$

Если взять конечный набор $\alpha_1, \ldots, \alpha_n$, то по вектору $\overline{m} = (m_{\alpha_1}, \ldots \ldots, m_{\alpha_n})$ и матрице $\overline{\mathbb{R}} = (r_{\alpha\beta})$, $\alpha, \beta = \alpha_1, \ldots, \alpha_n$, в R^n можно построить гауссовское распределение $F_{\alpha_1,\ldots,\alpha_n}(x_1, \ldots, x_n)$ с характеристической функцией

$$\varphi(t) = e^{i(t,\overline{m}) - \frac{1}{2}(\overline{\mathbb{R}}t,t)}, \qquad t = (t_{\alpha_1}, \ldots, t_{\alpha_n}).$$

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 $\mathbb{R} = \|r_{ij}\|$, $r_{ij} = \quad (\xi_i, \xi_j)$. 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 $\mathfrak{A} = [0, 1]$, $[0, \infty)$, $(-\infty, \infty)$, $\ldots$, 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 вероятностные свойства таких процессов полностью определяется видом «матрицы» ковариаций $\mathbb{R} = (r_{st})$, $s, t \in \mathfrak{A}$. Будем обозначать r_{st} через $r(s, t)$ и называть эту функцию от s и t *ковариационной функцией*.

Пример 1. Если $\mathfrak{A} = [0, \infty)$ и

$$r(s, t) = \min(s, t), \tag{25}$$

то гауссовский процесс $B = (B_t)_{t \geqslant 0}$ с такой функцией ковариаций (см. задачу 2) и $B_0 \equiv 0$ называется *процессом броуновского движения* или *винеровским процессом*.

Отметим, что этот процесс имеет *независимые приращения*, т. е. для любых $t_1 < t_2 < \ldots < t_n$ случайные величины

$$B_{t_2} - B_{t_1}, \ \ldots, \ B_{t_n} - B_{t_{n-1}}$$

являются независимыми. В самом деле, в силу гауссовости достаточно проверить лишь попарную некоррелированность приращений. Но если $s < t < u < v$, то

$$[B_t - B_s]\,[B_v - B_u] = \\ = [r(t, v) - r(t, u)] - [r(s, v) - r(s, u)] = (t - t) - (s - s) = 0.$$

Замечание. Приведенный в п. 4 § 9 пример процесса восстановления (конструктивно заданного по последовательности независимых одинаково распределенных величин $\sigma_1, \sigma_2, \ldots$) наводит на мысль о возможности построить аналогичным образом некоторую версию броуновского движения.

Такие конструкции, основанные на привлечении последовательности $\xi_1, \xi_2, \ldots$ независимых одинаково распределенных стандартных гауссовских случайных величин $\xi_i \sim \mathcal{N}(0, 1)$, действительно существуют.

Образуем, например, величины

$$B_t = \frac{\sqrt{2}}{\pi} \sum_{n=1}^{\infty} \frac{\xi_n}{n + 1/2} \sin((n + 1/2)\pi t), \qquad t \in [0, 1]. \tag{26}$$

Из приводимой далее теоремы о «двух рядах» (теорема 2 § 3 гл. IV) вытекает, что ряд, определяющий B_t, *сходится* (-п. н.) при каждом $t \in [0, 1]$. Более детальное рассмотрение показывает, что этот ряд сходится (-п. н.) *равномерно*, и тем самым процесс $B = (B_t)_{0 \leqslant t \leqslant 1}$ имеет (-п. н.) непрерывные траектории. Нормальность конечномерных распределений этого процесса следует из теоремы 1 b) и утверждения в п. 5 о сохранении гауссовости распределения предела по вероятности гауссовских случайных величин. Нетрудно убедиться также в том, что ковариационная функция $r(s, t) = \ B_s B_t = \min(s, t)$.

Таким образом, построенный в (26) процесс $B = (B_t)_{0 \leqslant t \leqslant 1}$ удовлетворяет всем требованиям в определении процесса броуновского движения, но, более того, этот процесс имеет (-п. н.) *непрерывные* траектории. Как правило, свойство непрерывности траекторий (которое является желательным и оправданным физическими применениями) включают в само определение броуновского движения. Как видим, такой процесс действительно существует.

Укажем еще на один известный способ построения броуновского движения, основанный на введенных в п. 5 § 11 *функциях Хаара* $H_n(x)$, $x \in [0, 1]$, $n = 1, 2, \ldots$

Построим по ним *функции Шаудера* $S_n(t)$, $t \in [0, 1]$, $n = 1, 2, \ldots$:

$$S_n(t) = \int_0^t H_n(x)\, dx. \tag{27}$$

Тогда если $\xi_0, \xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных стандартных, $\xi_i \sim \mathscr{N}(0, 1)$, случайных величин, то ряд

$$B_t = \sum_{n=1}^{\infty} \xi_n S_n(t) \tag{28}$$

сходится равномерно по $t \in [0, 1]$ с вероятностью единица. Процесс $B = (B_t)_{0 \leqslant t \leqslant 1}$ является броуновским движением.

Пример 2. Процесс $B^0 = (B_t^0)$, $t \in \mathfrak{A}$ с $\mathfrak{A} = [0, 1]$, $B_0^0 \equiv 0$ и

$$r(s, t) = \min(s, t) - st, \tag{29}$$

называется *условным винеровским процессом* или *броуновским мостом* (заметим, что поскольку $r(1, 1) = 0$, то $\ (B_1^0 = 0) = 1$).

Пример 3. Процесс $X = (X_t)$, $t \in \mathfrak{A}$ с $\mathfrak{A} = (-\infty, \infty)$ и

$$r(s, t) = e^{-|t-s|} \tag{30}$$

называют *гауссовско-марковским*.

8. Приведем одно интересное свойство броуновского движения, доказательство которого будет служить хорошей иллюстрацией применения леммы Бореля—Кантелли из § 10 (точнее — следствия 1 к ней).

Теорема 3. *Пусть* $B=(B_t)_{t\geqslant 0}$ — *стандартное броуновское движение. Тогда с вероятностью единица для всякого* $T>0$

$$\lim_{n\to\infty}\sum_{k=1}^{2^n T}[B_{k2^{-n}}-B_{(k-1)2^{-n}}]^2=T. \tag{31}$$

Доказательство. Без ограничения общности можно считать $T=1$. Пусть

$$A_n^\varepsilon=\left\{\omega\colon \left|\sum_{k=1}^{2^n}(B_{k2^{-n}}-B_{(k-1)2^{-n}})^2-1\right|\geqslant\varepsilon\right\}.$$

Поскольку величины $B_{k2^{-n}}-B_{(k-1)2^{-n}}$ являются гауссовскими с нулевыми средними и дисперсией, равной 2^{-n}, то

$$\left(\sum_{k=1}^{2^n}(B_{k2^{-n}}-B_{(k-1)2^{-n}})^2\right)=2^{-n+1};$$

значит, по неравенству Чебышева $\quad(A_n^\varepsilon)\leqslant\varepsilon^{-2}2^{-n+1}$, и тем самым

$$\sum_{n=1}^{\infty}\quad(A_n^\varepsilon)\leqslant\varepsilon^{-2}\sum_{n=1}^{\infty}2^{-n+1}=2\varepsilon^{-2}<\infty. \tag{32}$$

Требуемое утверждение (31) следует из этой оценки и следствия 1 к лемме Бореля—Кантелли (§ 10). □

9. Задачи.

1. Пусть ξ_1, ξ_2, ξ_3 — независимые гауссовские случайные величины, $\xi_i\sim\mathscr{N}(0,\,1)$. Показать, что

$$\frac{\xi_1+\xi_2\xi_3}{\sqrt{1+\xi_3^2}}\sim\mathscr{N}(0,\,1).$$

(В этой связи возникает интересная для исследования задача описания всех *нелинейных* преобразований от независимых гауссовских величин $\xi_1,\,\ldots,\,\xi_n$, распределение которых также является гауссовским.)

2. Доказать, что «матрицы» $\mathbb{R}=(r(s,\,t))_{s,t\in\mathfrak{A}}$, задаваемые функциями $r(s,\,t)$ из (25), (29) и (30), являются неотрицательно-определенными.

3. Пусть A — некоторая матрица порядка $m\times n$. Назовем матрицу $A^{\otimes}$ порядка $n\times m$ *псевдообратной* к матрице A, если найдутся такие матрицы U и V, что

$$AA^{\otimes}A=A,\qquad A^{\otimes}=UA^*=A^*V.$$

Показать, что матрица $A^{\otimes}$, 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 $_{\xi\xi}$, если в этих формулах вместо $^{-1}_{\xi\xi}$ рассматривать псевдообратную матрицу $^{\oplus}_{\xi\xi}$.

5. Пусть $(\theta, \xi) = (\theta_1, \ldots, \theta_k; \xi_1, \ldots, \xi_l)$ — гауссовский вектор с невырожденной матрицей $\Delta \equiv {}_{\theta\theta} - {}^{\oplus}_{\xi\xi}\, {}^{*}_{\theta\xi}$. Показать, что у функции распределения $(\theta \leqslant a \,|\, \xi) = (\theta_1 \leqslant a_1, \ldots, \theta_k \leqslant a_k \,|\, \xi)$ существует (-п. н.) плотность $p(a_1, \ldots, a_k \,|\, \xi)$, определяемая формулой

$$p(a_1, \ldots, a_k \,|\, \xi) = \frac{|\Delta|^{-1/2}}{(2\pi)^{k/2}} \exp\Big\{-\frac{1}{2}(a - (\theta \,|\, \xi))^* \Delta^{-1} (a - (\theta \,|\, \xi))\Big\}.$$

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.

7. *Теорема Мерсера.* Пусть $r = r(s, t)$ — непрерывная ковариационная функция на $[a, b] \times [a, b]$, где $-\infty < a < b < \infty$. Доказать, что уравнение

$$\lambda \int_a^b r(s, t) u(t)\, dt = u(s), \qquad a \leqslant s \leqslant b,$$

допускает бесконечное число значений $\lambda_k > 0$ и соответствующую систему непрерывных решений $\{u_k, k \geqslant 1\}$, образующих полную ортонормированную систему в $L^2(a, b)$ такую, что

$$r(s, t) = \sum_{k=1}^{\infty} \frac{u_k(s) u_k(t)}{\lambda_k},$$

где ряд сходится абсолютно и равномерно на $[a, b] \times [a, b]$.

8. Пусть $X = \{X_t, t \geqslant 0\}$ — гауссовский процесс с $X_t = 0$ и ковариационной функцией $r(s, t) = e^{-|t-s|}$, $s, t \geqslant 0$. Пусть $0 < t_1 < \ldots < t_n$ и $f_{t_1,\ldots,t_n}(x_1, \ldots, x_n)$ — плотность величин $X_{t_1}, \ldots, X_{t_n}$. Доказать, что

$$f_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \left[(2\pi)^n \prod_{i=2}^{n} \left(1 - e^{2(t_{i-1} - t_i)}\right)\right]^{-1/2} \times \exp\left\{-\frac{x_1^2}{2} - \frac{1}{2} \sum_{i=2}^{n} \frac{(x_i - e^{(t_{i-1} - t_i)} x_{i-1})^2}{1 - e^{2(t_{i-1} - t_i)}}\right\}.$$

9. Пусть $f = \{f_n, n \geqslant 1\} \subset L^2(0, 1)$ — полная ортонормированная система и (ξ_n) — независимые одинаково распределенные $\mathscr{N}(0, 1)$-величины. Показать, что процесс $B_t = \sum_{n \geqslant 1} \xi_n \int_0^t f_n(u)\, du$ есть броуновское движение.

10. Доказать, что в случае гауссовских систем $(\xi, \eta_1, \ldots, \eta_n)$ условные математические ожидания $(\xi \,|\, \eta_1, \ldots, \eta_n)$ совпадаютс математическими ожиданиями в широком смысле $\widehat{\ }(\xi \,|\, \eta_1, \ldots, \eta_n)$.

11. Пусть $(\xi, \eta_1, \ldots, \eta_k)$ — гауссовская система. Выяснить структуру условных математических ожиданий $(\xi^n \,|\, \eta_1, \ldots, \eta_k)$, $n \geqslant 1$ (как функций от $\eta_1, \ldots, \eta_k$).

12. Пусть $X = (X_k)_{1 \leqslant k \leqslant n}$ и $Y = (Y_k)_{1 \leqslant k \leqslant n}$ — две гауссовские случайные последовательности с $X_k = Y_k$, $X_k = Y_k$, $1 \leqslant k \leqslant n$, и

$$(X_k, X_l) \leqslant \quad (Y_k, Y_l), \quad 1 \leqslant k, l \leqslant n.$$

Доказать справедливость *неравенства Слепяна*: для всякого $x \in R$

$$\Big\{ \sup_{1 \leqslant k \leqslant n} X_k < x \Big\} \leqslant \Big\{ \sup_{1 \leqslant k \leqslant n} Y_k < x \Big\}.$$

13. Доказать, что если $B^{\circ} = (B^{\circ}_t)_{0 \leqslant t \leqslant 1}$ есть броуновский мост, то процесс $B = (B_t)_{t \geqslant 0}$ с $B_t = (1+t) B^{\circ}_{t/(1+t)}$ является броуновским движением.

14. Проверить, что для броуновского движения $B = (B_t)_{t \geqslant 0}$ следующие процессы также являются броуновскими движениями:

$$B^{(1)}_t = -B_t;$$
$$B^{(2)}_t = t B_{1/t}, \ \ t > 0, \quad \text{и} \quad B^{(2)}_0 = 0;$$
$$B^{(3)}_t = B_{t+s} - B_s, \quad s > 0;$$
$$B^{(4)}_t = B_T - B_{T-t} \quad \text{для } 0 \leqslant t \leqslant T, \ T > 0;$$
$$B^{(5)}_t = \frac{1}{a} B_{a^2 t}, \quad a > 0 \ \text{(свойство автомодельности)}.$$

15. Пусть $X = (X_k)_{1 \leqslant k \leqslant n}$ — гауссовская последовательность с

$$m = \max_{1 \leqslant k \leqslant n} \quad X_k, \quad \sigma^2 = \max_{1 \leqslant k \leqslant n} \quad X_k,$$

и

$$\Big\{ \max_{1 \leqslant k \leqslant n} (X_k - \quad X_k) \geqslant a \Big\} \leqslant 1/2 \quad \text{для некоторого } a.$$

Тогда имеет место следующее *неравенство Бореля*:

$$\Big\{ \max_{1 \leqslant k \leqslant n} X_k > x \Big\} \leqslant 2\Psi\Big(\frac{x - m - a}{\sigma} \Big),$$

где $\Psi(x) = (2\pi)^{-1/2} \int\limits_x^{\infty} e^{-y^2/2}\, dy$.

16. Пусть (X, Y) — двумерная гауссовская случайная величина с $X = Y = 0$, $X^2 > 0$, $Y^2 > 0$ и коэффициентом корреляции $\rho = \dfrac{XY}{\sqrt{\quad X^2 \quad Y^2}}$.

Показать, что

$$\{XY < 0\} = 1 - 2 \quad \{X > 0,\ Y > 0\} = \pi^{-1} \arccos \rho.$$

17. Пусть $Z = XY$, где $X \sim \mathscr{N}(0,\ 1)$ и $\quad \{Y = 1\} = \quad \{Y = -1\} = \dfrac{1}{2}$. Найти распределение пар $(X,\ Z)$ и $(Y,\ Z)$ и распределение $X + Z$. Показать, что $Z \sim \mathscr{N}(0,\ 1)$ и что X и Z некоррелированы, но зависимы.

18. Провести подробное доказательство того, что процессы $(B_t)_{0 \leqslant t \leqslant 1}$, определяемые в (26), (28), являются броуновскими движениями.

19. Пусть $B^{\mu} = (B_t + \mu t)_{t \geqslant 0}$ — броуновское движение со сносом.

(a) Найти распределение величин $B^{\mu}_{t_1} + B^{\mu}_{t_2}$, $t_1 < t_2$.

(b) Найти $\quad B^{\mu}_{t_0} B^{\mu}_{t_1}$ и $\quad B^{\mu}_{t_0} B^{\mu}_{t_1} B^{\mu}_{t_2}$ для $t_0 < t_1 < t_2$.

20. Для процесса B^{μ} из предыдущей задачи найти условные распределения

$$(B^{\mu}_{t_2} \in \cdot \mid B^{\mu}_{t_1})$$

для $t_1 < t_2$ и $t_1 > t_2$ и

$$(B^{\mu}_{t_2} \in \cdot \mid B^{\mu}_{t_0},\ B^{\mu}_{t_1})$$

для $t_0 < t_1 < t_2$.

Глава III

БЛИЗОСТЬ И СХОДИМОСТЬ ВЕРОЯТНОСТНЫХ МЕР. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

При формальном построении курса теории вероятностей предельные теоремы появляются в виде своего рода надстройки над элементарными главами теории вероятностей, в которых все задачи имеют конечный, чисто арифметический характер. В действительности, однако, познавательная ценность теории вероятностей раскрывается только предельными теоремами. Более того, без предельных теорем не может быть понято реальное содержание самого исходного понятия всей нашей науки — понятия вероятности.

Б. В. Гнеденко, А. Н. Колмогоров.
«Предельные распределения для сумм независимых случайных величин» [16]

§ 1. Слабая сходимость вероятностных мер и распределений

1. Многие из фундаментальных результатов теории вероятностей формулируются в виде *предельных теорем*. В форме предельных теорем были сформулированы закон больших чисел Я. Бернулли и теорема Муавра—Лапласа, положившие, собственно говоря, начало истинной теории вероятностей и, в частности, указавшие путь многочисленным исследованиям по выяснению условий справедливости разных форм «закона больших чисел» и «центральной предельной теоремы». В форме предельной теоремы была сформулирована *теорема Пуассона* об аппроксимации биномиального распределения «пуассоновским» в случае редких событий. Уже на примере этих утверждений, и также результатов о *скорости сходимости* в теоремах Муавра—Лапласа и Пуассона видно, что в теории вероятностей приходится иметь дело с разными видами сходимости распределений, а выяснение скорости сходимости требует введения тех или иных «естественных» мер близости между распределениями.

В настоящей главе будут рассматриваться некоторые общие аспекты *сходимости* вероятностных распределений и их *близости*. В данном параграфе рассматриваются вопросы общей теории *слабой сходимости вероятностных мер в метрических пространствах*. (Именно к кругу интересов этой теории относятся, в частности, как закон больших чисел Я. Бернулли, так и теорема Муавра—Лапласа — прародительница «центральной предельной теоремы».) Из § 3 станет ясно, что метод характеристических функций является одним из самых мощных средств доказательства предельных теорем о слабой сходимости вероятностных распределений в R^n. В § 7 будут рассмотрены вопросы «метризуемости» слабой сходимости. Затем в § 9 мы останавливаемся на другом виде сходимости распределений (более сильном, нежели слабая сходимость) — *сходимости по вариации*. Доказательству простейших результатов о скорости сходимости в центральной предельной теореме и теореме Пуассона отведены §§ 11, 12. В § 13 результаты о слабой сходимости из §§ 1, 2 применяются к некоторым (принципиально важным) задачам математической статистики.

2. 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 $\quad(\xi_i = 1) = p, \quad (\xi_i = 0) = q,\ p + q = 1$. Используя введенное в § 10 гл. II понятие сходимости по вероятности, закон больших чисел Я. Бернулли можно сформулировать в следующем виде:

$$\frac{S_n}{n} \to p, \qquad n \to \infty, \tag{1}$$

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

$$\begin{aligned} F_n(x) &= \quad \left\{ \frac{S_n}{n} \leqslant x \right\}, \\ F(x) &= \begin{cases} 1, & x \geqslant p, \\ 0, & x < p, \end{cases} \end{aligned} \tag{2}$$

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 гл. II сходимость по вероятности $\frac{S_n}{n} \to p$ влечет за собой сходимость по *распределению* $\frac{S_n}{n} \xrightarrow{d} p$, означающую, что

$$f\left(\frac{S_n}{n}\right) \to \quad f(p), \qquad n \to \infty, \tag{3}$$

для любой функции $f = f(x)$ из класса C *непрерывных ограниченных* функций на R.

Поскольку

$$f\left(\frac{S_n}{n}\right) = \int_R f(x)\, P_n(dx), \qquad f(p) = \int_R f(x)\, P(dx),$$

то (3) можно переписать в форме

$$\int_R f(x)\, P_n(dx) \to \int_R f(x)\, P(dx), \qquad f \in C, \tag{4}$$

или (в соответствии с обозначениями § 6 гл. II) — в форме

$$\int_R f(x)\, dF_n(x) \to \int_R f(x)\, dF(x), \qquad f \in C. \tag{5}$$

В анализе сходимость (4) называют *слабой сходимостью* (мер P_n к мере P, $n \to \infty$) и записывают в виде $P_n \xrightarrow{w} P$ (ср. с определением 2). Естественно и сходимость (5) также назвать слабой сходимостью функций распределения F_n к F и обозначить ее $F_n \xrightarrow{w} F$.

Итак, можно утверждать, что в схеме Бернулли

$$\frac{S_n}{n} \to p \;\Rightarrow\; F_n \xrightarrow{w} F. \tag{6}$$

Из (1) нетрудно также вывести, что для функций распределения, введенных в (2),

$$F_n(x) \to F(x), \qquad n \to \infty,$$

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 $F(x)$ при $n \to \infty$ *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 $n \to \infty$

$$F_n(x) \to F(x), \qquad x \in \mathbb{C}(F),$$

где $\mathbb{C}(F)$ — множество точек *непрерывности* предельной функции F.

В рассматриваемом случае схемы Бернулли функция $F = F(x)$ вырождена, и отсюда нетрудно вывести (см. задачу 7 к § 10 в гл. II), что

$$(F_n \Rightarrow F) \;\Rightarrow\; \left(\frac{S_n}{n} \to p\right).$$

Таким образом, с учетом приводимой ниже теоремы 2

$$\left(\frac{S_n}{n} \to p\right) \;\Rightarrow\; (F_n \xrightarrow{w} F) \;\Leftrightarrow\; (F_n \Rightarrow F) \;\Rightarrow\; \left(\frac{S_n}{n} \to p\right) \tag{7}$$

и, следовательно, утверждение закона больших чисел можно рассматривать как одно из утверждений *о слабой сходимости функций распределения*, определенных в (2).

Обозначим

$$\begin{aligned} F_n(x) &= \ \left\{\frac{S_n - np}{\sqrt{npq}} \leqslant x\right\}, \\ F(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du. \end{aligned} \tag{8}$$

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 распределения*, 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).

3. Пусть $(E, \mathscr{E}, \rho)$ — метрическое пространство с метрикой $\rho = \rho(x, y)$, σ-алгеброй $\mathscr{E}$ борелевских подмножеств, порожденных открытыми множествами, и пусть , $_1$, $_2$, … — вероятностные меры на $(E, \mathscr{E}, \rho)$.

Определение 2. Последовательность вероятностных мер $\{\ _n\}$ называется *слабо сходящейся* к вероятностной мере (обозначение: $_n \xrightarrow{w}$; w — от *weak* convergence — слабая сходимость), если

$$\int_E f(x) \quad _n(dx) \to \int_E f(x) \quad (dx) \tag{9}$$

для любой функции $f = f(x)$ из класса $C(E)$ непрерывных ограниченных функций на E.

Определение 3. Последовательность вероятностных мер $\{\ _n\}$ называется *сходящейся в основном* к вероятностной мере (обозначение: $_n \Rightarrow$), если

$$_n(A) \to \quad (A) \tag{10}$$

для любого множества A из $\mathscr{E}$ такого, что

$$(\partial A) = 0. \tag{11}$$

(Через ∂A обозначается граница множества A:

$$\partial A = [A] \cap [\bar{A}],$$

где $[A]$ — замыкание множества A.)

Приводимая далее 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) $\ _n \xrightarrow{w}\ $;

(II) $\overline{\lim}\ \ _n(A) \leqslant\ (A)$, A — *замкнутые множества*;

(III) $\underline{\lim}\ \ _n(A) \geqslant\ (A)$, A — *открытые множества*;

(IV) $\ _n \Rightarrow\ $.

Доказательство. (I) $\Rightarrow$ (II). Пусть A — замкнутое множество и

$$f_A^{\varepsilon}(x) = \left[1 - \frac{\rho(x, A)}{\varepsilon}\right]^+,$$

где

$$\rho(x, A) = \inf\{\rho(x, y)\colon\ y \in A\}, \qquad [x]^+ = \max[0, x].$$

Обозначим также

$$A^{\varepsilon} = \{x\colon\ \rho(x, A) < \varepsilon\}$$

и заметим, что $A^{\varepsilon} \downarrow A$, $\varepsilon \downarrow 0$.

Поскольку функции $f_A^{\varepsilon}(x)$ ограничены, непрерывны и

$$\ _n(A) = \int_E I_A(x)\ \ _n(dx) \leqslant \int_E f_A^{\varepsilon}(x)\ \ _n(dx),$$

то

$$\overline{\lim_n}\ \ _n(A) \leqslant \overline{\lim_n} \int_E f_A^{\varepsilon}(x)\ \ _n(dx) = \int_E f_A^{\varepsilon}(x)\ \ (dx) \leqslant\ (A^{\varepsilon}) \downarrow\ (A), \qquad \varepsilon \downarrow 0,$$

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 $\ (\partial A) = 0$

$$\overline{\lim_n}\ \ _n(A) \leqslant \overline{\lim_n}\ \ _n([A]) \leqslant\ ([A]) =\ (A),$$

$$\underline{\lim_n}\ \ _n(A) \geqslant \underline{\lim_n}\ \ _n(A^0) \geqslant\ (A^0) =\ (A),$$

и, значит, $\ _n(A) \to\ (A)$ для всякого A с $\ (\partial A) = 0$.

(IV) $\Rightarrow$ (I). Пусть $f = f(x)$ — непрерывная ограниченная функция с $|f(x)| < M$. Обозначим

$$D = \{t \in R\colon\ \ \{x\colon\ f(x) = t\} \neq 0\}$$

и рассмотрим разбиение $T_k = (t_0, t_1, \ldots, t_k)$ интервала $[-M, M]$:

$$-M = t_0 < t_1 < \ldots < t_k = M, \qquad k \geqslant 1,$$

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 конечна.)

Пусть $B_i = \{x\colon t_i \leqslant f(x) < t_{i+1}\}$. 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 $\ (\partial B_i) = 0$ и в силу (IV)

$$\sum_{i=0}^{k-1} t_i \ \ _n(B_i) \to \sum_{i=0}^{k-1} t_i \ \ (B_i). \tag{12}$$

Но

$$\left| \int_E f(x) \ \ _n(dx) - \int_E f(x) \ \ (dx) \right| \leqslant \left| \int_E f(x) \ \ _n(dx) - \sum_{i=0}^{k-1} t_i \ \ _n(B_i) \right| +$$

$$+ \left| \sum_{i=0}^{k-1} t_i \ \ _n(B_i) - \sum_{i=0}^{k-1} t_i \ \ (B_i) \right| + \left| \sum_{i=0}^{k-1} t_i \ \ (B_i) - \int_E f(x) \ \ (dx) \right| \leqslant$$

$$\leqslant 2 \max_{0 \leqslant i \leqslant k-1} (t_{i+1} - t_i) + \left| \sum_{i=0}^{k-1} t_i \ \ _n(B_i) - \sum_{i=0}^{k-1} t_i \ \ (B_i) \right|,$$

откуда в силу (12) и произвольности разбиений T_k, $k \geqslant 1$,

$$\lim_n \int_E f(x) \ \ _n(dx) = \int_E f(x) \ \ (dx). \qquad \square$$

Замечание 1. Участвующие в доказательстве импликации (I) $\Rightarrow$ (II) функции $f(x) = I_A(x)$ и $f_A^\varepsilon(x)$ являются соответственно *полунепрерывными сверху* и *равномерно непрерывными*. Учитывая это обстоятельство, нетрудно показать, что каждое из условий теоремы эквивалентно одному из нижеследующих условий:

(V) $\int_E f(x) \ \ _n(dx) \to \int_E f(x) \ \ (dx)$ *для всех ограниченных равномерно непрерывных функций* $f(x)$;

(VI) $\int_E f(x) \ \ _n(dx) \to \int_E f(x) \ \ (dx)$ *для всех ограниченных функций, удовлетворяющих условию Липшица* (см. лемму 2 в § 7);

(VII) $\varlimsup_n \int_E f(x) \ \ _n(dx) \leqslant \int_E f(x) \ \ (dx)$ *для всех ограниченных функций* $f(x)$, *являющихся полунепрерывными сверху* ($\varlimsup_n f(x_n) \leqslant f(x)$, $x_n \to x$);

(VIII) $\varliminf_n \int_E f(x) \ \ _n(dx) \geqslant \int_E f(x) \ \ (dx)$ *для всех ограниченных функций* $f(x)$, *являющихся полунепрерывными снизу* ($\varliminf_n f(x_n) \geqslant f(x)$, $x_n \to x$).

Замечание 2. Теорема 1 допускает естественное обобщение на тот случай, когда вместо вероятностных мер $\ $ и $\ _n$, заданных на $(E, \mathscr{E}, \rho)$, рассматриваются *произвольные* (не обязательно вероятностные) *конеч-*

ные 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*) $\overline{\lim\limits_n}\ \mu_n(A) \leqslant \mu(A)$, A — *замкнутые множества, и* $\mu_n(E) \to \mu(E)$;

(III*) $\underline{\lim\limits_n}\ \mu_n(A) \geqslant \mu(A)$, A — *открытые множества, и* $\mu_n(E) \to \mu(E)$;

(IV*) $\mu_n \Rightarrow \mu$.

Каждое из этих условий равносильно любому из условий (V*)—(VIII*), формулируемых, как и (V)—(VIII), с заменой мер ${}_n$ и на μ_n и μ соответственно.

4. 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) $P_n \xrightarrow{w} P$,

(2) $P_n \Rightarrow P$,

(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 $P_n \Rightarrow P$, то, в частности,

$$P_n(-\infty, x] \to P(-\infty, x]$$

для всех $x \in R$ таких, что $P\{x\} = 0$. А это и означает, что $F_n \Rightarrow F$.

Пусть теперь $F_n \Rightarrow F$. Для доказательства сходимости $P_n \Rightarrow P$ достаточно (в силу теоремы 1) показать, что $\underline{\lim\limits_n}\ P_n(A) \geqslant P(A)$ 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 $A = \sum\limits_{k=1}^{\infty} I_k$. 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 $P(I_k) \leqslant P(I_k') + \varepsilon 2^{-k}$. (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

$$\underline{\lim_n}\ P_n(A) = \underline{\lim_n} \sum_{k=1}^{\infty} P_n(I_k) \geqslant \sum_{k=1}^{\infty} \underline{\lim_n}\ P_n(I_k) \geqslant \sum_{k=1}^{\infty} \underline{\lim_n}\ P_n(I_k').$$

Но

$$P_n(I'_k) = F_n(b'_k) - F_n(a'_k) \to F(b'_k) - F(a'_k) = P(I'_k).$$

Поэтому

$$\varliminf_n P_n(A) \geqslant \sum_{k=1}^{\infty} P(I'_k) \geqslant \sum_{k=1}^{\infty} (P(I_k) - \varepsilon 2^{-k}) = P(A) - \varepsilon,$$

что в силу произвольности $\varepsilon > 0$ доказывает, что $\varliminf_n P_n(A) \geqslant P(A)$, если A — открытое множество. □

5. Пусть $(E, \mathscr{E})$ — измеримое пространство. Систему подмножеств $\mathscr{K}_0(E) \subseteq \mathscr{E}$ назовем *определяющим классом*, если для любых двух вероятностных мер   и , заданных на $(E, \mathscr{E})$, из равенства

$$(A) = \quad (A) \text{ для всех } A \in \mathscr{K}_0(E)$$

вытекает, что эти меры совпадают тождественно, т. е.

$$(A) = \quad (A) \text{ для всех } A \in \mathscr{E}.$$

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 , $_1$, $_2$, … из того, что

$$_n(A) \to \quad (A) \text{ для всех } A \in \mathscr{K}_1(E) \text{ с } \quad (\partial A) = 0$$

вытекает, что

$$_n(A) \to \quad (A) \text{ для всех } A \in \mathscr{E} \text{ с } \quad (\partial A) = 0.$$

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* 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).

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 $(C, \mathscr{B}(C), \rho)$ с равномерной метрикой ρ (см. п. 6 § 2 гл. II). Пусть — вероятностная мера, целиком сосредоточенная на функции $x(t) \equiv 0$, $0 \leqslant t \leqslant 1$, а $_n$ — вероятностные меры, $n \geqslant 1$, каждая из которых сосредоточена на функции $x_n = x_n(t)$, изображенной на рис. 35. Нетрудно убедиться, что $_n(A) \to (A)$ для всех цилиндрических множеств A с $(\partial A) = 0$. Но, если взять, например, множество

$$A = \left\{ \alpha \in C\colon\ |\alpha(t)| \leqslant \frac{1}{2},\ \ 0 \leqslant t \leqslant 1 \right\} \in \mathscr{B}_0(C),$$

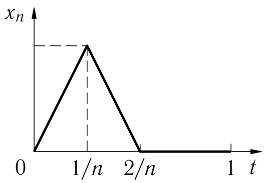


Рис. 35.

то $(\partial A) = 0$, $_n(A) = 0$, $(A) = 1$ и, следовательно, $_n \not\Rightarrow$.

Таким образом, класс цилиндрических множеств является определяющим классом, но не является классом, определяющим сходимость.

6. Задачи.

1. Будем говорить, что функция $F = F(x)$, заданная на R^m, *непрерывна в точке* $x \in R^m$, если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что $|F(x) - F(y)| < \varepsilon$ для всех $y \in R^m$, удовлетворяющих неравенству

$$x - \delta e < y < x + \delta e,$$

где $e = (1, \ldots, 1) \in R^m$. Будем говорить также, что последовательность функций распределения $\{F_n\}$ *сходится в основном* к функции распределения F (обозначение: $F_n \Rightarrow F$), если $F_n(x) \to F(x)$ для всех точек $x \in R^m$, где функция $F = F(x)$ непрерывна.

Показать, что утверждение теоремы 2 остается справедливым для R^m, $m > 1$. (См. замечание 1 к теореме 1.)

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 $\{ _n\}$ (заданных на σ-алгебре $\mathscr{E}$ борелевских множеств, порожденных открытыми множествами) сходится *в основном в смысле конечномерных распределений* к вероятностной мере (обозначение: $_n \overset{f}{\Rightarrow}$), если $_n(A) \to (A)$, $n \to \infty$, для всех *цилиндрических* множеств A с $(\partial A) = 0$.

Показать, что в случае пространства R^∞

$$(\ {}_n \xrightarrow{f}\ ) \Leftrightarrow (\ {}_n \Rightarrow\ ).$$

Верно ли это утверждение для пространств C и D?

4. Пусть F и G — функции распределения на числовой прямой и

$$L(F, G) = \inf\{h > 0\colon\ F(x-h) - h \leqslant G(x) \leqslant F(x+h) + h\}$$

— *расстояние Леви* (между F и G). Показать, что сходимость в основном эквивалентна сходимости в метрике Леви, определяемой расстоянием $L(\cdot, \cdot)$:

$$(F_n \Rightarrow F) \ \Leftrightarrow\ (L(F_n, F) \to 0).$$

5. Пусть $F_n \Rightarrow F$ и функция распределения F является непрерывной. Показать, что тогда сходимость $F_n(x)$ к $F(x)$ *равномерна*:

$$\sup_x |F_n(x) - F(x)| \to 0, \quad n \to \infty.$$

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 $\ {}_n \xrightarrow{w}\ $ тогда и только тогда, когда всякая подпоследовательность $\{\ {}_{n'}\}$ последовательности $\{\ {}_n\}$ содержит подпоследовательность $\{\ {}_{n''}\}$ такую, что $\ {}_{n''} \xrightarrow{w}\ $.

