ИЗДАТЕЛЬСТВО МИР

AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS

WILLIAM FELLER (1906—1970)
Eugene Higgins Professor of Mathematics
Princeton University

VOLUME I

Third Edition

Revised Printing

John Wiley & Sons
New York • Chichester • Brisbane • Toronto
1970

В. ФЕЛЛЕР

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

В 2-х томах

Том 1

Перевод с пересмотренного третьего английского издания
Ю. В. Прохорова
с предисловием
А. Н. Колмогорова

МОСКВА «МИР» 1984

ББК 22.171
Ф 30
УДК 519.24

Феллер В.

Ф 30 Введение в теорию вероятностей и ее приложения. В 2-х томах. Т. 1: Пер. с англ.— М.: Мир, 1984.—528 с., ил.

Перевод первого тома известного курса теории вероятностей, написанного выдающимся американским математиком, выполнен заново с пересмотренного третьего издания. Предыдущие издания (М.: ИЛ, 1952; М.: Мир, 1964; М.: Мир, 1967) быстро разошлись. Первый том содержит изложение той части теории вероятностей, которая имеет дело с дискретными распределениями. Такой отбор материала позволяет автору ввести читателя в круг основных идей теории вероятностей без применения сложного аналитического аппарата.

Для математиков разных уровней подготовки — от студентов до специалистов по теории вероятностей, для физиков и инженеров, а также для биологов, для которых вероятностные методы являются главными математическими методами.

Ф $\frac{1702060000-263}{041\ (01)-84}$ 31-84, ч. I

ББК 22.171
517.8

Редакция литературы по математическим наукам

ПРЕДИСЛОВИЕ ПЕРЕВОДЧИКА

Вниманию читателей предлагается новый перевод двухтомного курса В. Феллера «Введение в теорию вероятностей и ее приложения». В. Феллер (1906—1970) — один из выдающихся представителей современной теории вероятностей — родился в Загребе (Югославия). Математику изучал в Геттингенском университете в Германии, где получил в 1926 г. степень доктора. В 1933 г. покинул Германию. Работал в Копенгагене и Стокгольме и в 1939 г. переехал в США. Профессор Брауновского (1939—1945), Корнеллского (1945—1950) и Принстонского (1950—1970) университетов. Член ряда академий и научных обществ [1]).

Перевод выполнен с последних английских изданий (с пересмотренного третьего издания первого тома, 1970, и со второго издания второго тома, 1967). Феллер посвятил созданию, переработке и улучшению курса почти четверть века и делал это с неослабевающим энтузиазмом, никогда не уставая от этого занятия. Ни одна другая книга по теории вероятностей не может сравниться с этой — так удачно в ней соединены математическая строгость, совершенство доказательств и многочисленность рассматриваемых приложений. Излагая самые сложные математические вопросы, автор не упускает из виду тех явлений действительности, к которым может быть применена развиваемая теория. Характер курса таков, что он еще долго не устареет.

Воспроизведенные ниже предисловия к прежним изданиям первого тома (как к русским, так и к английским) позволяют мне быть более кратким. В этом томе в целом удачно демонстрируется тот факт, что сравнительно простые модели позволяют хотя бы в первом приближении правильно описать широкий круг практических задач (такими являются, например, модели размещений r шаров по n ящикам и урновые модели). Во многих случаях, особенно там, где интуиция не подсказывает правильного порядка соответствую-

---

[1]) См. Cramer H. William Feller.— Rev. Int. Stat. Inst., 1970, v. 38, No. 3, 435—436, а также статьи Doob J. L. William Feller and twentieth century probability и Kac M. William Feller, in memoriam.— В книге: Proc. Sixth Berkeley Symp. Math. Stat. and Prob.— vol. II.— Berkeley and Los Angeles: Univ. Calif. Press, 1972.

щих вероятностей, автор приводит численные результаты. Большое внимание уделено различным приближенным формулам. Их точность иллюстрируется примерами. Используемые при этом рассуждения «типичны для многих предельных теорем теории вероятностей» (с. 122).

Весьма полезны (и не только для начинающих читателей) приведенные в томе результаты «случайных экспериментов», которые создают представление о том, как выглядит «случайность» и сколь неожиданными могут оказаться отклонения от интуитивных представлений о ней.

В связи с результатами случайных экспериментов довольно рано ставится вопрос о способах проверки согласия модели с экспериментом и о способах оценки неизвестных вероятностей по данным опыта. Используется критерий χ^2, упоминается (по конкретному поводу) понятие оценки максимального правдоподобия (с. 66), указываются некоторые критерии случайности. Эта тенденция возможно более раннего показа типичных «статистических выводов», несомненно, целесообразна.

Следует отметить, что автор постоянно заботится о придании терминологии надлежащей точности (см., например, замечание на с. 322 о термине «рекуррентное событие»). Это очень важная сторона дела. Часто начинающие изучать теорию вероятностей запутываются, так как, скажем, слово «событие» на одних и тех же страницах учебников используют и в описательном «донаучном» смысле, и в смысле, предписываемом аксиоматической теорией. Точно так же выражение «произведем n независимых наблюдений случайной величины **X**» употребляют обычно без упоминания о том, что оно не имеет смысла в аксиоматической теории и служит лишь «разговорным вариантом» выражения «рассмотрим n независимых случайных величин, имеющих одно и то же распределение вероятностей».

Ограничение дискретными пространствами элементарных событий позволяет свести весь используемый аппарат к комбинаторике и к производящим функциям, а в случае, когда последние рациональны, к разложению их на простейшие дроби как методу исследования соответствующих распределений вероятностей. Убедительно показана мощь этих методов (гл. III, XII — XIV, XVI).

В качестве «дефекта» ограничения дискретным случаем автор отмечает, что оно «уменьшает изящество математических рассуждений» (с. 211). Можно добавить к этому, что в рамках теории, использующей дискретные пространства элементарных событий, оказывается невозможным отразить тесные и важные связи теории вероятностей с другими разделами современной математики. «Дискретный подход» позволяет прекрасно объяснить идею независимости, в меньшей степени идею марковской зависимости и совсем оставляет в стороне идею «спектрального анализа».

При подготовке к изданию перевода первого тома мне оказали большую помощь К. А. Боровков, А. М. Зубков, В. Г. Миранцев, В. В. Ульянов, А. П. Ушакова, В. Г. Ушаков, Н. Г. Ушаков и редактор издательства Г. М. Ильичева. Как и в работе над вторым томом, я пользовался советами А. В. Прохорова и В. В. Сазонова. Им всем я выражаю глубокую благодарность.

Москва, август 1983 *Ю. Прохоров*

ИЗ ПРЕДИСЛОВИЯ КО ВТОРОМУ РУССКОМУ ИЗДАНИЮ

Первое издание книги Феллера получило в СССР широкое признание. Сейчас вниманию читателей предлагается перевод второго английского издания, во многих деталях усовершенствованного автором. Во втором английском издании книга по-прежнему называется «Введение в теорию вероятностей и ее приложения», первый том двухтомного курса. Так как публикация второго тома вновь откладывается, то в русском издании сохранен подзаголовок [«дискретные распределения» — подзаголовок русского перевода первого издания.— *Ю. П.*], указывающий на принцип отбора материала, принятый автором для первого тома своего курса.

Именно этот принцип отбора материала позволяет книге Феллера занять самостоятельное место в литературе по теории вероятностей. Ограничиваясь дискретными распределениями, автор имеет возможность достигнуть вполне современной строгости и отчетливости изложения, не выходя за пределы элементарных чисто арифметических средств, и на твердой теоретической основе довести читателя до ряда важных принципиальных вопросов и большого числа практически интересных задач.

Несомненно, что при серьезном систематическом изучении теории вероятностей нельзя оставить в стороне непрерывные распределения. Но хорошо известно, что точное определение таких понятий, как условная функция распределения

$$F(x \mid y)=\mathbf{P}(\xi<x \mid \eta=y)$$

случайной величины ξ при заданном значении $\eta=y$ случайной величины η, в случае непрерывных распределений требует трудно воспринимаемых формальных конструкций, что строгое и в то же время общее изложение вопроса о суммировании произвольных независимых случайных величин требует хорошего владения теорией интеграла Стилтьеса и т. д. Ввиду практической важности непрерывных распределений часто обходятся более элементарными и не всегда строгими средствами. Но именно тем, кто для случая непрерывных распределений ограничится несколько кустарным или не вполне строгим изложением, будет особенно полезно проследить уже с полной отчетливостью параллельное развитие основных вероятностных идей для дискретного случая. Подробное изучение теории и применений производящих функций целочисленных случайных величин (см. гл. XI книги) является хорошим введением в более трудную и общую теорию характеристических функций произвольных случайных величин. Монографическое изучение марковских процессов с конечным или счетным числом состояний, данное в гл. XV—XVII, будет полезно многим читателям, предполагающим впоследствии изучать общую теорию случайных процессов.

Более квалифицированный читатель, для которого указанные преимущества первоочередного изучения дискретных распределений не существенны, заинтересуется книгой Феллера по преимуществу просто в качестве собрания большого числа частных задач и просчитанных до получения вполне конкретных результатов примеров. При разборе задач Феллер выдвигает на первый план решение их «прямыми», специфически вероятностными средствами. Эта тенденция видеть за аналитическими преобразованиями их «вероятностный» смысл принадлежит к числу наиболее ценных сторон книги Феллера. Заслуживает внимания также стремление автора книги на тщательно подобранных примерах наглядно показать характер действия вероятностных закономерностей. Во многих случаях автору удается ввести читателя в действительно интересные вопросы сопоставления статистических данных с вероятностной теорией явления.

1964 *А. Н. Колмогоров*

Нейгебауэру [1])

o et praesidium et dulce decus meum [2])

ПРЕДИСЛОВИЕ К ТРЕТЬЕМУ ИЗДАНИЮ

Когда эта книга впервые была задумана (более 25 лет тому назад) немногие математики за пределами Советского Союза видели в теории вероятностей полноправную ветвь математики. Приложения имели ограниченные цели, а исследования конкретных проблем часто приводили к непостижимым усложнениям. Если учесть эти обстоятельства, то задуманная книга была бы написана не для существовавшей тогда аудитории и не удовлетворяла бы осознанным в то время потребностям. Однако автор надеялся привлечь внимание к мало известным сторонам теории вероятностей, установить связи между отдельными ее частями, развить единые методы и указать на возможные или вероятные приложения. Благодаря возрастанию интереса к теории вероятностей книгу неожиданно часто использовали те, кто работал за пределами математических дисциплин. Ее широкое применение в те годы объяснимо, ибо принятая в ней точка зрения была новой, а изложенный материал был недоступен иным путем. Но книга остается популярной даже и теперь, когда содержание большей части ее глав можно найти в работах, посвященных отдельным областям и хорошо приспособленных для удовлетворения более специальных нужд. По этой причине и в новом издании характер книги остался тем же самым. Я надеюсь, что она будет продолжать обслуживать самые разнообразные требования и что по-прежнему найдутся читатели, которые прочтут ее просто для удовольствия и из любознательности.