9. Дать пример вероятностных мер P, P_n на $(R, \mathscr{B}(R))$, $n \geqslant 1$, таких, что $P_n \xrightarrow{w} P$, но для *всех* борелевских множеств $B \in \mathscr{B}(R)$ сходимости $P_n(B) \to P(B)$ может и не быть.

10. Привести пример функций распределения $F = F(x)$, $F_n = F_n(x)$, $n \geqslant 1$, таких, что $F_n \xrightarrow{w} F$, но $\sup\limits_x |F_n(x) - F(x)| \not\to 0$, $n \to \infty$.

11. Во многих руководствах по теории вероятностей утверждение (4) $\Rightarrow$ (3) теоремы 2 о сходимости функций распределения F_n, $n \geqslant 1$, к функции распределения F связывается с именами Хелли и Брэя. В этой связи предлагается передоказать следующие утверждения:

(а) *Лемма Хелли—Брэя*. Если $F_n \Rightarrow F$ (см. определение 1), то

$$\lim_n \int_a^b g(x)\, dF_n(x) = \int_a^b g(x)\, dF(x),$$

где a и b — точки из множества точек непрерывности функции распределения $F = F(x)$ и $g = g(x)$ — непрерывная функция на $[a, b]$.

(b) *Теорема Хелли—Брэя*. Если $F_n \Rightarrow F$ и $g = g(x)$ — непрерывная ограниченная функция на R, то

$$\lim_n \int_{-\infty}^{\infty} g(x)\, dF_n(x) = \int_{-\infty}^{\infty} g(x)\, dF(x).$$

12. Пусть $F_n \Rightarrow F$ и последовательность $\left(\int |x|^b\, dF_n(x)\right)_{n\geqslant 1}$ ограничена для некоторого $b > 0$. Показать, что тогда

$$\lim_n \int |x|^a\, dF_n(x) = \int |x|^a\, dF(x), \qquad 0 \leqslant a \leqslant b,$$

$$\lim_n \int x^k\, dF_n(x) = \int x^k\, dF(x) \quad \text{для всякого } k = 1, 2, \ldots, [b],\ k \neq b.$$

13. Пусть $F_n \Rightarrow F$ и $m = \mathrm{med}(F)$, $m_n = \mathrm{med}(F_n)$ — медианы F и F_n соответственно (см. задачу 5 в § 4 гл. I). Предположим, что медианы m_0 и m_n определены однозначно для всех $n \geqslant 1$. Доказать, что $m_n \to m$.

14. Пусть F — функция распределения, однозначно определяемая своими моментами $a_k = \int_{-\infty}^{\infty} x^k\, dF(x)$, $k = 1, 2, \ldots$ Пусть $(F_n)_{n\geqslant 1}$ — последовательность функций распределения такая, что моменты

$$a_{n,k} = \int_{-\infty}^{\infty} x^k\, dF_n(x) \to a_k = \int_{-\infty}^{\infty} x^k\, dF(x), \qquad k = 1, 2, \ldots$$

Показать, что тогда $F_n \Rightarrow F$.

15. Доказать следующую версию *закона больших чисел* (Хинчин): пусть $X_1, X_2, \ldots$ — попарно независимые одинаково распределенные случайные величины с конечным средним $\ X_1 = m$ и $S_n = X_1 + \ldots + X_n$, тогда $S_n/n \to m$.

§ 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 какой-либо мере или имеет ли она хотя бы одну сходящуюся подпоследовательность.

Так, например, последовательность $\{\ _n\}$, где $\ _{2n} =\ $, $\ _{2n+1} =\ $, а и — различные вероятностные меры, не является, очевидно, сходящейся, но имеет две сходящиеся подпоследовательности $\{\ _{2n}\}$ и $\{\ _{2n+1}\}$.

Совсем просто устроенная последовательность $\{\;_n\}$ вероятностных мер $\;_n$, $n \geqslant 1$, каждая из которых сосредоточена в точке $\{n\}$ ($\;_n(\{n\}) = 1$), не только не является сходящейся, но и не содержит ни одной сходящейся подпоследовательности. (Поскольку $\lim\limits_n \;_n(a, b] = 0$ для любых $a < b$, то предельная мера должна была бы быть тождественно равной нулю, а это противоречит тому, что $1 = \;_n(R) \not\to 0$, $n \to \infty$.) 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». (См. в этой связи задачу 7 в § 3.)

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 $\mathscr{P} = \{\;_\alpha;\ \alpha \in \mathfrak{A}\}$ 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 $\mathscr{P} = \{\;_\alpha;\ \alpha \in \mathfrak{A}\}$ называется *плотным*, если для каждого $\varepsilon > 0$ можно указать компакт $K \subseteq E$

такой, что

$$\sup_{\alpha\in\mathfrak{A}} \ {}_{\alpha}(E\setminus K)\leqslant\varepsilon. \tag{1}$$

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 $\mathscr{P}=\{\ {}_\alpha;\ \alpha\in\mathfrak{A}\}$, где ${}_\alpha$ — мера, построенная по F_α.

3. Следующий результат играет фундаментальную роль во всей проблематике слабой сходимости вероятностных мер.

Теорема 1 (теорема Ю. В. Прохорова). *Пусть* $\mathscr{P}=\{\ {}_\alpha;\ \alpha\in\mathfrak{A}\}$ — *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 теоремы будет приведено лишь для случая числовой прямой. (Это доказательство переносится ([55], [5]) 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 $\mathscr{P}=\{\ {}_\alpha;\ \alpha\in\mathfrak{A}\}$, заданных на $(R,\ \mathscr{B}(R))$, относительно компактно, но не плотно. Тогда найдется такое $\varepsilon>0$, что для любого компакта $K\subseteq R$

$$\sup_{\alpha} \ {}_{\alpha}(R\setminus K)>\varepsilon,$$

а значит, и для любого интервала $I=(a,\ b)$

$$\sup_{\alpha} \ {}_{\alpha}(R\setminus I)>\varepsilon.$$

Отсюда вытекает, что для каждого интервала $I_n=(-n,\ n)$, $n\geqslant 1$, найдется такая мера ${}_{\alpha_n}$, что

$${}_{\alpha_n}(R\setminus I_n)>\varepsilon.$$

Раз исходное семейство $\mathscr{P}$ относительно компактно, то из последовательности $\{\ {}_{\alpha_n}\}_{n\geqslant 1}$ можно извлечь подпоследовательность, скажем, $\{\ {}_{\alpha_{n_k}}\}$, такую, что ${}_{\alpha_{n_k}}\xrightarrow{w}\ $, где — некоторая вероятностная мера.

Тогда в силу эквивалентности условий (I) и (II) в теореме 1 из § 1 для всякого $n\geqslant 1$

$$\varlimsup_{k\to\infty} \ {}_{\alpha_{n_k}}(R\setminus I_n)\leqslant\ (R\setminus I_n). \tag{2}$$

Но $(R \setminus I_n) \downarrow 0$, $n \to \infty$, 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 $\mathscr{I} = \{G\}$ 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 $\mathscr{I}$ включает в себя класс функций распределения $\mathscr{F} = \{F\}$, для которых $F(-\infty) = 0$ и $F(+\infty) = 1$.

Теорема 2 (теорема Хелли). *Класс $\mathscr{I} = \{G\}$ обобщенных функций распределения является секвенциально компактным, т. е. для любой последовательности $\{G_n\}$ функций из $\mathscr{I}$ найдутся функция $G \in \mathscr{I}$ и подпоследовательность $\{n_k\} \subseteq \{n\}$ такие, что*

$$G_{n_k}(x) \to G(x), \quad k \to \infty,$$

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 $G_{n_i^{(1)}}(x_1)$ при $i \to \infty$ 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

$$G(x) = \inf\{G_T(y)\colon\ y \in T,\ y > x\}. \tag{3}$$

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 $\mathscr{I}$.

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

$$\varlimsup_m G_{n_m^{(m)}}(x^0) \leqslant \varlimsup_m G_{n_m^{(m)}}(y) = G_T(y),$$

откуда

$$\varlimsup_m G_{n_m^{(m)}}(x^0) \leqslant \inf\{G_T(y)\colon\ y \in T,\ y > x^0\} = G(x^0). \tag{4}$$

С другой стороны, пусть $x^1 < y < x^0$, $y \in T$. Тогда

$$G(x^1) \leqslant G_T(y) = \lim_m G_{n_m^{(m)}}(y) = \varliminf_m G_{n_m^{(m)}}(y) \leqslant \varliminf_m G_{n_m^{(m)}}(x^0).$$

Поэтому, полагая $x^1 \uparrow x^0$, получим, что

$$G(x^0-) \leqslant \varliminf_m G_{n_m^{(m)}}(x^0). \tag{5}$$

Но если $G(x^0-) = G(x^0)$, то тогда из (4) и (5) заключаем, что $G_{n_m^{(m)}}(x^0) \to G(x^0)$, $m \to \infty$. $\square$

Завершим теперь доказательство теоремы 1.

Достаточность. Пусть семейство $\mathscr{P}$ плотно и $\{\ {}_n\}$ — 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 $G \in \mathscr{I}$ 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

$$\sup_n \quad {}_n(R\setminus I)<\varepsilon,$$

или, что эквивалентно,

$$1-\varepsilon\leqslant \quad {}_n(a,\,b], \qquad n\geqslant 1.$$

Выберем точки a', $b'\in\mathbb{C}(G)$ такими, что $a'<a$, $b'>b$. Тогда

$$1-\varepsilon\leqslant \quad {}_{n_k}(a,\,b]\leqslant \quad {}_{n_k}(a',\,b']=F_{n_k}(b')-F_{n_k}(a')\to G(b')-G(a').$$

Отсюда следует, что $G(+\infty)-G(-\infty)=1$, и поскольку $0\leqslant G(-\infty)\leqslant$ $\leqslant G(+\infty)\leqslant 1$, 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 $\quad {}_{n_k}\xrightarrow{w}\quad$, где — вероятностная мера, построенная по функции распределения G. $\square$

4. Задачи.

1. Провести доказательство теорем 1 и 2 для пространств R^n, $n\geqslant 2$.

2. Пусть $\quad {}_\alpha$ — гауссовская мера на числовой прямой с параметрами m_α и σ_α^2, $\alpha\in\mathfrak{A}$. Показать, что семейство $\mathscr{P}=\{\quad {}_\alpha;\,\alpha\in\mathfrak{A}\}$ является плотным тогда и только тогда, когда существуют константы a и b такие, что

$$|m_\alpha|\leqslant a,\;\; \sigma_\alpha^2\leqslant b,\;\; \alpha\in\mathfrak{A}.$$

3. Привести примеры плотных и не плотных семейств вероятностных мер $\mathscr{P}=\{\quad {}_\alpha;\,\alpha\in\mathfrak{A}\}$, определенных на $(R^\infty,\,\mathscr{B}(R^\infty))$.

4. Пусть P есть вероятностная мера на метрическом пространстве $(E,\,\mathscr{E},\,\rho)$. Говорят (ср. с определением 2), что мера P является *плотной*, если для каждого $\varepsilon>0$ найдется компакт $K\subseteq E$ такой, что $P(K)\geqslant 1-\varepsilon$. Доказать следующий результат («теорема Улама»): *каждая вероятностная мера P на польском* (*т. е. полном сепарабельном метрическом*) *пространстве является плотной.*

5. Пусть $X=\{X_\alpha;\,\alpha\in\mathfrak{A}\}$ — некоторое семейство случайных векторов ($X_\alpha\in R^d$, $\alpha\in\mathfrak{A}$), при этом $\sup\limits_\alpha \quad \|X_\alpha\|^r<\infty$ для некоторого $r>0$. Показать, что семейство $\mathscr{P}=\{P_\alpha;\,\alpha\in\mathfrak{A}\}$ распределений $P_\alpha=\mathrm{Law}(X_\alpha)$ является плотным.

§ 3. Метод характеристических функций в доказательстве предельных теорем

1. Доказательство первых предельных теорем теории вероятностей — закона больших чисел, теорем Муавра—Лапласа и Пуассона для схемы Бернулли — основывалось на *прямом анализе* допредельных функций распределения 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», дало возможность не только 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 $\varepsilon > 0$

$$\left\{\left|\frac{S_n}{n} - \frac{\ S_n}{n}\right| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{1}$$

(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

$$\left\{\frac{S_n - \ S_n}{\sqrt{\ S_n}} \leqslant x\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\, du \ (= \Phi(x)), \tag{2}$$

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 оказывается 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* (φ_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 предел* $\lim\limits_n \varphi_n(t)$ *и функция* $\varphi(t) = \lim\limits_n \varphi(t)$ *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 $\mathrm{Re}\, e^{itx}$ и $\mathrm{Im}\, e^{itx}$.

Доказательству утверждения 2) предпошлем несколько вспомогательных предложений.

Лемма 1. *Пусть* $\{\ _n\}$ *— плотное семейство вероятностных мер. Предположим, что каждая слабо сходящаяся подпоследовательность* $\{\ _{n'}\}$ *последовательности* $\{\ _n\}$ *сходится к одной и той же вероятностной мере* . *Тогда и вся последовательность* $\{\ _n\}$ *слабо сходится к* .

Доказательство. Допустим, что $\ _n \overset{w}{\not\to}\ $. Тогда найдется такая ограниченная непрерывная функция $f = f(x)$, что

$$\int_R f(x)\ \ _n(dx) \not\to \int_R f(x)\ \ (dx).$$

Отсюда следует, что существуют $\varepsilon > 0$ и бесконечная последовательность чисел $\{n'\} \subseteq \{n\}$ такие, что

$$\left| \int_R f(x) \quad {}_{n'}(dx) - \int_R f(x) \quad (dx) \right| \geqslant \varepsilon > 0. \tag{3}$$

По теореме Прохорова (§ 2) из последовательности $\{\ {}_{n'}\}$ можно выбрать подпоследовательность $\{\ {}_{n''}\}$ такую, что $\ {}_{n''} \xrightarrow{w}\ $, где — некоторая вероятностная мера.

По предположению леммы $=$, и, значит,

$$\int_R f(x) \quad {}_{n''}(dx) \to \int_R f(x) \quad (dx),$$

что находится в противоречии с (3). □

Лемма 2. *Пусть* $\{\ {}_n\}$ — *плотное семейство вероятностных мер на* $(R, \mathscr{B}(R))$. *Последовательность* $\{\ {}_n\}$ *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* $\ {}_n$:

$$\varphi_n(t) = \int_R e^{itx} \quad {}_n(dx).$$

Доказательство. Если семейство $\{\ {}_n\}$ плотно, то по теореме Прохорова найдутся подпоследовательность $\{\ {}_{n'}\}$ и вероятностная мера такие, что $\ {}_{n'} \xrightarrow{w}\ $. Предположим, что вся последовательность $\{\ {}_n\}$ не сходится к ($\ {}_n \overset{w}{\not\to}\ $). Тогда в силу леммы 1 найдутся подпоследовательность $\{\ {}_{n''}\}$ и вероятностная мера такие, что $\ {}_{n''} \xrightarrow{w}\ $, причем $\neq$.

Воспользуемся теперь тем, что при каждом $t \in R$ существует $\lim\limits_n \varphi_n(t)$. Тогда

$$\lim_{n'} \int_R e^{itx} \quad {}_{n'}(dx) = \lim_{n''} \int_R e^{itx} \quad {}_{n''}(dx)$$

и, значит,

$$\int_R e^{itx} \quad (dx) = \int_R e^{itx} \quad (dx), \qquad t \in R.$$

Но характеристическая функция однозначно определяет распределение (теорема 2 § 12 гл. II). Поэтому $=$, что противоречит предположению $\ {}_n \overset{w}{\not\to}\ $.

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* $a>0$

$$\int\limits_{|x|\geqslant 1/a} dF(x) \leqslant \frac{K}{a}\int\limits_0^a [1-\operatorname{Re}\varphi(t)]\,dt. \tag{4}$$

Доказательство. Поскольку $\operatorname{Re}\varphi(t)=\int\limits_{-\infty}^{\infty}\cos tx\,dF(x)$, то, применяя теорему Фубини, находим, что

$$\frac{1}{a}\int\limits_0^a [1-\operatorname{Re}\varphi(t)]\,dt = \frac{1}{a}\int\limits_0^a \left[\int\limits_{-\infty}^{\infty}(1-\cos tx)\,dF(x)\right]dt =$$

$$= \int\limits_{-\infty}^{\infty}\left[\frac{1}{a}\int\limits_0^a (1-\cos tx)\,dt\right]dF(x) = \int\limits_{-\infty}^{\infty}\left(1-\frac{\sin ax}{ax}\right)dF(x) \geqslant$$

$$\geqslant \inf_{|y|\geqslant 1}\left(1-\frac{\sin y}{y}\right)\cdot\int\limits_{|ax|\geqslant 1} dF(x) = \frac{1}{K}\int\limits_{|x|\geqslant 1/a} dF(x),$$

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) заведомо справедливо с константой $K=7$. □

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 $\{\ _n\}$, где $_n$ — мера, соответствующая функции распределения F_n.

В силу (4) и теоремы о мажорируемой сходимости

$$_n\left\{R\setminus\left(-\frac{1}{a},\,\frac{1}{a}\right)\right\} = \int\limits_{|x|\geqslant\frac{1}{a}} dF_n(x) \leqslant$$

$$\leqslant \frac{K}{a}\int\limits_0^a [1-\operatorname{Re}\varphi_n(t)]\,dt \to \frac{K}{a}\int\limits_0^a [1-\operatorname{Re}\varphi(t)]\,dt$$

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

$$_n\left\{R\setminus\left(-\frac{1}{a},\,\frac{1}{a}\right)\right\}\leqslant\varepsilon \text{ для всех } n\geqslant 1.$$

Следовательно, семейство $\{\quad_n\}$ плотно, и в силу леммы 2 существует вероятностная мера такая, что

$$\quad_n \xrightarrow{w} \quad .$$

Отсюда

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} \quad_n(dx) \to \int_{-\infty}^{\infty} e^{itx} \quad (dx),$$

и в то же самое время $\varphi_n(t) \to \varphi(t)$. Поэтому $\varphi(t)$ является характеристической функцией вероятностной меры . □

Следствие. *Пусть (F_n) — последовательность функций распределения и (φ_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 4 § 10 гл. II тогда говорят, что *случайные величины* $\eta_1, \eta_2, \ldots$ *сходятся по распределению* к η, и записывают это в виде $\eta_n \xrightarrow{d} \eta$. Эта запись наглядна ($d$ — от *distribution*) 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; А. Я. Хинчин). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\quad|\xi_1| < \infty$, $S_n = \xi_1 + \ldots + \xi_n$ и $\quad\xi_1 = m$. Тогда $\dfrac{S_n}{n} \to m$, т. е. для всякого $\varepsilon > 0$*

$$\left\{\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty.$$

Доказательство. Пусть $\varphi(t) = \quad e^{it\xi_1}$ и $\varphi_{\frac{S_n}{n}}(t) = \quad e^{it\frac{S_n}{n}}$. Тогда в силу независимости случайных величин и формулы (6) § 12 гл. I

$$\varphi_{\frac{S_n}{n}}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n.$$

Но, согласно (14) § 12 гл. II,

$$\varphi(t) = 1 + itm + o(t), \quad t \to 0.$$

Значит, для всякого фиксированного $t \in R$

$$\varphi\left(\frac{t}{n}\right) = 1 + i\,\frac{t}{n}\,m + o\left(\frac{1}{n}\right), \quad n \to \infty,$$

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),

$$\frac{S_n}{n} \to m. \qquad \square$$

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 $\xi_1^2 < \infty$ и $S_n = \xi_1 + \ldots + \xi_n$. Тогда при $n \to \infty$*

$$\left\{\frac{S_n - \quad S_n}{\sqrt{\quad S_n}} \leqslant x\right\} \to \Phi(x), \quad x \in R, \tag{5}$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\,du.$$

Доказательство. Пусть $\xi_1 = m$, $\xi_1 = \sigma^2$ и

$$\varphi(t) = \quad e^{it(\xi_1 - m)}.$$

Тогда, если обозначить

$$\varphi_n(t) = \quad e^{it\frac{S_n - \quad S_n}{\sqrt{\quad S_n}}},$$

то получим, что

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

Но в силу (14) § 12 гл. II

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \quad t \to 0.$$

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:

$$\frac{S_n - \quad S_n}{\sqrt{\quad S_n}} \xrightarrow{d} \mathscr{N}(0,\,1). \tag{6}$$

□

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 $\xi_{n1}, \ldots, \xi_{nn}$. Иначе говоря, пусть задана треугольная таблица

$$\begin{pmatrix} \xi_{11} & & \\ \xi_{21} & \xi_{22} & \\ \xi_{31} & \xi_{32} & \xi_{33} \\ \ldots & \ldots & \ldots \end{pmatrix}$$

случайных величин, которые в каждой строчке независимы между собой. Положим $S_n = \xi_{n1} + \ldots + \xi_{nn}$.

Теорема 4 (теорема Пуассона)**.** *Пусть при каждом $n \geqslant 1$ независимые одинаково распределенные случайные величины $\xi_{n1}, \ldots, \xi_{nn}$ таковы, что*

$$\{\xi_{nk} = 1\} = p_{nk}, \quad \{\xi_{nk} = 0\} = q_{nk}, \ \ 1 \leqslant k \leqslant n,$$
$$p_{nk} + q_{nk} = 1, \ \ \max_{1 \leqslant k \leqslant n} p_{nk} \to 0, \ \ p_{n1} + \ldots + p_{nn} \to \lambda > 0, \ \ n \to \infty.$$

Тогда

$$\{S_n = m\} \to \frac{e^{-\lambda}\lambda^m}{m!}, \quad m = 0,\,1,\,\ldots \tag{7}$$

Доказательство. Поскольку для $1 \leqslant k \leqslant n$

$$e^{it\xi_{nk}} = p_{nk}e^{it} + q_{nk},$$

то

$$\varphi_{S_n}(t) = \quad e^{itS_n} = \prod_{k=1}^{n} (p_{nk}e^{it} + q_{nk}) = \prod_{k=1}^{n} (1 + p_{nk}(e^{it} - 1)) \to$$
$$\to \exp\{\lambda(e^{it} - 1)\}, \quad n \to \infty.$$

Функция $\varphi(t) = \exp\{\lambda(e^{it} - 1)\}$ является характеристической функцией пуассоновского распределения (пример 3 в п. 2 § 12 гл. II), 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:

$$S_n \xrightarrow{d} \pi(\lambda). \qquad \square$$

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 $\quad |\xi_n|$ и дисперсиями $\quad \xi_n$ такими, что $\quad \xi_n \leqslant K < \infty$, 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 $\varphi_{\xi_n}(t) \to 1$, $n \to \infty$, в некоторой окрестности точки $t = 0$.

5. Пусть $X_1, X_2, \ldots$ — последовательность независимых одинаково распределенных случайных векторов (со значения ми 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.)

6. Пусть $\xi_1, \xi_2, \ldots$ и $\eta_1, \eta_2, \ldots$ — две последовательности случайных величин такие, что ξ_n и η_n независимы при каждом n. Предположим, что $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} \eta$ при $n \to \infty$, где ξ и η независимы. Доказать, что последовательность двумерных случайных величин (ξ_n, η_n) сходится по распределению к (ξ, η).

Пусть $f = f(x, y)$ — непрерывная функция. Проверить, что последовательность $f(\xi_n, \eta_n)$ сходится по распределению к $f(\xi, \eta)$.

7. Привести пример, показывающий, что в утверждении 2) теоремы 1 условие непрерывности «предельной» характеристической функции $\varphi(t) = \lim\limits_n \varphi_n(t)$ в нуле, вообще говоря, не может быть ослаблено. (Иначе говоря, если $\varphi(t)$ — характеристическая функция распределения F — не непрерывна в нуле, то может случиться, что $\varphi_n(t) \to \varphi(t)$, но $F_n \overset{w}{\not\to} F$.) Убедиться на примере, что отсутствие непрерывности $\varphi(t)$ в нуле может привести к нарушению свойства плотности семейства вероятностных распределений P_n, $n \geqslant 1$, с характеристическими функциями $\varphi_n(t)$, $n \geqslant 1$.

§ 4. Центральная предельная теорема для сумм независимых случайных величин. I. Условие Линдеберга

1. В этом параграфе центральная предельная теорема для (нормированных и центрированных) сумм S_n независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, $n \geqslant 1$, будет доказываться при традиционном предположении выполнения *классического условия Линдеберга*. В следующем параграфе будет рассмотрена более общая ситуация: во-первых, центральная предельная теорема будет сразу формулироваться в «схеме серий», и, во-вторых, ее доказательство будет идти при выполнении так называемых *неклассических условий*.

Теорема 1. *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечными вторыми моментами. Пусть $m_k = \ \xi_k$, $\sigma_k^2 = \ \xi_k > 0$, $S_n = \xi_1 + \ldots + \xi_n$, $D_n^2 = \sum\limits_{k=1}^n \sigma_k^2$ и $F_k = F_k(x)$ — функция распределения случайной величины ξ_k.*

Предположим, что выполнено «условие Линдеберга»: для всякого $\varepsilon > 0$

$$\text{(L)} \qquad \frac{1}{D_n^2} \sum_{k=1}^n \int\limits_{\{x:\,|x-m_k| \geqslant \varepsilon D_n\}} (x - m_k)^2 \, dF_k(x) \to 0, \quad n \to \infty. \tag{1}$$

Тогда

$$\frac{S_n - \ S_n}{\sqrt{\ S_n}} \xrightarrow{d} \mathscr{N}(0, 1). \tag{2}$$

Доказательство. Без ограничения общности можно считать $m_k = 0$, $k \geqslant 1$. Обозначим $\varphi_k(t) = \ e^{it\xi_k}$, $T_n = \dfrac{S_n}{\sqrt{\ S_n}} = \dfrac{S_n}{D_n}$, $\varphi_{S_n}(t) = \ e^{itS_n}$, $\varphi_{T_n}(t) = \ e^{itT_n}$.

Тогда

$$\varphi_{T_n}(t) = \quad e^{itT_n} = \quad e^{i\frac{t}{D_n}S_n} = \varphi_{S_n}\left(\frac{t}{D_n}\right) = \prod_{k=1}^{n} \varphi_k\left(\frac{t}{D_n}\right) \tag{3}$$

и для доказательства (2) достаточно (в силу теоремы 1 из § 3) установить, что для каждого $t \in R$

$$\varphi_{T_n}(t) \to e^{-t^2/2}, \quad n \to \infty. \tag{4}$$

Возьмем некоторое $t \in R$ и будем считать его фиксированным на протяжении всего доказательства. В силу разложений

$$e^{iy} = 1 + iy + \frac{\theta_1 y^2}{2},$$
$$e^{iy} = 1 + iy - \frac{y^2}{2} + \frac{\theta_2 |y|^3}{3!},$$

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

$$\varphi_k(t) = \quad e^{it\xi_k} = \int_{-\infty}^{\infty} e^{itx}\, dF_k(x) = \int_{|x| \geqslant \varepsilon D_n} \left(1 + itx + \frac{\theta_1 (tx)^2}{2}\right) dF_k(x) +$$
$$+ \int_{|x| < \varepsilon D_n} \left(1 + itx - \frac{t^2 x^2}{2} + \frac{\theta_2 |tx|^3}{6}\right) dF_k(x) =$$
$$= 1 + \frac{t^2}{2} \int_{|x| \geqslant \varepsilon D_n} \theta_1 x^2\, dF_k(x) - \frac{t^2}{2} \int_{|x| < \varepsilon D_n} x^2\, dF_k(x) + \frac{|t|^3}{6} \int_{|x| < \varepsilon D_n} \theta_2 |x|^3\, dF_k(x)$$

(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

$$\frac{1}{2}\int\limits_{|x|\geqslant\varepsilon D_n}\theta_1 x^2\,dF_k(x)=\tilde\theta_1\int\limits_{|x|\geqslant\varepsilon D_n}x^2 dF_k(x), \tag{6}$$

где $\tilde\theta_1=\tilde\theta_1(t,\,k,\,n)$ и $|\tilde\theta_1|\leqslant 1/2$.

Точно так же

$$\left|\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\theta_2|x|^3\,dF_k(x)\right|\leqslant\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\frac{\varepsilon D_n}{|x|}|x|^3\,dF_k(x)\leqslant\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\varepsilon D_n x^2\,dF_k(x),$$

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

$$A_{kn}=\frac{1}{D_n^2}\int\limits_{|x|<\varepsilon D_n}x^2\,dF_k(x),\qquad B_{kn}=\frac{1}{D_n^2}\int\limits_{|x|\geqslant\varepsilon D_n}x^2\,dF_k(x).$$

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),

$$\sum_{k=1}^{n}B_{kn}\to 0,\quad n\to\infty. \tag{10}$$

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 комплексных чисел z с $|z|\leqslant 1/2$

$$\ln(1+z)=z+\theta|z|^2,$$

где $\theta = \theta(z)$ с $|\theta| \leqslant 1$ и ln обозначает *главное значение* логарифма ($\ln z = \ln |z| + i \arg z$, $-\pi < \arg z \leqslant \pi$). 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 такое $\varepsilon > 0$, 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)

$$\left|\sum_{k=1}^{n} \theta_{kn}|C_{kn}|^2\right| \leqslant \max_{1 \leqslant k \leqslant n} |C_{kn}| \cdot \sum_{k=1}^{n} |C_{kn}| \leqslant (t^2\varepsilon^2 + \varepsilon|t|^3)\,(t^2 + \varepsilon|t|^3).$$

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

$$\varphi_{T_n}(t) e^{t^2/2} \to 1, \quad n \to \infty,$$

и, значит,

$$\varphi_{T_n}(t) \to e^{-t^2/2}, \quad n \to \infty. \qquad \square$$

2. Остановимся на некоторых частных случаях, в которых выполнено *условие Линдеберга* (1) и, следовательно, справедлива центральная предельная теорема.

а) Пусть выполнено *условие Ляпунова*: для некоторого $\delta > 0$

$$\frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \quad |\xi_k - m_k|^{2+\delta} \to 0, \qquad n \to \infty. \tag{13}$$

Пусть $\varepsilon > 0$, тогда

$$|\xi_k - m_k|^{2+\delta} = \int_{-\infty}^{\infty} |x - m_k|^{2+\delta}\, dF_k(x) \geqslant$$
$$\geqslant \int_{\{x : |x - m_k| \geqslant \varepsilon D_n\}} |x - m_k|^{2+\delta}\, dF_k(x) \geqslant \varepsilon^\delta D_n^\delta \int_{\{x : |x - m_k| \geqslant \varepsilon D_n\}} (x - m_k)^2\, dF_k(x)$$

и, значит,

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x : |x - m_k| \geqslant \varepsilon D_n\}} (x - m_k)^2\, dF_k(x) \leqslant \frac{1}{\varepsilon^\delta} \cdot \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \quad |\xi_k - m_k|^{2+\delta}.$$

Следовательно, *условие Ляпунова* обеспечивает выполнение *условия Линдеберга.*

b) Пусть $\xi_1, \xi_2, \ldots$ — независимые *одинаково распределенные* случайные величины с $m = \quad \xi_1$ и дисперсией $0 < \sigma^2 \equiv \quad \xi_1 < \infty$. 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 $\sigma^2 = \quad |\xi_1 - m|^2 < \infty$.

Таким образом, *условие Линдеберга* выполнено и, следовательно, теорема 3 из § 3 вытекает из доказанной теоремы 1.

с) Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины такие, что для всех $n \geqslant 1$

$$|\xi_n| \leqslant K < \infty,$$

где K — некоторая постоянная, и $D_n \to \infty$, $n \to \infty$.

Тогда из неравенства Чебышева

$$\int_{\{x : |x - m_k| \geqslant \varepsilon D_n\}} |x - m_k|^2\, dF_k(x) = \quad [(\xi_k - m_k)^2 I(|\xi_k - m_k| \geqslant \varepsilon D_n)] \leqslant$$
$$\leqslant (2K)^2 \quad \{|\xi_k - m_k| \geqslant \varepsilon D_n\} \leqslant (2K)^2 \frac{\sigma_k^2}{\varepsilon^2 D_n^2}$$

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. Пусть $T_n=\dfrac{S_n-\ \ S_n}{D_n}$ и $F_{T_n}(x)=\ \ (T_n\leqslant x)$. 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):

$$\sup_{x\in R}|F_{T_n}(x)-\Phi(x)|\to 0,\qquad n\to\infty. \tag{14}$$

В частности, отсюда следует, что

$$\{S_n\leqslant x\}-\Phi\Big(\frac{x-\ \ S_n}{D_n}\Big)\to 0,\qquad n\to\infty.$$

Это утверждение часто выражают словами, что при достаточно большом n величина S_n примерно нормально распределена со средним $\ \ S_n$ и дисперсией $D_n^2\equiv\ \ S_n$.

Замечание 2. 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 (14). В том случае, когда величины $\xi_1, \xi_2, \ldots$ *независимы*, *одинаково распределены* и $\ \ |\xi_1|^3<\infty$, ответ на этот вопрос дается *теоремой* (*неравенством*) *Берри—Эссеена* (§ 11):

$$\sup_{x}|F_{T_n}(x)-\Phi(x)|\leqslant C\,\frac{\ \ |\xi_1-\ \ \xi_1|^3}{\sigma^3\sqrt{n}}, \tag{15}$$

где C — универсальная константа, *точное* значение которой до сих пор неизвестно. (В [90, гл. 5, § 4.3] для этой константы приводятся такие неравенства: $\dfrac{1}{\sqrt{2\pi}}\leqslant C\leqslant 0{,}7655$.)

Доказательство (15) дается в § 11.

Замечание 3. Придадим *условию Линдеберга* несколько иную (и даже более компактную) форму, особенно удобную в случае «схемы серий».

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, $m_k=\ \ \xi_k$, $\sigma_k^2=\ \ \xi_k$, $D_n^2=\sum\limits_{k=1}^{n}\sigma_k^2>0$, $n\geqslant 1$, и $\xi_{nk}=\dfrac{\xi_k-m_k}{D_n}$. С учетом этих обозначений условие (1) принимает следующий вид:

$$\text{(L)} \qquad \sum_{k=1}^{n} \quad [\xi_{nk}^2 I(|\xi_{nk}| \geqslant \varepsilon)] \to 0, \quad n \to \infty. \tag{16}$$

Если $S_n = \xi_{n1} + \ldots + \xi_{nn}$, то $\quad S_n = 1$ и теореме 1 можно придать такую форму: *если выполнено условие* (16), *то*

$$S_n \xrightarrow{d} \mathscr{N}(0, 1).$$

В таком виде центральная предельная теорема справедлива и без предположения о том, что величины ξ_{nk} имеют специальную форму $\dfrac{\xi_k - m_k}{D_n}$. Именно, имеет место следующий результат, доказываемый буквально так же, как теорема 1.

Теорема 2. *Пусть при каждом* $n \geqslant 1$

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn}$$

— последовательность независимых случайных величин таких, что $\quad \xi_{nk} = 0$ *и* $\quad S_n = 1$, *где* $S_n = \xi_{n1} + \ldots + \xi_{nn}$.

Тогда выполнение условия Линдеберга (16) *достаточно для сходимости* $S_n \xrightarrow{d} \mathscr{N}(0, 1)$.