На протяжении многих лет я с благодарностью получал многочисленные сообщения от тех, кто пользовался книгой, и эти сообщения были причиною различных усовершенствований. Многие параграфы были переписаны, чтобы облегчить изучение. Процесс чтения облегчен также за счет лучшего шрифта и благодаря превосходной редакторской работе, которую проделала г-жа Мак-Дугал, сохранившая сочувствие к требованиям читателей и доводам здравого смысла, несмотря на свою принадлежность к профессиональным редакторам.

Наибольшие изменения произведены в гл. III. Вопросы, входя-

---

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

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

щие в эту главу, были включены лишь во второе издание. Практической причиной этого явилось прежде всего неожиданное открытие, что ее увлекательные выводы можно получить элементарными методами. Старое изложение опиралось на остроумные комбинаторные соображения, которые теперь заменены более простыми и более естественными вероятностными соображениями. По существу эта глава написана заново.

Среди прочих добавлений наиболее заметными являются новые параграфы, посвященные ветвящимся процессам, цепям Маркова и теореме Муавра — Лапласа. Глава XIII перестроена; небольшие изменения сделаны во многих местах книги; появились новые примеры и задачи.

Я сожалею о том, что именной указатель может ввести в заблуждение, но я чувствовал себя обязанным давать точную ссылку каждый раз, когда идея или пример восходили к определенному источнику. Это означало, увы, что цитаты часто берутся из попутных замечаний и поэтому редко отражают истинный характер цитируемой статьи. Кроме того, многие примеры и задачи были навеяны чтением нематематических статей, в которых другими методами изучались сходные положения. (В написанных позже учебниках при цитировании этих нематематических статей считают, что они содержат мои примеры. Это показывает, как быстро развивается теория вероятностей, но одновременно показывает и ограниченную полезность цитат.) Недостаток места и умения помешали мне дать более подробные исторические указания и описать, каким образом теория вероятностей от полумистических рассуждений двадцатых годов перешла в современное состояние процветания.

На протяжении многих лет я имел благоприятную возможность работать со студентами и своими более молодыми коллегами. Я многим обязан их помощи и их воодушевлению. Возможность работать с ними явилась во многих отношениях следствием той поддержки, которую оказывало исследованиям по теории вероятностей, проводящимся в Принстонском университете, Научно-исследовательское управление Армии США.

Я выражаю особую благодарность Дж. Голдману за серьезный меморандум, описывающий его педагогический опыт, и Л. Питту за преданную помощь при чтении корректур.

Июль 1967 *Вильям Феллер*

ПРЕДИСЛОВИЕ К ПЕРЕСМОТРЕННОМУ ТРЕТЬЕМУ ИЗДАНИЮ

В отличие от первого издания книги третье ее издание содержало тревожащее число неточностей. В настоящем пересмотренном издании все обнаруженные неточности устранены. Многие формулировки улучшены, а также добавлены указания к задачам в тех случаях, когда это не требовало переверстки. Я благодарен издательству, разрешившему сделать эти дорогостоящие изменения, которые направлены к тому, чтобы книга читалась более легко.

Почти все изменения были предложены или профессором Р. Махолом и доктором Дж. Крофтом, работающими вместе в Чикаго, или подполковником (теперь в отставке) Королевской армии Дании Пребеном Кюлем. Они прочли книгу необычайно внимательно и с глубоким пониманием. Мне была весьма полезна последовавшая в результате этого приятная переписка.

Принстон, 1970

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Первоначально у автора было намерение написать книгу об аналитических методах теории вероятностей, которая трактовала бы эту теорию с точки зрения чистой математики. Изложение при этом было бы более единообразным и поэтому более удовлетворительным в эстетическом отношении; кроме того, оно было бы привлекательнее для чистых математиков. Однако щедрая поддержка, оказываемая Научно-исследовательским управлением Военно-морских сил США работам по теории вероятностей, ведущимся в Корнеллском университете, склонила автора к скорее ведущей к успеху, но менее благодарной попытке удовлетворить разнородные запросы более широкого круга читателей.

Целью этой книги является строгое изложение теории вероятностей как самостоятельной математической дисциплины, избегающее нематематических понятий. В то же время автор старался описать эмпирические основания теории вероятностей и развить у читателя ощущение всего многообразия ее приложений. Чтобы достигнуть второй цели, приведены многочисленные задачи, расчеты и примеры, которые прерывают основной ход изложения. Они четко выделены в тексте и отличаются более образным и менее формальным языком. В книгу включено некоторое количество специальных тем, чтобы показать силу общих методов и сделать книгу более полезной специалистам различных направлений. Чтобы облегчить чтение, отступления (detours) от основного пути отмечены звездочкой. Отмеченные разделы для понимания остальной части текста не обязательны.

Автором была предпринята серьезная попытка достичь единства методов. Специалисты найдут в книге много упрощений в существующих доказательствах, а также и новые результаты. В частности, для нужд этой книги была развита теория рекуррентных событий. Она позволяет по-новому изложить теорию цепей Маркова, что приводит к упрощениям даже в случае конечного числа состояний.

Примеры сопровождаются задачами (числом около 340), причем для большей части задач даны полные решения. Некоторые из этих задач — простые упражнения, но в основном они содержат добавочный иллюстративный материал или же так или иначе дополняют текст. Одно из назначений этих примеров и задач — развить вероятностную интуицию читателя и его умение формулировать вероятностные утверждения. Решение задач, кажущихся трудными, может потребовать совсем незначительных усилий, если их сформулировать естественным образом и включить в надлежащий контекст, а также предпослать им несколько разобранных примеров.

В преподавании теории вероятностей часто стремятся возможно быстрее сводить вероятностные задачи к задачам математического анализа, забывая при этом особенности самой этой теории. Такой подход основан на негодном определении случайной величины, которое обычно вводится в самом начале. В полную противоположность этому настоящая книга построена на понятии пространства элементарных событий. Вводить случайные величины без этого понятия — значит демонстрировать искусство вводить в заблуждение.

Чтобы отразить истинное положение вещей и избежать при этом теоретико-множественных и других чисто математических трудностей, в этом томе рассматриваются только *дискретные пространства элементарных событий*. Это жесткое ограничение, но его будут приветствовать читатели, не являющиеся математиками. Оно позволяет включить в книгу специальные вопросы, которые нелегко найти в литературе, и в то же время дает возможность, начав изложение элементарно, почти исчерпывающим образом рассмотреть такие глубокие темы, как теория случайных блужданий и теория цепей Маркова. Общая теория случайных величин и их распределений вероятностей, предельные теоремы, теория диффузионных процессов и некоторые другие темы отложены до следующего тома.

Эта книга не могла бы быть написана без поддержки Научно-исследовательского управления Военно-морских сил США. Как одно из следствий этой поддержки возникли постоянные личные контакты с Дж. Л. Дубом; его критические замечания и неизменно одобрительное отношение были неоценимы. Ему я выражаю свою особую благодарность. Затем я благодарю Дж. Риордана, который внимательно прочитал два варианта рукописи. Многие поправки и усовершенствования были предложены моей женой, которая прочитала и рукопись, и корректуру книги.

Автор обязан также Чжун Кайлаю, М. Донскеру и С. Голдбергу, которые прочитали рукопись и исправили некоторые неточности; решения большинства задач были подготовлены С. Голдбергом. В заключение я благодарю К. Холленбах, которая терпеливо и искусно перепечатала рукопись, а также Е. Ельяша, В. Хоффмана и Дж. Р. Кинни за помощь при чтении корректур.

Вильям Феллер

Корнеллский университет
Январь 1950 г.

КАК ПОЛЬЗОВАТЬСЯ ЭТОЙ КНИГОЙ[1])

В изложении имеется много отступлений, и оно не всегда идет от простого к сложному; сравнительно трудные с формальной точки зрения разделы встречаются в самом начале, а совсем легкие — в гл. XV и XVII. Неискушенный читатель отнюдь не должен следить за многочисленными побочными линиями, иначе он за деревьями не увидит леса. Вводные замечания к главам книги и звездочки, которыми помечены заголовки отдельных параграфов, должны облегчить ориентацию и выбор пропускаемых читателем разделов. Не помеченные звездочками разделы образуют единое целое, и для их понимания остальные разделы не нужны.

Первоначальное введение в основные понятия теории вероятностей содержится в главах I, V, VI, IX; начинающему следует изучить их с наименьшим возможным числом пропусков. Глава II предназначена для развития у читателя технических навыков и вероятностной интуиции; некоторое знакомство с ее содержанием желательно, но нет нужды изучать ее систематически: предпочтительнее, быть может, в дальнейшем возвращаться к этим простым иллюстрациям по мере надобности. Для первоначального ознакомления с элементарной теорией непрерывных распределений требуются лишь весьма немногие дополнительные разъяснения (соответствующий материал содержат элементарные главы тома 2).

Во вводном курсе можно от гл. IX сразу перейти к гл. XI, в которой производящие функции рассматриваются по образцу более общих преобразований. За гл. XI должны следовать какие-либо приложения из гл. XIII (рекуррентные события) или из гл. XII (цепные реакции, безгранично делимые распределения). Не используя понятия производящей функции, можно двигаться в одном из следующих направлений: предельные теоремы и теория флуктуаций (гл. VIII, X, III); стохастические процессы (гл. XVII); случайные блуждания (гл. III и основная часть гл. XIV). Главы, посвященные каждому из этих направлений, почти независимы одна от другой. Хотя теория цепей Маркова (гл. XV) опирается на понятия и факты теории рекуррентных событий, но ее можно изучать и независимо, если читатель пожелает принять без доказательства основную эргодическую теорему.

Глава III стоит в стороне от остальных глав. Ее содержание привлекательно само по себе, но, кроме того, она в высшей степени показательна с точки зрения новых взглядов и новых методов

---

[1]) Этот раздел отсутствовал в двух первых английских изданиях (во втором английском издании соответствующие указания были включены в предисловие автора). — *Прим. перев.*

теории вероятностей. Результаты, касающиеся флуктуаций, возникающих при последовательных бросаниях монеты, демонстрируют, что широко распространенные представления о действии закона больших чисел обманчивы. Эти результаты поразительны и резко расходятся с обычными представлениями; даже искушенные люди сомневались в том, что монеты в действительности ведут себя столь «неправильно», как предсказывает теория. Именно поэтому в § 6 включены результаты моделирования. Эта глава рассматривает лишь простейшую схему — бросание монеты, однако полученные выводы представительно отражают и значительно более общую ситуацию.

Знак ▶ указывает на завершение доказательства или на конец серии примеров.

Автор надеется, что подробный указатель поможет читателю согласовать при работе с книгой различные ее разделы.