4. Поскольку

$$\max_{1 \leqslant k \leqslant n} \quad \xi_{nk}^2 \leqslant \varepsilon^2 + \sum_{k=1}^{n} \quad [\xi_{nk}^2 I(|\xi_{nk}| \geqslant \varepsilon)],$$

то ясно, что из *условия Линдеберга* (16) вытекает, что

$$\max_{1 \leqslant k \leqslant n} \quad \xi_{kn}^2 \to 0, \quad n \to \infty. \tag{17}$$

Примечательно, что при выполнении этого условия из справедливости центральной предельной теоремы автоматически следует выполнение *условия Линдеберга.*

Теорема 3. *Пусть при каждом* $n \geqslant 1$

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn}$$

— последовательность независимых случайных величин таких, что $\quad \xi_{nk} = 0$ *и* $\quad S_n = 1$, *где* $S_n = \xi_{n1} + \ldots + \xi_{nn}$. *Пусть выполнено условие* (17). *Тогда условие Линдеберга является необходимым и достаточным для справедливости центральной предельной теоремы*, $S_n \to \mathscr{N}(0, 1)$.

Достаточность следует из теоремы 2. Для доказательства необходимости нам понадобится следующая лемма (ср. с леммой 3 в § 3).

Лемма. *Пусть ξ — случайная величина с функцией распределения $F = F(x)$, $\quad \xi = 0, \quad \xi = \gamma > 0$. Тогда для каждого $a > 0$*

$$\int\limits_{|x| \geqslant 1/a} x^2 \, dF(x) \leqslant \frac{1}{a^2} \, [\mathrm{Re} \; f(\sqrt{6}a) - 1 + 3\gamma a^2], \tag{18}$$

где $f(t) = \; e^{it\xi}$ — характеристическая функция ξ.

Доказательство. Имеем

$$\begin{aligned} \mathrm{Re} \; f(t) - 1 + \frac{1}{2}\gamma t^2 &= \frac{1}{2}\gamma t^2 - \int\limits_{-\infty}^{\infty} [1 - \cos tx] \, dF(x) = \\ &= \frac{1}{2}\gamma t^2 - \int\limits_{|x|<1/a} [1 - \cos tx] \, dF(x) - \int\limits_{|x| \geqslant 1/a} [1 - \cos tx] \, dF(x) \geqslant \\ &\geqslant \frac{1}{2}\gamma t^2 - \frac{1}{2} t^2 \int\limits_{|x|<1/a} x^2 \, dF(x) - 2a^2 \int\limits_{|x| \geqslant 1/a} x^2 dF(x) = \\ &= \left(\frac{1}{2} t^2 - 2a^2\right) \int\limits_{|x| \geqslant 1/a} x^2 \, dF(x). \end{aligned}$$

Полагая $t = \sqrt{6}a$, получаем требуемое неравенство (18). □

Перейдем теперь к доказательству необходимости в теореме 3.

Пусть

$$\begin{gathered} F_{nk}(x) = \; \{\xi_{nk} \leqslant x\}, \qquad f_{nk}(t) = \; e^{it\xi_{nk}}, \\ \xi_{nk} = 0, \qquad \xi_{nk} = \gamma_{nk} > 0, \\ \sum_{k=1}^{n} \gamma_{nk} = 1, \qquad \max_{1 \leqslant k \leqslant n} \gamma_{nk} \to 0, \;\; n \to \infty. \end{gathered} \tag{19}$$

Пусть $\ln z$ обозначает *главное значение* логарифма комплексного числа z (т. е. $\ln z = \ln |z| + i \arg z$, $-\pi < \arg z \leqslant \pi$).

Тогда

$$\ln \prod_{k=1}^{n} f_{nk}(t) = \sum_{k=1}^{n} \ln f_{nk}(t) + 2\pi i m,$$

где $m = m(n, t)$ — некоторое целое число. Следовательно,

$$\mathrm{Re} \ln \prod_{k=1}^{n} f_{nk}(t) = \mathrm{Re} \sum_{k=1}^{n} \ln f_{nk}(t). \tag{20}$$

Поскольку

$$\prod_{k=1}^{n} f_{nk}(t) \to e^{-\frac{1}{2}t^2},$$

то

$$\left|\prod_{k=1}^{n} f_{nk}(t)\right| \to e^{-\frac{1}{2}t^2}.$$

Тем самым

$$\operatorname{Re}\ln\prod_{k=1}^{n} f_{nk}(t) = \operatorname{Re}\ln\left|\prod_{k=1}^{n} f_{nk}(t)\right| \to -\frac{1}{2}\,t^2. \tag{21}$$

При $|z|<1$

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \ldots, \tag{22}$$

и при $|z| \leqslant \frac{1}{2}$

$$|\ln(1+z) - z| \leqslant |z|^2. \tag{23}$$

В силу (19) при каждом фиксированном t, всех достаточно больших n и всех $k = 1, 2, \ldots, n$ имеем

$$|f_{nk}(t) - 1| \leqslant \frac{1}{2}\,\gamma_{nk}t^2 \leqslant \frac{1}{2}. \tag{24}$$

Поэтому из (23), (24) получаем

$$\left|\sum_{k=1}^{n}\{\ln[1+(f_{nk}(t)-1)] - (f_{nk}(t)-1)\}\right| \leqslant \sum_{k=1}^{n} |f_{nk}(t)-1|^2 \leqslant$$

$$\leqslant \frac{t^4}{4}\max_{1\leqslant k\leqslant n}\gamma_{nk}\cdot\sum_{k=1}^{n}\gamma_{nk} = \frac{t^4}{4}\max_{1\leqslant k\leqslant n}\gamma_{nk} \to 0, \quad n\to\infty,$$

и, следовательно,

$$\left|\operatorname{Re}\sum_{k=1}^{n}\ln f_{nk}(t) - \operatorname{Re}\sum_{k=1}^{n}(f_{nk}(t)-1)\right| \to 0, \quad n\to\infty. \tag{25}$$

Из (20), (21), (25) вытекает, что

$$\operatorname{Re}\sum_{k=1}^{n}(f_{nk}(t)-1) + \frac{1}{2}\,t^2 = \sum_{k=1}^{n}\left[\operatorname{Re} f_{nk}(t) - 1 + \frac{1}{2}\,t^2\gamma_{nk}\right] \to 0, \quad n\to\infty.$$

Полагая $t=\sqrt{6}a$, находим, что при каждом $a>0$

$$\sum_{k=1}^{n}[\operatorname{Re} f_{nk}(\sqrt{6}a) - 1 + 3a^2\gamma_{nk}] \to 0, \quad n\to\infty. \tag{26}$$

Наконец, из (18) с $a = 1/\varepsilon$ и (26) получаем

$$\sum_{k=1}^{n} [\xi_{nk} I(|\xi_{nk}| \geqslant \varepsilon)] = \sum_{k=1}^{n} \int\limits_{|x| \geqslant \varepsilon} x^2 \, dF_{nk}(x) \leqslant$$
$$\leqslant \varepsilon^2 \sum_{k=1}^{n} [\mathrm{Re}\, f_{nk}(\sqrt{6}a) - 1 + 3a^2 \gamma_{nk}] \to 0, \quad n \to \infty,$$

что и доказывает выполнение *условия Линдеберга.* □

5. Задачи.

1. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с $\xi_1^2 < \infty$. Показать, что

$$\max\left(\frac{|\xi_1|}{\sqrt{n}}, \ldots, \frac{|\xi_n|}{\sqrt{n}}\right) \xrightarrow{d} 0, \quad n \to \infty.$$

2. Дать прямое доказательство того, что в схеме Бернулли величина $\sup\limits_x |F_{T_n}(x) - \Phi(x)|$ имеет порядок $\dfrac{1}{\sqrt{n}}$, $n \to \infty$.

3. Пусть $X_1, X_2, \ldots$ — последовательность перестановочных случайных величин (см. задачу 4 к § 5 гл. II) с $X_i = 0$, $X_i^2 = 1$ и

$$(X_1, X_2) = (X_1^2, X_2^2). \tag{27}$$

Доказать, что имеет место центральная предельная теорема:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \mathscr{N}(0, 1). \tag{28}$$

Обратно, если $X_n^2 < \infty$ и выполнено (28), то выполнено и (27).

4. *Локальная центральная предельная теорема.* Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины с $X_1 = 0$, $X_1^2 = 1$. Пусть их характеристическая функция $\varphi(t) = e^{itX_1}$ такова, что

$$\int\limits_{-\infty}^{\infty} |\varphi(t)|^r \, dt < \infty \quad \text{для некоторого } r \geqslant 1.$$

Показать, что плотность $f_n(x)$ распределения вероятностей величин $S_n/\sqrt{n}$ существует и

$$f_n(x) \to (2n)^{-1/2} e^{-x^2/2}, \quad n \to \infty, \quad \text{равномерно по } x \in R.$$

Как выглядит соответствующий результат для решетчатых случайных величин?

5. Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины с $X_1 = 0$, $X_1^2 = 1$. Пусть $d_1^2, d_2^2, \ldots$ — неотрицательные

константы такие, что $d_n = o(D_n)$, где $D_n^2 = \sum\limits_{k=1}^{n} d_k^2$. Показать, что последовательность взвешенных величин $d_1 X_1,\ d_2 X_2,\ \ldots$ удовлетворяет центральной предельной теореме:

$$\frac{1}{D_n} \sum_{k=1}^{n} d_k X_k \xrightarrow{d} \mathscr{N}(0,\ 1).$$

6. Пусть $\xi_1,\ \xi_2,\ \ldots$ — независимые одинаково распределенные случайные величины с $\xi_1 = 0$, $\xi_1^2 = 1$. Предположим, что $(\tau_n)_{n\geqslant 1}$ — последовательность случайных величин, принимающих значения из множества $\{1,\ 2,\ \ldots\}$, такая, что $\tau_n/n \to c$, где $c > 0$ — константа. Доказать, что $(S_n = \xi_1 + \ldots + \xi_n)$

$$\mathrm{Law}(\tau_n^{-1/2} S_{\tau_n}) \to \Phi$$

(т. е. $\tau_n^{-1/2} S_{\tau_n} \xrightarrow{d} \xi$, где $\xi \sim \mathscr{N}(0,\ 1)$). (Отметим, что независимость последовательностей $(\tau_n)_{n\geqslant 1}$ и $(\xi_n)_{n\geqslant 1}$ не предполагается.)

7. Пусть $\xi_1,\ \xi_2,\ \ldots$ — независимые одинаково распределенные случайные величины с $\xi_1 = 0$, $\xi_1^2 = 1$. Доказать, что

$$\mathrm{Law}\Big(n^{-1/2} \max_{1\leqslant m\leqslant n} S_m\Big) \to \mathrm{Law}(|\xi|), \qquad \text{где } \xi \sim \mathscr{N}(0,\ 1).$$

Иначе говоря, для $x > 0$

$$\Big\{n^{-1/2} \max_{1\leqslant m\leqslant n} S_m \leqslant x\Big\} \to \sqrt{\frac{2}{\pi}} \int_0^x e^{-y^2/2}\, dy \quad \Big(= \frac{1}{\sqrt{2}}\ \mathrm{erf}(x)\Big).$$

Указание: убедиться в справедливости сформулированного утверждения сначала для симметричных *бернуллиевских* случайных величин $\xi_1,\ \xi_2,\ \ldots$ с $\{\xi_n = \pm 1\} = 1/2$ и затем доказать, что вид предельного распределения будет тем же самым для любой последовательности $\xi_1,\ \xi_2,\ \ldots$ с определенными выше свойствами. (Отмеченная *независимость предельного распределения* от частного выбора последовательности $\xi_1,\ \xi_2,\ \ldots$, состоящей из независимых одинаково распределенных случайных величин с $\xi_n = 0$, $\xi_n^2 = 1$, носит название «принцип инвариантности»; ср. с § 7.)

8. В условиях предыдущей задачи доказать, что

$$\Big\{n^{-1/2} \max_{1\leqslant m\leqslant n} |S_m| \leqslant x\Big\} \to H(x), \qquad x > 0,$$

где

$$H(x) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\Big\{-\frac{(2k+1)^2\pi^2}{8x^2}\Big\}.$$

9. Пусть X_1, X_2, … — последовательность независимых случайных величин с

$$\{X_n = \pm n_\alpha\} = \frac{1}{2n^\beta}, \qquad \{X_n = 0\} = 1 - \frac{1}{n^\beta}, \quad \text{где } 2\alpha > \beta - 1.$$

Показать, что условие Линдеберга выполнено, если и только если $0 \leqslant \beta < 1$.

10. Пусть X_1, X_2, … — последовательность независимых случайных величин таких, что $|X_n| \leqslant C_n$ (-п. н.) и $C_n = o(D_n)$, где

$$D_n^2 = \sum_{k=1}^{n} \quad (X_k - \quad X_k)^2 \to \infty.$$

Показать, что

$$\frac{S_n - \quad S_n}{D_n} \to \mathscr{N}(0,\ 1), \quad \text{где } S_n = X_1 + \ldots + X_n.$$

11. Пусть X_1, X_2, … — последовательность независимых случайных величин с $\quad X_n = 0$, $\quad X_n^2 = \sigma_n^2$. Предположим, что для них выполняется центральная предельная теорема и

$$\left(D_n^{-1/2} \sum_{i=1}^{n} X_i\right)^k \to \frac{(2k)!}{2^k k!} \quad \text{для некоторого } k \geqslant 1.$$

Показать, что тогда выполнено условие Линдеберга порядка k, т. е.

$$\sum_{j=1}^{n} \int\limits_{\{|x|>\varepsilon\}} |x|^k \, dF_j(x) = o(D_n^k), \quad \varepsilon > 0.$$

(Обычное условие Линдеберга соответствует случаю $k = 2$; см. (1).)

12. Пусть $X = X(\lambda)$ и $Y = Y(\mu)$ — независимые пуассоновские случайные величины с параметрами $\lambda > 0$ и $\mu > 0$ соответственно. Показать, что

$$\frac{(X(\lambda) - \lambda) - (Y(\mu) - \mu)}{\sqrt{X(\lambda) + Y(\mu)}} \to \mathscr{N}(0,\ 1) \quad \text{при } \lambda \to \infty,\ \mu \to \infty.$$

13. Пусть при каждом $n \geqslant 1$ $(X_1^{(n)}, \ldots, X_{n+1}^{(n)})$ является $(n+1)$-мерным случайным вектором, равномерно распределенным на *единичной* сфере. Доказать справедливость следующей «теоремы Пуанкаре»:

$$\lim_{n\to\infty} \quad \{\sqrt{n}\, X_{n+1}^{(n)} \leqslant x\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} \, du.$$

§ 5. Центральная предельная теорема для сумм независимых случайных величин. II. Неклассические условия

1. В § 4 было показано, что *условие Линдеберга* (16) влечет выполнение условия

$$\max_{1\leqslant k\leqslant n} \quad \xi_{nk}^2 \to 0,$$

из которого в свою очередь вытекает так называемое условие *предельной пренебрегаемости* (*асимптотической малости*), состоящее в том, что для всякого $\varepsilon > 0$

$$\max_{1\leqslant k\leqslant n} \quad \{|\xi_{nk}| \geqslant \varepsilon\} \to 0, \qquad n \to \infty.$$

Таким образом, можно сказать, что теоремы 1 и 2 из § 4 дают условия выполнимости центральной предельной теоремы для сумм независимых случайных величин в предположении *предельной пренебрегаемости*. Предельные теоремы, в которых на отдельные слагаемые наложены условия их предельной пренебрегаемости, принято называть теоремами в «классической постановке». Нетрудно однако привести примеры невырожденных случайных величии, для которых не выполнено ни *условие Линдеберга*, ни *условие предельной пренебрегаемости*, но тем не менее центральная предельная теорема справедлива. Вот простейший пример.

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых нормально распределенных случайных величин с $\quad \xi_n = 0, \quad \xi_1 = 1, \quad \xi_k = 2^{k-2}$, $k \geqslant 2$. Положим $S_n = \xi_{n1} + \ldots + \xi_{nn}$ с

$$\xi_{nk} = \frac{\xi_k}{\sqrt{\sum\limits_{i=1}^{n} \quad \xi_i}}.$$

Нетрудно проверить, что здесь не выполнено ни *условие Линдеберга*, ни *условие предельной пренебрегаемости*, хотя справедливость центральной предельной теоремы очевидна, поскольку S_n распределены нормально с $\quad S_n = 0, \quad S_n = 1$.

Приводимая далее теорема 1 дает достаточное (и необходимое) условие справедливости центральной предельной теоремы без предположения «классического» условия предельной пренебрегаемости. В этом смысле формулируемое ниже условие (Λ) является примером «неклассических» условий, что и отражено в заголовке настоящего параграфа.

2. Будем предполагать, что для каждого $n \geqslant 1$ задана последовательность («схема серий») независимых случайных величин

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn}$$

с $\xi_{nk} = 0$, $\xi_{nk} = \sigma^2_{nk} > 0$, $\sum\limits_{k=1}^{n} \sigma^2_{nk} = 1$. Пусть $S_n = \xi_{n1} + \ldots + \xi_{nn}$, $F_{nk}(x) =$ $= \{\xi_{nk} \leqslant x\}$, $\Phi(x) = (2\pi)^{-1/2} \int\limits_{-\infty}^{x} e^{-y^2/2}\, dy$, $\Phi_{nk}(x) = \Phi\Big(\dfrac{x}{\sigma_{nk}}\Big)$.

Теорема 1. *Для того чтобы*

$$S_n \xrightarrow{d} \mathcal{N}(0,\ 1), \tag{1}$$

достаточно (*и необходимо*) *выполнение для каждого* $\varepsilon > 0$ *условия*

$$(\Lambda) \qquad \sum_{k=1}^{n} \int\limits_{|x|>\varepsilon} |x|\, |F_{nk}(x) - \Phi_{nk}(x)|\, dx \to 0, \quad n \to \infty. \tag{2}$$

Следующая теорема проясняет связь между условием (Λ) и классическим *условием Линдеберга*

$$(\mathrm{L}) \qquad \sum_{k=1}^{n} \int\limits_{|x|>\varepsilon} x^2\, dF_{nk}(x) \to 0, \quad n \to \infty. \tag{3}$$

Теорема 2. 1. *Условие Линдеберга обеспечивает выполнение условия* (Λ):

$$(\mathrm{L}) \ \Rightarrow\ (\Lambda).$$

2. *Если* $\max\limits_{1\leqslant k\leqslant n} \xi^2_{nk} \to 0$, $n \to \infty$, *то условие* (Λ) *обеспечивает выполнение условия Линдеберга* (L):

$$(\Lambda) \ \Rightarrow\ (\mathrm{L}).$$

Доказательство теоремы 1. Доказательство необходимости условия (Λ) довольно сложно ([88], [91], [96]). Приведем здесь лишь доказательство его достаточности.

Пусть

$$f_{nk}(t) = e^{it\xi_{nk}}, \qquad f_n(t) = e^{itS_n},$$

$$\varphi_{nk}(t) = \int\limits_{-\infty}^{\infty} e^{itx}\, d\Phi_{nk}(x), \qquad \varphi(t) = \int\limits_{-\infty}^{\infty} e^{itx}\, d\Phi(x).$$

Из § 12 гл. II следует, что

$$\varphi_{nk}(t) = e^{-\frac{t^2\sigma^2_{nk}}{2}}, \qquad \varphi(t) = e^{-\frac{t^2}{2}}.$$

Согласно следствию к теореме 1 из § 3, $S_n \xrightarrow{d} \mathcal{N}(0,\ 1)$ тогда и только тогда, когда $f_n(t) \to \varphi(t)$, $n \to \infty$, для всякого действительного t.

Имеем
$$f_n(t)-\varphi(t)=\prod_{k=1}^{n} f_{nk}(t)-\prod_{k=1}^{n}\varphi_{nk}(t).$$

Поскольку $|f_{nk}(t)|\leqslant 1$, $|\varphi_{nk}(t)|\leqslant 1$, то
$$|f_n(t)-\varphi(t)|=\left|\prod_{k=1}^{n} f_{nk}(t)-\prod_{k=1}^{n}\varphi_{nk}(t)\right|\leqslant\sum_{k=1}^{n}|f_{nk}(t)-\varphi_{nk}(t)|=$$
$$=\sum_{k=1}^{n}\left|\int_{-\infty}^{\infty} e^{itx}\,d(F_{nk}-\Phi_{nk})\right|=\sum_{k=1}^{n}\left|\int_{-\infty}^{\infty}\left(e^{itx}-itx+\frac{t^2x^2}{2}\right)d(F_{nk}-\Phi_{nk})\right|, \quad (4)$$
где мы воспользовались тем, что для $k=1,\,2$
$$\int_{-\infty}^{\infty} x^k\,dF_{nk}=\int_{-\infty}^{\infty} x^k\,d\Phi_{nk}.$$

Применяя формулу интегрирования по частям (теорема 11 в § 6 гл. II) к интегралам
$$\int_a^b\left(e^{itx}-itx+\frac{t^2x^2}{2}\right)d(F_{nk}-\Phi_{nk}),$$
получаем (с учетом того, что $x^2[1-F_{nk}(x)+F_{nk}(-x)]\to 0$, $x^2[1-\Phi_{nk}(x)+\Phi_{nk}(-x)]\to 0$, $x\to\infty$)
$$\int_{-\infty}^{\infty}\left(e^{itx}-itx+\frac{t^2x^2}{2}\right)d(F_{nk}-\Phi_{nk})=$$
$$=-it\int_{-\infty}^{\infty}(e^{itx}-1-itx)(F_{nk}(x)-\Phi_{nk}(x))\,dx. \quad (5)$$

Из (4) и (5) имеем
$$|f_n(t)-\varphi(t)|\leqslant\sum_{k=1}^{n}\left|t\int_{-\infty}^{\infty}(e^{itx}-1-itx)(F_{nk}(x)-\Phi_{nk}(x))\,dx\right|\leqslant$$
$$\leqslant\frac{|t|^3}{2}\,\varepsilon\sum_{k=1}^{n}\int_{|x|\leqslant\varepsilon}|x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx+$$
$$+2t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}|x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx\leqslant$$
$$\leqslant\varepsilon|t|^3\sum_{k=1}^{n}\sigma_{nk}^2+2t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}|x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx, \quad (6)$$

при этом мы воспользовались неравенством

$$\int_{|x|\leqslant\varepsilon} |x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx \leqslant 2\sigma_{nk}^2, \tag{7}$$

справедливость которого легко установить, опираясь на формулу (71) из § 6 гл. II.

Из (6) в силу произвольности $\varepsilon>0$ и условия (Λ) следует, что $f_n(t)\to\varphi(t)$, $n\to\infty$. □

Доказательство теоремы 2. 1. Согласно § 4, *условие Линдеберга* (L) влечет условие $\max\limits_{1\leqslant k\leqslant n}\sigma_{nk}^2\to 0$. Поэтому, учитывая, что $\sum\limits_{k=1}^n \sigma_{nk}^2=1$, получаем

$$\sum_{k=1}^n \int_{|x|>\varepsilon} x^2\,d\Phi_{nk}(x) \leqslant \int_{|x|>\varepsilon/\sqrt{\max\limits_{1\leqslant k\leqslant n}\sigma_{nk}^2}} x^2\,d\Phi(x)\to 0, \qquad n\to\infty. \tag{8}$$

Вместе с условием (L) это дает, что для всякого $\varepsilon>0$

$$\sum_{k=1}^n \int_{|x|>\varepsilon} x^2\,d\,[F_{nk}(x)+\Phi_{nk}(x)]\to 0, \qquad n\to\infty. \tag{9}$$

Зафиксируем $\varepsilon>0$. Тогда найдется такая непрерывно дифференцируемая четная функция $h=h(x)$, что $|h(x)|\leqslant x^2$, $|h'(x)|\leqslant 4|x|$,

$$h(x)=\begin{cases} x^2, & |x|>2\varepsilon,\\ 0, & |x|\leqslant\varepsilon.\end{cases}$$

Для такой функции $h(x)$ в силу (9)

$$\sum_{k=1}^n \int_{|x|>\varepsilon} h(x)\,d[F_{nk}(x)+\Phi_{nk}(x)]\to 0, \qquad n\to\infty. \tag{10}$$

С помощью интегрирования по частям из (10) находим:

$$\sum_{k=1}^n \int_{x\geqslant\varepsilon} h'(x)[(1-F_{nk}(x))+(1-\Phi_{nk}(x))]\,dx = \sum_{k=1}^n \int_{x\geqslant\varepsilon} h(x)\,d[F_{nk}+\Phi_{nk}]\to 0,$$

$$\sum_{k=1}^n \int_{x\leqslant-\varepsilon} h'(x)[F_{nk}(x)+\Phi_{nk}(x)]\,dx = \sum_{k=1}^n \int_{x\leqslant-\varepsilon} h(x)\,d[F_{nk}+\Phi_{nk}]\to 0.$$

Поскольку $h'(x)=2x$ при $|x|\geqslant 2\varepsilon$, то

$$\sum_{k=1}^n \int_{|x|\geqslant 2\varepsilon} |x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx\to 0, \qquad n\to\infty.$$

Таким образом, в силу произвольности $\varepsilon > 0$, (L) $\Rightarrow$ (Λ).

2. В силу условия $\max\limits_{1 \leqslant k \leqslant n} \sigma_{nk}^2 \to 0$ и (8) для введенной выше функции $h = h(x)$ получаем, что

$$\sum_{k=1}^{n} \int\limits_{|x|>\varepsilon} h(x)\, d\Phi_{nk}(x) \leqslant \sum_{k=1}^{n} \int\limits_{|x|>\varepsilon} x^2\, d\Phi_{nk}(x) \to 0, \quad n \to \infty. \tag{11}$$

Далее, с учетом интегрирования по частям, находим:

$$\begin{aligned}
&\left|\sum_{k=1}^{n} \int\limits_{|x|\geqslant\varepsilon} h(x)\, d[F_{nk} - \Phi_{nk}]\right| \leqslant \\
&\leqslant \left|\sum_{k=1}^{n} \int\limits_{x\geqslant\varepsilon} h(x)\, d[(1 - F_{nk}) - (1 - \Phi_{nk})]\right| + \left|\sum_{k=1}^{n} \int\limits_{x\leqslant-\varepsilon} h(x)\, d[F_{nk} - \Phi_{nk}]\right| \leqslant \\
&\leqslant \sum_{k=1}^{n} \int\limits_{x\geqslant\varepsilon} |h'(x)|\, |(1 - F_{nk}) - (1 - \Phi_{nk})|\, dx + \sum_{k=1}^{n} \int\limits_{x\leqslant-\varepsilon} |h'(x)|\, |F_{nk} - \Phi_{nk}|\, dx \leqslant \\
&\qquad\qquad \leqslant 4 \sum_{k=1}^{n} \int\limits_{|x|\geqslant\varepsilon} |x|\, |F_{nk}(x) - \Phi_{nk}(x)|\, dx.
\end{aligned} \tag{12}$$

Из (11) и (12) следует, что

$$\sum_{k=1}^{n} \int\limits_{|x|\geqslant 2\varepsilon} x^2\, dF_{nk}(x) \leqslant \sum_{k=1}^{n} \int\limits_{|x|\geqslant\varepsilon} h(x)\, dF_{nk}(x) \to 0, \quad n \to \infty,$$

т. е. выполнено *условие Линдеберга* (L). □

3. Задачи.

1. Доказать справедливость формулы (5).

2. Проверить справедливость соотношений (10), (12).

3. Пусть $N = (N_t)_{t\geqslant 0}$ — процесс восстановления, введенный в п. 4 § 9 гл. II ($N_t = \sum\limits_{n=1}^{\infty} I(T_n \leqslant t)$, $T_n = \sigma_1 + \ldots + \sigma_n$, где σ_1, σ_2, … — последовательность независимых одинаково распределенных положительных случайных величин). Предполагая, что $\mu = \ \sigma_1 < \infty$, $0 < \ \sigma_1 < \infty$, доказать справедливость центральной предельной теоремы:

$$\frac{N_t - t\mu^{-1}}{\sqrt{t\mu^{-3}\ \ \sigma_1}} \xrightarrow{d} \mathcal{N}(0,\ 1),$$

где $\mathcal{N}(0, 1)$ — стандартно распределенная нормальная случайная величина с нулевым средним и единичной дисперсией.

§ 6. 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

$$\xi_{n,k} = \frac{\xi_k - \quad \xi_k}{D_n}, \qquad D_n^2 = \quad S_n,$$

то тогда

$$T_n = \sum_{k=1}^{n} \xi_{n,k} = \frac{S_n - \quad S_n}{D_n}.$$

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 в таком-то смысле также суммой независимых одинаково распределенных случайных величин. Имея это в виду, введем такое определение.

Определение 1. Случайная величина T (а также ее функция распределения F_T и ее характеристическая функция φ_T) называется *безгранично делимой*, если для каждого $n \geqslant 1$ на вероятностном пространстве $(\Omega, \mathscr{F},\ \ )$ 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 $\varphi_T = (\varphi_{\eta_1})^n$).

Замечание 1. Если исходное вероятностное пространство, на котором задана случайная величина T, достаточно «бедное», то может случиться, что функция распределения F_T и ее характеристическая функция φ_T *допускают* при любом $n \geqslant 1$ представления $F_T = F^{(n)} * \ldots * F^{(n)}$ (n раз) и $\varphi_T = {\varphi^{(n)}}^n$ с некоторыми функциями распределения $F^{(n)}$ и их характеристическими функциями $\varphi^{(n)}$, хотя, в то же самое время, представление $T \overset{d}{=} \eta_1 + \ldots + \eta_n$ *невозможно*. Дж. Дубу принадлежит как раз пример (см. [103]) «бедного» вероятностного пространства, на котором определена случайная величина T, имеющая распределение Пуассона с параметром $\lambda = 1$ (которое является безгранично делимым: $F_T = F^{(n)} * \ldots * F^{(n)}$ с функциями распределения $F^{(n)}$, отвечающими пуассоновскому распределению с параметром $\lambda = 1/n$), но отсутствуют случайные величины η_1 и η_2, имеющие распределение Пуассона с параметром $\lambda = 1/2$.

Имея в виду сказанное, подчеркнем, что данное выше определение 1, в сущности, неявно предполагает, что исходное вероятностное пространство $(\Omega, \mathscr{F},\ \ )$ уже достаточно «богато», настолько, чтобы избежать эффектов, отмеченных Дж. Дубом (задача 11).

Теорема 1. *Случайная величина T может быть пределом по распределению сумм $T_n = \sum_{i=1}^{n} \xi_{n,i}$ в том и только том случае, когда T безгранично делима.*

Доказательство. Если T безгранично делима, то для каждого $n \geqslant 1$ существуют независимые одинаково распределенные случайные величины $\xi_{n,1}, \ldots, \xi_{n,n}$ такие, что $T \overset{d}{=} \xi_{n,1} + \ldots + \xi_{n,k}$, а это и означает, что $T \overset{d}{=} T_n$, $n \geqslant 1$.

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 $T_{nk} = \sum_{i=1}^{nk} \xi_{nk,i}$ в виде $\zeta_n^{(1)} + \ldots + \zeta_n^{(k)}$, где

$$\zeta_n^{(1)} = \xi_{nk,1} + \ldots + \xi_{nk,n}, \quad \ldots, \quad \zeta_n^{(k)} = \xi_{nk,n(k-1)+1} + \ldots + \xi_{nk,nk}.$$

---

*) Запись $\xi \overset{d}{=} \eta$ означает, что случайные величины ξ и η совпадают (конгруэнтны) *по распределению*, т. е. $F_\xi(x) = F_\eta(x)$, $x \in R$, где $F_\xi(x)$ и $F_\eta(x)$ — функции распределения ξ и η.

Поскольку $T_{nk} \xrightarrow{d} T$, $n \to \infty$, то последовательность функций распределения, соответствующих случайным величинам T_{nk}, $n \geqslant 1$, относительно компактна и, значит, по теореме Прохорова *плотна* (§ 2). Далее,

$$[\quad\{\zeta_n^{(1)} > z\}]^k = \quad\{\zeta_n^{(1)} > z, \ldots, \zeta_n^{(k)} > z\} \leqslant \quad\{T_{nk} > kz\}$$

и

$$[\quad\{\zeta_n^{(1)} < -z\}]^k = \quad\{\zeta_n^{(1)} < -z, \ldots, \zeta_n^{(k)} < -z\} \leqslant \quad\{T_{nk} < -kz\}.$$

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 найдутся подпоследовательность $\{n_i\} \subset \{n\}$ и случайный вектор $(\eta_1, \ldots, \eta_k)$, который без ограничения общности можно считать определенным на исходном («богатом») вероятностном пространстве, такие, что

$$(\zeta_{n_j}^{(1)}, \ldots, \zeta_{n_j}^{(k)}) \xrightarrow{d} (\eta_1, \ldots, \eta_k),$$

или эквивалентно, что для любых $\lambda_1, \ldots, \lambda_k \in R$

$$\quad e^{i(\lambda_1 \zeta_{n_j}^{(1)} + \ldots + \lambda_k \zeta_{n_j}^{(k)})} \to \quad e^{i(\lambda_1 \eta_1 + \ldots + \lambda_k \eta_k)}.$$

В силу независимости величин $\zeta_{n_j}^{(1)}, \ldots, \zeta_{n_j}^{(k)}$

$$\quad e^{i(\lambda_1 \zeta_{n_j}^{(1)} + \ldots + \lambda_k \zeta_{n_j}^{(k)})} = \quad e^{i\lambda_1 \zeta_{n_j}^{(1)}} \ldots \quad e^{i\lambda_k \zeta_{n_j}^{(k)}} \to \quad e^{i\lambda_1 \eta_1} \ldots \quad e^{i\lambda_k \eta_k}.$$

Значит,

$$\quad e^{i(\lambda_1 \eta_1 + \ldots + \lambda_k \eta_k)} = \quad e^{i\lambda_1 \eta_1} \ldots \quad e^{i\lambda_k \eta_k}$$

и в силу теоремы 4 из § 12 гл. II величины $\eta_1, \ldots, \eta_k$ независимы. Ясно также, что они имеют одно и то же распределение.

Далее,

$$T_{n_j k} = \zeta_{n_j}^{(1)} + \ldots + \zeta_{n_j}^{(k)} \xrightarrow{d} \eta_1 + \ldots + \eta_k$$

и к тому же $T_{n_j k} \xrightarrow{d} T$. Поэтому (задача 1)

$$T \overset{d}{=} \eta_1 + \ldots + \eta_k. \qquad \square$$

Замечание 2. Утверждение теоремы остается в силе, если рассмотренное в начале параграфа условие, что при каждом $n \geqslant 1$ величины $\xi_{n,1}, \ldots, \xi_{n,n}$ *одинаково распределены*, заменить на условие их *асимптотической малости* $\max\limits_{1 \leqslant k \leqslant n} \quad\{|\xi_{nk}| \geqslant \varepsilon\} \to 0$.

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 Γ-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 (табл. 5, § 12 гл. II) 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 Колмогорова—Леви—Хинчина)**.** *Случайная величина T является безгранично делимой тогда и только тогда, когда ее характеристическая функция $\varphi(t)$ имеет вид $\varphi(t)=\exp\psi(t)$ 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 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 $0 < \sigma^2 \equiv \quad \xi_1 < \infty$, то, полагая $b_n = n \quad \xi_1$ 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, $\varphi_{S_n/n}(t) = (e^{-\frac{\theta}{n}|t|})^n = e^{-\theta|t|}$, т. е. величина S_n/n также имеет 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

$$\frac{S_n - b_n}{a_n} = \sum_{k=1}^{n} \xi_{n,k} \ (= T_n).$$

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 (с учетом замечания 1) 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

$$a_n T + b_n \overset{d}{=} \xi_1 + \ldots + \xi_n, \tag{4}$$

или, что то же самое, $F\left(\dfrac{x - b_n}{a_n}\right) = F * \ldots * F(x)$, или

$$[\varphi(t)]^n = [\varphi(a_n t)] e^{ib_n t}. \tag{5}$$

Теорема 3. *Случайная величина T может быть пределом по распределению случайных величин $\dfrac{S_n - b_n}{a_n}$, $a_n > 0$, тогда и только тогда, когда T является устойчивой.*

Доказательство. Если T устойчива, то, согласно (4),

$$T \overset{d}{=} \frac{S_n - b_n}{a_n},$$

где $S_n = \xi_1 + \ldots + \xi_n$, и, следовательно, $\dfrac{S_n - b_n}{a_n} \xrightarrow{d} T$.

Обратно, пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, $S_n = \xi_1 + \ldots + \xi_n$ и $\dfrac{S_n - b_n}{a_n} \xrightarrow{d} T$, $a_n > 0$. 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

$$S_n^{(1)} = \xi_1 + \ldots + \xi_n, \quad \ldots, \quad S_n^{(k)} = \xi_{(k-1)n+1} + \ldots + \xi_{kn},$$
$$T_n^{(1)} = \frac{S_n^{(1)} - b_n}{a_n}, \quad \ldots, \quad T_n^{(k)} = \frac{S_n^{(k)} - b_n}{a_n}.$$

Ясно, что по распределению все величины $T_n^{(1)}, \ldots, T_n^{(k)}$ совпадают и

$$T_n^{(i)} \xrightarrow{d} T, \quad n \to \infty, \ \ i = 1, \ldots, k.$$

Обозначим

$$U_n^{(k)} = T_n^{(1)} + \ldots + T_n^{(k)}.$$

Тогда так же, как и при доказательстве теоремы 1, получаем, что

$$U_n^{(k)} \xrightarrow{d} T^{(1)} + \ldots + T^{(k)},$$

где $T^{(i)}$, $1 \leqslant i \leqslant k$, независимы и $T^{(1)} \overset{d}{=} \ldots \overset{d}{=} T^{(k)} \overset{d}{=} T$.

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

$$\alpha_n^{(k)} = \frac{a_{kn}}{a_n}, \qquad \beta_n^{(k)} = \frac{b_{kn} - kb_n}{a_n}$$

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 $\alpha_n^{(k)}\to\alpha^{(k)}$, $\beta_n^{(k)}\to\beta^{(k)}$, $n\to\infty$, и

$$T\overset{d}{=}\frac{T^{(1)}+\ldots+T^{(k)}-\beta^{(k)}}{\alpha^{(k)}},$$

что и доказывает устойчивость случайной величины T. □

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 $e^{itb_n}\varphi_n(a_nt)$, и, согласно следствию к теореме 1 и задаче 3 из § 3,

$$e^{itb_n}\varphi_n(a_ny)\to\tilde{\varphi}(t), \tag{7}$$

$$\varphi_n(t)\to\varphi(t) \tag{8}$$

равномерно на каждом конечном интервале изменения t.

Пусть $\{n_i\}$ — подпоследовательность $\{n\}$ такая, что $a_{n_i}\to a$. Покажем прежде всего, что $a<\infty$. Предположим, что $a=\infty$. В силу (7) для любого $c>0$

$$\sup_{|t|\leqslant c}\big||\varphi_n(a_nt)|-|\tilde{\varphi}(t)|\big|\to 0,\qquad n\to\infty.$$

Возьмем вместо t величину $t_{n_i}=\dfrac{t_0}{a_{n_i}}$. Тогда поскольку $a_{n_i}\to\infty$, то

$$\left|\varphi_{n_i}\Big(a_{n_i}\frac{t_0}{a_{n_i}}\Big)\right|-\left|\tilde{\varphi}\Big(\frac{t_0}{a_{n_i}}\Big)\right|\to 0$$

и, значит,

$$|\varphi_{n_i}(t_0)|\to|\tilde{\varphi}(0)|=1.$$

Но $|\varphi_{n_i}(t_0)| \to |\varphi(t_0)|$. Потому $|\varphi(t_0)| = 1$ для любого $t_0 \in R$, и, следовательно, согласно теореме 5 из § 12 гл. II, случайная величина 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)

$$|\varphi_{n_i}(a_{n_i} t)| \to |\varphi(at)|, \quad |\varphi_{n_i}(a_{n_i} t)| \to |\tilde{\varphi}(t)|$$

и

$$|\varphi_{n_i'}(a_{n_i'} t)| \to |\varphi(a' t)|, \quad |\varphi_{n_i'}((a_{n_i'} t)| \to |\tilde{\varphi}(t)|.$$

Следовательно,

$$|\varphi(at)| = |\varphi(a' t)|,$$

и, значит, для любого $t \in R$

$$|\varphi(t)| = \left|\varphi\left(\frac{a'}{a} t\right)\right| = \ldots = \left|\varphi\left(\left(\frac{a'}{a}\right)^n t\right)\right| \to 1, \quad n \to \infty.$$

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 $\lim e^{itb_n}$. Отсюда можно вывести (задача 9), что тогда $\overline{\lim}\, |b_n| < \infty$.

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 $\tilde{\xi}$ невырождена, то $a > 0$. □

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 $\varphi(t)=\exp\psi(t)$ с*

$$\psi(t)=it\beta-d|t|^{\alpha}\Big(1+i\theta\frac{t}{|t|}G(t,\alpha)\Big), \tag{9}$$

где $0<\alpha<2$, $\beta\in R$, $d\geqslant 0$, $|\theta|\leqslant 1$, $\dfrac{t}{|t|}=0$ при $t=0$ и

$$G(t,\alpha)=\begin{cases}\operatorname{tg}\dfrac{\pi}{2}\alpha, & \text{если } \alpha\neq 1,\\ \dfrac{2}{\pi}\ln|t|, & \text{если } \alpha=1.\end{cases} \tag{10}$$

Отметим, что особо просто устроены характеристические функции *симметричных* устойчивых распределений:

$$\varphi(t)=e^{-d|t|^{\alpha}}, \tag{11}$$

где $0<\alpha\leqslant 2$, $d\geqslant 0$.

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.

3. Пусть φ_n — безгранично делимые характеристические функции и $\varphi_n(t)\to\varphi(t)$ для каждого $t\in R$, где $\varphi(t)$ — некоторая характеристическая функция. Показать, что $\varphi(t)$ безгранично делима.

4. Показать, что характеристическая функция безгранично делимого распределения не обращается в нуль.

5. Привести пример случайной величины, являющейся безгранично делимой, но не устойчивой.

6. Показать, что для устойчивой случайной величины ξ математическое ожидание $|\xi|^r<\infty$ для всех $r\in(0,\alpha)$, $0<\alpha<2$.

7. Показать, что если ξ — устойчивая случайная величина с параметром $0<\alpha\leqslant 1$, то $\varphi(t)$ не дифференцируема при $t=0$.

8. Дать прямое доказательство того, что функция $e^{-d|t|^{\alpha}}$ с $d\geqslant 0$, $0<\alpha\leqslant 2$ является характеристической.

9. Пусть $(b_n)_{n\geqslant 1}$ — числовая последовательность такая, что для всех $|t|<\delta$, $\delta>0$, существует $\lim\limits_n e^{itb_n}$. Показать, что тогда $\overline{\lim\limits_n}\,|b_n|<\infty$.

10. Показать, что биномиальное и равномерное распределения *не являются* безгранично делимыми.

11. Пусть функция распределения F и ее характеристическая функция φ допускают представления $F=F^{(n)}*\ldots*F^{(n)}$ (n раз), $\varphi=[\varphi^{(n)}]^n$

с некоторыми функциями распределения $F^{(n)}$ и их характеристическими функциями $\varphi^{(n)}$, $n \geqslant 1$. Показать, что можно найти (достаточно «богатое») вероятностное пространство $(\Omega, \mathscr{F}, \ )$ и определенные на нем случайные величины T и $(\eta_k^n)_{k \leqslant n}$, $n \geqslant 1$ (T имеет распределение F, $\eta_1^{(n)}, \ldots, \eta_n^{(n)}$ независимы и одинаково распределены с распределением $F^{(n)}$) , такие, что $T \overset{d}{=} \eta_1^{(n)} + \ldots + \eta_n^{(n)}$, $n \geqslant 1$.

12. Привести пример случайной величины, не являющейся безгранично делимой, характеристическая функция которой, тем не менее, в нуль не обращается.

§ 7. «Метризуемость» слабой сходимости

1. Пусть $(E, \mathscr{E}, \rho)$ — метрическое пространство и $\mathscr{P}(E) = \{P\}$ — семейство вероятностных мер на $(E, \mathscr{E})$. Естественно поставить вопрос о том, нельзя ли «метризовать» рассмотренную в § 1 слабую сходимость $P_n \xrightarrow{w} P$, т. е. нельзя ли ввести такое расстояние $\delta(P, \widetilde{P})$ между любыми двумя мерами P и $\widetilde{P}$ из $\mathscr{P}(E)$, чтобы сходимость $\delta(P_n, P) \to 0$ была равносильна сходимости $P_n \xrightarrow{w} P$.

В связи с такой постановкой вопроса полезно отметить, что *сходимость* случайных величин *по вероятности*, $\xi_n \to \xi$, *может быть метризована* с помощью, например, расстояния

$$d\ \ (\xi, \eta) = \inf\{\varepsilon > 0 \colon \quad \{|\xi - \eta| \geqslant \varepsilon\} \leqslant \varepsilon\}$$

или расстояний $d(\xi, \eta) = \quad \min(1, |\xi - \eta|)$, $d(\xi, \eta) = \quad \dfrac{|\xi - \eta|}{1 + |\xi - \eta|}$. (Более общим образом, можно положить $d(\xi, \eta) = \quad g(|\xi - \eta|)$, где в качестве функции $g = g(x)$, $x \geqslant 0$, можно взять любую борелевскую неотрицательную возрастающую функцию, непрерывную в нуле и такую, что $g(x + y) \leqslant g(x) + g(y)$ для всех $x \geqslant 0$, $y \geqslant 0$, $g(0) = 0$, $g(x) > 0$ для $x > 0$.) Но в то же самое время в пространстве всех случайных величин на $(\Omega, \mathscr{F}, \ )$ *не существует* расстояния $d(\xi, \eta)$ такого, что $d(\xi_n, \xi) \to 0$ тогда и только тогда, когда ξ_n сходится к ξ *с вероятностью единица*. (В этом легко убедиться, взяв последовательность случайных величин ξ_n, $n \geqslant 1$, сходящихся по вероятности к ξ, но не сходящихся с вероятностью единица.) Иначе говоря, *сходимость с вероятностью единица не метризуема*. (См. утверждения задач 11 и 12 к § 10 гл. II.)

Цель настоящего параграфа — установить, конкретно указав метрики ($L(P, \widetilde{P})$ и $\|P - \widetilde{P}\|_{BL}^*$), в пространстве мер $\mathscr{P}(E)$, *метризуемость* слабой сходимости:

$$P_n \xrightarrow{w} P \ \Leftrightarrow \ L(P_n, P) \to 0 \ \Leftrightarrow \ \|P_n - P\|_{BL}^* \to 0. \tag{1}$$

2. Метрика Леви—Прохорова $L(P, \widetilde{P})$**.** Пусть

$$\rho(x, A) = \inf\{\rho(x, y)\colon\ y \in A\},$$
$$A^{\varepsilon} = \{x \in E\colon\ \rho(x, A) < \varepsilon\}, \quad A \in \mathscr{E}.$$

Для любых двух мер $P, \widetilde{P} \in \mathscr{P}(\mathscr{E})$ положим

$$\sigma(P, \widetilde{P}) = \inf\{\varepsilon > 0\colon\ P(F) \leqslant \widetilde{P}(F^{\varepsilon}) + \varepsilon \text{ для всех замкнутых } F \in \mathscr{E}\} \tag{2}$$

и

$$L(P, \widetilde{P}) = \max[\sigma(P, \widetilde{P}), \sigma(\widetilde{P}, P)]. \tag{3}$$

Следующая лемма показывает, что так определенная функция $L(P, \widetilde{P})$, $P, \widetilde{P} \in \mathscr{P}(\mathscr{E})$, называемая *метрикой Леви—Прохорова*, действительно является метрикой.

Лемма 1. *Функция* $L(P, \widetilde{P})$ *обладает свойствами расстояния*:

a) $L(P, \widetilde{P}) = L(\widetilde{P}, P)$ $(= \sigma(P, \widetilde{P}) = \sigma(\widetilde{P}, P))$,

b) $L(P, \widetilde{P}) \leqslant L(P, \widehat{P}) + L(\widehat{P}, \widetilde{P})$,

c) $L(P, \widetilde{P}) = 0$ *тогда и только тогда*, *когда* $\widetilde{P} = P$.

Доказательство. a) Достаточно показать, что ($\alpha > 0$, $\beta > 0$)

$$\text{«}P(F) \leqslant \widetilde{P}(F^{\alpha}) + \beta \text{ для всех замкнутых } F \in \mathscr{E}\text{»} \tag{4}$$

тогда и только тогда, когда

$$\text{«}\widetilde{P}(F) \leqslant P(F^{\alpha}) + \beta \text{ для всех замкнутых } F \in \mathscr{E}\text{»}. \tag{5}$$

Пусть T — замкнутое множество из $\mathscr{E}$. Тогда множество T^{α} открыто и нетрудно проверить, что $T \subseteq E \setminus (E \setminus T^{\alpha})^{\alpha}$. Если выполнено (4), то тогда, в частности,

$$P(E \setminus T^{\alpha}) \leqslant \widetilde{P}((E \setminus T^{\alpha})^{\alpha}) + \beta$$

и, тем самым,

$$\widetilde{P}(T) \leqslant \widetilde{P}(E \setminus (E \setminus T^{\alpha})^{\alpha}) \leqslant P(T^{\alpha}) + \beta,$$

что и показывает равносильность (4) и (5). Отсюда следует, что

$$\sigma(P, \widetilde{P}) = \sigma(\widetilde{P}, P) \tag{6}$$

и, тем самым,

$$L(P, \widetilde{P}) = \sigma(P, \widetilde{P}) = \sigma(\widetilde{P}, P) = L(\widetilde{P}, P). \tag{7}$$

b) Пусть $L(P, \widehat{P}) < \delta_1$, $L(\widehat{P}, \widetilde{P}) < \delta_2$. Тогда для каждого замкнутого $F \in \mathscr{E}$

$$\widetilde{P}(F) \leqslant \widehat{P}(F^{\delta_2}) + \delta_2 \leqslant P((F^{\delta_2})^{\delta_1}) + \delta_1 + \delta_2 \leqslant P(F^{\delta_1 + \delta_2}) + \delta_1 + \delta_2$$

и поэтому $L(P,\,\widetilde{P}) \leqslant \delta_1 + \delta_2$. Отсюда следует, что

$$L(P,\,\widetilde{P}) \leqslant L(P,\,\widehat{P}) + L(\widehat{P},\,\widetilde{P}).$$

c) Если $L(P,\,\widetilde{P}) = 0$, то тогда для каждого замкнутого $F \in \mathscr{E}$ и любого $\alpha > 0$

$$P(F) \leqslant \widetilde{P}(F^\alpha) + \alpha. \tag{8}$$

Поскольку $F^\alpha \downarrow F$, $\alpha \downarrow 0$, то из (8) предельным переходом по $\alpha \downarrow 0$ находим, что $P(F) \leqslant \widetilde{P}(F)$, и по симметрии $\widetilde{P}(F) \leqslant P(F)$. Тем самым, $P(F) = \widetilde{P}(F)$ для всех замкнутых $F \in \mathscr{E}$. Для каждого борелевского множества $A \in \mathscr{E}$ и всякого $\varepsilon > 0$ найдутся такие открытое множество $G_\varepsilon \supseteq A$ и замкнутое множество $F_\varepsilon \subseteq A$, что $P(G_\varepsilon \setminus F_\varepsilon) \leqslant \varepsilon$. Отсюда следует, что всякая вероятностная мера P на метрическом пространстве $(E,\,\mathscr{E},\,\rho)$ полностью определяется своими значениями на замкнутых множествах. Следовательно, из $\widetilde{P}(F) = P(F)$ для всех замкнутых $F \in \mathscr{E}$ вытекает, что $\widetilde{P}(A) = P(A)$ для всех борелевских $A \in \mathscr{E}$. □

Теорема 1. *Метрика Леви—Прохорова* $L(P,\,\widetilde{P})$ *метризует слабую сходимость*:

$$L(P_n,\,P) \to 0 \;\Leftrightarrow\; P_n \xrightarrow{w} P. \tag{9}$$

Доказательство. ($\Rightarrow$) Пусть $L(P_n,\,P) \to 0$, $n \to \infty$. Тогда для всякого фиксированного замкнутого множества $F \in \mathscr{E}$ и любого $\varepsilon > 0$, согласно (2) и утверждению a) леммы 1,

$$\varlimsup_n P_n(F) \leqslant P(F^\varepsilon) + \varepsilon. \tag{10}$$

Полагая здесь $\varepsilon \downarrow 0$, находим, что

$$\varlimsup_n P_n(F) \leqslant P(F).$$

Согласно теореме 1 из § 1, отсюда следует, что

$$P_n \xrightarrow{w} P. \tag{11}$$

Доказательство импликации ($\Leftarrow$) будет опираться на ряд глубоких и полезных фактов, дополнительно проливающих свет как на само содержание понятия слабой сходимости, так и на методы ее установления и методы изучения «скорости» сходимости.

Итак, пусть $P_n \xrightarrow{w} P$. Это означает, что для любой *непрерывной ограниченной* функции $f = f(x)$

$$\int_E f(x)\,P_n(dx) \to \int_E f(x)\,P(dx). \tag{12}$$

Предположим теперь, что $\mathscr{G}$ — некоторый класс *равностепенно непрерывных* функций $g = g(x)$ (для всякого $\varepsilon > 0$ найдется такое $\delta > 0$, что $|g(y)-g(x)|<\varepsilon$, если $\rho(x, y)<\delta$ для всех $g\in\mathscr{G}$), таких, что $|g(x)|\leqslant C$ с одной и той же константой $C>0$ (для всех $x\in E$ и $g\in\mathscr{G}$). Согласно теореме 3 § 8, для класса $\mathscr{G}$ имеет место следующее усиление свойства (12):

$$P_n \xrightarrow{w} P \;\Rightarrow\; \sup_{g\in\mathscr{G}} \left| \int_E g(x)\, P_n(dx) - \int_E g(x)\, P(dx) \right| \to 0. \tag{13}$$

Для каждого $A\in\mathscr{E}$ и $\varepsilon>0$ положим (как в теореме 1 из § 1)

$$f_A^\varepsilon(x) = \left[1 - \frac{\rho(x, A)}{\varepsilon}\right]^+. \tag{14}$$

Ясно, что

$$I_A(x) \leqslant f_A^\varepsilon(x) \leqslant I_{A^\varepsilon}(x) \tag{15}$$

и

$$|f_A^\varepsilon(x) - f_A^\varepsilon(y)| \leqslant \varepsilon^{-1}|\rho(x, A) - \rho(y, A)| \leqslant \varepsilon^{-1}\rho(x, y).$$

Тем самым для класса $\mathscr{G}^\varepsilon = \{f_A^\varepsilon(x),\ A\in\mathscr{E}\}$ имеет место (13). Значит,

$$\Delta_n \equiv \sup_{A\in\mathscr{E}} \left| \int_E f_A^\varepsilon(x)\, P_n(dx) - \int_E f_A^\varepsilon(x)\, P(dx) \right| \to 0, \quad n\to\infty. \tag{16}$$

Отсюда и из (15) заключаем, что для всякого замкнутого множества $A\in\mathscr{E}$ и $\varepsilon>0$

$$P(A^\varepsilon) \geqslant \int_E f_A^\varepsilon(x)\, dP \geqslant \int_E f_A^\varepsilon(x)\, dP_n - \Delta_n \geqslant P_n(A) - \Delta_n. \tag{17}$$

Выберем $n(\varepsilon)$ так, что $\Delta_n\leqslant\varepsilon$ для всех $n\geqslant n(\varepsilon)$. Тогда из (17) для $n\geqslant n(\varepsilon)$

$$P(A^\varepsilon) \geqslant P_n(A) - \varepsilon. \tag{18}$$

Отсюда в силу определений (2), (3) вытекает, что $L(P_n, P)\leqslant\varepsilon$, коль скоро $n\geqslant n(\varepsilon)$. Тем самым

$$P_n \xrightarrow{w} P \;\Rightarrow\; \Delta_n\to 0 \;\Rightarrow\; L(P_n, P)\to 0.$$

Теорема (с точностью до утверждения (13)) доказана. □

3. Метрика $\|P-\widetilde{P}\|_{BL}^*$. Обозначим BL множество всех непрерывных ограниченных функций $f=f(x)$, $x\in E$ (с $\|f\|_\infty = \sup_x |f(x)|<\infty$), удовлетворяющих к тому же условию Липшица:

$$\|f\|_L = \sup_{x\neq y} \frac{|f(x)-f(y)|}{\rho(x, y)} < \infty.$$

Положим $\|f\|_{BL} = \|f\|_\infty + \|f\|_L$. Пространство BL с нормой $\|\cdot\|_{BL}$, является банаховым пространством.

Определим метрику $\|P-\widetilde{P}\|_{BL}^{*}$, положив

$$\|P-\widetilde{P}\|_{BL}^{*}=\sup_{f\in BL}\left\{\left|\int f\,d(P-\widetilde{P})\right|:\ \|f\|_{BL}\leqslant 1\right\}. \tag{19}$$

(Можно проверить, что действительно $\|P-\widetilde{P}\|_{BL}^{*}$ удовлетворяет всем требованиям, предъявляемым к метрике; задача 2.)

Теорема 2. *Метрика $\|P-\widetilde{P}\|_{BL}^{*}$ метризует слабую сходимость*:

$$\|P_n-P\|_{BL}^{*}\to 0 \ \Leftrightarrow\ P_n\xrightarrow{w}P.$$

Доказательство. Импликация ($\Leftarrow$) вытекает немедленно из (13). Для доказательства ($\Rightarrow$) достаточно показать, что в определении слабой сходимости $P_n\xrightarrow{w}P$ как выполнения свойства (12) для любой непрерывной ограниченной функции $f=f(x)$ достаточно ограничиться рассмотрением лишь класса ограниченных функций удовлетворяющих условию Липшица. Иначе говоря, импликация ($\Rightarrow$) будет доказана, если установить справедливость следующего результата.

Лемма 2. *Слабая сходимость $P_n\xrightarrow{w}P$ имеет место тогда и только тогда, когда свойство* (12) *выполнено для любой функции $f=f(x)$ из класса BL.*

Доказательство. В одну сторону доказательство очевидно. Рассмотрим теперь функции $f_A^{\varepsilon}=f_A^{\varepsilon}(x)$, определенные в (14). Как было установлено выше при доказательстве теоремы 1, при каждом $\varepsilon>0$ класс $\mathscr{G}^{\varepsilon}=\{f_A^{\varepsilon}(x),\ A\in\mathscr{E}\}\subseteq BL$. Если теперь проанализировать доказательство импликации (I) $\Rightarrow$ (II) в теореме 1 из § 1, то можно заметить, что в действительности в ее доказательстве выполнение свойства (12) использовалось *не для всех* ограниченнных непрерывных функций, а лишь для функций из классов $\mathscr{G}^{\varepsilon}$, $\varepsilon>0$. Поскольку $\mathscr{G}^{\varepsilon}\subseteq BL$, $\varepsilon>0$, то заведомо верно, что из выполнения свойства (12) для функций из класса BL следует утверждение II теоремы 1 § 1, которое равносильно (в силу той же теоремы 1 § 1) слабой сходимости $P_n\xrightarrow{w}P$. □

Замечание. Утверждение теоремы 2 можно было бы вывести из теоремы 1 (так же как и наоборот), если воспользоваться следующими неравенствами между метриками $L(P,\widetilde{P})$ и $\|P-\widetilde{P}\|_{BL}^{*}$, справедливыми для случая сепарабельных метрических пространств $(E,\mathscr{E},\rho)$:

$$\|P-\widetilde{P}\|_{BL}^{*}\leqslant 2L(P,\widetilde{P}), \tag{20}$$

$$\varphi(L(P,\widetilde{P}))\leqslant\|P-\widetilde{P}\|_{BL}^{*},\qquad \text{где } \varphi(x)=\frac{2x^2}{2+x}. \tag{21}$$

Замечая, что $0 \leqslant \varphi(x) \leqslant \dfrac{2}{3}$ для $x \geqslant 0$ тогда и только тогда, когда $x \leqslant 1$, и $\dfrac{2}{3}x^2 \leqslant \varphi(x)$ для $0 \leqslant x \leqslant 1$, из (20) и (21) выводим, что если $L(P, \widetilde{P}) \leqslant 1$ или $\|P - \widetilde{P}\|^*_{BL} \leqslant \dfrac{2}{3}$, то

$$\frac{2}{3}\, L^2(P, \widetilde{P}) \leqslant \|P - \widetilde{P}\|^*_{BL} \leqslant 2L(P, \widetilde{P}). \tag{22}$$

4. Задачи.

1. Показать, что в случае $E = R$ метрика Леви—Прохорова $L(P, \widetilde{P})$ между распределениями вероятностей P и $\widetilde{P}$ *не меньше* расстояния Леви $L(F, \widetilde{F})$ между функциями распределения F и $\widetilde{F}$, соответствующими P и $\widetilde{P}$ (см. задачу 4 в § 1). Привести пример выполнения строгого неравенства между этими метриками.

2. Показать, что формула (19) определяет метрику в пространстве BL.

3. Доказать справедливость неравенств (20), (21) и (22).

4. Пусть $F = F(x)$ и $G = G(x)$ — две функции распределения, P_c и Q_c — точки их пересечения прямой $x + y = c$. Показать, что расстояние Леви (см. задачу 4 в §1)

$$L(F, G) = \sup_c \frac{\overline{P_c Q_c}}{\sqrt{2}},$$

где $\overline{P_c Q_c}$ — длина отрезка между точками P_c и Q_c.

5. Показать, что множество всех функций распределения с метрикой Леви есть полное пространство.

§ 8. О связи слабой сходимости мер со сходимостью случайных элементов почти наверное («метод одного вероятностного пространства»)

1. Предположим, что на вероятностном пространстве $(\Omega, \mathscr{F}, \;\;)$ заданы случайные элементы $X = X(\omega)$, $X_n = X_n(\omega)$, $n \geqslant 1$, принимающие значения в метрическом пространстве $(E, \mathscr{E}, \rho)$; см. § 5 гл. II. Обозначим P и P_n распределения вероятностей X и X_n, т. е. пусть

$$P(A) = \;\; \{\omega\colon X(\omega) \in A\}, \qquad P_n(A) = \;\; \{\omega\colon X_n(\omega) \in A\}, \qquad A \in \mathscr{E}.$$

Обобщая понятие сходимости случайных величин по распределению (см. § 10 гл. II), введем такое

Определение 1. Последовательность случайных элементов X_n, $n \geqslant 1$, называется сходящейся *по распределению* или *по закону* (обозначения: $X_n \xrightarrow{\mathscr{D}} X$, или $X_n \xrightarrow{\mathscr{L}} X$, или $X_n \xrightarrow{Law} X$), если $P_n \xrightarrow{w} P$.

По аналогии с определением сходимости случайных величин по вероятности и с вероятностью единица (§ 10 гл. II) естественны следующие определения.

Определение 2. Последовательность случайных элементов X_n, $n \geqslant 1$, называется сходящейся *по вероятности* к X $(X_n \to X)$, если

$$\{\omega\colon\ \rho(X_n(\omega),\ X(\omega)) \geqslant \varepsilon\} \to 0, \qquad n \to \infty. \tag{1}$$

Определение 3. Последовательность случайных элементов X_n, $n \geqslant 1$, называется сходящейся к X *с вероятностью единица* (*почти наверное*, *почти всюду*; $X_n \xrightarrow{\text{п. н.}} X$, $X_n \xrightarrow{\text{п.в.}} X$), если $\rho(X_n(\omega),\ X(\omega)) \xrightarrow{\text{п. н.}} 0$, $n \to \infty$.

Замечание 1. Оба последних определения имеют смысл, конечно, если только $\rho(X_n(\omega),\ X(\omega))$ как функции от $\omega \in \Omega$ являются случайными величинами, т. е. $\mathscr{F}$-измеримыми функциями. Это будет заведомо так, если пространство $(E,\ \mathscr{E},\ \rho)$ является сепарабельным (задача 1).

Замечание 2. В связи с определением 2 отметим, что введенная сходимость по вероятности метризуется следующей *метрикой Ки Фан* между случайными элементами X и Y (определенными на $(\Omega,\ \mathscr{F},\quad)$ и принимающие значения в E; задача 2):

$$d\ \ (X,\ Y) = \inf\{\varepsilon > 0\colon\quad \{\rho(X(\omega),\ Y(\omega)) \geqslant \varepsilon\} \leqslant \varepsilon\}. \tag{2}$$

Замечание 3. Если определения сходимости по вероятности и с вероятностью единица требуют задания случайных элементов на одном и том же вероятностном пространстве, то определение сходимости по распределению $X_n \xrightarrow{\mathscr{D}} X$ связано лишь со сходимостью распределений, и, следовательно, можно считать, что $X(\omega)$, $X_1(\omega)$, $X_2(\omega)$, $\ldots$ принимают значения в одном и том же пространстве E, но могут быть заданы на «*своих*» вероятностных пространствах $(\Omega,\ \mathscr{F},\quad)$, $(\Omega_1,\ \mathscr{F}_1,\quad_1)$, $(\Omega_2,\ \mathscr{F}_2,\quad_2)$, $\ldots$ Без ограничения общности, однако, всегда можно считать их заданными на одном и том же вероятностном пространстве, беря в качестве такового прямое произведение указанных пространств и определяя $X(\omega,\ \omega_1,\ \omega_2,\ \ldots) = X(\omega)$, $X_1(\omega,\ \omega_1,\ \omega_2,\ \ldots) = X_1(\omega_1)$, $\ldots$

2. Согласно определению 1 и теореме о замене переменных под знаком интеграла Лебега (теорема 7 § 6 гл. II),

$$X_n \xrightarrow{\mathscr{D}} X \iff \quad f(X_n) \to \quad f(X) \tag{3}$$

для всякой непрерывной ограниченной функции $f = f(x)$, $x \in E$.

Из (3) видно, что на основании теоремы Лебега о мажорируемой сходимости (теорема 3 § 6 гл. II) из сходимости $X_n \xrightarrow{\text{п. н.}} X$ сразу вытекает сходимость $X_n \xrightarrow{\mathscr{D}} X$, что вовсе и не удивительно, если иметь в виду ситуацию, когда X и X_n — случайные величины (теорема 2 § 10 гл. II). Более неожиданно, что в некотором смысле имеет место обратный результат, к точным формулировкам, а затем и применениям которого мы сейчас и переходим.

Предварительно введем такое

Определение 4. Случайные элементы $X = X(\omega')$ и $Y = Y(\omega'')$, заданные на вероятностных пространствах $(\Omega', \mathscr{F}', \quad ')$ и $(\Omega'', \mathscr{F}'', \quad '')$ соответственно и принимающие значения в одном и том же пространстве E, называются *совпадающими* (*эквивалентными*, *конгруэнтными*) *по распределению* (обозначение: $X \overset{\mathscr{D}}{=} Y$), если они имеют совпадающие распределения вероятностей.

Теорема 1. *Пусть $(E, \mathscr{E}, \rho)$ — сепарабельное метрическое пространство.*

1. *Пусть случайные элементы X, X_n, $n \geqslant 1$, заданные на вероятностном пространстве $(\Omega, \mathscr{F}, \quad)$ и со значениями в E, таковы, что $X_n \xrightarrow{\mathscr{D}} X$. Тогда можно найти вероятностное пространство $(\Omega^*, \mathscr{F}^*, \quad ^*)$ и определенные на нем случайные элементы X^*, X_n^*, $n \geqslant 1$, со значениями в E такие, что*

$$X_n^* \xrightarrow{\text{п. н.}} X^*$$

и

$$X^* \overset{\mathscr{D}}{=} X, \quad X_n^* \overset{\mathscr{D}}{=} X_n, \quad n \geqslant 1.$$

2. *Пусть P, P_n, $n \geqslant 1$, — вероятностные меры на $(E, \mathscr{E}, \rho)$ такие, что $P_n \xrightarrow{w} P$. Тогда найдутся вероятностное пространство $(\Omega^*, \mathscr{F}^*, \quad ^*)$ и определенные на нем случайные элементы X^*, X_n^*, $n \geqslant 1$, со значениями в E такие, что*

$$X_n^* \xrightarrow{\text{п. н.}} X^*$$

и

$$P^* = P, \quad P_n^* = P_n, \quad n \geqslant 1,$$

где P^* и P_n^* — распределения вероятностей X^* и X_n^*.

Прежде чем переходить к доказательствам, заметим, во-первых, что достаточно доказать лишь второе утверждение, поскольку первое следует из него, если взять в качестве P и P_n распределения X и X_n. Соответствующим образом и второе утверждение следует из первого. Во-вторых, отметим, что доказательство этой теоремы в ее полной общности технически довольно сложно. Именно поэтому мы приводим здесь доказательство лишь для случая $E = R$. Это доказательство довольно прозрачно и к тому же дает простую явную конструкцию искомых объектов. (К сожалению, эта конструкция не «работает» в общем случае, даже для $E = R^2$.)

Доказательство теоремы в случае $E = R$. Пусть $F = F(x)$ и $F_n = F_n(x)$ — функции распределения, соответствующие мерам P и P_n на $(R, \mathscr{B}(R))$. Свяжем с функцией $F = F(x)$ соответствующую ей *квантильную функцию* $Q = Q(u)$, однозначно определяемую формулой

$$Q(u) = \inf\{x : F(x) \geqslant u\}, \quad 0 < u < 1. \tag{4}$$

Нетрудно проверить, что

$$F(x) \geqslant u \ \Leftrightarrow \ Q(u) \leqslant x. \tag{5}$$

Возьмем теперь $\Omega^* = (0, 1)$, $\mathscr{F}^* = \mathscr{B}(0, 1)$ и в качестве $\ ^*$ — меру Лебега, $\ ^*(d\omega^*) = d\omega^*$. Положим также $X^*(\omega^*) = Q(\omega^*)$, $\omega^* \in \Omega^*$. Тогда

$$\ ^*\{\omega^* : X^*(\omega^*) \leqslant x\} = \ ^*\{\omega^* : Q(\omega^*) \leqslant x\} = \ ^*\{\omega^* : \omega^* \leqslant F(x)\} = F(x),$$

т. е. распределение построенной случайной величины $X^*(\omega^*) = Q(\omega^*)$ в точности совпадает с P. Аналогично, распределение величин $X_n^*(\omega^*) = Q_n(\omega^*)$ совпадает с P_n.