ПРИРОДА ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. ИСХОДНЫЕ ПРЕДСТАВЛЕНИЯ

Теория вероятностей — математическая дисциплина, родственная таким дисциплинам, как, например, геометрия или теоретическая механика. В каждой дисциплине мы должны заботиться о различении трех сторон теории: а) формального логического содержания, б) интуитивных представлений, в) приложений. Не рассматривая этих трех сторон в их взаимосвязи, нельзя правильно оценить отличительные черты теории в целом и ее привлекательность.

а. Формальное логическое содержание

С точки зрения аксиоматического подхода математика занимается исключительно соотношениями между неопределяемыми объектами. Эту сторону дела хорошо поясняет пример игры в шахматы. Невозможно «определить» шахматы иначе, как сформулировав систему правил игры. Можно до некоторой степени описать условную форму фигур, но не всегда ясно, например, какая из них является королем. Шахматная доска и фигуры полезны, но можно обойтись и без них. Суть дела состоит в том, как ходят и действуют фигуры, и бессмысленно говорить об «определении» или «истинной природе» пешки или короля.

Аналогично этому геометрия не беспокоится о том, чем «на самом деле» являются точки и прямые. Они остаются неопределяемыми понятиями, и аксиомы геометрии лишь устанавливают связи между ними: две точки определяют прямую и т. п. Это «правила игры», и в них нет ничего неприкосновенного. Различные геометрии основаны на различных системах аксиом, и логическая структура каждой из неевклидовых геометрий не зависит от ее отношения к действительности. Физики исследовали движение тел в предположении, что закон тяготения имеет форму, отличную от открытой Ньютоном, и такое изучение имеет смысл, даже если считать, что в природе действует закон Ньютона.

б. Интуитивные представления

В отличие от шахматных правил аксиомы геометрии и теоретической механики опираются на интуитивные представления. В самом деле, геометрическая интуиция столь сильна, что склонна опережать логическое рассуждение. Вопрос о том до какой степени взаимозависимы логика, интуиция и физический опыт,— это проблема, в которую нам незачем вникать. Несомненно, что интуицию можно совершенствовать упражнениями и развивать. Теряющийся в шахматной игре новичок делает ходы осторожно, вспоминая отдельные правила, в то время как опытный игрок с одного взгляда оценивает сложную позицию и не может объяснить свою интуицию разумными доводами. Подобным же образом и математическая интуиция растет вместе с опытностью, и можно развить, скажем, интуитивное восприятие соотношений в четырехмерном пространстве.

Создается впечатление, что и общая интуиция человечества усиливается. Ньютоновские понятия поля сил и действия на расстоянии, а также максвелловская идея электромагнитных волн сначала открыто осуждались и считались «невообразимыми» и «противоречащими интуиции». Современная техника и радиоприемники в домах сделали эти понятия столь распространенными, что они стали частью повседневного словаря. Аналогично этому современный студент не в состоянии оценить способы рассуждений, предрассудки и прочие трудности, с которыми приходилось бороться теории вероятностей в первое время ее существования. В наши дни газеты сообщают о выборочных обследованиях общественного мнения, и магия статистики охватывает все стороны жизни в такой степени, что молодые девушки следят за статистикой, оценивая свои шансы выйти замуж. Поэтому каждый приобретает интуитивное представление о смысле таких утверждений, как «за это событие — три шанса из пяти». Эта интуиция (хотя и расплывчатая) является достаточной предпосылкой для первых шагов. Она будет развиваться по мере изучения теории и ознакомления с некоторыми весьма непростыми ее приложениями.

в. Приложения

В приложениях геометрии и теоретической механики теоретические понятия отождествляются с некоторыми физическими объектами, но способ этого отождествления гибок и меняется от случая к случаю, так что нельзя дать общих правил. Понятие твердого тела — одно из основных и полезных понятий теоретической механики, и все же ни один физический объект не является твердым. Можно ли данное тело рассматривать как твердое, зависит от обстоятельств и требуемой точности приближения к действительности. Резина не может быть примером твердого тела, однако, обсуждая движение автомобиля по льду, учебники обычно рассматривают резиновые покрышки как твердое тело. В зависимости от целей теории мы можем пренебрегать атомной структурой вещества; мы можем также рассматривать Солнце то как шар из непрерывного вещества, то как материальную точку.

В приложениях абстрактные модели служат лишь орудием, и одно и то же явление, наблюдаемое опытным путем, могут описывать различные модели. *Способ применения математических теорий не* *обусловливается какими-либо заранее сложившимися мнениями; это направленное к определенной цели умение, зависящее от опытности и меняющееся вместе с ней.* Вполне уместен философский анализ этого умения, но такой анализ находится вне области математики, физики или статистики. Философское рассмотрение оснований теории вероятностей должно быть отделено от математической теории вероятностей и математической статистики в такой же мере, как рассмотрение наших интуитивных представлений о пространстве отделяется теперь от геометрии.

§ 2. СПОСОБ ИЗЛОЖЕНИЯ

История теории вероятностей (и математики вообще) свидетельствует о стимулирующем взаимодействии теории и ее приложений: достижения теории открывают новое поле приложений, а приложения в свою очередь приводят к новым проблемам и плодотворно влияют на направление исследований. В настоящее время теорию вероятностей применяют во многих далеких друг от друга областях, и общая теория должна быть достаточно гибкой, чтобы разработать средства подхода ко множеству разнообразных проблем. Мы должны противостоять соблазну (к которому присоединяется давление обстоятельств) разработать теорию, терминологию и математический аппарат, слишком приближая их к какой-либо одной частной области приложений. Вместо этого мы намерены развить общую теорию, родственную тем, которые привели к успеху геометрию и теоретическую механику.

Мы начнем с простейших опытов, таких, как бросание монеты или игральной кости, где все утверждения имеют очевидный интуитивный смысл. Эти интуитивные соображения будут переведены на язык некоторой абстрактной модели, которая будет обобщаться постепенно, шаг за шагом. Иллюстративные примеры будут объяснять эмпирические предпосылки теории и развивать интуицию читателя, но сама теория будет иметь математический характер. Мы будем пытаться объяснить «истинный смысл» вероятностей не больше чем современный физик останавливается на «действительном смысле» массы и энергии или геометр объясняет природу точки. Вместо этого мы будем доказывать точные теоремы и приводить примеры их применения.

История показывает, что первоначально теория вероятностей развивалась для описания очень ограниченного круга опытов, связанных с азартными играми, и основные усилия были направлены на вычисление определенных вероятностей. В соответствии с этим и мы в нескольких первых главах вычислим некоторые типичные вероятности. При этом следует иметь в виду, что вовсе не отыскание этих численных значений вероятностей является целью общей теории. Объектом последней является раскрытие общих законов и зависимостей, а также построение абстрактных моделей, которые могут в удовлетворительной степени описывать физические явления.

Вероятности играют для нас ту же роль, что и массы в теоретической механике: можно обсуждать движение планетной системы, не зная масс отдельных планет и не рассматривая методов их действительного измерения. Можно также с пользой (для прояснения сути дела) изучать гипотетическое движение несуществующей планетной системы. Точно так же и *вероятностные модели могут быть полезны даже в том случае, когда они описывают объекты, которые не могут наблюдаться или не заслуживают наблюдения.* Например, миллиарды долларов вкладывают в развитие систем автоматической телефонной связи. При этом используют простые вероятностные модели, позволяющие сравнивать различные осуществимые системы. Лучшая в соответствии с этой теорией система принимается, другие же никогда не будут реализованы. Аналогично в страховом деле теорию вероятностей используют для подсчета вероятности разорения страховой компании. На этой основе теория дает рекомендации, как избежать определенных нежелательных ситуаций, которые в результате никогда не будут наблюдаться. Теория вероятностей была бы действенной и полезной, даже если было бы нелегко найти хотя бы один численный результат.

§ 3. «СТАТИСТИЧЕСКАЯ» ВЕРОЯТНОСТЬ

Успех современной математической теории вероятностей приобретен следующей ценой: теория ограничивается лишь одной стороной «случайности». Интуитивное понятие вероятности связано с индуктивными умозаключениями и суждениями вроде следующих: «Павел, вероятно, счастливый человек», «вероятно, эта книга будет неудачной», «гипотеза Ферма, вероятно, ошибочна». Суждения такого рода интересны философам и логикам и являются также законным объектом математической теории [1]). Следует подчеркнуть, однако, что мы будем иметь дело не с модальностями индуктивных умозаключений, а с тем, что может быть названо физической или *статистической вероятностью.* Грубо говоря, мы можем охарактеризовать это понятие, сказав, что наши вероятности относятся не к мнениям, а к возможным исходам *мыслимого эксперимента.* Прежде чем говорить о вероятностях, мы должны условиться насчет идеализированной модели рассматриваемого мыслимого эксперимента, подобного бросанию монеты, составлению выборки из совокупности кенгуру, обитающих на Луне, наблюдению движения частицы при

---

[1]) Koopman B. O., The axioms and algebra of intuitive probability, Ann. of Math., 41 (1940), 269—292; The bases of probability, Bull. AMS, 46 (1940), 763—774.

В качестве современного курса, основанного на понятии субъективных вероятностей, можно рекомендовать Savage L. J., The foundations of statistics, New York, John Wiley, 1954.

диффузии или подсчету числа телефонных вызовов. С самого начала мы должны условиться о том, что представляют собой возможные исходы такого эксперимента (их совокупность будет нашим пространством элементарных событий [1])) и каковы соответствующие им вероятности. Это аналогично обычному образу действий в теоретической механике, когда вводят воображаемую модель, включающую две, три или семнадцать материальных точек, и эти точки лишаются своих индивидуальных свойств. Подобным же образом, анализируя результаты бросаний монеты в игре, мы отвлекаемся от несущественных черт реального опыта и принимаем в качестве объекта нашей теории последовательности символов типа «герб, герб, решетка, герб,...». В нашей системе нет места для догадок о вероятности того, что завтра взойдет Солнце. Прежде чем говорить о такой вероятности, мы должны были бы условиться о (идеализированной) модели опыта, описание которой предположительно начиналось бы так: «случайным образом выбирается один из бесконечного множества миров...». Для того чтобы построить такую модель, не требуется особого воображения, но она представляется и неинтересной, и незначительной.

Астрономы говорят об измерении температуры в центре Солнца или о путешествии на Сириус. Эти действия кажутся невозможными, и все-таки размышления о них не бессодержательны. По таким же причинам мы не будем беспокоиться о том, выполним или нет наш мыслимый эксперимент; мы будем анализировать абстрактные модели. В глубине нашей души мы сохраняем интуитивное истолкование смысла вероятности, и это истолкование в ряде приложений приобретает и практическое значение. Мы *представляем себе* эксперимент, выполняемый очень много раз. Следует ожидать, что событие, имеющее скажем, вероятность 0,6, будет происходить с частотой шестьдесят случаев из ста. Это описание умышленно неопределенно, но оно создает образные интуитивные представления, достаточные для некоторых более простых приложений. По мере развития и усложнения теории практическое истолкование вероятности и интуитивное представление о ней будут принимать более определенные формы.