Далее, несложно показать, что из сходимости $F_n(x)$ к $F(x)$ в каждой точке непрерывности предельной функции $F = F(x)$ (равносильной в случае $E = R$ сходимости $P_n \xrightarrow{w} P$; см. теорему 1 в § 1) вытекает, что последовательность квантильных функций $Q_n(u)$, $n \geqslant 1$, также сходится к $Q(u)$ в каждой точке непрерывности предельной функции $Q = Q(u)$. Поскольку множество точек разрыва функции $Q = Q(u)$, $u \in (0, 1)$, не более чем счетно, то его мера Лебега $\ ^*$ равна нулю и тем самым

$$X_n^*(\omega^*) = Q_n(\omega^*) \xrightarrow{\text{п. н.}} X^*(\omega^*) = Q(\omega^*).$$

Теорема (в случае $E = R$) доказана. □

Описываемая теоремой 1 конструкция перехода от заданных случайных элементов X и X_n к новым X^* и X_n^*, определяемым на одном и том же вероятностном пространстве, объясняет вынесенное в заголовок параграфа название «*метод одного вероятностного пространства*».

Остановимся теперь на ряде утверждений, которые проще всего устанавливать, пользуясь этим методом.

3. Предположим, что случайные элементы X, X_n, $n \geqslant 1$, заданные, скажем, на одном вероятностном пространстве $(\Omega, \mathscr{F}, \ )$ и принимающие значения в сепарабельном метрическом пространстве $(E, \mathscr{E}, \rho)$, таковы, что $X_n \xrightarrow{\mathscr{D}} X$. Пусть также $h = h(x)$, $x \in E$, — измеримое отображение $(E, \mathscr{E}, \rho)$ в некоторое другое сепарабельное метрическое пространство $(E', \mathscr{E}', \rho')$. В теории вероятностей и математической статистике часто приходится сталкиваться с вопросом о том, при каких условиях на $h = h(x)$ можно утверждать, что из сходимости $X_n \xrightarrow{\mathscr{D}} X$ следует сходимость $h(X_n) \xrightarrow{\mathscr{D}} h(X)$.

Например, пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $\quad \xi_1 = m, \quad \xi_1 = \sigma^2 > 0$. Пусть

$$\bar{X}_n = \frac{\xi_1 + \ldots + \xi_n}{n}.$$

Центральная предельная теорема устанавливает, что

$$\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \xrightarrow{d} \mathscr{N}(0, 1).$$

Спрашивается, для каких функций $h = h(x)$ можно гарантировать, что

$$h\Big(\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}\Big) \xrightarrow{d} h(\mathscr{N}(0, 1))?$$

(Известна *теорема Манна—Вальда*, которая утверждает применительно к данному случаю, что это заведомо выполнено для непрерывных функций $h = h(x)$, и, следовательно, сразу можно утверждать, что

$$\frac{n(\bar{X}_n - m)^2}{\sigma^2} \xrightarrow{d} \chi_1^2,$$

где χ_1^2 — случайная величина, имеющая распределение хи-квадрат с одной степенью свободы; см. табл. 2 в § 3 гл. I.)

Другой пример. Если $X = X(t, \omega)$, $X_n = X_n(t, \omega)$, $t \in T$, — случайные процессы (см. § 5 гл. II) и

$$h(X) = \sup_{t \in T} |X(t, \omega)|, \qquad h(X_n) = \sup_{t \in T} |X_n(t, \omega)|,$$

то сформулированный вопрос означает следующее: при каких условиях из сходимости по распределению процессов, $X_n \xrightarrow{\mathscr{D}} X$, следует сходимость по распределению их супремумов, $h(X_n) \xrightarrow{\mathscr{D}} h(X)$?

Одно такое простое условие, обеспечивающее справедливость импликации

$$X_n \xrightarrow{\mathscr{D}} X \ \Rightarrow\ h(X_n) \xrightarrow{\mathscr{D}} h(X),$$

состоит в том, что отображение $h = h(x)$ непрерывно. Действительно, если $f = f(x')$ — *непрерывная ограниченная* функция на E', то функция $f(h(x))$ будет также непрерывной ограниченной функцией на E. Следовательно,

$$X_n \xrightarrow{\mathscr{D}} X \ \Rightarrow \quad f(h(X_n)) \to \quad f(h(X)).$$

Приводимая далее теорема показывает, что на самом деле требование непрерывности функции $h = h(x)$ можно несколько ослабить, учитывая свойства предельного случайного элемента X.

Обозначим $\Delta_h = \{x \in E: h(x)$ не ρ-непрерывна в точке $x\}$, иначе говоря, пусть Δ_h — множество точек разрыва функции $h = h(x)$. Заметим, что $\Delta_h \in \mathscr{E}$ (задача 4).

Теорема 2. 1. *Пусть* $(E, \mathscr{E}, \rho)$ *и* $(E', \mathscr{E}', \rho')$ — *сепарабельные метрические пространства,* $X_n \xrightarrow{\mathscr{D}} X$. *Пусть отображение* $h = h(x)$, $x \in$, *таково, что*

$$\{\omega\colon X(\omega) \in \Delta_h\} = 0. \tag{6}$$

Тогда $h(X_n) \xrightarrow{\mathscr{D}} h(X)$.

2. *Пусть* P, P_n, $n \geqslant 1$, — *вероятностные распределения на сепарабельном метрическом пространстве* $(E, \mathscr{E}, \rho)$ *такие, что* $P \xrightarrow{w} P$ *и* $h = h(x)$ — *измеримое отображение* $(E, \mathscr{E}, \rho)$ *в сепарабельное метрическое пространство* $(E', \mathscr{E}', \rho')$. *Пусть*

$$P\{x\colon x \in \Delta_h\} = 0.$$

Тогда $P_n^h \xrightarrow{w} P^h$, *где* $P_n^h(A) = P_n\{h(x) \in A\}$, $P^h(A) = P\{h(x) \in A\}$, $A \in \mathscr{E}'$.

Доказательство. Как и в теореме 1, достаточно доказать справедливость лишь, скажем, первого утверждения.

Пусть X^* и X_n^*, $n \geqslant 1$, — случайные элементы, построенные «методом одного вероятностного пространства», такие, что $X^* \overset{\mathscr{D}}{=} X$, $X_n^* \overset{\mathscr{D}}{=} X_n$, $n \geqslant 1$, и $X_n^* \xrightarrow{\text{п. н.}} X^*$. Пусть

$$A^* = \{\omega^*\colon \rho(X_n^*, X^*) \not\to 0\}, \qquad B^* = \{\omega^*\colon X^*(\omega^*) \in \Delta_n\}.$$

Тогда $\ ^*(A^* \cup B^*) = 0$ и для $\omega^* \notin A^* \cup B^*$

$$h(X_n^*(\omega^*)) \to h(X^*(\omega^*)),$$

а это означает, что $h(X_n^*) \xrightarrow{\text{п. н.}} h(X^*)$. Как уже отмечалось в п. 1, отсюда следует, что $h(X_n^*) \xrightarrow{\mathscr{D}} h(X^*)$. Но

$$h(X_n^*) \overset{\mathscr{D}}{=} h(X_n), \qquad h(X^*) \overset{\mathscr{D}}{=} h(X).$$

Поэтому $h(X_n) \xrightarrow{\mathscr{D}} h(X)$. □

4. В § 7 при доказательстве импликации ($\Leftarrow$) теоремы 1 было использовано утверждение (13). Приведем сейчас его доказательство, снова обращаясь к «методу одного вероятностного пространства».

Пусть $(E, \mathscr{E}, \rho)$ — сепарабельное метрическое пространство и $\mathscr{G}$ — класс равностепенно непрерывных функций $g = g(x)$ таких, что $|g(x)| \leqslant C$ для всех $x \in E$, $g \in \mathscr{G}$ и некоторой константы C.

Теорема 3. *Пусть* P, P_n, $n \geqslant 1$, — *вероятностные меры на* $(E, \mathscr{E}, \rho)$ *такие, что* $P_n \xrightarrow{w} P$. *Тогда*

$$\sup_{g \in \mathscr{G}} \left| \int_E g(x)\, P_n(dx) - \int_E g(x)\, P(dx) \right| \to 0, \qquad n \to \infty. \tag{7}$$

Доказательство. Пусть (7) не имеет места. Тогда существуют $a > 0$ и функции $g_1,\ g_2,\ \ldots$ из $\mathscr{G}$ такие, что

$$\left| \int_E g_n(x)\, P_n(dx) - \int_E g_n(x)\, P(dx) \right| \geqslant a > 0 \tag{8}$$

для бесконечно многих n. Переходя «методом одного вероятностного пространства» к случайным элементам X^* и X_n^* (см. теорему 1), перепишем (8) в виде

$$|\ \ {}^* g_n(X_n^*) - \ \ {}^* g_n(X^*)| \geqslant a > 0 \tag{9}$$

для бесконечно многих n. Но в силу свойств класса $\mathscr{G}$ для всякого $\varepsilon > 0$ найдется такое $\delta > 0$, что $|g(y) - g(x)| < \varepsilon$ для всех $g \in \mathscr{G}$, если $\rho(x,\ y) < \delta$. Кроме того, $|g(x)| \leqslant C$ для всех $x \in E$ и $g \in \mathscr{G}$. Поэтому

$$\begin{aligned} |\ \ {}^* g_n(X_n^*) - \ \ {}^* g_n(X^*)| \leqslant \ \ {}^*\{|g_n(X_n^*) - g_n(X^*)|;\ \rho(X_n^*,\ X^*) > \delta\} + \\ + \ \ {}^*\{|g_n(X_n^*) - g_n(X^*)|;\ \rho(X_n^*,\ X^*) \leqslant \delta\} \leqslant 2C\ \ {}^*\{\rho(X_n^*,\ X^*) > \delta\} + \varepsilon. \end{aligned}$$

Поскольку $X_n^* \xrightarrow{\text{п. н.}} X^*$, то $\ \ {}^*\{\rho(X_n^*,\ X^*) > \delta\} \to 0,\ n \to \infty$. Следовательно, в силу произвольности $\varepsilon > 0$

$$\overline{\lim_n}\ |\ \ {}^* g_n(X_n^*) - \ \ {}^* g_n(X^*)| = 0,$$

что противоречит (9). $\square$

5. В этом пункте идеи «метода одного вероятностного пространства», использованные в теореме 1, будут применены к оцениванию *сверху* значения метрики Леви—Прохорова $L(P,\ \widetilde{P})$ между двумя распределениями вероятностей на сепарабельном метрическом пространстве $(E,\ \mathscr{E},\ \rho)$.

Теорема 4. *Для каждой пары мер P, $\widetilde{P}$ можно найти вероятностное пространство $(\Omega^*,\ \mathscr{F}^*,\ \ {}^*)$ и случайные элементы X и $\widetilde{X}$ на нем со значениями в E такие, что распределения X и $\widetilde{X}$ совпадают с P и $\widetilde{P}$ соответственно и*

$$L(P,\ \widetilde{P}) \leqslant d_{\ \ *}(X,\ \widetilde{X}) = \inf\{\varepsilon > 0:\ \ \ {}^*(\rho(X,\ \widetilde{X}) \geqslant \varepsilon) \leqslant \varepsilon\}. \tag{10}$$

Доказательство. Согласно теореме 1, действительно можно найти вероятностное пространство $(\Omega^*,\ \mathscr{F}^*,\ \ {}^*)$ и случайные элементы X и $\widetilde{X}$ такие, что

$${}^*\{X \in A\} = P(A), \qquad {}^*\{\widetilde{X} \in A\} = \widetilde{P}(A), \qquad A \in \mathscr{E}.$$

Пусть $\varepsilon > 0$ таково, что

$${}^*\{\rho(X,\ \widetilde{X}) \geqslant \varepsilon\} \leqslant \varepsilon. \tag{11}$$

Тогда для всякого $A \in \mathscr{E}$

$$\widetilde{P}(A) = \mathsf{P}^*\{\widetilde{X} \in A\} = \mathsf{P}^*\{\widetilde{X} \in A,\ X \in A^\varepsilon\} + \mathsf{P}^*\{\widetilde{X} \in A,\ X \notin A^\varepsilon\} \leqslant$$
$$\leqslant \mathsf{P}^*\{X \in A^\varepsilon\} + \mathsf{P}^*\{\rho(X, \widetilde{X}) \geqslant \varepsilon\} \leqslant P(A^\varepsilon) + \varepsilon,$$

где $A^\varepsilon = \{x \in E : \rho(x, A) < \varepsilon\}$. Поэтому по определению метрики Леви—Прохорова (п. 2 § 7)

$$L(P, \widetilde{P}) \leqslant \varepsilon. \tag{12}$$

Из (11) и (12), переходя к inf по $\varepsilon > 0$, получаем требуемое утверждение (10). □

Следствие. *Пусть X и $\widetilde{X}$ — случайные элементы, заданные на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ и принимающие значения в E. Пусть P_X и $P_{\widetilde{X}}$ — их распределения вероятностей. Тогда*

$$L(P_X, P_{\widetilde{X}}) \leqslant d_{\mathsf{P}}(X, \widetilde{X}).$$

Замечание 1. Проведенное доказательство показывает, что на самом деле (10) справедливо всякий раз, когда на *каком-либо* вероятностном пространстве $(\Omega^*, \mathscr{F}^*, \mathsf{P}^*)$ можно указать случайные элементы X и $\widetilde{X}$ со значениями в E, распределения которых совпадают с P и $\widetilde{P}$ и для которых множество $\{\omega^* : \rho(X(\omega^*), \widetilde{X}(\omega^*)) \geqslant \varepsilon\} \in \mathscr{F}^*$, $\varepsilon > 0$. Тем самым качество оценки (10) существенным образом зависит от того, насколько удачно по мерам P и $\widetilde{P}$ построены объекты $(\Omega^*, \mathscr{F}^*, \mathsf{P}^*)$ и X, $\widetilde{X}$. (Процедуру построения Ω^*, $\mathscr{F}^*$, P^* и X, $\widetilde{X}$ по P и $\widetilde{P}$ называют каплингом — от англ. *coupling* — соединение, сцепление.) Можно, например, взять «каплинг-меру» P^* равной прямому произведению мер P и $\widetilde{P}$, но такой выбор не приводит, как правило, к хорошей оценке (10).

Замечание 2. Естественно поставить вопрос о том, когда в (10) имеет место знак равенства. На этот счет приведем без доказательства следующий результат: *пусть P и $\widetilde{P}$ — две вероятностные меры на сепарабельном метрическом пространстве $(E, \mathscr{E}, \rho)$; тогда существуют такие $(\Omega^*, \mathscr{F}^*, \mathsf{P}^*)$ и X, $\widetilde{X}$, что*

$$L(P, \widetilde{P}) = d_{\mathsf{P}^*}(X, \widetilde{X}) = \inf\{\varepsilon > 0 : \mathsf{P}^*\{\rho(X, \widetilde{X}) \geqslant \varepsilon\} \leqslant \varepsilon\}.$$

6. Задачи.

1. Показать, что в случае сепарабельных метрических пространств действительная функция $\rho(X(\omega), Y(\omega))$ является случайной величиной для любых случайных элементов $X(\omega)$ и $Y(\omega)$, определенных на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$.

2. Доказать, что функция $d_{\mathsf{P}}(X, Y)$, определенная в (2), является метрикой в пространстве случайных элементов со значениями в E.

3. Доказать справедливость (5).

4. Показать, что множество $\Delta_h = \{x \in E : h(x)$ не ρ-непрерывна в точке $x\} \in \mathscr{E}$.

§ 9. Расстояние по вариации между вероятностными мерами. Расстояние Какутани—Хеллингера и интегралы Хеллингера. Применение к абсолютной непрерывности и сингулярности мер

1. Пусть $(\Omega, \mathscr{F})$ — измеримое пространство и $\mathscr{P} = \{P\}$ — семейство вероятностных мер на нем.

Определение 1. *Расстоянием по вариации* между мерами P и $\widetilde{P}$ из $\mathscr{P}$ (обозначение: $\|P - \widetilde{P}\|$) называется *полная вариация* меры (со знаком) $P - \widetilde{P}$, т. е. величина

$$\operatorname{Var}(P - \widetilde{P}) \equiv \sup \left| \int_\Omega \varphi(\omega)\, d(P - \widetilde{P}) \right|, \tag{1}$$

где sup берется по классу всех $\mathscr{F}$-измеримых функций $\varphi(\omega)$, удовлетворяющих условию $|\varphi(\omega)| \leqslant 1$.

Лемма 1. *Расстояние по вариации*

$$\|P - \widetilde{P}\| = 2 \sup_{A \in \mathscr{F}} |P(A) - \widetilde{P}(A)|. \tag{2}$$

Доказательство. Поскольку для всякого $A \in \mathscr{F}$

$$P(A) - \widetilde{P}(A) = \widetilde{P}(\bar{A}) - P(\bar{A}),$$

то

$$2|P(A) - \widetilde{P}(A)| = |P(A) - \widetilde{P}(A)| + |P(\bar{A}) - \widetilde{P}(\bar{A})| \leqslant \|P - \widetilde{P}\|,$$

где последнее неравенство следует из (1).

Для доказательства обратного неравентсва обратимся к разложению Хана (см., например, [33, § 5 гл. VI] или [70, с. 121]) *меры со знаком* $\mu \equiv P - \widetilde{P}$. Согласно этому разложению, мера μ представима в виде $\mu = \mu_+ - \mu_-$, где неотрицательные меры μ_+ и μ_- (верхняя и нижняя вариации меры μ) имеют вид

$$\mu_+(A) = \int_{A \cap M} d\mu, \quad \mu_-(A) = -\int_{A \cap \bar{M}} d\mu, \quad A \in \mathscr{F},$$

где M — некоторое множество из $\mathscr{F}$. При этом

$$\operatorname{Var} \mu = \operatorname{Var} \mu_+ + \operatorname{Var} \mu_- = \mu_+(\Omega) + \mu_-(\Omega).$$

Поскольку

$$\mu_+(\Omega) = P(M) - \widetilde{P}(M), \qquad \mu_-(\Omega) = \widetilde{P}(\bar{M}) - P(\bar{M}),$$

то

$$\|P - \widetilde{P}\| = (P(M) - \widetilde{P}(M)) + (\widetilde{P}(\bar{M}) - P(\bar{M})) \leqslant 2 \sup_{A \in \mathscr{F}} |P(A) - \widetilde{P}(A)|. \quad \square$$

Определение 2. Последовательность вероятностных мер $(P_n)_{n \geqslant 1}$ называется *сходящейся по вариации* к мере P (обозначение: $P_n \xrightarrow{var} P$), если

$$\|P_n - P\| \to 0, \qquad n \to \infty. \tag{3}$$

Из этого определения и теоремы 1 § 1 нетрудно видеть, что сходимость по вариации вероятностных мер, заданных на метрическом пространстве $(\Omega, \mathscr{F}, \rho)$, влечет их слабую сходимость.

Близость распределений по вариации является, пожалуй, наиболее сильной формой близости вероятностных распределений, поскольку если два распределения близки по вариации, то практически в конкретных ситуациях их можно считать неразличимыми. В этой связи может создаться впечатление, что исследование расстояния по вариации не представляет большого вероятностного интереса. Однако, например, в теореме Пуассона (§ 6 гл. I) сходимость биномиального распределения к пуассоновскому имеет место в смысле сходимости к нулю расстояния по вариации. (Ниже, в § 12, будет дана оценка сверху для этого расстояния.)

Приведем также пример из области математической статистики, где необходимость нахождения *расстояния по вариации* между мерами P и $\widetilde{P}$ возникает естественным образом в связи с задачей различения по результатам наблюдений двух статистических гипотез H (истинное распределение есть P) и $\widetilde{H}$ (истинное распределение есть $\widetilde{P}$) относительно того, какая из вероятностных моделей $(\Omega, \mathscr{F}, P)$ или $(\Omega, \mathscr{F}, \widetilde{P})$ более соответствует «статистике» результатов наблюдений. Если $\omega \in \Omega$ трактовать как результат наблюдения, то под *тестом* (для различения гипотез H и $\widetilde{H}$) понимают любую $\mathscr{F}$-измеримую функцию $\varphi = \varphi(\omega)$ со значениями в $[0, 1]$, статистический смысл которой состоит в том, что $\varphi(\omega)$ есть «вероятность, с которой принимается гипотеза $\widetilde{H}$, когда результат наблюдения есть ω».

Будем качество различения гипотез H и $\widetilde{H}$ характеризовать *вероятностями ошибок первого и второго рода*:

$$\alpha(\varphi) = E\varphi(\omega) \quad (= \text{«}\textit{вероятность принять } \widetilde{H}\textit{, если верна } H\text{»}),$$

$$\beta(\varphi) = \widetilde{E}(1 - \varphi(\omega)) \quad (= \text{«}\textit{вероятность принять } H\textit{, если верна } \widetilde{H}\text{»}),$$

где E и $\widetilde{E}$ — усреднения по мерам P и $\widetilde{P}$. В том случае, когда гипотезы H и $\widetilde{H}$ *равноправны*, *оптимальным* естественно считать тест $\varphi^* = \varphi^*(\omega)$ (если таковой существует), который минимизирует сумму ошибок $\alpha(\varphi)+\beta(\varphi)$.

Положим

$$\mathscr{E}r(P,\widetilde{P})=\inf_{\varphi}[\alpha(\varphi)+\beta(\varphi)]. \tag{4}$$

Пусть $Q=\frac{1}{2}(P+\widetilde{P})$ и $z=\frac{dP}{dQ}$, $\tilde{z}=\frac{d\widetilde{P}}{dQ}$. Тогда

$$\mathscr{E}r(P,\widetilde{P})=\inf_{\varphi}[E\varphi+\widetilde{E}(1-\varphi)]=\inf_{\varphi}E_Q[z\varphi+\tilde{z}(1-\varphi)]=1+\inf_{\varphi}E_Q[\varphi(z-\tilde{z})]$$

(здесь и далее E_Q — математическое ожидание по мере Q).

Нетрудно видеть, что inf достигается на функции

$$\varphi^*(\omega)=I\{\tilde{z}<z\},$$

и поскольку $E_Q(z-\tilde{z})=0$, то

$$\mathscr{E}r(P,\widetilde{P})=1-\frac{1}{2}E_Q|z-\tilde{z}|=1-\frac{1}{2}\|P-\widetilde{P}\|, \tag{5}$$

где последнее равенство следует из приводимой ниже леммы 2. Таким образом, из (5) ясно, что качество различения гипотез, характеризуемое функцией $\mathscr{E}r(P,\widetilde{P})$, зависит от степени близости мер P и $\widetilde{P}$ именно в смысле *расстояния по вариации*.

Лемма 2. *Пусть Q — некоторая σ-конечная мера такая, что $P\ll Q$, $\widetilde{P}\ll Q$ и $z=\frac{dP}{dQ}$, $\tilde{z}=\frac{d\widetilde{P}}{dQ}$ — производные Радона—Никодима мер P и $\widetilde{P}$ относительно Q. Тогда*

$$\|P-\widetilde{P}\|=E_Q|z-\tilde{z}| \tag{6}$$

и если $Q=\frac{1}{2}(P+\widetilde{P})$, то

$$\|P-\widetilde{P}\|=E_Q|z-\tilde{z}|=2E_Q|1-z|=2E_Q|1-\tilde{z}|. \tag{7}$$

Доказательство. Для всех $\mathscr{F}$-измеримых функций $\psi=\psi(\omega)$ таких, что $|\psi(\omega)|\leqslant 1$, по определению z и $\tilde{z}$ находим

$$|E\psi-\widetilde{E}\psi|=|E_Q\psi(z-\tilde{z})|\leqslant E_Q|\psi|\,|z-\tilde{z}|\leqslant E_Q|z-\tilde{z}|. \tag{8}$$

Поэтому

$$\|P-\widetilde{P}\|\leqslant E_Q|z-\tilde{z}|. \tag{9}$$

Но для функции

$$\psi=\operatorname{sign}(\tilde{z}-z)=\begin{cases}1, & \tilde{z}\geqslant z,\\ -1, & \tilde{z}<z,\end{cases}$$

имеем

$$|E\psi - \widetilde{E}\psi| = E_Q|z - \tilde{z}|. \tag{10}$$

Из (9) и (10) получаем требуемое равенство (6). Соотношения (7) следуют из (6), поскольку $z + \tilde{z} = 2$ (Q-п. н.). □

Следствие 1. *Пусть P и $\widetilde{P}$ — два распределения вероятностей на $(R, \mathscr{B}(R))$* с плотностями *вероятностей* (*относительно меры Лебега dx*) *$p(x)$ и $\tilde{p}(x)$, $x \in R$. Тогда*

$$\|P - \widetilde{P}\| = \int_{-\infty}^{\infty} |p(x) - \tilde{p}(x)|\, dx. \tag{11}$$

(*В качестве меры Q надо взять меру Лебега на $(R, \mathscr{B}(R))$.*)

Следствие 2. *Пусть P и $\widetilde{P}$ — две* дискретные *меры, $P = (p_1, p_2, \ldots)$, $\widetilde{P} = (\tilde{p}_1, \tilde{p}_2, \ldots)$, сосредоточенные в счетном числе точек $x_1, x_2, \ldots$ Тогда*

$$\|P - \widetilde{P}\| = \sum_{i=1}^{\infty} |p_i - \tilde{p}_i|. \tag{12}$$

(*В качестве меры Q надо взять считающую меру, т. е. такую, что $Q(\{x_i\}) = 1$, $i = 1, 2, \ldots$*)

2. Обратимся теперь к еще одной мере близости вероятностными мерами, во многом (как это будет следовать из дальнейшего) родственной близости мер по вариации.

Пусть P и $\widetilde{P}$ — вероятностные меры на $(\Omega, \mathscr{F})$ и Q — третья вероятностная мера, *доминирующая* меры P и $\widetilde{P}$, т. е. такая, что $P \ll Q$, $\widetilde{P} \ll Q$. Снова обозначим

$$z = \frac{dP}{dQ}, \qquad \tilde{z} = \frac{d\widetilde{P}}{dQ}.$$

Определение 3. *Расстоянием Какутани—Хеллингера* между мерами P и $\widetilde{P}$ называется неотрицательная величина $\rho(P, \widetilde{P})$ такая, что

$$\rho^2(P, \widetilde{P}) = \frac{1}{2} E_Q[\sqrt{z} - \sqrt{\tilde{z}}]^2. \tag{13}$$

Поскольку

$$E_Q[\sqrt{z} - \sqrt{\tilde{z}}]^2 = \int_{\Omega} \left[\sqrt{\frac{dP}{dQ}} - \sqrt{\frac{d\widetilde{P}}{dQ}}\right]^2 dQ, \tag{14}$$

то становится понятной *символическая* запись величины $\rho^2(P, \widetilde{P})$ в виде

$$\rho^2(P, \widetilde{P}) = \frac{1}{2} \int_{\Omega} [\sqrt{dP} - \sqrt{d\widetilde{P}}\,]^2. \tag{15}$$

Если положить

$$H(P, \widetilde{P}) = E_Q\sqrt{z\tilde{z}}, \tag{16}$$

то по аналогии с (15) можно символически записать

$$H(P, \widetilde{P}) = \int_\Omega \sqrt{dP\, d\widetilde{P}}. \tag{17}$$

Из (13), (16), а также из (15), (17) понятно, что

$$\rho^2(P, \widetilde{P}) = 1 - H(P, \widetilde{P}). \tag{18}$$

Величина $H(P, \widetilde{P})$ называется *интегралом Хеллингера* порядка 1/2. Для многих целей удобным оказывается рассмотрение *интегралов Хеллингера* $H(\alpha; P, \widetilde{P})$ *порядка* $\alpha \in (0, 1)$, определяемых формулой

$$H(\alpha; P, \widetilde{P}) = E_Q z^\alpha \tilde{z}^{1-\alpha}, \tag{19}$$

или, символически,

$$H(\alpha; P, \widetilde{P}) = \int_\Omega (dP)^\alpha (d\widetilde{P})^{1-\alpha}. \tag{20}$$

Ясно, что $H\left(\frac{1}{2}; P, \widetilde{P}\right) = H(P, \widetilde{P})$.

Чтобы определение 3 было корректным, надо показать, что величина $\rho^2(P, \widetilde{P})$ не зависит от выбора доминирующей меры и что действительно $\rho(P, \widetilde{P})$ удовлетворяет требованиям, предъявляемым к «расстоянию».

Лемма 3. 1. *Интеграл Хеллингера порядка* $\alpha \in (0, 1)$ (*а следовательно, и* $\rho(P, \widetilde{P})$) *не зависит от выбора доминирующей меры* Q.

2. *Функция* ρ, *определенная в* (13), *является метрикой на множестве вероятностных мер.*

Доказательство. 1. Если мера Q' доминирует P и $\widetilde{P}$, то Q' доминирует и меру $Q = \dfrac{P + \widetilde{P}}{2}$. Поэтому достаточно доказать, что если $Q \ll Q'$, то

$$E_Q(z^\alpha \tilde{z}^{1-\alpha}) = E_{Q'}(z')^\alpha (\tilde{z}')^{1-\alpha},$$

где $z' = \dfrac{dP}{dQ'}$, $\tilde{z}' = \dfrac{d\widetilde{P}}{dQ'}$.

Обозначим $v = \dfrac{dQ}{dQ'}$. Тогда $z' = zv$, $\tilde{z}' = \tilde{z}v$ и

$$E_Q(z^\alpha \tilde{z}^{1-\alpha}) = E_{Q'}(vz^\alpha \tilde{z}^{1-\alpha}) = E_{Q'}(z')^\alpha (\tilde{z}')^{1-\alpha},$$

что и доказывает первое утверждение.

2. Если $\rho(P, \widetilde{P}) = 0$, то $z = \tilde{z}$ (Q-п. н.), откуда $P = \widetilde{P}$. Симметричность $\rho(P, \widetilde{P}) = \rho(\widetilde{P}, P)$ очевидна. Наконец, пусть P, P', P'' — три меры, $P \ll Q$,

$P' \ll Q$ и $P'' \ll Q$ с $z = \dfrac{dP}{dQ}$, $z' = \dfrac{dP'}{dQ}$, $z'' = \dfrac{dP''}{dQ}$. Используя справедливость неравенства треугольника для нормы в $L^2(\Omega, \mathscr{F}, Q)$, получаем

$$[E_Q(\sqrt{z} - \sqrt{z''})^2]^{1/2} \leqslant [E_Q(\sqrt{z} - \sqrt{z'})^2]^{1/2} + [E_Q(\sqrt{z'} - \sqrt{z''})^2]^{1/2},$$

т. е.

$$\rho(P, P'') \leqslant \rho(P, P') + \rho(P', P''). \qquad \square$$

Из определения (19) и теоремы Фубини (§ 6 гл. II) непосредственно вытекает, что в том случае, когда меры P и $\widetilde{P}$ являются *прямыми произведениями* мер, $P = P_1 \times \ldots \times P_n$, $\widetilde{P} = \widetilde{P}_1 \times \ldots \times \widetilde{P}_n$ (см. п. 10 § 6 гл. II), интеграл Хеллингера между мерами P и $\widetilde{P}$ равен произведению соответствующих интегралов Хеллингера:

$$H(\alpha; P, \widetilde{P}) = \prod_{i=1}^{n} H(\alpha; P_i, \widetilde{P}_i).$$

Следующая теорема раскрывает связь между расстоянием по вариации и расстоянием Какутани—Хеллингера (или, эквивалентно, интегралом Хеллингера). В частности, она показывает, что эти расстояния определяют *одну и ту же топологию* в пространстве вероятностных мер на $(\Omega, \mathscr{F})$.

Теорема 1. *Имеют место следующие неравенства*:

$$2[1 - H(P, \widetilde{P})] \leqslant \|P - \widetilde{P}\| \leqslant \sqrt{8[1 - H(P, \widetilde{P})]}, \tag{21}$$

$$\|P - \widetilde{P}\| \leqslant 2\sqrt{1 - H^2(P, \widetilde{P})}. \tag{22}$$

В частности,

$$2\rho^2(P, \widetilde{P}) \leqslant \|P - \widetilde{P}\| \leqslant \sqrt{8}\,\rho(P, \widetilde{P}). \tag{23}$$

Доказательство. Поскольку $H(P, \widetilde{P}) \leqslant 1$ и $1 - x^2 \leqslant 2(1 - x)$ для $0 \leqslant x \leqslant 1$, то правое неравенство в (21) вытекает из (22), доказательство которого дается следующей цепочкой неравенств (с $Q = \frac{1}{2}(P + \widetilde{P})$):

$$\frac{1}{2}\|P - \widetilde{P}\| = E_Q|1 - z| \leqslant \sqrt{E_Q|1 - z|^2} = \sqrt{1 - E_Q z(2 - z)} =$$
$$= \sqrt{1 - E_Q z\tilde{z}} = \sqrt{1 - E_Q(\sqrt{z\tilde{z}})^2} \leqslant \sqrt{1 - (E_Q\sqrt{z\tilde{z}})^2} = \sqrt{1 - H^2(P, \widetilde{P})}.$$

Наконец, первое неравенство в (21) следует из того, что в силу неравенства

$$\frac{1}{2}[\sqrt{z} - \sqrt{2 - z}]^2 \leqslant |z - 1|, \quad z \in [0, 2],$$

имеем $(Q = \frac{1}{2}(P + \widetilde{P}))$

$$1 - H(P, \widetilde{P}) = \rho^2(P, \widetilde{P}) = \frac{1}{2} E_Q[\sqrt{z} - \sqrt{2-z}]^2 \leqslant E_Q|z-1| = \frac{1}{2}\|P - \widetilde{P}\|. \quad \square$$

Замечание. Сходным образом показывается, что для всякого $\alpha \in (0, 1)$

$$2[1 - H(\alpha;\, P, \widetilde{P})] \leqslant \|P - \widetilde{P}\| \leqslant \sqrt{c_\alpha(1 - H(\alpha;\, P, \widetilde{P}))}, \tag{24}$$

где c_α — некоторая константа.

Следствие 3. *Пусть* P *и* P^n, $n \geqslant 1$, — *вероятностные меры на* $(\Omega, \mathscr{F})$. *Тогда* $(n \to \infty)$

$$\|P^n - P\| \to 0 \;\Leftrightarrow\; H(P^n, P) \to 1 \;\Leftrightarrow\; \rho(P^n, P) \to 0,$$
$$\|P^n - P\| \to 2 \;\Leftrightarrow\; H(P^n, P) \to 0 \;\Leftrightarrow\; \rho(P^n, P) \to 1.$$

Следствие 4. *Поскольку в силу* (5)

$$\mathscr{E}r(P, \widetilde{P}) = 1 - \frac{1}{2}\|P - \widetilde{P}\|,$$

то из (21) *и* (22) *имеем*

$$\frac{1}{2} H^2(P, \widetilde{P}) \leqslant 1 - \sqrt{1 - H^2(P, \widetilde{P})} \leqslant \mathscr{E}r(P, \widetilde{P}) \leqslant H(P, \widetilde{P}). \tag{25}$$

В частности, пусть

$$P^n = \underbrace{P \times \ldots \times P}_{n}, \qquad \widetilde{P}^n = \underbrace{\widetilde{P} \times \ldots \times \widetilde{P}}_{n}$$

— *прямые произведения мер. Тогда, поскольку*

$$H(P^n, \widetilde{P}^n) = [H(P, \widetilde{P})]^n = e^{-\lambda n}, \qquad \lambda = -\ln H(P, \widetilde{P}) \geqslant \rho^2(P, \widetilde{P}),$$

то из (25) *получаем*

$$\frac{1}{2}\, e^{-2\lambda n} \leqslant \mathscr{E}r(P^n, \widetilde{P}^n) \leqslant e^{-\lambda n} \leqslant e^{-n\rho^2(P, \widetilde{P})}. \tag{26}$$

Применительно к рассмотренной выше задаче *различения двух статистических гипотез* из этих неравенств вытекает следующий результат.

Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные элементы, имеющие распределение вероятностей или P (гипотеза H), или $\widetilde{P}$ (гипотеза $\widetilde{H}$), причем $\widetilde{P} \neq P$ и, значит, $\rho^2(P, \widetilde{P}) > 0$. Тогда при $n \to \infty$ функция $\mathscr{E}r(P^n, \widetilde{P}^n)$, характеризующая качество оптимального различения гипотез H и $\widetilde{H}$ по наблюдениям $\xi_1, \ldots, \xi_n$, убывает к нулю *экспоненциально быстро*.

3. С помощью введенных выше интегралов Хеллингера порядка α удобно выражать условия *абсолютной непрерывности* и *сингулярности* вероятностных мер.

Пусть P и $\widetilde{P}$ — две вероятностные меры, заданные на измеримом пространстве $(\Omega, \mathscr{F})$. Напомним, что мера $\widetilde{P}$ *абсолютно непрерывна* относительно меры P (обозначение: $\widetilde{P} \ll P$), если $\widetilde{P}(A)=0$ всякий раз, когда $P(A)=0$ для $A \in \mathscr{F}$. Если $\widetilde{P} \ll P$ и $P \ll \widetilde{P}$, то говорят, что меры P и $\widetilde{P}$ *эквивалентны* ($\widetilde{P} \sim P$). Меры P и $\widetilde{P}$ называются *сингулярными* или *ортогональными* ($\widetilde{P} \perp P$), если существует $A \in \mathscr{F}$ такое, что $P(A)=1$ и $\widetilde{P}(\bar{A})=1$ (т. е. меры P и $\widetilde{P}$ «сидят» на разных множествах).

Пусть Q — вероятностная мера, $P \ll Q$, $\widetilde{P} \ll Q$, $z=\dfrac{dP}{dQ}$, $\tilde{z}=\dfrac{d\widetilde{P}}{dQ}$.

Теорема 2. *Следующие условия являются эквивалентными*:

(a) $\widetilde{P} \ll P$,

(b) $\widetilde{P}(z>0)=1$,

(c) $H(\alpha;\, P, \widetilde{P}) \to 1,\ \alpha \downarrow 0$.

Теорема 3. *Следующие условия являются эквивалентными*:

(a) $\widetilde{P} \perp P$,

(b) $\widetilde{P}(z>0)=0$,

(c) $H(\alpha;\, P, \widetilde{P}) \to 0,\ \alpha \downarrow 0$,

(d) $H(\alpha;\, P, \widetilde{P})=0$ *для всех* $\alpha \in (0, 1)$,

(e) $H(\alpha;\, P, \widetilde{P})=0$ *для некоторого* $\alpha \in (0, 1)$.

Доказательство обеих теорем будем проводить одновременно. Согласно определению z и $\tilde{z}$,

$$P(z=0)=E_Q[zI(z=0)]=0, \tag{27}$$

$$\begin{aligned}\widetilde{P}(A \cap \{z>0\}) &= E_Q[\tilde{z}I(A \cap \{z>0\})] = \\ &= E_Q\left[\tilde{z}\,\frac{z}{z}\,I(A \cap \{z>0\})\right] = E\left[\frac{\tilde{z}}{z}\,I(A \cap \{z>0\})\right] = E\left[\frac{\tilde{z}}{z}\,I(A)\right].\end{aligned} \tag{28}$$

Следовательно, имеет место **«разложение Лебега»**:

$$\widetilde{P}(A)=E\left[\frac{\tilde{z}}{z}\,I(A)\right]+\widetilde{P}(A \cap \{z=0\}), \quad A \in \mathscr{F}, \tag{29}$$

в котором величина $Z=\dfrac{\tilde{z}}{z}$ называется *производной Лебега* («*абсолютно непрерывной компоненты*») *меры* $\widetilde{P}$ *по* (*отношению к*) *мере* P и обозначается $\dfrac{d\widetilde{P}}{dP}$ (ср. с замечанием к теореме Радона—Никодима, § 6 гл. II).

Отсюда сразу получаем эквивалентность (a) и (b) в обеих теоремах.

Далее, поскольку

$$z^{\alpha}\tilde{z}^{1-\alpha} \to \tilde{z} I(z>0), \quad \alpha\downarrow 0,$$

и для $\alpha\in(0,\,1)$

$$0\leqslant z^{\alpha}\tilde{z}^{1-\alpha}\leqslant \alpha z+(1-\alpha)\tilde{z}\leqslant z+\tilde{z}$$

с $E_Q(z+\tilde{z}\,)=2$, то по теореме Лебега о мажорируемой сходимости

$$\lim_{\alpha\downarrow 0} H(\alpha;\,P,\,\widetilde{P})=E_Q\tilde{z}I(z>0)=\widetilde{P}(z>0)$$

и, значит, (b) $\Leftrightarrow$ (c) в обеих теоремах.

Наконец, покажем, что во второй теореме (c) $\Leftrightarrow$ (d) $\Leftrightarrow$ (e). Для этого надо лишь заметить, что $H(\alpha;\,P,\,\widetilde{P})=\widetilde{E}\left(\frac{z}{\tilde{z}}\right)^{\alpha}I(\tilde{z}>0)$ и $\widetilde{P}(\tilde{z}>0)=1$. Значит, для каждого $\alpha\in(0,\,1)$ $\widetilde{P}(z>0)=0\Leftrightarrow H(\alpha;\,P,\,\widetilde{P})=0$, откуда и следуют импликации (c) $\Leftrightarrow$ (d) $\Leftrightarrow$ (e). $\square$

Пример 1. Пусть $P=P_1\times P_2\times\ldots$, $\widetilde{P}=\widetilde{P}_1\times\widetilde{P}_2\times\ldots$, где P_k и $\widetilde{P}_k$ — гауссовские меры на $(R,\,\mathscr{B}(R))$ с плотностями

$$p_k(x)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{(x-a_k)^2}{2}}, \qquad \tilde{p}_k(x)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{(x-\tilde{a}_k)^2}{2}}.$$

Поскольку, как показывает простой подсчет,

$$H(\alpha;\,P,\,\widetilde{P})=\prod_{k=1}^{\infty} H(\alpha;\,P_k,\,\widetilde{P}_k),$$

где

$$H(\alpha;\,P_k,\,\widetilde{P}_k)=\int_{-\infty}^{\infty} p_k^{\alpha}(x)\,\tilde{p}_k^{1-\alpha}(x)\,dx=e^{-\frac{\alpha(1-\alpha)}{2}(a_k-\tilde{a}_k)^2},$$

то

$$H(\alpha;\,P,\,\widetilde{P})=e^{-\frac{\alpha(1-\alpha)}{2}\sum\limits_{k=1}^{\infty}(a_k-\tilde{a}_k)^2}.$$

Из теорем 2 и 3 получаем, что

$$\widetilde{P}\ll P \;\Leftrightarrow\; P\ll\widetilde{P} \;\Leftrightarrow\; \widetilde{P}\sim P \;\Leftrightarrow\; \sum_{k=1}^{\infty}(a_k-\tilde{a}_k)^2<\infty,$$

$$\widetilde{P}\perp P \;\Leftrightarrow\; \sum_{k=1}^{\infty}(a_k-\tilde{a}_k)^2=\infty.$$

Пример 2. Снова пусть $P = P_1 \times P_2 \times \ldots$, $\widetilde{P} = \widetilde{P}_1 \times \widetilde{P}_2 \times \ldots$, где P_k и $\widetilde{P}_k$ — распределения Пуассона с параметрами $\lambda_k > 0$ и $\widetilde{\lambda}_k > 0$ соответственно. Тогда нетрудно показать, что

$$\begin{aligned} \widetilde{P} \ll P \;\Leftrightarrow\; P \ll \widetilde{P} \;\Leftrightarrow\; \widetilde{P} \sim P \;&\Leftrightarrow\; \sum_{k=1}^{\infty} (\sqrt{\lambda_k} - \sqrt{\widetilde{\lambda}_k})^2 < \infty, \\ \widetilde{P} \perp P \;&\Leftrightarrow\; \sum_{k=1}^{\infty} (\sqrt{\lambda_k} - \sqrt{\widetilde{\lambda}_k})^2 = \infty. \end{aligned} \tag{30}$$

4. Задачи.

1. В обозначениях леммы 2 положим

$$P \wedge \widetilde{P} = E_Q(z \wedge \tilde{z}),$$

где $z \wedge \tilde{z} = \min(z, \tilde{z})$. Показать, что

$$\|P - \widetilde{P}\| = 2(1 - P \wedge \widetilde{P})$$

(и, следовательно, $\mathscr{E}r(P, \widetilde{P}) = P \wedge \widetilde{P}$).

2. Пусть $P, P_n, n \geqslant 1,$ — вероятностные меры на $(R, \mathscr{B}(R))$ с плотностями (относительно лебеговской меры) $p(x)$, $p_n(x)$, $n \geqslant 1$. Пусть $p_n(x) \to p(x)$ для почти всех x по мере Лебега. Показать, что тогда

$$\|P - P_n\| = \int_{-\infty}^{\infty} |p(x) - p_n(x)|\, dx \to 0, \quad n \to \infty$$

(ср. с задачей 17 в § 6 гл. II).

3. Пусть P и $\widetilde{P}$ — две вероятностные меры. Определим *информацию Кульбака* $K(P, \widetilde{P})$ — информацию в пользу P против $\widetilde{P}$ — равенством

$$K(P, \widetilde{P}) = \begin{cases} E \ln \dfrac{dP}{d\widetilde{P}}, & \text{если } P \ll \widetilde{P}, \\ \infty & \text{в противном случае.} \end{cases}$$

Показать, что

$$K(P, \widetilde{P}) \geqslant -2 \ln(1 - \rho^2(P, \widetilde{P})) \geqslant 2\rho^2(P, \widetilde{P}).$$

4. Доказать формулы (11), (12).

5. Доказать неравенства (24).

6. Пусть P, $\widetilde{P}$, Q — вероятностные меры на $(R, \mathscr{B}(R))$, $P * Q$ и $\widetilde{P} * Q$ — их свертка (см. п. 4 § 8 гл. II). Тогда

$$\|P * Q - \widetilde{P} * Q\| \leqslant \|P - \widetilde{P}\|.$$

7. Доказать (30).

8. Пусть ξ и η — случайные элементы на $(\Omega, \mathscr{F}, \ )$ со значениями в измеримом пространстве $(E, \mathscr{E})$. Показать, что

$$|\ \{\xi \in A\} - \ \{\eta \in A\}| \leqslant \ (\xi \neq \eta), \quad A \in \mathscr{E}.$$

§ 10. Контигуальность (сближаемость) и полная асимптотическая разделимость вероятностных мер

1. Эти понятия играют фундаментальную роль в *асимптотической* теории математической статистики, являясь естественным распространением понятий абсолютной непрерывности и сингулярности двух мер на случай *последовательностей* пар мер.

Начнем с определений.

Пусть $(\Omega^n, \mathscr{F}^n)_{n\geqslant 1}$ — некоторая последовательность измеримых пространств и $(P^n)_{n\geqslant 1}$, $(\widetilde{P}^n)_{n\geqslant 1}$ — последовательности вероятностных мер, где P^n и $\widetilde{P}^n$ определены на $(\Omega^n, \mathscr{F}^n)$, $n \geqslant 1$.

Определение 1. Говорят, что последовательность мер $(\widetilde{P}^n)$ *контигуальна* последовательности (P^n) (обозначение: $(\widetilde{P}^n) \lhd (P^n)$), если для всех $A^n \in \mathscr{F}^n$ таких, что $P^n(A^n) \to 0$, $n \to \infty$, имеем $\widetilde{P}^n(A^n) \to 0$, $n \to \infty$.

Определение 2. Говорят, что последовательности мер $(\widetilde{P}^n)$ и (P^n) *полностью асимптотически разделимы* (для краткости — разделимы; обозначение: $(\widetilde{P}^n) \vartriangle (P^n)$), если существуют подпоследовательность $n_k \uparrow \infty$, $k \to \infty$, и множества $A^{n_k} \in \mathscr{F}^{n_k}$ такие, что

$$P^{n_k}(A^{n_k}) \to 1 \text{ и } \widetilde{P}^{n_k}(A^{n_k}) \to 0, \quad k \to \infty.$$

Сразу отметим, что разделимость есть понятие *симметричное*: $(\widetilde{P}^n) \vartriangle (P^n) \Leftrightarrow (P^n) \vartriangle (\widetilde{P}^n)$. Контигуальность этим свойством не обладает. Если $(\widetilde{P}^n) \lhd (P^n)$ и $(P^n) \rhd (\widetilde{P}^n)$, то пишут $(\widetilde{P}^n) \lhd\rhd (P^n)$ и говорят, что последовательности мер (P^n) и $(\widetilde{P}^n)$ *взаимно контигуальны*.

Отметим, что в случае, когда $(\Omega^n, \mathscr{F}^n) = (\Omega, \mathscr{F})$, $P^n = P$, $\widetilde{P}^n = \widetilde{P}$ для всех $n \geqslant 1$, имеем

$$(\widetilde{P}^n) \lhd (P^n) \Leftrightarrow \widetilde{P} \ll P, \tag{1}$$

$$(\widetilde{P}^n) \lhd\rhd (P^n) \Leftrightarrow \widetilde{P} \sim P, \tag{2}$$

$$(\widetilde{P}^n) \vartriangle (P^n) \Leftrightarrow \widetilde{P} \perp P. \tag{3}$$

Эти свойства и данные выше определения объясняют, почему контигуальность и полная асимптотическая разделимость часто трактуются как «асимптотическая абсолютная непрерывность» и «асимптотическая сингулярность» для последовательностей $(\widetilde{P}^n)$ и (P^n).

2. Приводимые ниже теоремы 1 и 2 являются естественным распространением теорем 2 и 3 из § 9 на случай последовательностей мер.

Пусть $(\Omega^n, \mathscr{F}^n)_{n\geqslant 1}$ — последовательность измеримых пространств, Q^n — вероятностная мера на $(\Omega^n, \mathscr{F}^n)$, $n\geqslant 1$, и ξ^n — случайные величины (вообще говоря, расширенные; см. § 4 гл. II) на $(\Omega^n, \mathscr{F}^n)$, $n\geqslant 1$.

Определение 3. Последовательность случайных величин (ξ^n) называется *плотной* относительно последовательности мер (Q^n) (обозначение: $(\xi^n\,|\,Q^n)$ *плотно*), если

$$\lim_{N\uparrow\infty}\varlimsup_{n} Q^n(|\xi^n|>N)=0. \tag{4}$$

(Ср. с соответствующим определением плотности семейства вероятностных мер в § 2.)

Далее всюду будем полагать

$$Q^n=\frac{P^n+\widetilde{P}^n}{2},\qquad z^n=\frac{dP^n}{dQ^n},\qquad \tilde{z}^n=\frac{d\widetilde{P}^n}{dQ^n}.$$

Будем обозначать также

$$Z^n=\frac{\tilde{z}^n}{z^n} \tag{5}$$

производную Лебега меры $\widetilde{P}^n$ относительно меры P^n (см. формулу (29) в § 9), считая $\dfrac{2}{0}=\infty$. Заметим, что если $\widetilde{P}^n\ll P^n$, то Z^n есть в точности один из вариантов плотности $\dfrac{d\widetilde{P}^n}{dP^n}$ меры $\widetilde{P}^n$ относительно P^n (см. § 6 гл. II).

Для дальнейшего полезно отметить, что поскольку

$$P^n\Big(z^n\leqslant\frac{1}{N}\Big)=E_{Q^n}\Big(z^n I\Big(z^n\leqslant\frac{1}{N}\Big)\Big)\leqslant\frac{1}{N} \tag{6}$$

и $Z^n\leqslant\dfrac{2}{z^n}$, то

$$\Big(\frac{1}{z^n}\,\Big|\,P^n\Big)\ \text{плотно},\ \ (Z^n\,|\,P^n)\ \text{плотно}. \tag{7}$$

Теорема 1. *Следующие условия являются эквивалентными*:
(a) $(\widetilde{P}^n)\lhd(P^n)$,
(b) $\Big(\dfrac{1}{z^n}\,\Big|\,\widetilde{P}^n\Big)$ *плотно*,
(b′) $(Z^n\,|\,\widetilde{P}^n)$ *плотно*,
(c) $\lim\limits_{\alpha\downarrow 0}\varliminf\limits_{n} H(\alpha;\,P^n,\,\widetilde{P}^n)=1$.

Теорема 2. *Следующие условия являются эквивалентными*:
(a) $(\widetilde{P}^n)\,\triangle\,(P^n)$,
(b) $\varliminf\limits_{n}\widetilde{P}^n(z^n\geqslant\varepsilon)=0$ *для всякого* $\varepsilon>0$,

(b′) $\overline{\lim\limits_n}\, \widetilde{P}^n(Z^n \leqslant N) = 0$ *для всякого* $N > 0$,

(c) $\lim\limits_{\alpha\downarrow 0} \underline{\lim\limits_n}\, H(\alpha;\, P^n,\, \widetilde{P}^n) = 0$,

(d) $\underline{\lim\limits_n}\, H(\alpha;\, P^n,\, \widetilde{P}^n) = 0$ *для всех* $\alpha \in (0,\, 1)$,

(e) $\underline{\lim\limits_n}\, H(\alpha;\, P^n,\, \widetilde{P}^n) = 0$ *для некоторого* $\alpha \in (0,\, 1)$.

Доказательство теоремы 1.

(a) $\Rightarrow$ (b). Если (b) не выполнено, то существуют $\varepsilon > 0$ и последовательность $n_k \uparrow \infty$ такие, что $\widetilde{P}^{n_k}\Big(z^{n_k} < \dfrac{1}{n_k}\Big) \geqslant \varepsilon$. Но в силу (6) $P^{n_k}\Big(z^{n_k} < \dfrac{1}{n_k}\Big) \leqslant \leqslant \dfrac{1}{n_k} \to 0$, $k \to \infty$, что противоречит предположению $(\widetilde{P}^n) \lhd (P^n)$.

(b) $\Leftrightarrow$ (b′). Достаточно лишь заметить, что $Z^n = \dfrac{2}{z^n} - 1$.

(b) $\Rightarrow$ (a). Пусть $A^n \in \mathscr{F}^n$ и $P^n(A^n) \to 0$, $n \to \infty$. Имеем

$$\widetilde{P}^n(A^n) \leqslant \widetilde{P}^n(z^n \leqslant \varepsilon) + E_{Q^n}(\tilde{z}^n I(A^n \cap \{z^n > \varepsilon\})) \leqslant$$
$$\leqslant \widetilde{P}^n(z^n \leqslant \varepsilon) + \frac{2}{\varepsilon}\, E_{Q^n}(z^n I(A^n)) = \widetilde{P}^n(z^n \leqslant \varepsilon) + \frac{2}{\varepsilon}\, P^n(A^n).$$

Значит,

$$\overline{\lim_n}\, \widetilde{P}^n(A^n) \leqslant \overline{\lim_n}\, \widetilde{P}^n(z^n \leqslant \varepsilon), \qquad \varepsilon > 0.$$

Утверждение (b) равносильно тому, что $\lim\limits_{\varepsilon\downarrow 0} \overline{\lim\limits_n}\, \widetilde{P}^n(z^n \leqslant \varepsilon) = 0$. Тем самым $\widetilde{P}^n(A^n) \to 0$, т. е. (b) $\Rightarrow$ (a).

(b) $\Rightarrow$ (c). Пусть $\varepsilon > 0$. Тогда

$$H(\alpha;\, P^n,\, \widetilde{P}^n) = E_{Q^n}[(z^n)^\alpha (\tilde{z}^n)^{1-\alpha}] \geqslant E_{Q^n}\Big[\Big(\frac{z^n}{\tilde{z}^n}\Big)^\alpha I(z^n \geqslant \varepsilon) I(\tilde{z}^n > 0)\tilde{z}^n\Big] =$$
$$= E_{\widetilde{P}^n}\Big[\Big(\frac{z^n}{\tilde{z}^n}\Big)^\alpha I(z^n \geqslant \varepsilon)\Big] \geqslant \Big(\frac{\varepsilon}{2}\Big)^\alpha \widetilde{P}^n(z^n \geqslant \varepsilon), \quad (8)$$

поскольку $z^n + \tilde{z}^n = 2$. Значит, для $\varepsilon > 0$

$$\underline{\lim_{\alpha\downarrow 0}}\, \underline{\lim_n}\, H(\alpha;\, P^n,\, \widetilde{P}^n) \geqslant \underline{\lim_{\alpha\downarrow 0}} \Big(\frac{\varepsilon}{2}\Big)^\alpha\, \underline{\lim_n}\, \widetilde{P}(z^n \geqslant \varepsilon) = \underline{\lim_n}\, \widetilde{P}^n(z^n \geqslant \varepsilon). \qquad (9)$$

В силу (b) $\underline{\lim\limits_{\varepsilon\downarrow 0}}\, \underline{\lim\limits_n}\, \widetilde{P}^n(z^n \geqslant \varepsilon) = 1$. Поэтому из (9) и того, что $H(\alpha;\, P^n,\, \widetilde{P}^n) \leqslant 1$, следует утверждение (c).

(c) $\Rightarrow$ (b). Пусть $\delta \in (0,\ 1)$. Тогда

$$
\begin{aligned}
&H(\alpha;\ P^n,\ \widetilde{P}^n) = E_{Q^n}[(z^n)^\alpha(\tilde{z}^n)^{1-\alpha}I(z^n<\varepsilon)] + \\
&+E_{Q^n}[(z^n)^\alpha(\tilde{z}^n)^{1-\alpha}I(z^n\geqslant\varepsilon,\ \tilde{z}^n\leqslant\delta)] + E_{Q^n}[(z^n)^\alpha(\tilde{z}^n)^{1-\alpha}I(z^n\geqslant\varepsilon,\ \tilde{z}^n>\delta)] \leqslant \\
&\qquad\leqslant 2\varepsilon^\alpha + 2\delta^{1-\alpha} + \quad {}_{Q^n}\left[\tilde{z}^n\left(\frac{z^n}{\tilde{z}^n}\right)^\alpha I(z^n\geqslant\varepsilon,\ \tilde{z}^n>\delta)\right] \leqslant \\
&\qquad\qquad\leqslant 2\varepsilon^\alpha + 2\delta^{1-\alpha} + \left(\frac{2}{\delta}\right)^\alpha \widetilde{P}^n(z^n\geqslant\varepsilon).
\end{aligned} \tag{10}
$$

Поэтому

$$
\varliminf_{\varepsilon\downarrow 0}\ \varliminf_{n}\ \widetilde{P}^n(z^n\geqslant\varepsilon) \geqslant \left(\frac{\delta}{2}\right)^\alpha\ \varliminf_{n} H(\alpha;\ P^n,\ \widetilde{P}^n) - \frac{2}{2^\alpha}\delta
$$

для всех $\alpha\in(0,\ 1)$, $\delta\in(0,\ 1)$. Полагая сначала $\alpha\downarrow 0$, используя (c) и затем беря $\delta\downarrow 0$, получаем

$$
\varliminf_{\varepsilon\downarrow 0}\ \varliminf_{n}\ \widetilde{P}^n(z^n\geqslant\varepsilon)\geqslant 1,
$$

откуда вытекает справедливость (b). $\square$

Доказательство теоремы 2.

(a) $\Rightarrow$ (b). Пусть $(\widetilde{P}^n)\vartriangle(P^n)$, $n_k\uparrow\infty$, и $A^{n_k}\in\mathscr{F}^{n_k}$ таковы, что $P^{n_k}(A^{n_k})\to 1$ и $\widetilde{P}^{n_k}(A^{n_k})\to 0$. Тогда с учетом того, что $z^n+\tilde{z}^n=2$, имеем

$$
\begin{aligned}
\widetilde{P}^{n_k}(z^{n_k}\geqslant\varepsilon) &\leqslant \widetilde{P}^{n_k}(A^{n_k}) + E_{Q^{n_k}}\left\{z^{n_k}\cdot\frac{\tilde{z}^{n_k}}{z^{n_k}}\ I(\bar{A}^{n_k})I(z^{n_k}\geqslant\varepsilon)\right\} = \\
&= \widetilde{P}^{n_k}(A^{n_k}) + M_{P^{n_k}}\left\{\frac{\tilde{z}^{n_k}}{z^{n_k}}\ I(\bar{A}^{n_k})I(z^{n_k}\geqslant\varepsilon)\right\} \leqslant \widetilde{P}^{n_k}(A^{n_k}) + \frac{2}{\varepsilon}\ P^{n_k}(\bar{A}^{n_k}).
\end{aligned}
$$

Следовательно, $\widetilde{P}^{n_k}(z^{n_k}\geqslant\varepsilon)\to 0$ и, значит, выполнено (b).

(b) $\Rightarrow$ (a). Если выполнено (b), то существует последовательность $n_k\uparrow\infty$ такая, что

$$
\widetilde{P}^{n_k}\left(z^{n_k}\geqslant\frac{1}{k}\right)\leqslant\frac{1}{k}\to 0, \qquad k\to\infty.
$$

Отсюда, заметив (см. (6)), что $P^{n_k}\ \left(z^{n_k}\geqslant\dfrac{1}{k}\right)\geqslant 1-\dfrac{1}{k}$, получаем утверждение (a).

(b) $\Leftrightarrow$ (b$'$). Достаточно лишь заметить, что $Z^n=\dfrac{2}{z^n}-1$.

(b) $\Rightarrow$ (d). В силу (10) и (b)

$$
\varliminf_{n} H(\alpha;\ P^n,\ \widetilde{P}^n)\leqslant 2\varepsilon^\alpha+2\delta^{1-\alpha}
$$

для произвольных ε и δ из интервала (0, 1). Поэтому (d) имеет место.

(d) $\Rightarrow$ (c) и (d) $\Rightarrow$ (e) очевидны.

Наконец, из (8) имеем

$$\varliminf_n \widetilde{P}^n(z^n \geqslant \varepsilon) \leqslant \left(\frac{2}{\varepsilon}\right)^{\alpha} \varliminf_n H(\alpha;\, P^n,\, \widetilde{P}^n).$$

Поэтому (c) $\Rightarrow$ (b) и (e) $\Rightarrow$ (b), поскольку $\left(\dfrac{2}{\varepsilon}\right)^{\alpha} \to 1$, $\alpha \downarrow 0$. $\square$

3. Рассмотрим один частный случай, соответствующий схеме *независимых* наблюдений, где вычисление интегралов $H(\alpha;\, P^n,\, \widetilde{P}^n)$ и применение теорем 1 и 2 не представляет больших трудностей.

Предположим, что меры P^n и $\widetilde{P}^n$ есть *прямые произведения* мер:

$$P^n = P_1 \times \ldots \times P_n, \quad \widetilde{P}^n = \widetilde{P}_1 \times \ldots \times \widetilde{P}_n, \quad n \geqslant 1,$$

где P_k и $\widetilde{P}_k$ заданы на $(\Omega_k,\, \mathscr{F}_k)$, $k \geqslant 1$.

Поскольку в этом случае

$$H(\alpha;\, P^n,\, \widetilde{P}^n) = \prod_{k=1}^n H(\alpha;\, P_k,\, \widetilde{P}_k) = e^{\sum_{k=1}^n \ln[1-(1-H(\alpha; P_k, \widetilde{P}_k))]},$$

то из теорем 1 и 2 получаем следующий результат:

$$(\widetilde{P}^n) \lhd (P^n) \;\Leftrightarrow\; \lim_{\alpha\downarrow 0} \varlimsup_n \sum_{k=1}^n [1 - H(\alpha;\, P_k,\, \widetilde{P}_k)] = 0, \tag{11}$$

$$(\widetilde{P}^n) \vartriangle (P^n) \;\Leftrightarrow\; \varlimsup_n \sum_{k=1}^n [1 - H(\alpha;\, P_k,\, \widetilde{P}_k)] = \infty. \tag{12}$$

Пример. Пусть $(\Omega_k,\, \mathscr{F}_k) = (R,\, \mathscr{B}(R))$, $a_k \in [0,\, 1)$,

$$P_k(dx) = I_{[0,1]}(x)\, dx, \quad \widetilde{P}_k(dx) = \frac{1}{1-a_k}\, I_{[a_k,1]}(x)\, dx.$$

Поскольку здесь $H(\alpha;\, P_k,\, \widetilde{P}_k) = (1-a_k)^{\alpha}$, $\alpha \in (0,\, 1)$, то из (11) и того факта, что $H(\alpha;\, P_k,\, \widetilde{P}_k) = H(1-\alpha;\, \widetilde{P}_k,\, P_k)$, находим

$$(\widetilde{P}^n) \lhd (P^n) \;\Leftrightarrow\; \varlimsup_n n a_n < \infty, \text{ т. е. } a_n = O\left(\frac{1}{n}\right),$$

$$(P^n) \lhd (\widetilde{P}^n) \;\Leftrightarrow\; \varlimsup_n n a_n = 0, \text{ т. е. } a_n = o\left(\frac{1}{n}\right),$$

$$(\widetilde{P}^n) \vartriangle (P^n) \;\Leftrightarrow\; \varlimsup_n n a_n = \infty.$$

4. Задачи.

1. Пусть $P^n = P_1^n \times \ldots \times P_n^n$, $\widetilde{P}^n = \widetilde{P}_1^n \times \ldots \times \widetilde{P}_n^n$, $n \geqslant 1$, где P_k^n и $\widetilde{P}_k^n$ — *гауссовские меры* с параметрами $(a_k^n,\, 1)$ и $(\tilde{a}_k^n,\, 1)$. Найти условия на (a_k^n) и $(\tilde{a}_k^n)$, при которых $(\widetilde{P}^n) \lhd (P^n)$, $(\widetilde{P}^n) \vartriangle (P^n)$.

2. Пусть $P^n = P_1^n \times \ldots \times P_n^n$, $\widetilde{P}^n = \widetilde{P}_1^n \times \ldots \times \widetilde{P}_n^n$, где P_k^n и $\widetilde{P}_k^n$ — вероятностные меры на $(R, \mathscr{B}(R))$ такие, что $P_k^n(dx) = I_{[0,1]}(x)\,dx$, $\widetilde{P}_k^n(dx) = I_{[a_n, 1+a_n]}(x)\,dx$, $0 \leqslant a_n \leqslant 1$. Показать, что $H(\alpha;\, P_k^n, \widetilde{P}_k^n) = 1 - a_n$ и

$$(\widetilde{P}^n) \lhd (P^n) \;\Leftrightarrow\; (P^n) \lhd (\widetilde{P}^n) \;\Leftrightarrow\; \varlimsup_n n a_n = 0,$$
$$(\widetilde{P}^n) \bigtriangleup (P^n) \;\Leftrightarrow\; \varlimsup_n n a_n = \infty.$$

3. Пусть $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0})$ — фильтрованное измеримое пространство, т. е. измеримое пространство $(\Omega, \mathscr{F})$ с введенным на нем потоком σ-алгебр $(\mathscr{F}_n)_{n\geqslant 0}$ таких, что $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \ldots \subseteq \mathscr{F}$. Предположим, что $\mathscr{F} = \sigma(\bigcup_n \mathscr{F}_n)$. Пусть P и $\widetilde{P}$ — две вероятностные меры на $(\Omega, \mathscr{F})$ и $P_n = P|\mathscr{F}_n$, $\widetilde{P}_n = \widetilde{P}|\mathscr{F}_n$ — их сужения на $\mathscr{F}_n$. Показать, что

$$(P_n) \lhd (P_n) \;\Leftrightarrow\; P \ll P,$$
$$(\widetilde{P}_n) \lhd\rhd (P_n) \;\Leftrightarrow\; \widetilde{P} \sim P,$$
$$(\widetilde{P}_n) \bigtriangleup (P_n) \;\Leftrightarrow\; \widetilde{P} \perp P.$$

§ 11. О скорости сходимости в центральной предельной теореме

1. Пусть $\xi_{n1}, \ldots, \xi_{nn}$ — последовательность независимых случайных величин, $S_n = \xi_{n1} + \ldots + \xi_{nn}$, $F_n(x) = \ \{S_n \leqslant x\}$, $n \geqslant 1$. Если $S_n \xrightarrow{d} \mathscr{N}(0, 1)$, то для всякого $x \in R$ $F_n(x) \to \Phi(x)$. Поскольку функция $\Phi(x)$ *непрерывна*, то на самом деле сходимость здесь *равномерная* (задача 5 в § 1):

$$\sup_x |F_n(x) - \Phi(x)| \to 0, \quad n \to \infty. \tag{1}$$

Естественно поставить вопрос о *скорости сходимости* в (1). Приведем соответствующий результат для того случая, когда

$$S_n = \frac{\xi_1 + \ldots + \xi_n}{\sigma\sqrt{n}}, \quad n \geqslant 1,$$

где $\xi_1, \xi_2, \ldots$ — последовательность *независимых одинаково распределенных* случайных величин с $\ \xi_k = 0$, $\ \xi_k = \sigma^2 > 0$ и $\ |\xi_1|^3 < \infty$.

Теорема (Берри и Эссеен). *Имеет место оценка*

$$\sup_x |F_n(x) - \Phi(x)| \leqslant \frac{C\ \ |\xi_1|^3}{\sigma^3\sqrt{n}}, \tag{2}$$

где C — абсолютная константа $((2\pi)^{-1/2} \leqslant C < 0{,}7655)$.

Доказательство. Для простоты пусть $\sigma^2 = 1$ и $\beta_3 = \quad |\xi_1|^3$. Согласно *неравенству Эссеена* (п. 10 в § 12 гл. II),

$$\sup_x |F_n(x) - \Phi(x)| \leqslant \frac{2}{\pi} \int_0^T \left| \frac{f_n(t) - \varphi(t)}{t} \right| dt + \frac{24}{\pi T} \cdot \frac{1}{\sqrt{2\pi}}, \tag{3}$$

где $\varphi(t) = e^{-\frac{t^2}{2}}$ и

$$f_n(t) = \left[f\left(\frac{t}{\sqrt{n}} \right) \right]^n$$

с $f(t) = \quad e^{it\xi_1}$.

В (3) положительное число T можно выбирать произвольным образом. Возьмем

$$T = \frac{\sqrt{n}}{5\beta_3}.$$

Как будет показано ниже, при таком выборе T

$$|f_n(t) - \varphi(t)| \leqslant \frac{7}{6} \frac{\beta_3}{\sqrt{n}} |t|^3 e^{-\frac{t^2}{4}}, \quad |t| \leqslant T. \tag{4}$$

С учетом этого неравенства из (3) сразу получаем требуемую оценку (2), где C — некоторая абсолютная константа. (Более тонкие рассмотрения показывают, что ее значение меньше 0,7655; см. [88, гл. 5, § 4.3].)

Итак, перейдем к доказательству неравенства (4).

Согласно формуле (18) из § 12 гл. II ($n = 3$, $\quad \xi_1 = 0$, $\quad \xi_1^2 = 1$, $\quad |\xi_1|^3 < \infty$), имеем

$$f(t) = \quad e^{it\xi_1} = 1 - \frac{t^2}{2} + \frac{(it)^3}{6} [\quad \xi_1^3 (\cos \theta_1 t \xi_1 + i \sin \theta_2 t \xi_1)], \tag{5}$$

где $|\theta_1| \leqslant 1$, $|\theta_2| \leqslant 1$. Поэтому

$$f\left(\frac{t}{\sqrt{n}} \right) = 1 - \frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}} \left[\quad \xi_1^3 \left(\cos \theta_1 \frac{t}{\sqrt{n}} \xi_1 + i \sin \theta_2 \frac{t}{\sqrt{n}} \xi_1 \right) \right].$$

Если $|t| \leqslant T = \sqrt{n}/(5\beta_3)$, то с учетом неравенства $\beta_3 \geqslant \sigma^3 = 1$ (см. (28) § 6 гл. II) находим, что

$$1 - \left| f\left(\frac{t}{\sqrt{n}} \right) \right| \leqslant \left| 1 - f\left(\frac{t}{\sqrt{n}} \right) \right| \leqslant \frac{t^2}{2n} + \frac{|t|^3 \beta_3}{3n^{3/2}} \leqslant \frac{1}{25}.$$

Следовательно, при $|t| \leqslant T$ возможно представление

$$\left[f\left(\frac{t}{\sqrt{n}} \right) \right]^n = e^{n \ln f\left(\frac{t}{\sqrt{n}} \right)}, \tag{6}$$

где под $\ln z$ понимается *главное значение* логарифма комплексного числа z ($\ln z = \ln |z| + i \arg z$, $-\pi < \arg z \leqslant \pi$).