§ 4. РЕЗЮМЕ

Мы будем иметь дело с теоретическими моделями, в которые вероятности входят в качестве свободных параметров, подобно массам в теоретической механике. Эти модели применяют многими способами, которые также подвержены изменению. Умение применять теорию и интуиция развиваются одновременно с развитием теории.

[1]) В оригинале sample space, т. е. пространство, образуемое возможными результатами выборки (см. пример гл. I, 2, г)). Подробное разъяснение терминологии можно найти в гл. 25 книги Крамер Г. Математические методы статистики. — 2-е изд. — М.: Мир, 1975. — *Прим. перев.*

Это обычный путь, принятый и в других математических дисциплинах и оказавшийся плодотворным. Никто не предложил другого пути, который мог бы в достаточной мере удовлетворить *все* разнообразные нужды и потребности как растущего организма, называемого теорией вероятностей, так и приложений этой теории.

Можно сетовать на то, что интуитивное понятие вероятности недостаточно для научной теории, но это исторический факт. В примере гл. I, 6, б) мы исследуем случайное размещение частиц по ячейкам. Соответствующее «естественное» распределение вероятностей представлялось совершенно очевидным каждому и принималось без колебаний физиками. Однако оказалось, что физические частицы не обладают «здравым смыслом», и «естественное» распределение Больцмана пришлось в одних случаях заменить на распределение Бозе — Эйнштейна, а в других — на распределение Ферми — Дирака. Не существовало никаких интуитивных доводов, почему фотоны ведут себя иначе, чем протоны и почему частицы обоих типов не подчиняются «априорным» законам. *Если бы* теперь и удалось обосновать эти факты, то это свидетельствовало бы только о том, что интуиция развивается вместе с теорией. Во всяком случае, даже для приложений существенны свобода и гибкость теории, и было бы пагубной ошибкой сковывать ее слишком жесткими ограничениями.

Утверждали также, что современная теория вероятностей слишком абстрактна и слишком обща, чтобы быть полезной. Столь же воинственный крик поднимался в свое время практически мыслящими людьми против максвелловской теории поля. Чтобы опровергнуть эти доводы, достаточно было бы указать на неожиданные новые применения, появившиеся благодаря абстрактной теории случайных процессов или на новые достижения современной теории флуктуаций, которые противоречат наивной интуиции и приводят к пересмотру некоторых практических рекомендаций. Однако этот спор просто бесполезен: осуждать слишком легко. Те вещи, которые стали сегодня практически важными, еще вчера порицались как непрактичные, и теории, которые станут практически важными завтра, практические люди сегодняшнего дня всегда будут клеймить и называть ничего не стоящей абстрактной игрой.

§ 5. ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ

Статистический или эмпирический подход к вероятности был развит главным образом Фишером [1]) и Мизесом [2]). Понятие прост-

[1]) Фишер Роналд Эймлер (Fisher R. A.; 1890—1962) — английский статистик и генетик, один из основателей математической статистики и математической популяционной генетики. В русском переводе издана книга Фишер Р. Э. Статистические методы для исследователей.— М.: Статиздат, 1958.— *Прим. перев.*

[2]) Мизес Рихард (von Mises R.; 1883—1953) — немецкий математик и механик. Основные работы относятся к теории вероятностей, гидродинамике и прикладной механике. В 1933 г. эмигрировал из Германии, с 1939 г. жил в США.— *Прим. перев.*

ранства элементарных событий [1]) идет от Мизеса. Это понятие сделало возможным построение строгой математической теории вероятностей на основе теории меры. Такой подход развивался постепенно в течение 20-х годов под влиянием многих авторов. Аксиоматический подход на современном уровне был разработан А. Н. Колмогоровым [2]). Мы будем следовать этому направлению, хотя термин «аксиоматическое построение», может быть, звучит слишком торжественно, ибо в этом томе речь идет только о простом случае дискретных вероятностей.

---

[1]) Немецкий термин *Merkmalraum* (пространство меток). Основной курс Мизеса Wahrscheinlichkeitsrechnung появился в 1931 г. Обновленный вариант (изданный и дополненный Хильдой Гейрингер) опубликован в 1964 г. под названием Mathematical theory of probability and statistics, New York, Academic Press. Философские взгляды Мизеса получили наибольшую известность благодаря его более ранней книге, изданной в 1928 г. (пересмотренное Х. Гейрингер издание Probability, statistics and truth, London, Macmillan, 1957). [Имеется русский перевод издания 1928 г.: Мизес Р. Вероятность и статистика / Под редакцией и с предисловием А. Я. Хинчина.— М. — Л.: ГИЗ, 1930.— *Перев.*]

[2]) Колмогоров А. Н., Основные понятия теории вероятностей.— М.— Л.: ОНТИ, 1936.— 2-е изд.— М.: Наука, 1974; первоначальное издание — Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, Springer, 1933.

используя логарифмы. Для малых положительных x $\log(1-x)\approx\approx -x$, и, таким образом, из (3.2) имеем

$$\log p \approx -\frac{1+2+\ldots+(r-1)}{365} = -\frac{r(r-1)}{730}. \tag{3.4}$$

Для $r=30$ это приводит к приближенному значению 0,3037, в то время как точное значение $p=0{,}294$. Для $r\leqslant 40$ погрешность в (3.4) не превосходит 0,08 (к этим вопросам мы вернемся в § 7; см. также ответ к задаче 44 § 10).

§ 4. ПОДМНОЖЕСТВА И РАЗБИЕНИЯ

Как и ранее, будем использовать термин *генеральная совокупность объема n* для обозначения множества из n элементов *безотносительно к их порядку*. Две генеральные совокупности считаются различными только тогда, когда одна из них содержит элемент, не содержащийся в другой.

Рассмотрим подмножество объема r из заданной генеральной совокупности, состоящей из n элементов. Произвольная нумерация элементов этого подмножества превращает его в упорядоченную выборку объема r, и обратно, каждая такая выборка может быть получена указанным путем. Так как r элементов можно занумеровать $r!$ различными способами, отсюда следует, что число упорядоченных выборок объема r в $r!$ раз больше, чем число подмножеств объема r. Поэтому число подмножеств объема r равно $(n)_r/r!$. Выражения такого вида известны как *биномиальные коэффициенты* и обычно обозначаются так:

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n(n-1)\cdot\ldots\cdot(n-r+1)}{1\cdot 2\cdot\ldots\cdot(r-1)\cdot r}. \tag{4.1}$$

Таким образом, доказана следующая теорема.

Теорема 1. *Генеральная совокупность из n элементов имеет $\binom{n}{r}$ различных подмножеств объема $r\leqslant n$!*

Иначе говоря, подмножество из r элементов может быть выбрано $\binom{n}{r}$ различными способами. Такое подмножество однозначно определяется $n-r$ элементами, не принадлежащими ему и образующими подмножество объема $n-r$. Отсюда следует, что подмножеств объема $n-r$ существует ровно столько же, сколько и подмножеств объема r, и, следовательно, для $1\leqslant r\leqslant n$ должно быть

$$\binom{n}{r} = \binom{n}{n-r}. \tag{4.2}$$

Для того чтобы непосредственно доказать (4.2), заметим, что биномиальный коэффициент (4.1) можно записать в виде

$$\binom{n}{r}=\frac{n!}{r!\,(n-r)!} \tag{4.3}$$

(для этого достаточно умножить числитель и знаменатель (4.1) на $(n-r)!$). Заметим, что для $r=0$ левая часть (4.2) не определена, а правая определена. Для того чтобы (4.2) было справедливо для всех целых r, таких, что $0\leqslant r\leqslant n$, *мы положим*

$$\binom{n}{0}=1, \quad 0!=1 \tag{4.4}$$

и $(n)_0=1$.

Примеры. а) *Бридж и покер* (см. примечание к § 1 гл. I). Порядок карт игрока несуществен, и, следовательно, имеется $\binom{52}{13}=$ $=635\,013\,559\,600$ различных комбинаций карт у одного игрока при игре в бридж и $\binom{52}{5}=2\,598\,960$ при игре в покер. Найдем вероятность x того, что при игре в покер фиксированный игрок имеет пять различных по значению карт. Значения карт могут быть выбраны $\binom{13}{5}$ способами, и для каждого значения мы свободны в выборе любой из четырех мастей. Отсюда следует, что $x=$ $=4^5\cdot\binom{13}{5}\binom{52}{5}^{-1}$, что приблизительно равно 0,5071. Для бриджа вероятность тринадцати различных значений карт равна $4^{13}\binom{52}{13}^{-1}$, или приблизительно 0,0001057.

б) Каждый из 50 штатов представлен двумя сенаторами. Рассмотрим события, состоящие в том, что в комитете из 50 случайно выбранных сенаторов: 1) представлен данный штат, 2) представлены все штаты.

В первом случае проще найти вероятность q противоположного события, а именно что данный штат *не* представлен. Всего имеется 100 сенаторов и 98 из них не из данного штата. Следовательно,

$$q=\binom{98}{50}\binom{100}{50}^{-1}=\frac{50\cdot 49}{100\cdot 99}=0{,}24747\ldots.$$

Далее, из теоремы § 2 следует, что комитет, содержащий по одному сенатору из каждого штата, может быть образован 2^{50} различными способами. Поэтому вероятность того, что *все* штаты представлены в комитете, равна $p=2^{50}\binom{100}{50}^{-1}$. Используя формулу Стирлинга (см. § 9), можно показать, что $p\approx\sqrt{2\pi}\cdot 5\cdot 2^{-50}\approx 4{,}126\cdot 10^{-14}$.

в) *Задача о размещении.* Еще раз рассмотрим случайное раз-

..., или (5, 5)» на девять элементарных событий. Заметим, что если в результате нашего эксперимента мы получили (3, 3), то *одни и те же бросания* привели и к событию «сумма очков равна шести», и к событию «выпали две грани с нечетным числом очков». Эти события не являются взаимно исключающими и поэтому могут происходить одновременно. В качестве второго примера рассмотрим возраст человека. Каждое частное значение x представляет *элементарное* событие, тогда как утверждение о том, что данному человеку пошел шестой десяток, описывает составное событие «$50 < x < 60$». Итак, каждое составное событие может быть разложено на элементарные события; иначе говоря, составное событие есть *совокупность элементарных событий*.

Если мы хотим говорить об «опытах» и «наблюдениях» в рамках нашей теории и без каких бы то ни было неясностей, то мы должны прежде всего условиться, каковы элементарные (неразложимые далее) события, представляющие собой мыслимые исходы опыта или наблюдения; *они определяют идеализированный опыт* (заметим, что термин «элементарное (или неразложимое) событие» остается столь же неопределенным, что и термины «точка» или «прямая» в геометрии). По определению *каждый неразложимый исход (идеализированного) опыта представляется одним и только одним элементарным событием*. Совокупность всех элементарных событий будем называть *пространством элементарных событий* [1]), а сами элементарные события — *точками* этого пространства. Все события, связанные с данным (идеализированным) опытом, могут быть описаны как совокупности элементарных событий.

Прежде чем формализовать это основное соглашение, мы обсудим несколько типичных примеров, которые будут играть определенную роль и в дальнейшем.