Поскольку $\beta_3 < \infty$, то по формуле Тейлора с остаточным членом в форме Лагранжа (ср. также с (35) из § 12 гл. II) видим, что

$$\ln f\left(\frac{t}{\sqrt{n}}\right) = \frac{it}{\sqrt{n}} s^{(1)}_{\xi_1} + \frac{(it)^2}{2n} s^{(2)}_{\xi_1} + \frac{(it)^3}{6n^{3/2}} (\ln f)'''\left(\theta \frac{t}{\sqrt{n}}\right) =$$

$$= -\frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}} (\ln f)'''\left(\theta \frac{t}{\sqrt{n}}\right), \quad |\theta| \leqslant 1, \quad (7)$$

так как семиинварианты $s^{(1)}_{\xi_1} = \quad \xi_1 = 0$, $s^{(2)}_{\xi_1} = \sigma^2 = 1$.

Далее,

$$(\ln f(s))''' = \frac{f'''(s) f^2(s) - 3f''(s) f'(s) f(s) + 2(f'(s))^3}{f^3(s)} =$$

$$= \frac{\quad [(i\xi_1)^3 e^{i\xi_1 s}] f^2(s) - 3 \quad [(i\xi_1)^2 e^{i\xi_1 s}] \quad [(i\xi_1) e^{i\xi_1 s}] f(s) + 2 \quad [(i\xi_1) e^{i\xi_1 s}]^3}{f^3(s)},$$

Откуда с учетом того, что $\left| f\left(\frac{t}{\sqrt{n}}\right)\right| \geqslant \frac{24}{25}$ при $|t| \leqslant T$ и $|f(s)| \leqslant 1$, находим

$$\left| (\ln f)'''\left(\theta \frac{t}{\sqrt{n}}\right)\right| \leqslant \frac{\beta_3 + 3\beta_1\beta_2 + 2\beta_1^3}{\left(\frac{24}{25}\right)^3} \leqslant 7\beta_3 \quad (8)$$

($\beta_k = \quad |\xi_1|^k$, $k = 1, 2, 3$; $\beta_1 \leqslant \beta_2^{1/2} \leqslant \beta_3^{1/3}$; см. (28) § 6 гл. II).

Из (6)—(8), используя неравенство $|e^z - 1| \leqslant |z| e^{|z|}$, получаем для $|t| \leqslant T = \frac{\sqrt{n}}{5\beta_3}$, что

$$\left| \left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n - e^{-\frac{t^2}{2}} \right| = \left| e^{n \ln f\left(\frac{t}{\sqrt{n}}\right)} - e^{-\frac{t^2}{2}} \right| \leqslant$$

$$\leqslant \frac{7}{6} \frac{\beta_3 |t|^3}{\sqrt{n}} \exp\left\{ -\frac{t^2}{2} + \frac{7}{6} |t|^3 \frac{\beta_3}{\sqrt{n}} \right\} \leqslant \frac{7}{6} \frac{\beta_3 |t|^3}{\sqrt{n}} e^{-\frac{t^2}{4}}. \quad \square$$

Замечание. Отметим, что без дополнительных предположений о природе суммируемых случайных величин порядок оценки (2) и оценка $C \geqslant (2\pi)^{-1/2}$ *не* могут быть улучшены. Действительно, пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные бернуллиевские случайные величины с

$$\{\xi_k = +1\} = \quad \{\xi_k = -1\} = \frac{1}{2}.$$

В силу симметрии очевидно, что

$$2 \quad \left\{ \sum_{k=1}^{2n} \xi_k < 0 \right\} + \quad \left\{ \sum_{k=1}^{2n} \xi_k = 0 \right\} = 1$$

и, значит, по формуле Стирлинга ((6) § 2 гл. I)

$$\left| \quad \left\{\sum_{k=1}^{2n} \xi_k < 0\right\} - \frac{1}{2}\right| = \frac{1}{2} \quad \left\{\sum_{k=1}^{2n} \xi_k = 0\right\} = \frac{1}{2} \cdot C_{2n}^{n} \cdot 2^{-2n} \sim \frac{1}{2\sqrt{\pi n}} = \frac{1}{\sqrt{(2\pi)(2n)}}.$$

Отсюда, в частности, следует, что константа C, входящая в (2), не меньше чем $(2\pi)^{-1/2}$, и

$$\left\{\sum_{k=1}^{2n} \xi_k = 0\right\} \sim \frac{1}{\sqrt{\pi n}}, \qquad n \to \infty. \tag{9}$$

2. Задачи.

1. Доказать формулу (8).

2. Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с $\quad \xi_k = 0, \quad \xi_k = \sigma^2$ и $\quad |\xi_1|^3 < \infty$. Ивестно, что тогда справедлива следующая *неравномерная* оценка: для всех $-\infty < x < \infty$

$$|F_n(x) - \Phi(x)| \leqslant \frac{C \quad |\xi_1|^3}{\sigma^3 \sqrt{n}} \cdot \frac{1}{(1 + |x|)^3}.$$

Дать доказательство этого результата, по крайней мере, для бернуллиевских случайных величин.

3. Пусть $(\xi_k)_{k \geqslant 1}$ — последовательность независимых одинаково распределенных случайных величин, принимающих значения ± 1 с вероятностями $1/2$. Пусть $\varphi_2(t) = \quad e^{it\xi_1} = \frac{1}{2}(e^{it} + e^{-it})$. Показать, следуя Лапласу, что $(S_k = \xi_1 + \ldots + \xi_k)$

$$\{S_{2n} = 0\} = \frac{1}{\pi} \int_0^{\pi} \varphi_2^n(t)\, dt \sim \frac{1}{\sqrt{\pi n}}, \qquad n \to \infty.$$

4. Пусть $(\xi_k)_{k \geqslant 0}$ — последовательность независимых одинаково распределенных случайных величин, принимающих $2a + 1$ целочисленных значения $0, \pm 1, \ldots, \pm a$. Пусть $\varphi_{2a+1}(t) = \quad e^{it\xi_1} = \frac{1}{1+2a}\left(1 + 2\sum_{k=1}^{a} \cos tk\right)$.

Также, следуя Лапласу, показать, что

$$\{S_n = 0\} = \frac{1}{\pi} \int_0^{\pi} \varphi_{2a+1}^n(t)\, dt \sim \frac{\sqrt{3}}{\sqrt{2\pi(a+1)n}}, \qquad n \to \infty.$$

В частности, для $a = 1$, т. е. случая, когда ξ_k принимают *три* значения $-1, 0, 1$,

$$\{S_n = 0\} \sim \frac{\sqrt{3}}{2\sqrt{\pi n}}, \qquad n \to \infty.$$

§ 12. О скорости сходимости в теореме Пуассона

1. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые бернуллиевские случайные величины, принимающие значения 1 и 0 с вероятностями

$$\{\xi_k = 1\} = p_k, \qquad \{\xi_k = 0\} = q_k \ (= 1 - p_k), \quad 1 \leqslant k \leqslant n.$$

Обозначим $S = \xi_1 + \ldots + \xi_n$, $B = (B_0, B_1, \ldots, B_n)$ — биномиальное распределение вероятностей суммы S, где $B_k = \ \{S = k\}$. Пусть также $\Pi = (\Pi_0, \Pi_1, \ldots)$ — пуассоновское распределение с параметром λ, где

$$\Pi_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k \geqslant 0.$$

В п. 4 § 6 гл. I отмечалось, что если

$$p_1 = \ldots = p_n, \quad \lambda = np, \tag{1}$$

то имеет место следующая оценка (Ю. В. Прохоров) для *расстояния по вариации* между мерами B и Π ($B_{n+1} = B_{n+2} = \ldots = 0$):

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |B_k - \Pi_k| \leqslant C_1(\lambda) p = C_1(\lambda) \cdot \frac{\lambda}{n}, \tag{2}$$

где

$$C_1(\lambda) = 2 \min(2, \lambda). \tag{3}$$

Для случая не обязательно равных p_k таких, что $\sum\limits_{k=1}^{n} p_k = \lambda$, Л. Ле Кам показал, что

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |B_k - \Pi_k| \leqslant C_2(\lambda) \max_{1 \leqslant k \leqslant n} p_k, \tag{4}$$

где

$$C_2(\lambda) = 2 \min(9, \lambda). \tag{5}$$

Из приводимой ниже теоремы будет следовать оценка

$$\|B - \Pi\| \leqslant C_3(\lambda) \max_{1 \leqslant k \leqslant n} p_k, \tag{6}$$

в которой

$$C_3(\lambda) = 2\lambda. \tag{7}$$

Хотя $C_2(\lambda) < C_3(\lambda)$ при $\lambda > 9$, т. е. оценка (6) хуже оценки (4), мы, однако, предпочитаем дать доказательство оценки (6), поскольку оно, по существу, элементарно, в то время как стремление получить «хорошую» константу $C_2(\lambda)$ в (4) сильно усложняет технику доказательства.

2. Теорема. *Пусть* $\lambda = \sum\limits_{k=1}^{n} p_k$. *Тогда*

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |B_k - \Pi_k| \leqslant 2 \sum_{k=1}^{n} p_k^2. \tag{8}$$

Доказательство. Воспользуемся тем обстоятельством, что каждое из распределений B и Π есть *свертка* распределений:

$$\begin{aligned} B &= B(p_1) * B(p_2) * \ldots * B(p_n), \\ \Pi &= \Pi(p_1) * \Pi(p_2) * \ldots * \Pi(p_n), \end{aligned} \tag{9}$$

понимаемых как свертка соответствующих функций распределения (см. п. 4 § 8 гл. II), где $B(p_k) = (1 - p_k,\ p_k)$ — бернуллиевское распределение в точках 0 и 1, а $\Pi(p_k)$ — пуассоновское распределение, сосредоточенное в точках 0, 1, …, с параметром p_k.

Нетрудно показать, что разность $B - \Pi$ может быть представлена в виде

$$B - \Pi = R_1 + \ldots + R_n, \tag{10}$$

где

$$R_k = (B(p_k) - \Pi(p_k)) * F_k \tag{11}$$

с

$$\begin{aligned} &F_1 = \Pi(p_2) * \ldots * \Pi(p_n), \\ &F_k = B(p_1) * \ldots * B(p_{k-1}) * \Pi(p_{k+1}) * \ldots * \Pi(p_n), \quad 2 \leqslant k \leqslant n-1, \\ &F_n = B(p_1) * \ldots * B(p_{n-1}). \end{aligned}$$

В силу задачи 6 из § 9 $\|R_k\| \leqslant \|B(p_k) - \Pi(p_k)\|$. Поэтому из (10) сразу находим, что

$$\|B - \Pi\| \leqslant \sum_{k=1}^{n} \|B(p_k) - \Pi(p_k)\|. \tag{12}$$

Учитывая формулу (12) из § 9, видим, что подсчет вариаций $\|B(p_k) - \Pi(p_k)\|$ не представляет сложностей:

$$\|B(p_k) - \Pi(p_k)\| = |(1 - p_k) - e^{-p_k}| + |p_k - p_k e^{-p_k}| + \sum_{j \geqslant 2} \frac{p_k^j e^{-p_k}}{j!} =$$
$$= |(1 - p) - e^{-p_k}| + |p_k - p_k e^{-p_k}| + 1 - e^{-p_k} - p_k e^{-p_k} = 2p_k(1 - e^{-p_k}) \leqslant 2p_k^2.$$

Отсюда и из (12) получаем требуемое неравенство (8). $\square$

Следствие. *Поскольку*

$$\sum_{k=1}^{n} p_k^2 \leqslant \lambda \max_{1\leqslant k\leqslant n} p_k,$$

то имеет место оценка (6).

3. Задачи.

1. Показать, что при $\lambda_k = -\ln(1-p_k)$

$$\|B(p_k)-\Pi(\lambda_k)\| = 2(1-e^{-\lambda_k}-\lambda_k e^{-\lambda_k}) \leqslant \lambda_k^2$$

и, следовательно, $\|B-\Pi\| \leqslant \sum_{k=1}^{n} \lambda_k^2$.

2. Доказать справедливость представлений (9) и (10).

§ 13. Фундаментальные теоремы математической статистики

1. В § 7 главы I были рассмотрены некоторые задачи оценивания и построения доверительных интервалов для вероятности «успеха» по наблюдениям над случайными величинами в схеме Бернулли. Эти задачи являются типичными для *математической статистики*, занимающейся в определенном смысле *обратными* задачами теории вероятностей. Так, если в теории вероятностей основной интерес состоит в том, чтобы для заданной вероятностной модели рассчитывать те или иные вероятностные показатели (вероятности событий, распределения вероятностей и их характеристики случайных элементов, и т. д.), то в математической статистике мы интересуемся тем, как по имеющемуся *статистическому сырью* выявить (с определенной степенью надежности) ту вероятностную модель, в рамках которой *статистические* свойства эмпирических данных наиболее всего согласуются с *вероятностными* свойствами случайного механизма, порождающего эти данные.

Приводимые ниже результаты (Гливенко, Кантелли, Колмогоров и Смирнов) по праву могут быть названы *фундаментальными теоремами математической статистики* — они не только устанавливают *принципиальную возможность* извлечения из статистического сырья вероятностной информации (о функции распределения случайных величин, над которыми производятся наблюдения), но и дают возможность *оценить степень согласия* эмпирических данных с той или иной вероятностной моделью.

2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность *независимых одинаково распределенных* случайных величин, заданных на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \quad)$ и $F=F(x)$, $x\in R$, есть их функция распределения ($F(x)= \quad \{\xi_k\leqslant x\}$). Для каждого $N\geqslant 1$ введем *эмпирические функции распределения*

$$F_N(x;\,\omega)=\frac{1}{N}\sum_{k=1}^{N} I(\xi_k(\omega)\leqslant x),\qquad x\in R. \tag{1}$$

В соответствии с законом больших чисел (§ 3, теорема 2) для *каждого* $x\in R$

$$F_N(x;\,\omega)\to F(x),\qquad N\to\infty, \tag{2}$$

т. е. имеет место сходимость по -вероятности.

Из теорем 1 и 2 § 3 главы IV следует также, что для *каждого* $x\in R$ имеет место и сходимость *с вероятностью единица*: при $N\to\infty$

$$F_N(x;\,\omega)\to F(x)\quad(\quad\text{-п. н.}). \tag{3}$$

Весьма замечательно, что имеет место и более сильный результат о том, что сходимость в (3) *равномерна* по x.

Теорема 1 (Гливенко и Кантелли). *В сформулированных условиях* (*случайные*) *величины*

$$D_N(\omega)=\sup_{x\in R}|F_N(x;\,\omega)-F(x)| \tag{4}$$

сходятся к нулю с вероятностью единица:

$$(\lim_N D_N(\omega)=0)=1. \tag{5}$$

Доказательство. Пусть Q — множество рациональных чисел на R. Понятно, что

$$\sup_{r\in Q}|F_N(r;\,\omega)-F(r)|$$

есть случайная величина. И поскольку

$$D_N(\omega)=\sup_{x\in R}|F_N(x;\,\omega)-F(x)|=\sup_{r\in Q}|F_N(r;\,\omega)-F(r)|,$$

то статистика $D_N(\omega)$ также есть *случайная величина* и, следовательно, можно говорить о ее распределении.

Пусть целое $M\geqslant 2$ и $k=1, 2, \ldots, M-1$. Определим последовательность чисел

$$x_{M,k}=\min\{x\in R:\ k/M\leqslant F(x)\},$$

полагая также $x_{M,0}=-\infty$, $x_{M,M}=+\infty$.

Пусть интервал $[x_{M,k},\, x_{M,k+1}) \neq \varnothing$ и x принадлежит этому интервалу. Тогда очевидным образом

$$
\begin{aligned}
&F_N(x;\,\omega) - F(x) \leqslant F_N(x_{M,k+1}-0) - F(x_{M,k}) = \\
&\qquad = [F_N(x_{M,k+1}-0) - F(x_{M,k+1}-0)] + [F(x_{M,k+1}-0) - F(x_{M,k})] \leqslant \\
&\qquad\qquad \leqslant F_N(x_{M,k+1}-0) - F(x_{M,k+1}-0) + 1/\nu.
\end{aligned}
$$

Аналогично, снова предполагая, что $x \in [x_{M,k},\, x_{M,k+1})$, находим, что

$$F_N(x;\,\omega) - F(x) \geqslant F_N(x_{M,k};\,\omega) - F(x_{M,k}) - 1/M.$$

Тем самым, для каждого $x \in R$

$$
\begin{aligned}
&|F_N(x;\,\omega) - F(x)| \leqslant \\
&\quad \leqslant \max_{\substack{1\leqslant k\leqslant M-1\\ 1\leqslant l\leqslant M-1}} \{|F_N(x_{M,k};\,\omega) - F(x_{M,k})|,\, |F_N(x_{M,l}-0;\,\omega) - F(x_{M,l}-0)|\} + 1/M,
\end{aligned}
$$

и, значит, -п. н.

$$\varlimsup_{n\to\infty} \sup_x |F_N(x;\,\omega) - F(x)| \leqslant 1/M.$$

Отсюда в силу произвольности M получаем требуемое утверждение (5). $\square$

Теорема Гливенко и Кантелли, являющаяся одной из *фундаментальных* теорем математической статистики, утверждает, как уже отмечалось, принципиальную возможность проверки на основании наблюдений над (*независимыми одинаково распределенными*) величинами $\xi_1, \xi_2, \ldots$ того, что функцией распределения этих величин является именно функция $F = F(x)$. Иначе говоря, эта теорема гарантирует возможность установления согласия «теории и эксперимента».

3. Несмотря на отмеченную принципиальную важность этой теоремы, она не отвечает на вопрос о *скорости* сходимости величины отклонения $D_N(\omega)$ к нулю при $N \to \infty$ и, тем самым, не позволяет судить о *степени правдоподобности того*, что результаты наблюдений «идут» от (независимых) случайных величин, имеющих своей гипотетической функцией распределения функцию $F = F(x)$, $x \in R$.

Из центральной предельной теоремы следует, что для каждого *фиксированного* $x \in R$

$$\sqrt{N}(F_N(x;\,\omega) - F(x)) \xrightarrow{law} \mathscr{N}(0,\, F(x)[1 - F(x)]), \tag{6}$$

что означает сходимость распределений величин $\sqrt{N}\,F_N(x;\,\omega) - F(x))$ к нормальному распределению с нулевым средним и дисперсией $F(x)\,(1 - F(x))$.

Более интересно, конечно, получить предельное распределение для (равномерной) статистики

$$D_N(\omega) = \sup_x |F_N(x;\,\omega) - F(x)|$$

или, скажем, родственной статистики

$$D_N^+(\omega) = \sup_x (F_N(x;\,\omega) - F(x)). \tag{7}$$

Следующее наблюдение (А. Н. Колмогоров) является ключевым при отыскании предельных распределений для этих статистик.

Лемма 1. *Пусть $\mathbb{F}$ — класс непрерывных функций распределения $F = F(x)$.*

Для каждого $N \geqslant 1$ распределение вероятностей статистик $D_N(\omega)$ одно и то же для всех $F \in \mathbb{F}$. Аналогичное справедливо и для статистики $D_N^+(\omega)$.

Доказательство. Пусть $\eta_1, \eta_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, имеющих равномерное $(U = U(x))$ распределение на $[0,\,1]$: $\{\eta_1 \leqslant x\} = x$, $0 \leqslant x \leqslant 1$.

Утверждение леммы будет следовать из того, что, как оказывается, для каждой *непрерывной* функции $F = F(x)$ распределение статистики $\sup_x |F_N(x;\,\omega) - F(x)|$ совпадает с распределением $\sup_x |U_N(x;\,\omega) - U(x)|$, где функция

$$U_N(x;\,\omega) = N^{-1} \sum_{k=1}^{N} I(\eta_k(\omega) \leqslant x)$$

есть эмпирическая функция распределения величин $\eta_1, \ldots, \eta_N$.

Обозначим через A множество интервалов $I = [a,\,b]$, $-\infty < a < b < \infty$, *постоянства* функции распределения $F = F(x)$, т. е. пусть $\{\xi_1 \in I\} = 0$.

Тогда

$$D_N(\omega) = \sup_{x \in R} |F_N(x;\,\omega) - F(x)| = \sup_{x \in \bar{A}} |F_N(x;\,\omega) - F(x)|.$$

Вводя величины $\tilde{\eta}_k = F(\xi_k)$ и эмпирические функции распределения

$$U_N(x;\,\omega) = \frac{1}{N} \sum_{k=1}^{N} I(\tilde{\eta}_k(\omega) \leqslant x),$$

находим, что для $x \in \bar{A}$

$$U_N(F(x);\,\omega) = \frac{1}{N} \sum_{k=1}^{N} I(F(\xi_k(\omega)) \leqslant F(x)) = \frac{1}{N} \sum_{k=1}^{N} I(\xi_k(\omega) \leqslant x) = F_N(x;\,\omega),$$

поскольку для таких x $\{\omega\colon \xi_k(\omega)\leqslant x\}=\{\omega\colon F(\xi_k(\omega))\leqslant F(x)\}$.

Таким образом,

$$D_N(\omega)=\sup_{x\in\bar{A}}|F_N(x;\,\omega)-F(x)|=\sup_{x\in\bar{A}}|U_N(F(x);\,\omega)-F(x)|=$$

$$=\sup_{x\in R}|U_N(F(x);\,\omega)-F(x)|=\sup_{y\in(0,1)}|U_N(y;\,\omega)-y|\overset{\text{-п. н.}}{=}\sup_{y\in[0,1]}|U_N(y;\,\omega)-y|,$$

где последнее равенство ($\overset{\text{-п. н.}}{=}$) следует из того, что

$$\{\tilde\eta_1=0\}=\quad\{\tilde\eta_1=1\}=0. \tag{8}$$

Покажем теперь, что случайные величины $\tilde\eta_k$ имеют *равномерное* (на $[0,\,1]$) распределение. С этой целью обозначим (для $y\in(0,\,1)$)

$$x(y)=\inf\{x\in R\colon\ F(x)\geqslant y\}.$$

Тогда находим, что $F(x(y))=y$, $y\in(0,\,1)$, и

$$\{\tilde\eta_1\leqslant y\}=\quad\{F(\xi_1)\leqslant y\}=\quad\{F(\xi_1)\leqslant F(x(y))\}=\quad\{\xi_1\leqslant x(y)\}=F(x(y))=y.$$

Вместе с (8) это доказывает, что случайная величина $\tilde\eta_1$ (а значит, и каждая из величин $\tilde\eta_2,\,\tilde\eta_3,\,\ldots$) имеет равномерное распределение на $[0,\,1]$. $\square$

4. Установленная лемма показывает, что для отыскания предельного при $N\to\infty$ распределения статистик D_N и D_N^+ (в классе $\mathbb{F}$ непрерывных функций распределения $F=F(x)$ для наблюдений $\xi_1,\,\xi_2,\,\ldots$) достаточно сразу предположить, что $\xi_1,\,\xi_2,\,\ldots$ — это последовательность независимых одинаково распределенных случайных величин с *равномерным* распеределением на $[0,\,1]$.

Зафиксируем некоторое $N\geqslant 1$ и упорядочим (для каждого $\omega\in\Omega$) величины $\xi_1,\,\xi_2,\,\ldots,\,\xi_N$ в порядке их *возрастания*, обозначая эти новые величины (*порядковые статистики*) через $\xi_1^{(N)},\,\xi_2^{(N)},\,\ldots,\,\xi_N^{(N)}$, где

$$\xi_1^{(N)}=\min(\xi_1,\,\xi_2,\,\ldots,\,\xi_N),\ \ldots,\ \xi_N^{(N)}=\max(\xi_1,\,\xi_2,\,\ldots,\,\xi_N).$$

(Вероятность совпадения двух таких упорядоченных величин равна нулю.)

Полагая

$$U_N(y;\,\omega)=\frac{1}{N}\sum_{k=1}^{N}I(\xi_k(\omega)\leqslant y), \tag{9}$$

имеем

$$D_N(\omega)=\max_{y\in R}|U_N(y;\,\omega)-y|. \tag{10}$$

Нетрудно видеть, что *максимальное* значение правой части может достигаться лишь в точках *скачков* функции распределения $U_N(y;\,\omega)$, т. е. в точках $\xi_1^{(N)},\,\xi_2^{(N)},\,\ldots,\,\xi_N^{(N)}$. Тем самым,

$$D_N(\omega)=\max_{k\leqslant N}\left|\frac{k}{N}-\xi_k^{(N)}\right|. \tag{11}$$

Отсюда следует, что для отыскания распределения статистики $D_N(\omega)$ надо знать *совместное распределение* порядковых статистик $\xi_1^{(N)},\,\xi_2^{(N)},\,\ldots,\,\xi_N^{(N)}$.

С целью отыскания этого распределения рассмотрим последовательность $\zeta_1,\,\zeta_2,\,\ldots$ независимых экспоненциально распределенных ($\{\zeta_k>x\}=$ $=e^{-x}$) случайных величин, и пусть $S_n=\zeta_1+\zeta_2+\ldots+\zeta_n$, $n\geqslant 1$.

Лемма 2. *Совместное распределение вектора* $(\xi_1^{(N)},\,\xi_2^{(N)},\,\ldots,\,\xi_N^{(N)})$ *совпадает с совместным распределением вектора*

$$\left(\frac{S_1}{S_{N+1}},\,\frac{S_2}{S_{N+1}},\,\ldots,\,\frac{S_N}{S_{N+1}}\right).$$

Доказательство. С одной стороны, для $0\leqslant y_1\leqslant\ldots\leqslant y_N\leqslant 1$

$$\{\xi_1^{(N)}\in dy_1,\,\ldots,\,\xi_N^{(N)}\in dy_N\}=\sum\quad\{\xi_{k_1}\in dy_1,\,\ldots,\,\xi_{k_N}\in dy_N\}, \tag{12}$$

где суммирование производится по всем перестановкам $(k_1,\,\ldots,\,k_N)$ чисел $(1,\,\ldots,\,N)$. (В (12) и также в дальнейшем используется несколько вольная «дифференциальная» форма записи, которой нетрудно придать точный смысл.)

В правой части (12) величины $\xi_{k_1},\,\ldots,\,\xi_{k_N}$ независимы, равномерно распределены на $[0,\,1]$. Число перестановок чисел $(1,\,\ldots,\,N)$ равно $N!$, поэтому правая часть в (12) равна $N!\,dy_1\ldots dy_N$. Следовательно, если $(y_1,\,\ldots,\,y_n)\in\Delta_N$, где $\Delta_N=\{y_1\in[0,\,1],\,\ldots,\,y_N\in[0,\,1]:\,0\leqslant y_1\leqslant\ldots\leqslant y_N\leqslant$ $\leqslant 1\}$, то

$$\{(\xi_1^{(N)}\in dy_1,\,\ldots,\,\xi_N^{(N)}\in dy_N\}=N!\,dy_1\ldots dy_N. \tag{13}$$

Если же $(y_1,\,\ldots,\,y_n)\notin\Delta_N$, то

$$\{\xi_1^{(N)}\in dy_1,\,\ldots,\,\xi_N^{(N)}\in dy_N\}=0. \tag{14}$$

Покажем теперь, что, с другой стороны, точно так же

$$\left\{\frac{S_1}{S_{N+1}}\in dy_1,\,\ldots,\,\frac{S_N}{S_{N+1}}\in dy_N\right\}= \\ =\begin{cases}N!\,dy_1\ldots dy_N, & (y_1,\,\ldots,\,y_N)\in\Delta_N,\\ 0, & (y_1,\,\ldots,\,y_N)\notin\Delta_N.\end{cases} \tag{15}$$

С этой целью рассмотрим совместное распределение случайных величин $S_1, \ldots, S_{N+1}$.

Если $(x_1, \ldots, x_{N+1}) \in \Delta_{N+1}$, то находим, что

$$\begin{aligned}\{S_1 \in dx_1,\, S_2 \in dx_2,\, \ldots,\, S_{N+1} \in dx_{N+1}\} &= \\ = \{\zeta_1 \in dx_1,\, \zeta_2 \in dx_2 - x_1 \ldots,\, \zeta_{N+1} \in dx_{N+1} - x_N\} &= \\ = e^{-x_1} e^{-(x_2 - x_1)} \ldots e^{-(x_{N+1} - x_N)}\, dx_1 \ldots dx_{N+1} = e^{-x_N + 1} dx_1 \ldots dx_{N+1}. &\quad (16)\end{aligned}$$

Если же $(x_1, \ldots, x_{N+1}) \notin \Delta_{N+1}$, то

$$\{S_1 \in dx_1,\, S_2 \in dx_2,\, \ldots,\, S_{N+1} \in dx_{N+1}\} = 0. \tag{17}$$

Поскольку $S_{N+1} = \zeta_1 + \ldots + \zeta_{N+1}$, где $\zeta_1, \ldots, \zeta_{N+1}$ — независимые экспоненциально распределенные случайные величины, то (задача 3)

$$\{S_{N+1} \in dx_{N+1}\} = \frac{x_{N+1}^N e^{-x_{N+1}}}{N!}\, dx_{N+1}. \tag{18}$$

Из (16) и (18) находим, что для $(x_1, \ldots, x_{N+1}) \in \Delta_{N+1}$

$$(S_1 \in dx_1, \ldots, S_N \in dx_N \mid S_{N+1} \in dx_{N+1}) = N!\, x_{N+1}^{-N}\, dx_1 \ldots dx_N. \tag{19}$$

Если же $(x_1, \ldots, x_{N+1}) \notin \Delta_{N+1}$, то левая часть в (19) равна нулю.

В результате видим, что

$$\left(\frac{S_1}{S_{N+1}} \in dy_1, \ldots, \frac{S_N}{S_{N+1}} \in dy_N \,\middle|\, S_{N+1} = x_{N+1}\right) = \\ = \begin{cases} N!\, dy_1 \ldots dy_N, & (y_1, \ldots, y_N) \in \Delta_N, \\ 0, & (y_1, \ldots, y_N) \notin \Delta_N, \end{cases}$$

и в силу независимости правой части от x_{N+1} приходим к справедливости представления (15), сопоставление которого с (13), (14) доказывает утверждение леммы. □

Из этой леммы вытекает, что

$$\sqrt{N} D_N(\omega) \overset{d}{=} \sqrt{N} \max_{k \leqslant N} \left|\frac{S_k}{S_{N+1}} - \frac{k}{N}\right|, \tag{20}$$

где равенство $\overset{d}{=}$ означает совпадение *по распределению* величин в левой и правой частях.

Из (20)

$$\sqrt{N} D_N(\omega) \overset{d}{=} \frac{N}{S_{N+1}} \max_{k \leqslant N} \left|\frac{S_k - k}{\sqrt{N}} - \frac{k}{N} \frac{S_{N+1} - N}{\sqrt{N}}\right|. \tag{21}$$

Это соотношение весьма удобно для исследования предельного поведения величин $\sqrt{N} D_N(\omega)$ при $N \to \infty$.

Поскольку $\dfrac{N}{S_{N+1}} \to 1$ (-п. н.), то, полагая для $t \in [0,\ 1]$

$$X_t^{(N)} = \frac{S_{[Nt]} - [Nt]}{\sqrt{N}},$$

находим, что

$$\lim_{N\to\infty} \quad \{\sqrt{N} D_N(\omega) \leqslant y\} = \lim_{N\to\infty} \quad \Big(\sup_{0\leqslant t\leqslant 1} |X_t^{(N)} - tX_1^{(N)}| \leqslant y\Big).$$

Тем самым, вопрос о предельном распределении величин $\sqrt{N} D_N(\omega)$ сводится к изучению предельного поведения (по распределению) величин $\sup\limits_{0\leqslant t\leqslant 1} |X_t^{(N)} - tX_1^{(N)}|$ при $N \to \infty$.

При каждом *фиксированном* $t \in [0,\ 1]$ предельное распределение величин $X_t^{(N)}$ в соответствии с центральной предельной теоремой (теорема 1 в § 4) совпадает с распределением величины B_t, являющейся гауссовской с $\ \ B_t = 0$ и $\ \ B_t^2 = t$.

На самом же деле можно утверждать больше. Именно, рассмотрим введенный в § 13 главы II процесс *броуновского движения* (*винеровский процесс*) $B = (B_t)_{0\leqslant t\leqslant 1}$. Такой процесс был там определен как гауссовский процесс с $B_0 = 0$, $\ \ B_t = 0$ и ковариационной функцией $\ \ B_s B_t = \min(s,\ t)$. Так вот, оказывается, что в соответствии с замечанием 3 к теореме 1 в § 8 гл. VII *совместное распределение* $P_{t_1,\ldots,t_k}^{(N)}$ величин $(X_{t_1}^{(N)},\ \ldots,\ X_{t_k}^{(N)})$ слабо сходится к совместному распределению $P_{t_1,\ldots,t_k}$ величин $(B_{t_1},\ \ldots,\ B_{t_k})$. (О слабой сходимости см. §§ 1, 2.) Поэтому естественно ожидать, что распределение и величин $\sup\limits_{0\leqslant t\leqslant 1} |X_t^{(N)} - tX_1^{(N)}|$ сходится к распределению величины $\sup\limits_{0\leqslant t\leqslant 1} |B_t - tB_1|$.

Вообще говоря, слабой сходимости всех конечномерных распределений $P_{t_1,\ldots,t_k}^{(N)} \to P_{t_1,\ldots,t_k}$, $0 \leqslant t_1 < t_2 < \ldots < t_k = 1$, $k \geqslant 1$ (т. е. сходимости $P^{(N)} \overset{f}{\Rightarrow} P$ в обозначениях задачи 3 в § 1) еще не достаточно (см. п. 5 в § 1) для сходимости распределений функционалов $f(X^{(N)}) = \sup\limits_{0\leqslant t\leqslant 1} |X_t^{(N)} - tX_1^{(N)}|$ к функционалу $f(X) = \sup\limits_{0\leqslant t\leqslant 1} |X_t - tX_1|$, где $X_t = B_t$ — броуновское движение, $0 \leqslant t \leqslant 1$.

Однако в рассматриваемом случае это действительно так и вытекает из следующих рассмотрений.

Траектории процессов $X^{(N)}$ принадлежат пространству $D = D[0,\ 1]$, а траектории X — пространству $C = C[0,\ 1] \subseteq D$ (см. пп. 6 и 7 § 2 гл. II). В пространстве D можно ввести «метрику Прохорова» ρ (в [5, гл. 3] ρ обозначено d_0, в [87, гл. VI] $\rho = \delta$), относительно которой метрическое пространство $(D,\ \mathscr{D},\ \rho)$ становится польским, т. е. полным сепарабельным пространством. (Пространство $(D, \mathscr{D}, \rho)$ с «метрикой Скорохода» d, определенной в п. 7 § 2 гл. II, будет только сепарабельным, а для последующего применения *теоремы Прохорова* из § 2 гл. III нужна еще и «плотность».)