§ 2. ПРИМЕРЫ

а) *Размещение трех шаров по трем ящикам*. Табл. 1 содержит все возможные исходы «опыта», состоящего в размещении трех шаров по трем ящикам.

Каждое из этих размещений представляет неразложимый исход эксперимента, т. е. элементарное событие. Событие A «существует ящик, содержащий не менее двух шаров», происходит при размещениях с номерами 1—21, и мы выражаем это, говоря, что событие A есть множество элементарных событий 1—21. Аналогично этому событие B «первый ящик не пуст» можно описать как множество точек (элементарных событий) 1, 4—15, 22—27. Событие C «произошло и A, и B» есть множество, состоящее из тринадцати элементарных событий (1, 4—15). При выбранных A и B оказывается, что каждая из 27 точек принадлежит или A, или B (или одновременно

[1]) См. примечание на с. 21.— *Прим. перев.*

Таблица 1

1. $\{abc \mid - \mid -\}$	10. $\{a \mid bc \mid -\}$	19. $\{- \mid a \mid bc\}$
2. $\{- \mid abc \mid -\}$	11. $\{b \mid a\ c \mid -\}$	20. $\{- \mid b \mid a\ c\}$
3. $\{- \mid - \mid abc\}$	12. $\{c \mid ab \mid -\}$	21. $\{- \mid c \mid ab\}$
4. $\{ab \mid c \mid -\}$	13. $\{a \mid - \mid bc\}$	22. $\{a \mid b \mid c\}$
5. $\{a\ c \mid b \mid -\}$	14. $\{b \mid - \mid a\ c\}$	23. $\{a \mid c \mid b\}$
6. $\{bc \mid a \mid -\}$	15. $\{c \mid - \mid ab\}$	24. $\{b \mid a \mid c\}$
7. $\{ab \mid - \mid c\}$	16. $\{- \mid ab \mid c\}$	25. $\{b \mid c \mid a\}$
8. $\{a\ c \mid - \mid b\}$	17. $\{- \mid a\ c \mid b\}$	26. $\{c \mid a \mid b\}$
9. $\{bc \mid - \mid a\}$	18. $\{- \mid bc \mid a\}$	27. $\{c \mid b \mid a\}$.

и A, и B); поэтому событие «произошло или A, или B, или одновременно и A и B» совпадает со всем пространством элементарных событий и происходит с абсолютной достоверностью. Событие D «A не произошло» состоит из точек 22—27 и может быть описано условием, что нет пустых ящиков. Наконец, событие «первый ящик пуст, и не существует ящика, содержащего более одного шара», невозможно (не может произойти), так как не существует элементарных событий, удовлетворяющих этим условиям.

б) *Размещение r шаров по n ящикам.* Более общий случай размещения r шаров по n ящикам может быть рассмотрен аналогичным образом, однако число возможных исходов эксперимента быстро возрастает с ростом r и n. Для $r=3$ шаров и $n=4$ ящиков соответствующее пространство элементарных событий состоит уже из 81 точки, а для $r=n=10$ это число точек равно 10^{10}; полная таблица заняла бы примерно сто тысяч больших томов.

Мы воспользуемся примером, чтобы разъяснить тот важный факт, что природа пространства элементарных событий не играет роли для нашей теории. Для нас это пространство (вместе с заданным на нем распределением вероятностей) *определяет* идеализированный эксперимент. Мы пользуемся образным языком шаров и ящиков, но то же самое пространство допускает большое число интерпретаций. Чтобы пояснить нашу точку зрения, а также для удобства дальнейших ссылок *мы приведем ряд схем, внешне весьма различных, но по существу эквивалентных абстрактной схеме размещения r шаров по n ящикам в том смысле, что соответствующие исходы отличаются лишь словесным их описанием.* Вероятности, приписываемые элементарным событиям, могут при этом быть различными в различных примерах, и эта сторона дела будет обсуждаться впоследствии.

(б.1) *Дни рождения.* Распределение дней рождения r человек соответствует размещению r шаров по $n=365$ ящикам (полагаем, что в году 365 дней).

(б.2) *Несчастные случаи.* Разделение несчастных случаев на группы по дням недели, в которые они происходят, равносильно размещению r шаров по $n=7$ ящикам.

(6.3) *При стрельбе* по n мишеням пули соответствуют шарам, мишени — ящикам.

(6.4) *Выборочное обследование.* Пусть группа из r человек разбивается на классы, скажем, по возрасту или по профессии. Классы играют роль ящиков, люди — шаров.

(6.5) *Облучение в биологии.* Когда сетчатка глаза подвергается воздействию света, кванты света играют роль шаров, а клетки сетчатки соответствуют ящикам. Аналогично при исследовании генетического эффекта облучения хромосомы соответствуют ящикам, а α-частицы — шарам [1]).

(6.6) *При экспериментах с космическими лучами* частицы, попадающие в сетчики Гейгера, играют роль шаров, а сами счетчики — ящиков.

(6.7) *Лифт* отправляется с r пассажирами и останавливается на n этажах. Распределение пассажиров по группам в зависимости от этажа, на котором они выйдут, соответствует размещению r шаров по n ящикам.

(6.8) *Игра в кости.* Возможному исходу эксперимента, состоящего в бросании r игральных костей, соответствует распределение r шаров по $n=6$ ящикам. Если *бросают монеты*, то имеют дело с $n=2$.

(6.9) *Случайные цифры.* Каждой последовательности из r случайных цифр соответствует размещение r шаров (десятичных разрядов) по $n=10$ ящикам с номерами 0, 1, 2, . . ., 9.

(6.10) *Распределение r человек по признаку пола.* В этом случае имеется $n=2$ ящика и r шаров.

(6. 11) *Собирание купонов* [2]). Различные типы купонов соответствуют ящикам, все собранные купоны — шарам.

(6.12). *Распределение тузов между игроками при игре в бридж.* Четыре игрока соответствуют четырем ящикам, четыре туза — шарам.

(6.13) *Распределение генов.* Каждый потомок некоторой особи (человека, животного или растения) наследует определенные гены родителя. Если некоторый ген может находиться в одной из n форм $A_1, \ldots, A_n$, то потомков можно классифицировать по формам данного гена. Потомки особи соответствуют шарам, генотипы $A_1, \ldots, A_n$ — ящикам.

(6.14) *Химические реакции.* Предположим, что молекулярные цепочки некоторого полимера взаимодействуют с кислородом. Каждая цепочка может прореагировать с 0, 1, 2, . . . молекулами

[1]) Здесь (как и в других примерах, связанных с генетикой) используется терминология, принятая в книге Мюнтцинг А. Генетика. Общая и прикладная.— М.: Мир, 1967.— *Прим. перев.*

[2]) Под купоном здесь понимается торговый ярлык, фирменная обертка, этикетка или аналогичное свидетельство сделанной покупки. Некоторое количество собранных купонов дает право на известные льготы при приобретении товаров.— *Прим. перев.*

кислорода. Реагирующие молекулы кислорода играют роль шаров, а цепочки полимера — роль ящиков, в которых размещают шары.

(6.15) *Теория светочувствительных материалов.* Фотографическая пластинка покрыта слоем светочувствительных зерен, причем каждое зерно реагирует, если в него ударяется определенное число r квантов. Для теории черно-белой фотографии важно знать число прореагировавших зерен (т. е. тех, в которые попало r квантов). В этом случае мы получаем типичную задачу о размещении, где зерна эмульсии соответствуют ящикам, а кванты света — шарам. (В действительности дело обстоит сложнее, так как пластинка обычно бывает покрыта зернами различной чувствительности.)

(6.16) *Опечатки.* Возможные распределения r опечаток в книге, содержащей n страниц, соответствуют размещениям r шаров по n ящикам, если только r меньше, чем число печатных знаков на странице.

в) *Случай неразличимых шаров.* Вернемся к примеру а) и предположим теперь, что все три шара одинаковы. Это означает, что мы больше не делаем различия между такими размещениями, как 4, 5, 6 и т. п. В этом случае табл. 1 сводится к табл. 2, которая определяет новое пространство элементарных событий. Соответствующий мыслимый эксперимент мы назовем *размещением трех неразличимых шаров по трем ящикам.* Аналогичные рассуждения применимы и к более общему случаю r шаров и n ящиков.

Таблица 2

<table>
<tr><td>1. {***| — | — }</td><td>6. { * |**|—}</td></tr>
<tr><td>2. { — |***| — }</td><td>7. { * |—|**}</td></tr>
<tr><td>3. { — | — |***}</td><td>8. {—|**| *}</td></tr>
<tr><td>4. {** | * | — }</td><td>9. {—| *|**}</td></tr>
<tr><td>5. {** | — | * }</td><td>10. { *| *| *}.</td></tr>
</table>

Различимы ли шары на самом деле, для нашей теории несущественно. Если это даже и так, мы можем условиться считать их неразличимыми. Тузы в бридже [пример (6.12)] или люди в лифте [пример (6.7)], конечно, различимы, и тем не менее часто предпочтительнее считать их неразличимыми. Игральные кости в примере (6.8) можно окрасить и сделать их тем самым различимыми, но выберем ли мы при решении какой-либо данной задачи схему с различимыми или схему с неразличимыми шарами, определяется поставленными целями и достигаемыми преимуществами. Характер задачи может предписать определенный выбор, но в любом случае нашу теорию можно развивать только после того как соответствующая модель выбрана, т. е. после того как определено пространство элементарных событий.

В приведенном выше примере мы рассматривали неразличимые шары, но в табл. 2 еще различаются первый, второй и третий ящики, и их порядок существен. Мы можем пойти еще дальше и считать, что даже ящики неразличимы (например, ящики можно выбирать наудачу независимо от их содержимого). Если и шары, и ящики неразличимы, то возможны только три размещения, а именно

$$\{***|-|-\},\quad \{**|*|-\},\quad \{*|*|*\}.$$

г) *Выборочное обследование*. Предположим, что с целью оценить число курящих выбрана группа в 100 человек. При этом единственное интересующее нас свойство данной выборки — число x курящих; оно может быть равно любому целому числу от 0 до 100. В этом случае мы можем принять, что наше пространство элементарных событий состоит из 101 «точки» $x=0, 1, \ldots, 100$. Итог каждой отдельной выборки (или наблюдения) полностью описывается заданием соответствующей точки x. Примером составного события может служить следующее: «большинство людей в данной выборке — курящие». Это означает, что опыт заканчивается одним из 50 элементарных событий $x=51, 52, \ldots, 100$ (безразлично, каким именно). Аналогично и любое другое свойство выборки можно описать, перечислив соответствующие случаи или элементарные события. Для единства терминологии мы предпочитаем говорить о событиях, а не о свойствах выборки. С математической точки зрения событие является просто множеством соответствующих точек пространства элементарных событий.

д) *Выборочное обследование (продолжение)*. Допустим теперь, что выбранные нами 100 человек делятся на классы не только по признаку «курящие» или «некурящие», но и по полу. Выбранная группа может быть теперь охарактеризована четверкой чисел $М_к$, $Ж_к$, $М_н$, $Ж_н$, означающих по порядку число курящих мужчин, число курящих женщин, число некурящих мужчин и число некурящих женщин. В качестве элементарных событий мы возьмем четверки целых чисел, принимающих значения от 0 до 100 и дающих в сумме 100. Таких четверок существует 176 851; они и образуют пространство элементарных событий (см. гл. II, 5). Утверждение, что в выборке «среди мужчин доля курящих больше, чем среди женщин», означает, что для нашей выборки отношение $М_к/М_н$ больше отношения $Ж_к/Ж_н$. Точка (73, 2, 8, 17) этим свойством обладает, а точка (0, 1, 50, 49) не обладает. По существу наше событие может быть описано перечислением всех четверок чисел, обладающих заданным свойством.

е) *Бросание монеты*. Если монета бросается три раза, то пространство элементарных событий состоит из восьми точек, которые удобно обозначить следующим образом: *ГГГ*, *ГГР*, *ГРГ*, *РГГ*, *ГРР*, *РГР*, *РРГ*, *РРР* [*Г* означает выпадение герба при соответствующем бросании, *Р* — решетки.— *Перев.*]. Событие *А* «выпало не менее двух гербов» совпадает с множеством первых четырех точек.

Событие B «выпала ровно одна решетка» означает или $ГГР$, или $ГРГ$, или $РГГ$; мы говорим, что B содержит эти три точки.

ж) *Возраст супругов.* Страховые компании интересуются распределением возрастов супругов. Пусть x означает возраст мужа, а y — возраст жены. Каждое наблюдение дает пару чисел (x, y). В качестве пространства элементарных событий мы берем первый квадрант плоскости x, y, так что каждая точка $x > 0$, $y > 0$ будет элементарным событием. Событие A «мужу свыше 40 лет» представляется всеми точками, лежащими справа от прямой $x=40$; событие B «муж старше жены» представляется областью, лежащей между осью x и прямой $y=x$, т. е. множеством точек, для которых $x > y$; событие C «жене свыше 40 лет» представляется точками первого квадранта, расположенными выше прямой $y=40$. Для геометрического представления возрастов двух супружеских пар нам потребовалось бы четырехмерное пространство.

з) *Фазовое пространство.* В статистической механике каждое возможное «состояние» системы называют «точкой фазового пространства». Отличие здесь только в терминологии. Фазовое пространство есть просто наше пространство элементарных событий, а его точки — наши элементарные события [1]).

§ 3. ПРОСТРАНСТВО ЭЛЕМЕНТАРНЫХ СОБЫТИЙ. СОБЫТИЯ

Из предыдущего должно быть ясно, что мы никогда не будем говорить о вероятностях вне связи с каким-либо пространством элементарных событий (или, говоря на физическом языке, вне связи с некоторым мыслимым опытом). *Мы начинаем с понятия пространства элементарных событий и его точек; впредь они будут рассматриваться как данные. Они являются первоначальными и неопределяемыми понятиями теории*, так же как понятие «точка» и «прямая» остаются неопределяемыми при аксиоматическом построении евклидовой геометрии. Природа элементарных событий не играет роли в нашей теории. Пространство элементарных событий служит моделью идеализированного опыта в том смысле, что по определению любой мыслимый исход опыта полностью описывается одной и только одной точкой этого пространства. О каком-либо событии A имеет смысл говорить только тогда, когда для *каждого* исхода опыта известно, произошло или не произошло событие A. Совокупность точек, представляющих все те исходы, при которых происходит событие A, полностью описывает это событие. Обратно, произвольно заданное множество A, содержащее одну или более точек нашего пространства, можно назвать событием; оно происходит или не происходит в зависимости от того, принадлежит или не принадлежит множеству A точка, представляющая исход опыта. Для нас

---

[1]) См., например, Хинчин А. Я. Математические основания статистической механики.— М.— Л.: ОГИЗ, 1943.— *Прим. перев.*

поэтому слово *событие будет означать то же самое, что некоторое множество элементарных событий*. Мы будем говорить, что *событие A состоит из определенных точек (или содержит эти точки)*, а именно из точек, представляющих те исходы идеализированного опыта, при которых происходит событие A.

Пример. В пространстве элементарных событий примера 2, а) рассмотрим событие U, состоящее из точек с номерами 1, 7, 13 (см. табл. 1). Это определение является формальным и самым простым, но U можно описать и многими другими способами. Например, U можно определить как событие, состоящее в одновременном выполнении следующих трех условий: (1) второй ящик пуст, (2) шар a находится в первом ящике, (3) номер ящика, где находится шар b, не превышает номера ящика, где находится шар c. Каждое из этих трех условий в свою очередь определяет событие. Событие U_1, определяемое только условием (1), состоит из точек 1, 3, 7—9, 13—15. Событие U_2, определяемое условием (2), состоит из точек 1, 4, 5, 7, 8, 10, 13, 22, 23, и, наконец, событие U_3, определяемое условием (3), состоит из точек 1—4, 6, 7, 9—11, 14, 16, 18—20, 22, 24, 25. Событие U можно теперь описать как одновременное осуществление всех трех событий U_1, U_2, U_3. ▶

Хотя термины «событие» и «элементарное событие» вызывают определенные интуитивные представления, для нас они будут равносильны терминам «точечное множество» и «точка», как их понимают во всех разделах математики.

Предыдущий пример и пример 2, а) показывают, что по данным двум или более событиям можно определить новые события. Опираясь на подобные примеры, мы введем теперь систему обозначений, относящихся к формальной *алгебре событий* (т. е. алгебре точечных множеств).

§ 4. ОТНОШЕНИЯ МЕЖДУ СОБЫТИЯМИ

Теперь мы предположим, что задано произвольное, но фиксированное пространство элементарных событий $\mathfrak{S}$. Мы будем использовать заглавные латинские буквы для обозначения *событий*, т. е. множеств элементарных событий. Тот факт, что точка x содержится в событии A, обозначим символом $x \in A$. Так, для любой точки x имеем $x \in \mathfrak{S}$. Мы пишем $A = B$ в том и только том случае, когда эти события состоят из одних и тех же точек.

События будут, вообще говоря, определяться некоторыми условиями, налагаемыми на их точки. Удобно ввести символ, выражающий тот факт, что ни одна из точек не удовлетворяет определенному набору условий. Этой цели служит следующее определение.

Определение 1. *Мы будем пользоваться записью $A=0$ для выражения того, что событие A не содержит элементарных событий* *(событие A невозможно).* Нуль следует понимать символически, а не как цифру [1].

Каждому событию A соответствует некоторое другое событие, определяемое условием «событие A не произошло». Оно содержит все точки, не содержащиеся в A.

Определение 2. *Событие, состоящее из всех точек, не содержащихся в событии A, называют событием, противоположным событию A (или отрицанием A), и обозначают через A'* [2]. *В частности,* $\mathfrak{S}'=0$.

С любыми двумя событиями A и B можно связать два новых события, определенных условиями *«имеют место и A, и B»* и *«имеют место или A, или B, или и A, и B»*. Эти события будут обозначаться соответственно AB и $A \cup B$. Событие AB содержит все точки, общие событиям A и B. Если события A и B взаимно исключают друг друга, то они не имеют общих точек и событие AB невозможно; аналитически это описывают формулой

$$AB=0, \tag{4.1}$$

которая должна читаться так: «события A и B *несовместны*». Событие AB' означает, что произошло и A, и B', т. е. что событие A произошло, а событие B не произошло. Аналогично $A'B'$ означает, что ни A, ни B не произошли. Событие $A \cup B$ означает, что хотя бы одно из событий A или B произошло; оно содержит все точки, за исключением точек, не принадлежащих ни A, ни B.

С точки зрения теории вероятностей мы можем охарактеризовать событие AB как одновременное осуществление событий A и B. По стандартной математической терминологии событие AB называют (логическим) пересечением событий A и B. Подобным же образом событие $A \cup B$ называют объединением событий A и B. Введенные нами понятия переносятся и на случай многих событий A, B, C, D, ...

Определение 3. *Любой совокупности событий A, B, C, ... мы сопоставляем два новых события следующим образом. Множество, состоящее из элементарных событий (точек), принадлежащих одновременно всем заданным событиям, будет обозначаться через ABC... и называться пересечением (или одновременным осуществлением) событий A, B, C, ...* [3] *(см. рис. 1). Множество, состоящее из элементарных событий, каждое из которых принадлежит хотя бы одному из заданных событий, будет обозначаться через $A \cup B \cup C$ ...*

---

[1] Часто предпочитают другой символ, а именно $\varnothing$.— *Прим. перев.*

[2] Такое событие иногда называют дополнительным событием (дополнением A, complementary event) и обозначают также через $\bar{A}$ или A^c.— *Прим. перев.*

[3] Стандартное математическое обозначение для пересечения двух или более множеств таково: $A \cap B$, $A \cap B \cap C$, ... Это обозначение в определенных случаях удобнее (см. гл. IV, 1 тома 2). Здесь мы используем обозначение AB, ABC, ... для того, чтобы облегчить труд наборщика.

и называться объединением заданных событий (или осуществлением хотя бы одного из них). События A, B, C, . . . попарно несовместны (взаимно исключают друг друга), если никакие два из них не имеют общих точек, т. е. если $AB=0$, $AC=0$, . . ., $BC=0$, ...

Мы нуждаемся еще в символе для обозначения того, что событие A не может произойти, если не произошло B, или того, что событие B является следствием события A. Это означает, что каждая точка события A содержится в событии B. Примером может служить множество всех матерей, являющееся частью множества всех женщин: все матери суть женщины, но не все женщины суть матери.

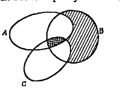

Рис. 1. Область, ограниченная внешним контуром, является объединением $A\cup B\cup C$. «Треугольная» (дважды заштрихованная) область является пересечением ABC. Область в форме полумесяца (заштрихована) является пересечением B и события, противоположного $A\cup C$.

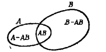

Рис. 2. Пересечение и разности событий.

Определение 4. *Символы $A\subset B$ и $B\supset A$ означают одно и то же, а именно что каждая точка A содержится в B; словесно это выражают так: «событие A влечет за собой событие B» или соответственно «событие B является следствием события A». В этом случае мы будем также писать $B-A$ вместо BA' для обозначения того, что событие B произошло, а событие A не произошло (см. рис. 2).*

Событие $B-A$ содержит все точки события B, не являющиеся точками события A [1]). Пользуясь этим обозначением, напишем

$$A'=\mathfrak{S}-A \quad \text{и} \quad A-A=0.$$

Примеры. а) Если события A и B несовместны, то осуществление события A влечет за собой осуществление события B', и наоборот. Следовательно, $AB=0$ означает то же самое, что $A\subset B'$ и $B\subset A'$.