Относительно метрики ρ функционал

$$f(x) = \sup_{0 \leqslant t \leqslant 1} |x_t - tx_1|$$

(совпадающий, очевидно, с $\sup\limits_{0 \leqslant t < 1} |x_t - tx_1|$), где $x = (x_t)_{0 \leqslant t \leqslant 1} \in D$, является непрерывным, и поэтому для доказательства сходимости $f(X^{(N)}) \xrightarrow{d} f(X)$ (т. е. сходимости по распределению) достаточно (задача 2) доказать лишь слабую сходимость $P^{(N)} \to P$ (в $(D, \mathscr{D}, \rho)$).

Обычная процедура установления этой сходимости, непосредственно инициируемая теоремой Прохорова, основана на справедливости (задача 3; см. также обозначения в задаче 3 § 1) следующей импликации:

$$\big([P^{(N)} \overset{f}{\Rightarrow} P] \oplus [\text{плотность } \{P^{(N)}\}] \oplus \\ \oplus [\text{класс } \mathscr{K}_0(D) \text{ является определяющим}]\big) \ \Rightarrow \ (P^{(N)} \xrightarrow{w} P). \quad (22)$$

(Здесь $\mathscr{K}_0(D)$ — класс цилиндрических множеств; п. 5 § 1.)

В рассматриваемом случае имеется сходимость конечномерных распределений $P^{(N)} \overset{f}{\Rightarrow} P$ и класс цилиндрических множеств $\mathscr{K}_0(D)$ действительно является определяющим (п. 7 § 2 гл. II). Самое трудное, конечно, в применении импликации (22) — это проверка того, что рассматриваемое семейство мер $\{P^{(N)}\}$ плотно. Эта проверка основывается на характеризации *компактных* множеств в D, входящих в определение (формула (1) в § 2) плотности семейства мер, что выходит за рамки настоящей книги. (Соответствующие доказательства см. в [5, теорема 15.2] и [87, гл. VI, 3.21].)

В заключение этих рассмотрений заметим, что существует и другой путь установления сходимости $f(X^{(N)}) \xrightarrow{d} f(X)$. Состоит он в следующем. Разрывные процессы $X^{(N)}$ можно аппроксимировать *непрерывными* процессами $\widetilde{X}^{(N)}$ такими, что

$$\sup_{0 \leqslant t \leqslant 1} |X_t^{(N)}(\omega) - \widetilde{X}_t^{(N)}(\omega)| \leqslant \varepsilon$$

для всякого $\varepsilon > 0$ при всех достаточно больших N и всех $\omega \in \Omega$. Поэтому достаточно доказывать лишь сходимость $f(\widetilde{X}^{(N)}) \xrightarrow{d} f(X)$, что несколько проще, поскольку тогда можно оперировать не с пространством D, а с более простым пространством C, в котором критерии «плотности» ([5, гл. 2], [87, гл. VI], по крайней мере в рассматриваемом случае, легко проверяемы.

5. Положим $B_t^\circ = B_t - tB_1$, $0 \leqslant t \leqslant 1$. Этот процесс является гауссовским, $B_0^\circ = 0$, $B_t^\circ = 0$ и $B_s^\circ B_t^\circ = \min(s, t) - st$. В п. 7 § 13 гл. II этот процесс был назван *условным винеровским процессом* или *броуновским мостом*.

Итак, из предшествующих рассмотрений приходим к такому заключению:

$$\lim_{N\to\infty} \{\sqrt{N} D_N(\omega) \leqslant y\} = \Big\{\sup_{0\leqslant t\leqslant 1} |B_t^\circ| \leqslant y\Big\}, \tag{23}$$

где $B^\circ = (B_t^\circ)_{0\leqslant t\leqslant 1}$ — броуновский мост.

Из теории броуновского движения известно (см., например, [5, гл. 2, § 11]), что распределение

$$K(y) = \Big\{\sup_{0\leqslant t\leqslant 1} |B_t^\circ| \leqslant y\Big\},$$

называемое *распределением Колмогорова*, определяется следующим образом:

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2}, \quad y \geqslant 0. \tag{24}$$

Таким образом, справедлив следующий результат (А. Н. Колмогоров):

$$\lim_{N\to\infty} \{\sqrt{N} D_N(\omega) \leqslant y\} = K(y), \quad y \geqslant 0. \tag{25}$$

Аналогично изложенному показывается также, что

$$\lim_{N\to\infty} \{\sqrt{N} D_N^+(\omega) \leqslant y\} = \Big\{\sup_{0\leqslant t\leqslant 1} B_t^\circ \leqslant y\Big\}. \tag{26}$$

Распределение величины $\sup\limits_{0\leqslant t\leqslant 1} B_t^\circ$ проще, нежели распределение величины $\sup\limits_{0\leqslant t\leqslant 1} |B_t^\circ|$, и имеет следующий вид (см. [5, гл. 2, § 11]):

$$\Big\{\sup_{0\leqslant t\leqslant 1} B_t^\circ \leqslant y\Big\} = 1 - e^{-2y^2}, \quad y \geqslant 0. \tag{27}$$

Тем самым, имеет место следующий результат (Н. В. Смирнов):

$$\lim_{N\to\infty} \{\sqrt{N} D_N^+(\omega) \leqslant y\} = 1 - e^{-2y^2}, \quad y \geqslant 0. \tag{28}$$

6. Остановимся на том, как, скажем, знание результата (25), где $K(y)$ определяется формулой (24), позволяет дать *критерий согласия эксперимента с теорией*. С этой целью приведем сначала небольшую таблицу значений функции распределения $K(y)$:

y	$K(y)$	y	$K(y)$	y	$K(y)$
		1,10	0,822282	2,10	0,999705
0,28	0,000001	1,20	0,887750	2,20	0,999874
0,30	0,000009	1,30	0,931908	2,30	0,999949
0,40	0,002808	1,40	0,960318	2,40	0,999980
0,50	0,036055	1,50	0,977782	2,50	0,9999925
0,60	0,135718	1,60	0,988048	2,60	0,9999974
0,70	0,288765	1,70	0,993828	2,70	0,9999990
0,80	0,455857	1,80	0,996932	2,80	0,9999997
0,90	0,607270	1,90	0,998536	2,90	0,99999990
1,00	0,730000	2,00	0,999329	3,00	0,99999997

Если N достаточно велико, то можно считать, что $K(y)$ дает хорошее приближение для значений $\{\sqrt{N}D_N(\omega)\leqslant y\}$.

Естественно, что если величины $\sqrt{N}D_N(\omega)$, подсчитанные по эмпирическим значениям $\xi_1(\omega), \ldots, \xi_N(\omega)$, оказываются *большими*, то гипотезу о том, что (гипотетическое) распределение вероятностей этих величин есть именно (непрерывная) функция $F=F(x)$, *надо отклонить*.

Приведенная таблица дает представление о *степени надежности* сделанных таким образом выводов. Если, скажем, $\sqrt{N}D_N(\omega)>1{,}80$, то (поскольку $K(1{,}80)=0{,}996932$) можно утверждать, что такое событие имеет вероятность, примерно равную $0{,}003068$ $(=1{,}000000-0{,}996932)$. Считая, что события с такой малой вероятностью $(=0{,}003068)$, практически малореализуемы, мы приходим к выводу, что *гипотезу о том, что распределение* $\{\xi_1\leqslant x\}=F(x)$, где $F(x)$ — именно та функция, которая участвовала в подсчете величин $D_N(\omega)$, надо *отклонить*. Если же $\sqrt{N}D_N(\omega)\leqslant 1{,}80$, то мы можем сказать (опираясь на закон больших чисел), что согласие «эксперимента с теорией» будет выполняться в 996 932 таких случаях из 1 000 000.

Замечание. Важно подчеркнуть, что при применении критериев согласия, основанных на использовании распределения Колмогорова или распределения Смирнова, предполагается, что (тестируемая) функция распределения $F=F(x)$ *полностью* специфицирована. Эти критерии «не работают», если, скажем, известно лишь только то, что функция распределения $F=F(x)$ принадлежит *некоторому параметрическому семейству* $\mathbb{G}=\{G=G(x;\theta);\ \theta\in\Theta\}$ функций распределения $G(x;\theta)$, зависящих от параметра $\theta\in\Theta$. (При каждом θ функция $G(x;\theta)$ считается полностью определенной.) В этом случае напрашивается следующий путь проверки того, что эмпирические данные согласуются с тем, что истинная функция распределения $F \in \mathbb{F}$: сначала по N наблюдениям построить оценку $\hat{\theta}_N = \hat{\theta}_N(\omega)$ параметра θ, затем образовать величину $\sqrt{N} \sup\limits_{x \in R} |F_N(x;\, \omega) - G(x;\, \hat{\theta}_N(\omega))|$ и принимать решение так, как было проделано в приведенном выше примере. К сожалению, функция распределения $G(x;\, \hat{\theta}_N(\omega))$ будет *случайной* и распределение статистики $\sqrt{N} \sup\limits_{x \in R} |F_N(x;\, \omega) - G(x;\, \hat{\theta}_N(\omega))|$ не будет, вообще говоря, даваться распределением Колмогорова $K = K(y)$.

По поводу того, как же все-таки проверять гипотезу $F \in \mathbb{G}$, см. [132].

7. Задачи.

1. Доказать формулу (18).

2. Доказать, что $P^{(N)} \xrightarrow{w} P$ (в $(D, \mathscr{D}, \rho)$) влечет сходимость $f(X^{(N)}) \xrightarrow{d} f(X)$.

3. Доказать справедливость импликации (22).

Библиографическая справка (главы I—III)

ВВЕДЕНИЕ

История теории вероятностей до Лапласа изложена в монографии И. Тодхантера [68]. Период от Лапласа до конца XIX в. освещен в статье Б. В. Гнеденко и О. В. Шейнина, опубликованной в сборнике [45]. В книге С. Стиглера [122] дается весьма подробное изложение истории теории вероятностей и математической статистики до 1900 г. В книге Д. Е. Майстрова [44] история теории вероятностей изложена от ее возникновения до 30-х годов прошлого 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 [83], справочное пособие Ю. В. Прохорова и Ю. А. Розанова [56], справочник [65] и книги В. Феллера [69], 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], П. Биллингсли [106], Р. Эша [80], [81], Р. Эша и М. Гарднера [82], Р. Даррета [108]—[110] и Дж. Ламперти [37], являющиеся (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, в Математической Энциклопедии [121] и в энциклопедии «Вероятность и математическая статистика» [125].

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]. При современном широком использовании компьютерной техники полезно обращение к статистическим пакетам MINITAB, SPSS, … Среди пользователей весьма популярны пакеты «Mathematica®» (см. книгу [126]).

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 А. Исихары [25].

§ 3. Формула и теорема Байеса лежат в основе так называемого байесовского подхода в математической статистике. См., например, книги М. Де Гроота [18] и Ш. 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, т. 1]. 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 Б. Л. 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 в сборнике [45]. Обширный материал на эту тему содержится в книге В. Феллера [69, т. 1].

§ 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, т. 1].

§ 12. Обширный материал по марковским цепям содержится в книгах: В. Феллер [69, т. 1], Е. Б. Дынкин [21], Е. Б. Дынкин и А. А. Юшкевич [102], Чжун Кай-Лай [75], [120], Д. Ревюз [117], Дж. Кемени и Дж. Снелл [27], Т. А. Сарымсаков [61], С. Х. Сираждинов [64]. Теории ветвящихся процессов посвящена монография Б. А. Севастьянова [62].

ГЛАВА II

§ 1. Аксиоматика Колмогорова изложена в его книге [32].

§ 2. Материал об алгебрах и σ-алгебрах см. также в книгах А. Н. Колмогорова и С. В. Фомина [33], Ж. Невё [49], Л. Бреймана [8], Р. Эша [81].

§ 3. Доказательство теоремы Каратеодори см. в книгах М. Лоэва [42], П. Халмоша [70].

§ 4—5. Большой материал об измеримых функциях можно найти в книге П. Халмоша [70].

§ 6. См. также книги А. Н. Колмогорова и С. В. Фомина [33], П. Халмоша [70], Р. Эша [81]. В этих книгах содержится и доказательство теоремы Радона—Никодима. Иногда неравенством Чебышева называют неравенство

$$\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\xi^2}{\varepsilon^2},$$

а неравенство

$$\{|\xi| \geqslant \varepsilon\} \leqslant \frac{|\xi|^r}{\varepsilon^r}, \quad r > 0,$$

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 [81].

§ 8. См. также книги А. А. Боровкова [7], Р. Эша [81], 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 [81]. Приводимое здесь доказательство следует [81].

§ 10—11. См. также книги А. Н. Колмогорова и С. В. Фомина [33], Р. Эша [81], 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], Б. 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], Дж. Гриммета и Д. Стирзакера [105], Дж. Ламперти [37].

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, разнораспределены: $\{\xi_i = 1\} = p_i$, $\{\xi_i = 0\} = 1 - p_i$, $i \geqslant 1$.

§ 4. Здесь приводится традиционное доказательство центральной предельной теоремы для сумм независимых случайных величин при выполнении условия Линдеберга. Ср. с [16], [92].

§ 5. Вопросы справедливости центральной предельной теоремы без условия предельной пренебрегаемости привлекали внимание еще П. Леви. В монографии В. М. Золотарева [88] содержится подробное изложение современного состояния теории предельных теорем в *неклассической постановке*. Приводимая формулировка и доказательство теоремы 1 даны В. И. Ротарем [96].

§ 6. Изложение использует материал книг Б. В. Гнеденко и А. Н. Колмогорова [16], Р. Эша [81] и В. В. Петрова [53], [92].

§ 7. Метрика Леви—Прохорова введена в известной работе Ю. В. Прохорова [55], которому принадлежат и результаты относительно метризуемости слабой сходимости мер, заданных на метрических пространствах. По поводу метрики $\|P - \tilde{P}\|^*_{BL}$ см. статью Р. Дадли [85] и книгу Д. Полларда [93].

§ 8. Теорема 1 принадлежит А. В. Скороходу. Полезный материал относительно метода одного вероятностного пространства можно найти в учебном пособии 1999 г. А. А. Боровкова [7] и монографии Д. Полларда [93].

§ 9—10. Укажем ряд книг, в которых содержится большой материал, касающийся затрагиваемых вопросов: Ж. Жакод и А. Н. Ширяев [87], Л. Ле Кам [89], П. Е. Гринвуд и А. Н. Ширяев [84], Ф. Лизе и И. Вайда [90].

§ 11. Большой материал относительно оценок скорости сходимости в центральной предельной теореме содержится в монографии В. В. Петрова [92]. Приводимое доказательство теоремы Берри и Эссеена содержится в книге Б. В. Гнеденко и А. Н. Колмогорова [16].

§ 12. Доказательство следует статье Э. Л. Пресмана [94].

§ 13. Дополнительный материал о фундаментальных теоремах математической статистики см. в [8], [35], [58], [106], [107].

Список литературы

[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 // 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. — 3-е изд. — М.: Наука, 1983.

[7] Боровков А. А. Теория вероятностей. — 3-е изд. — М.: УРСС, 1999.

[8] Брейман (Breiman L.). Probability. — Reading, MA: 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, 1975.

[12] Гарсиа (Garsia A. M.). A simple proof of E. Hopf's maximal ergodic theorem // Journal of Mathematics and Mechanics. — 1965. — V. 14, № 3. — P. 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: В 3 т. — М.: Наука, 1971—1975.

[15] Гнеденко Б. В. Курс теории вероятностей. — 6-е изд. — М.: Наука, 1988.

[16] Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — М.; Л.: Гостехиздат, 1949.

[17] Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей. — 9-е изд. — М.: Наука, 1982.

[18] Де Гроот М. Оптимальные статистические решения. — М.: Мир, 1974.

[19] Дохерти (Doherty M.). An amusing proof in fluctuation theory // Combinatorial Mathematics, III: Proceedings of the Third Australian Conference, Univ. Queensland, St. Lucia, 1974. — Berlin etc.: Springer-Verlag, 1975. — P. 101—104. — (Lecture Notes in Mathematics; V. 452.)

[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.

Под номерами 85—101 идет литература, добавленная во втором издании к той, которая была приведена в первом издании книги. Под номерами 102—136 идет литература, добавленная в третьем издании. Под номером 137 идет книга, добавленная в настоящее издание.

[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 // Математический сборник. — 1977. — Т. 104, № 2. — С. 227—247.

[27] Кемени Дж., Снелл Дж. Конечные цепи Маркова. — М.: Наука, 1970.

[28] Колмогоров А. Н. Цепи Маркова со счетным числом возможных состояний // Бюллетень МГУ. — 1937. — Т. 1, № 3. — С. 1—16.

[29] Колмогоров А. Н. Стационарные последовательности в гильбертовском пространстве // Бюллетень МГУ. — 1941. — Т. 2, № 6. — С. 1—40.

[30] Колмогоров А. Н. Роль русской науки в развитии теории вероятностей // Ученые записки МГУ. — 1947. — Вып. 91. — С. 53—64.

[31] Колмогоров А. Н. Теория вероятностей // Математика, ее содержание, методы и значение. — М.: Изд-во АН СССР, 1956. — Т. II. — С. 252—284.

[32] Колмогоров А. Н. Основные понятия теории вероятностей. — М.; Л.: ОНТИ, 1936; 2-е изд. М.: Наука, 1974; 3-е изд. М.: Фазис, 1998.

[33] Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — 6-е изд. — М.: Наука, 1989.

[34] Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. Случайные размещения. — М.: Наука, 1976.

[35] Крамер Г. Математические методы статистики. — 2-е изд. — М.: Мир, 1976.

[36] Кубилюс Й. Вероятностные методы в теории чисел. — Вильнюс: Гос. изд-во полит. и науч. лит. ЛитССР, 1959.

[37] Ламперти Дж. Вероятность. — М.: Наука, 1973.

[38] Ламперти (Lamperti J.). Stochastic Processes. — Aarhus Univ., 1974. — (Lecture Notes Series; № 38).

[39] Ленгляр (Lenglart E.). Relation de domination entre deux processus // Annales de l'Institut H. Poincaré Sect. B. (N. S.). — 1977. — V. 13, № 2. — P. 171—179.

[40] Леонов В. П., Ширяев А. Н. К технике вычисления семиинвариантов // Теория вероятностей и ее применения. — 1959. — Т. IV, вып. 2. — С. 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. — Berlin etc.: Springer-Verlag, 1972. — (Lecture Notes in Mathematics; V. 284).
[48] Мейер П.-А. Вероятность и потенциалы. — М.: Мир, 1973.
[49] Неве Ж. Математические основы теории вероятностей. — М.: Мир, 1969.
[50] Невё (Neveu J.). Discrete-Parameter Martingales. — Amsterdam etc.: North-Holland, 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 // Известия АН СССР. Серия математическая. — 1980. — Т. 40, вып. 4. — С. 868—885.
[53] Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972.
[54] Прохоров Ю. В. Асимптотическое поведение биномиального распределения // Успехи математических наук. — 1953. — Т. VIII, вып. 3 (55). — С. 135—142.
[55] Прохоров Ю. В. Сходимость случайных процессов и предельные теоремы теории вероятностей // Теория вероятностей и ее применения. — 1956. — Т. I, вып. 2. — С. 177—238.
[56] Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. — 2-е изд. — М.: Наука, 1973.
[57] Рамачандран Б. Теория характеристических функций. — М.: Наука, 1975.
[58] Реньи (Rényi A.) Probability Theory. — Amsterdam: North-Holland, 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 Ереван. 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 etc.: Academic Press, 1974.
[67] Теорія имовірностей. — 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: Macmillan, 1865.

[69] Феллер В. Введение в теорию вероятностей и ее приложения: В 2-х т.— М.: Мир, 1984.

[70] Халмош П. Теория меры. — М.: ИЛ, 1953.

[71] Хеннан Э. Анализ временных рядов. — М.: Наука, 1964.

[72] Хеннан Э. Многомерные временные ряды. — М.: Мир, 1974.

[73] Чао, Тейчер (Chow Y. S., Teicher H.). Probability Theory. Independence, Interchangeability, Martingales. — 3rd ed. — New York: Springer-Verlag, 1997.

[74] 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.

[75] Чжун Кай-лай. Однородные цепи Маркова. — М.: Мир, 1964.

[76] Ширяев А. Н. Случайные процессы. — М.: Изд-во МГУ, 1972.

[77] Ширяев А. Н. Вероятность, статистика, случайные процессы: В 2-х т. — М.: Изд-во МГУ, 1973—1974.

[78] Ширяев А. Н. Статистический последовательный анализ. — 2-е изд. — М.: Наука, 1976.

[79] Энгельберт, Ширяев (Engelbert H.-J., Shiryaev A. N.). On the sets of convergence of generalized submartingales // Stochastics. — 1979. — V. 2, № 3. — P. 155—166.

[80] Эш (Ash R. B.). Basic Probability Theory. — New York etc.: Wiley, 1970.

[81] Эш (Ash R. B.). Real Analysis and Probability. — New York etc.: Academic Press, 1972.

[82] Эш, Гарднер (Ash R. B., Gardner M. F.). Topics in Stochastic Processes. — New York etc.: Academic Press, 1975.

[83] Яглом А. М., Яглом И. М. Вероятность и информация. — 3-е изд. — М.: Наука, 1973.

[84] Гринвуд, Ширяев (Greenwood P. E., Shiryayev A. N.). Contiguity and the Statistical Invariance Principle. — London: Gordon & Breach, 1985.

[85] Дадли (Dudley R. M.) Distances of probability measures and random variables // Annals of Mathematical Statistics. — 1968. — V. 39, № 5. — P. 1563—1572.

[86] Дакуна-Кастелль, Дюфло (Dacunha-Castelle D., Duflo M.). Probabilités et statistiques: 1, 2. — Paris: Masson. — 1: Problèmes à temps fixe. — 1982; — 2: Problèmes à temps mobile. — 1983. — Перев. на англ. яз.: Probability and Statistics: V. I, II. — Berlin etc.: Springer-Verlag, 1986.

[87] Жакод Ж., Ширяев А. Н. Предельные теоремы для случайных процессов: В 2-х т. — М.: Физматлит, 1994.

[88] Золотарев В. М. Современная теория суммирования независимых случайных величин. — М.: Наука, 1986.

[89] Ле Кам (Le Cam L.). Asymptotic Methods in Statistical Decision Theory. — Berlin etc.: Springer-Verlag, 1986.

[90] Лизе, Вайда (Liese F., Vajda I.). Convex Statistical Distances. — Leipzig: Teubner, 1987.

[91] Липцер Р. Ш., Ширяев А. Н. Теория мартингалов. — М.: Наука, 1986.

[92] Петров В. В. Предельные теоремы для сумм независимых случайных величин. — М.: Наука, 1987.

[93] Поллард (Pollard D.). Convergence of Stochastic Processes. — Berlin etc.: Springer-Verlag, 1984.

[94] Пресман Э. Л. О сближении по вариации распределения суммы независимых бернуллиевских величин с пуассоновским законом // Теория вероятностей и ее применения. — 1985. — Т. XXX, вып. 2. — С. 391—396.

[95] Розанов Ю. А. Теория вероятностей, случайные процессы и математическая статистика. — М.: Наука, 1985.

[96] Ротарь В. И. К обобщению теоремы Линдеберга—Феллера // Математические заметки. — 1975. — Т. 18, вып. 1. — С. 129— 135.

[97] Севастьянов Б. А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982.

[98] Ширяев (Shiryayev A. N.) Probability. — 2nd ed. — Berlin etc.: Springer-Verlag, 1995.

[99] Ширяев (Shirjayev A. N.) Wahrscheinlichkeit. — Berlin: VEB Deutscher Verlag der Wissenschaften, 1988.

[100] Ширяев А. Н. Основы стохастической финансовой математики: В 2-х т. — М.: ФАЗИС, 1998.

[101] Фёллмер, Проттер, Ширяев (Föllmer H., Protter Ph., Shiryaev A. N.). Quadratic covariation and an extension of Itô's formula // Bernoulli. — 1995. — V. 1, № 1/2. — P. 149—170.

[102] Дынкин Е. Б., Юшкевич А. А. Теоремы и задачи о процессах Маркова. — М.: Наука, 1967.

[103] Гнеденко, Колмогоров (Gnedenko B. V., Kolmogorov A. N). Limit Distributions for Sums of Independent Random Variables. — Reading, MA, etc.: Addison-Wesley, 1954.

[104] Боровков А. А. Эргодичность и устойчивость случайных процессов. — М.: УРСС, 1999.

[105] Гриммет, Стирзакер (Grimmet G. R., Stirzaker D. R.). Probability and Random Processes. — Oxford: Clarendon Press, 1993.

[106] Биллингсли (Billingsley P.). Probability and Measure. — 3rd ed. — New York: Wiley, 1995.

[107] Боровков А. А. Математическая статистика. — М.: Наука, 1984.

[108] Дарретт (Durrett R.). Probability: Theory and Examples. — Pacific Grove, CA: Wadsworth & Brooks/Cole, 1991.

[109] Дарретт (Durrett R.). Stochastic Calculus. — Boca Raton, FL: CRC Press, 1996.

[110] Дарретт (Durrett R.). Brownian Motion and Martingales in Analysis. — Belmont, CA: Wadsworth International Group, 1984.

[111] Калленберг (Kallenberg O.). Foundations of Modern Probability. — 2nd ed. — New York: Springer-Verlag, 2002.
[112] Карлин, Тейлор (Karlin S., Taylor H. M.). A First Course in Stochastic Processes. — 2nd ed. — New York etc.: Academic Press, 1975.
[113] Кашин Б. С., Саакян А. А. Ортогональные ряды. — 2-е изд. — М.: АФЦ, 1999.
[114] Жакод, Проттер (Jacod J., Protter Ph.). Probability Essentials. — Berlin etc.: Springer-Verlag, 2000.
[115] Нётс (Neuts V. F.) Probability. — Boston, MA: Allyn & Bacon, 1973.
[116] Плато (Plato J.). Creating Modern Probability. — Cambridge: Cambridge Univ. Press, 1998.
[117] Ревюз Д. Цепи Маркова. — М.: РФФИ, 1997.
[118] Уильямс (Williams D.). Probability with Martingales. — Cambridge: Cambridge Univ. Press, 1991.
[119] Холл, Хейде (Hall P., Heyde C. C.). Martingale Limit Theory and Its Applications. — New York etc.: Academic Press, 1980.
[120] Чжун Кай-Лай (Chung Kai Lai). Elementary Probability Theory with Stochastic Processes. — 3rd ed. — Berlin etc.: Springer-Verlag, 1979.
[121] Математическая энциклопедия: В 5 т. / Гл. ред. И. М. Виноградов. — М.: Советская энциклопедия, 1977—1985.
[122] Стиглер (Stigler S. M.). The History of Statistics: The Measurement of Uncertainty Before 1900. — Cambridge: Belknap Press of Harvard Univ. Press, 1986.
[123] Гербер Х. Математика страхования жизни. — М.: Мир, 1995.
[124] Эренфесты П. и Т. (Ehrenfest P., Ehrenfest T.). Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem // Physikalische Zeitschrift. — 1907. — V. 8. — P. 311—314.
[125] Теория вероятностей и математическая статистика: энциклопедия / Гл. ред. Ю. В. Прохоров. — М.: Большая Российская энциклопедия, 1999.
[126] Вольфрам (Wolfram S.). The Mathematica® Book. — 4th ed. — Champaign; Cambridge: Wolfram Media; Cambridge Univ. Press, 1999.
[127] Дуб (Doob J. L.). What is a martingale? // The American Mathematical Monthly. — 1971. — V. 78. — P. 451—463.
[128] Синай Я. Г. Курс теории вероятностей. — М.: Изд-во МГУ, 1985. — 2-е изд., 1986.
[129] Синай (Sinaĭ Ya. G.). Topics in Ergodic Theory. — Princeton, NJ: Princeton Univ. Press, 1999. — (Princeton Mathematical Series; V. 44.)
[130] Вальтерс (Walters P.) An Introduction to Ergodic Theory. — New York etc.: Springer-Verlag, 1982.
[131] Булинский А. В., Ширяев А. Н. Теория случайных процессов. — М.: Физматлит, 2003.
[132] Хмаладзе Э. В. Мартингальный подход в теории непараметрических критериев согласия // Теория вероятностей и ее применения. — 1981. — Т. XXVI, вып. 2. — С. 246—265.

[133] Гамильтон (Hamilton J. B.). Time Series Analysis. — Princeton, NJ: Princeton Univ. Press, 1994.

[134] Бернулли Я. О законе больших чисел. — Ч. 4: Искусство предположений. — М.: Наука, 1986.

[135] Лукач Е. Характеристические функции. — М.: Наука, 1979.

[136] Хренников (Khrennikov A.). Interpretations of Probability. — Utrecht: VSP, 1999.

[137] Ширяев А. Н. Задачи по теории вероятностей. М.: МЦНМО, 2006.

Предметный указатель

К

Л

М

Н

О

П

Р

С

Х

Ц

Ч

Э

Я

Указатель обозначений

Некоторые общематематические обозначения

$R = (-\infty,\ \infty)$ — множество действительных чисел, действительная прямая, евклидово одномерное пространство

$R_+ = [0,\ \infty)$

$\bar{R} = [-\infty,\ \infty]$ — расширенная действительная прямая: $\bar{R} = R \cup \{-\infty\} \cup \{\infty\}$

$\bar{R}_+ = [0,\ \infty]$

Q — множество рациональных чисел

$Q_+ = Q \cap R_+$

R^d — евклидово d-мерное пространство

N — натуральные числа: или $\{0,\ 1,\ 2,\ \ldots\}$, или $\{1,\ 2,\ \ldots\}$

Z — множество целых чисел: $\{0,\ \pm 1,\ \pm 2,\ \ldots\}$

C — множество комплексных чисел

$(a,\ b) = \{x \in \bar{R} : a < x < b\}$, $\quad [a,\ b] = \{x \in \bar{R} : a \leqslant x \leqslant b\}$

$(a,\ b] = \{x \in \bar{R} : a < x \leqslant b\}$, $\quad [a,\ b) = \{x \in \bar{R} : a \leqslant x < b\}$

$\inf X$ — нижняя грань множества $X \subseteq \bar{R}$

$\sup X$ — верхняя грань множества $X \subseteq \bar{R}$

$\inf\limits_{n \geqslant m} x_n$ — нижняя грань множества $X = \{x_m,\ x_{m+1},\ \ldots\}$

$\sup\limits_{n \geqslant m} x_n$ — верхняя грань множества $X = \{x_m,\ x_{m+1},\ \ldots\}$

Если $x_n \in \bar{R}$, $n \geqslant 1$, то

$$\liminf_{n\to\infty} x_n = \varliminf_n x_n \equiv \sup_{m\geqslant 1} \inf_{n\geqslant m} x_n, \quad \limsup_{n\to\infty} x_n = \varlimsup_n x_n \equiv \inf_{m\geqslant 1} \sup_{n\geqslant m} x_n,$$

$$\lim x_n = x \ \Leftrightarrow\ \varliminf x_n = \varlimsup x_n = x \ \Leftrightarrow\ \varliminf x_n \geqslant x \geqslant \varlimsup x_n.$$

Для действительных чисел:

$x^+ = \max(x,\ 0), \quad x^- = -\min(x,\ 0)$

$$x^{\oplus} = \begin{cases} x^{-1}, & x \neq 0, \\ 0, & x = 0 \end{cases}$$

$x \vee y = \max(x,\ y), \quad x \wedge y = \min(x,\ y)$

$[x]$ или $\lfloor x \rfloor$ — наибольшее целое число, не превосходящее x

$\lceil x \rceil$ — наименьшее целое, большее или равное x

$\operatorname{sign} x$ — знак числа x: $\quad \operatorname{sign} x = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0 \end{cases}$

(иногда $\operatorname{sign} x$ определяется как 1, если $x \geqslant 0$, и -1, если $x < 0$)

$x_n \to x$, где $n \in \{1, 2, \ldots\}$, означает, что $\lim\limits_n x_n = x$

$x_n \uparrow$ означает, что $x_1 \leqslant x_2 \leqslant \ldots$; $x_n \uparrow x$ означает, что $x_n \uparrow$ и $\lim\limits_n x_n = x$

$x_n \downarrow$ означает, что $x_1 \geqslant x_2 \geqslant \ldots$; $x_n \downarrow x$ означает, что $x_n \downarrow$ и $\lim\limits_n x_n = x$

Для комплексных чисел $z = a + ib$,
где $a, b \in R$ и $i = \sqrt{-1}$ — мнимая единица:

$\bar{z} = a - ib$ — число, сопряженное с z

$|z|$ — модуль числа z $(= (a^2 + b^2)^{1/2})$

$\operatorname{Re} z$, $\operatorname{Im} z$ — действительная и мнимая части z: $\operatorname{Re} z = a$, $\operatorname{Im} z = b$

Для евклидова d-мерного пространства R^d:

$|x|$ — евклидова норма $x = (x_1, \ldots, x_d)$, т. е. $(x_1^2 + \ldots + x_d^2)^{1/2}$

$x \cdot y$ или (x, y) — скалярное произведение $x = (x_1, \ldots, x_d)$ и $y = (y_1, \ldots, y_d)$, т. е. $x_1 y_1 + \ldots + x_d y_d$

В теории множеств:

$A_n \uparrow$ означает, что $A_1 \subseteq A_2 \subseteq \ldots$; $A_n \uparrow A$ означает, что $A_n \uparrow$ и $\bigcup A_n = A$

$A_n \downarrow$ означает, что $A_1 \supseteq A_2 \supseteq \ldots$; $A_n \downarrow A$ означает, что $A_n \downarrow$ и $\bigcap A_n = A$

$\limsup A_n$, или $\overline{\lim}\, A_n$, или $\{A_n \text{ б. ч.}\}$ означает $\bigcap\limits_{m \geqslant 1} \left(\bigcup\limits_{n \geqslant m} A_n \right)$ — множество точек, принадлежащих бесконечному числу множеств A_n, $n \geqslant 1$

$\liminf A_n$ или $\underline{\lim}\, A_n$ означает $\bigcup\limits_{m \geqslant 1} \left(\bigcap\limits_{n \geqslant m} A_n \right)$ — множество точек, принадлежащих всем A_n, $n \geqslant 1$, за исключением, быть может, их конечного числа

I_A или $I(A)$ — индикатор множества A

$\{\ldots\}$ — множество

Математические символы:

$\ll$ — абсолютная непрерывность

$\sim$ — эквивалентность

$\perp$ — ортогональность

$f(x) = O(g(x))$, $x \to \infty \Leftrightarrow \exists C > 0$ и x_0 такие, что $|f(x)| \leqslant C|g(x)|$, $x \geqslant x_0$

$f(x) = o(g(x))$, $x \to \infty \Leftrightarrow \forall \varepsilon > 0\ \exists x_0(\varepsilon)$ такое, что $|f(x)| \leqslant \varepsilon |g(x)|$ для всех $x \geqslant x_0(\varepsilon)$

$f(x) \sim g(x)$, $x \to \infty \Leftrightarrow f(x) = g(x) + o(g(x))$

$(f \circ g)(x) = f(g(x))$ — композиция f и g

$\equiv$ — тождественно равно; по определению

Альберт Николаевич Ширяев

ВЕРОЯТНОСТЬ — 1

Редактор Т. Толозова

Подписано в печать 6.06.2007 г. Формат 60×90 $^1/_{16}$. Бумага офсетная №1.
Печать офсетная. Печ. л. 34,5. Тираж 2000 экз. Заказ №

Издательство Московского центра непрерывного математического образования.
119002, Москва, Большой Власьевский пер., 11. Тел. (495) 241—74—83.

Отпечатано с готовых диапозитивов в ППП «Типография „Наука“».
121099, Москва, Шубинский пер., 6.