---

[1]) Именно так определяют разность множеств B и A, не предполагая, что $A\subset B$ (другое обозначение $B\setminus A$). Наряду с так определенной разностью множеств B и A определяют и их симметрическую разность $(B\setminus A)\cup(A\setminus B)$ (обозначения $A+B$ или $A\Delta B$). Вводя в дополнение к так определенной операции «сложения» операцию «умножения» и полагая $A\cdot B=AB$, нетрудно показать, что по отношению к операциям $+$ и $\cdot$ класс всех подмножеств $\mathfrak{S}$ будет кольцом (в алгебраическом смысле); так как $A\cdot A=A$, то оно будет булевым кольцом с единицей $\mathfrak{S}$, т. е. булевой алгеброй.— *Прим. перев.*

б) Событие $A-AB$ означает, что произошло событие A, но не произошли одновременно оба события A и B. Поэтому $A-AB=$ $=AB'$.

в) В примере 2, ж) событие AB означает, что мужу свыше 40 лет и что он старше своей жены, тогда как AB' означает, что ему свыше 40 лет, но он *не* старше своей жены. Событие AB поэтому изображается неограниченной трапециевидной областью, лежащей между осью x и прямыми $x=40$ и $y=x$. Событие AB' изображается неограниченной треугольной областью, лежащей между прямыми $x=40$ и $y=x$, с включенной второй границей. Событие AC означает, что каждому из супругов свыше 40 лет. Событие $A \cup C$ означает, что хотя бы одному из них свыше 40 лет, тогда как $A \cup B$ означает, что либо мужу свыше 40 лет, либо, если это не так, он хотя бы старше своей жены (выражаясь юридически «возраст мужа превосходит 40 лет или возраст жены, смотря по тому, что из них меньше» [1])).

г) Пусть в примере 2, а) E_i означает событие, состоящее в том, что ящик с номером i пуст ($i=1, 2, 3$). Аналогично пусть S_i, D_i, T_i соответственно означают события, состоящие в том, что ящик с номером i содержит один шар, два шара, три шара. Тогда $E_1E_2=$ $=T_3$, $S_1S_2 \subset S_3$ и $D_1D_2=0$. Заметим также, что $T_1 \subset E_2$ и т. д. Событие $D_1 \cup D_2 \cup D_3$ определяется условием, что существует хотя бы один ящик, содержащий два шара.

д) *Бридж* (см. примечание к § 1). Пусть A, B, C, D означают события, состоящие соответственно в том, что при сдаче колоды карт для игры в бридж игрок «Север», «Юг», «Восток», «Запад» получил по крайней мере одного туза. Ясно, что хотя бы один из игроков имеет туза, так что по крайней мере одно из четырех событий должно иметь место. Следовательно, $A \cup B \cup C \cup D=\mathfrak{S}$ есть все пространство элементарных событий. Событие $ABCD$ происходит тогда и только тогда, когда каждый игрок имеет туза. Событие «игрок «Восток» получил всех четырех тузов» означает, что ни одно из событий A, B, C не произошло, или, что то же самое, что одновременно осуществились события A', B' и C', т. е. произошло событие $A'B'C'$.

е) В примере 2, ж) мы имеем $BC \subset A$, т. е. «если муж старше жены (B) и жене свыше 40 лет (C), то мужу свыше 40 лет (A)». Как словесно описать событие $A-BC$? ▶

§ 5. ДИСКРЕТНЫЕ ПРОСТРАНСТВА ЭЛЕМЕНТАРНЫХ СОБЫТИЙ

Простейшими пространствами элементарных событий являются те, которые содержат конечное число n точек. Если n достаточно мало (как в случае бросания нескольких монет), то это пространст-

---

[1]) В подлиннике фраза имитирует стиль английских юридических документов.— *Прим. перев.*

во легко себе представить. Пространство элементарных событий при сдаче карт для игры в бридж гораздо сложнее. Тем не менее можно представить себе, что каждое элементарное событие изображается какой-то фишкой, и затем рассматривать множество этих фишек как изображение пространства элементарных событий. Каждое событие A (например, «игрок «Север» получил двух тузов») изображается группой фишек; событие A' изображается остальными фишками. Остается сделать один шаг для того, чтобы представить себе ящик с бесконечным числом фишек или пространство элементарных событий, состоящее из бесконечной последовательности точек E_1, E_2, E_3, ...

Примеры. а) Условимся бросать монету до тех пор, пока не выпадет герб. Тогда элементарными событиями будут: $E_1=\Gamma$, $E_2=P\Gamma$, $E_3=PP\Gamma$, $E_4=PPP\Gamma$ и т. д. Мы можем допустить, что герб никогда не выпадет, или не допускать такой возможности. Допуская эту возможность, мы должны представить ее дополнительной точкой E_0.

б) Три игрока a, b и c участвуют в игре (подобной игре в шахматы) по следующей системе. В первом туре играют a и b, а игрок c свободен. Проигравший заменяется игроком c, и во втором туре играют победитель и c, а игрок, потерпевший поражение в первом туре, свободен. Соревнование продолжается таким образом до тех пор, пока один из игроков не выиграет двух партий подряд, и в этом случае его объявляют победителем. Для простоты мы исключаем возможность ничьей в отдельной партии. Возможные исходы соревнования описываются тогда следующей схемой:

$$\begin{array}{l} aa,\ acc,\ acbb,\ acbaa,\ acbacc,\ acbacbb,\ acbacbaa, \ldots, \\ bb,\ bcc,\ bcaa,\ bcabb,\ bcabcc,\ bcabcaa,\ bcabcabb,\ \ldots \end{array} \qquad (*)$$

Наряду с этим вполне мыслимо, что ни один из игроков не выиграет двух партий подряд, т. е. что соревнование будет продолжаться неограниченно долго по одной из двух следующих схем:

$$acbacbacb\ldots,\quad bcabcabca\ \ldots . \qquad (**)$$

Пространство элементарных событий для нашего мыслимого опыта определяется формулами (*) и (**) и является бесконечным. Однако ясно, что точки этого пространства можно занумеровать в простую последовательность. Для этого достаточно, например, на первые два места поставить точки (**), а на последующие — точки (*) в таком порядке: aa, bb, acc, bcc, ... (Продолжение этого примера см. в задачах 5 и 6, примере гл. V, 2,а) и задаче 5 гл. XV, 14.) ►

Определение. *Пространство элементарных событий называется* **дискретным**, *если оно состоит лишь из конечного числа точек или из бесконечного числа точек, которые могут быть занумерованы в простую последовательность* E_1, E_2, ...

Не каждое пространство элементарных событий дискретно. Известна теорема (принадлежащая Г. Кантору), утверждающая, что пространство элементарных событий, состоящее из всех положительных чисел, не дискретно. Здесь мы сталкиваемся с разграничением, известным и в теоретической механике, где обычно сначала рассматривают системы, составленные из отдельных материальных точек, каждая из которых имеет положительную массу, а затем переходят к случаю непрерывного распределения массы, когда каждая отдельная точка имеет массу, равную нулю. В первом случае масса системы получается попросту сложением масс отдельных точек, во втором случае она вычисляется интегрированием плотности. Совершенно аналогично вероятности событий в дискретном пространстве элементарных событий получаются просто сложением, тогда как в других пространствах необходимо интегрирование. Кроме аналитических средств, которые приходится привлекать, эти два случая ничем существенным не отличаются. Желая изложить собственно вероятностные рассуждения, избегая при этом трудностей технического характера, мы сначала займемся лишь дискретными пространствами элементарных событий. Мы увидим, что даже этот частный случай приводит ко многим интересным и важным результатам.

В этом томе мы будем рассматривать только дискретные пространства элементарных событий.

§ 6. ВЕРОЯТНОСТИ В ДИСКРЕТНЫХ ПРОСТРАНСТВАХ ЭЛЕМЕНТАРНЫХ СОБЫТИЙ; ПОДГОТОВИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Вероятности событий — числа той же природы, что и расстояния в геометрии или массы в теоретической механике. Теория предполагает, что они заданы, и не нуждается ни в каких предположениях об их действительном численном значении или о способе их измерения на практике. Некоторые из наиболее важных приложений носят качественный характер и не зависят от численных значений вероятностей событий. В тех сравнительно немногих случаях, когда требуется знать численные значения вероятностей некоторых событий, вычислительные приемы столь же разнообразны, сколь и методы определения расстояний. Когда плотник, землемер, лоцман и астроном измеряют расстояния, то в их действиях мало общего. В этом же контексте мы можем рассмотреть коэффициент диффузии. Это понятие теории вероятностей. Чтобы найти численное значение этого коэффициента, требуются физические соображения, связывающие его с другими теориями; прямое же измерение невозможно. Статистические таблицы смертности, наоборот, составляются на основании наблюдений. В наиболее существенных приложениях измерение вероятностей событий или сравнение результатов теории с данными наблюдений требуют применения довольно сложных статистических методов, опирающихся в свою очередь на развитую вероятностную теорию. Иначе говоря, хотя наглядный смысл понятия вероятности и ясен, но лишь по мере развития теории мы сумеем увидеть, как следует применять это понятие. Все возможные «определения» понятия вероятности весьма неполно отражают реальную практику.

Бросая «правильную» монету, мы не колеблемся связать вероятность 1/2 с выпадением герба или решетки. Это приводит к выводу, что при n бросаниях монеты все 2^n возможных случаев равновероятны. С теоретической точки зрения, это — *допущение*. Часто говорят, что такое допущение — логически неизбежное и единственно возможное. Однако были философы и статистики, которые отвергали это допущение и исходили из предположений, полностью его исключающих (в этом отражаются различные точки зрения на законы природы) [1]).

Часто утверждают также, что значение 1/2 для вероятности получается из опыта. На самом деле применение утонченных статистических методов к фактическим результатам опытов с бросанием монет неизменно показывало, что выпадение герба и выпадение решетки *не* являются одинаково вероятными событиями. Тем не менее мы придерживаемся нашей модели «идеальной» монеты, хотя на самом деле правильных монет не существует. Мы сохраняем эту модель не только из-за ее логической простоты, но в основном из-за ее полезности и применимости. Во многих приложениях эта модель достаточно точно описывает действительность. Еще важнее тот извлекаемый из опыта факт, что отклонения от нашей схемы всегда связаны с такими явлениями, как, например, несовпадение центра тяжести монеты с ее геометрическим центром.

Таким образом, наши идеализированные модели могут быть в высшей степени полезными, даже несмотря на то, что они никогда не являются вполне точными. Например, в современном статистическом контроле качества продукции [2]) идеализированные вероятностные модели используются для выявления «объяснимых причин» [3]) бросающихся в глаза отклонений от этих моделей и последующего быстрейшего устранения неисправностей в работе машин и неправильностей течения производственного процесса.

Подобные замечания относятся и к другим случаям. Число возможных распределений карт между игроками при игре в бридж равно почти 10^{30}. Обычно мы соглашаемся считать эти распределения карт

---

[1]) В оригинале в скобках стоят слова uniformity or non-uniformity in nature. Термин uniformity in nature в английском толковом словаре Вебстера (Webster's third new international dictionary.— Springfield: C. G. Merriam Co., 1965) объясняется так: «доктрина или принцип, утверждающие неизменность или правильность в природе; в более специальном понимании — принцип, утверждающий, что идентичные исходные состояния или идентичные причины неизменно имеют следствием идентичные эффекты».— *Прим. перев.*

[2]) В оригинале «основанном на методах Шухарта»; см. разд. 6 гл. II, 5.— *Прим. перев.*

[3]) В оригинале «assignable causes», т. е. соображения, которые могут быть приведены в качестве причин.— *Прим. перев.*

равновероятными. Для проверки этого допущения потребовалось бы более 10^{30} опытов, т. е. тысячи миллиардов лет круглосуточной игры всех живущих на Земле людей, если считать продолжительность игры равной одной секунде. Однако следствия нашего предположения могут быть проверены экспериментально посредством наблюдения, например, за частотой появления нескольких тузов у одного игрока. Оказывается, что идеализированная модель описывает опыт с достаточной для предварительных целей точностью (в предположении, что карты тасуют лучше, чем это обычно делается). Еще важнее то, что в случаях когда наблюдаемые результаты не согласуются с идеализированной схемой, на основе этой схемы можно обнаружить «объяснимые причины» расхождения, например изменение способа тасовки карт. Эти примеры имеют ограниченную ценность, но они свидетельствуют о полезности принятых моделей. Более интересные примеры встретятся прн дальнейшем развитии теории.

Примеры. а) *Различимые шары.* В примере 2,а) представляется естественным предположение о том, что все элементарные события *равновероятны*, т. е. *что каждое из них имеет вероятность* $1/27$. Мы можем, отправляясь от этого *определения*, изучать его следствия. Будет или не будет наша модель достаточно точно описывать действительность, зависит от типа явлений, к которым ее будут применять. В одних приложениях равновероятность исходов является следствием физических соображений. В других приложениях основанную на равновероятности исходов простейшую модель используют для общей ориентировки, даже если совершенно очевидно, что это лишь самое первое, грубое приближение. [Сказанное хорошо иллюстрируют примеры 2, (6.1), дни рождения; 2, (6.7), лифт; 2, (6.11), собирание купонов.]

б) *Неразличимые шары; статистика Бозе — Эйнштейна.* Обратимся теперь к примеру 2, в), связанному с размещением трех неразличимых шаров по трем ящикам. Можно рассуждать так: невозможность различать шары не отражается на сущности физического эксперимента, и остается по-прежнему 27 исходов, хотя только 10 из них оказываются различимыми. Эти рассуждения показывают, что десяти точкам табл. 2 надлежит приписать следующие вероятности:

Номер точки	1	2	3	4	5	6	7	8	9	10
Вероятность	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$

Следует отметить, что по отношению к бо́льшей части приложений, перечисленных в примере 2, б), такое рассуждение звучит убедительно, и это оправдывает указанный способ задания вероятностей, соответствующих точкам табл. 2. Исторически это рассуждение долгое время принималось как безусловно верное и в статистической механике служило основанием *статистики Максвелла — Больцмана* для размещения r частиц (шаров) по n ячейкам (ящикам). Тем больше было удивление, когда Бозе и Эйнштейн показали, что определенные типы частиц подчиняются *статистике Бозе — Эйнштейна* (подробнее см. в гл. II, 5). В рассматриваемом случае $r=n=3$ модель Бозе — Эйнштейна сопоставляет *каждому из десяти элементарных событий вероятность* 1/10.

Этот пример показывает, что в одном и том же пространстве элементарным событиям вероятность можно приписывать по-разному. Он поясняет трудное для понимания взаимное действие друг на друга теории и практики и учит нас, в частности, не надеяться особенно на априорные аргументы и быть готовыми принять новые и непредвиденные схемы.

в) *Бросание монеты.* Для частотного истолкования постулата о равенстве вероятностей элементарных событий необходимы данные о результатах проведенных экспериментов. Бросание любой настоящей монеты приводит к искаженным результатам, из-за действия не поддающихся учету факторов. Можно, однако, осуществить физический эксперимент, результаты которого оказываются значительно ближе к идеальной «модели бросания монеты», чем для любой настоящей монеты. Для того чтобы дать представление о случайных колебаниях (флуктуациях), которые можно ожидать, мы приводим запись результатов подобного эксперимента, соответствующую 10 000 бросаний монеты. Табл. 3 содержит числа «появлений герба» в ста последовательных сериях, каждая из которых насчитывает сто «бросаний». Суммарное число «появлений герба» равно 4979[1]). Поглядев на эту таблицу, читатель, вероятно, скажет

Таблица 3

Номера испытаний	Число «появлений герба»										Сумма по строке
1— 1000	54	46	53	55	46	54	41	48	51	53	501
1001— 2000	48	46	40	53	49	49	48	54	53	45	485
2001— 3000	43	52	58	51	51	50	52	50	53	49	509
3001— 4000	58	60	54	55	50	48	47	57	52	55	536
4001— 5000	48	51	51	49	44	52	50	46	53	41	485
5001— 6000	49	50	45	52	52	48	47	47	47	51	488
6001— 7000	45	47	41	51	49	59	50	55	53	50	500
7001— 8000	53	52	46	52	44	51	48	51	46	54	497
8001— 9000	45	47	46	52	47	48	59	57	45	48	494
9001—10000	47	41	51	48	59	51	52	55	39	41	484

[1]) На самом деле приведены числа появления четных цифр в одном из разделов таблицы случайных чисел A million random digits with 100 000 normal deviates, RAND Corporation, Glencoe, Illinois, Free Press, 1955. [Для моделирования «бросания монеты» при этом используют генераторы случайных шумов; о других методах см., например, Ермаков С. М. Метод Монте-Карло и смежные вопросы.— М.: Наука, 1975.— *Перев.*]

с неопределенным чувством: «Ну и что же?». По правде говоря, судить о том, в какой степени такие эмпирические данные согласуются с нашей абстрактной моделью, можно лишь при помощи достаточно развитой теории. (Между прочим, мы вернемся к этому кругу вопросов в гл. III, 6.) ►

§ 7. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И СООТНОШЕНИЯ

Основное положение[1]). *Пусть дано дискретное пространство элементарных событий $\mathfrak{S}$ с точками $E_1, E_2, \ldots$. Мы полагаем, что с каждой точкой E_j связано число, называемое вероятностью E_j и обозначаемое $\mathbf{P}\{E_j\}$. Эти числа должны быть неотрицательны и таковы, что*

$$\mathbf{P}\{E_1\}+\mathbf{P}\{E_2\}+\ldots=1. \tag{7.1}$$

Заметим, что мы не исключаем возможность равенства нулю вероятностей отдельных элементарных событий. Это допущение может показаться неестественным, но оно необходимо для того, чтобы избежать осложнений. В случае дискретного пространства элементарных событий равенство нулю вероятности на практике рассматривают как невозможность и элементарное событие, имеющее вероятность, равную нулю, можно безнаказанно изъять из пространства элементарных событий. Однако часто численные значения вероятностей заранее неизвестны, и без дополнительных, порой сложных, рассуждений невозможно решить, равна или не равна нулю вероятность некоторой точки.

Определение. *Вероятность $\mathbf{P}\{A\}$ любого события A есть сумма вероятностей элементарных событий, из которых оно состоит.*

По формуле (7.1) вероятность всего пространства $\mathfrak{S}$ элементарных событий равна единице: $\mathbf{P}\{\mathfrak{S}\}=1$. Далее, для любого события

$$0\leqslant\mathbf{P}\{A\}\leqslant1. \tag{7.2}$$

Рассмотрим теперь два произвольных события A_1 и A_2. Чтобы вычислить вероятность $\mathbf{P}\{A_1\cup A_2\}$ того, что имеет место либо событие A_1, либо событие A_2, либо оба эти события, мы должны сложить вероятности всех точек, содержащихся в A_1, и всех точек, содержащихся в A_2, считая, однако, каждую точку по одному разу. Мы имеем поэтому

$$\mathbf{P}\{A_1\cup A_2\}\leqslant\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}. \tag{7.3}$$

Если теперь E — любая точка, содержащаяся и в A_1, и в A_2, то $\mathbf{P}\{E\}$ входит два раза в правую и один раз в левую часть неравенства

[1]) В оригинале Fundamental convention. К нему с учетом определения вероятности (см. ниже) по существу сводятся аксиомы Колмогорова (см. книгу А. Н. Колмогорова, цитированную на с. 23) в случае дискретного пространства элементарных событий.— *Прим. перев.*

(7.3). Поэтому правая часть превосходит левую на $\mathbf{P}\{A_1A_2\}$, и мы получаем простую, но имеющую полезные следствия теорему.

Теорема. *Для любых двух событий A_1 и A_2 вероятность того, что имеет место либо событие A_1, либо событие A_2, либо оба эти события, дается формулой*

$$\mathbf{P}\{A_1 \cup A_2\}=\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}-\mathbf{P}\{A_1A_2\}. \tag{7.4}$$

Если $A_1A_2=0$, т. е. если события A_1 и A_2 несовместны, то (7.4) *сводится к следующему виду:*

$$\mathbf{P}\{A_1 \cup A_2\}=\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}. \tag{7.5}$$

Пример. Монета бросается два раза. В качестве пространства элементарных событий возьмем четыре точки $ГГ$, $ГР$, $РГ$, $РР$ и свяжем с каждой вероятность 1/4. Пусть A_1 и A_2 означают соответственно события «при первом бросании выпал герб» и «при втором бросании выпал герб». Тогда событие A_1 состоит из точек $ГГ$, $ГР$ и событие A_2 состоит из точек $РГ$, $ГГ$. Далее событие $A_1 \cup A_2$ содержит три точки $ГГ$, $ГР$, $РГ$, в то время как событие A_1A_2 состоит из единственной точки $ГГ$. Таким образом,

$$\mathbf{P}\{A_1A_2\}=1/2+1/2-1/4=3/4.$$ ▶

Вероятность $\mathbf{P}\{A_1 \cup A_2 \cup \ldots \cup A_n\}$ осуществления хотя бы одного из n событий может быть вычислена по формуле, аналогичной формуле (7.4); она будет получена в гл. IV, 1. Здесь мы лишь заметим, что рассуждения, приводящие к (7.3), пригодны в случае любого числа объединяемых событий. Таким образом, *для произвольно взятых событий $A_1, A_2, \ldots$ выполняется неравенство*

$$\mathbf{P}\{A_1 \cup A_2 \cup \ldots\} \leqslant \mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}+\ldots. \tag{7.6}$$

В частном случае, когда события $A_1, A_2, \ldots$ попарно несовместны,

$$\mathbf{P}\{A_1 \cup A_2 \cup \ldots\}=\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}+\ldots. \tag{7.7}$$

Иногда неравенство (7.6) называют *неравенством Буля*[1]).

Мы исследуем сначала [в гл. II и III.— *Перев.*] частный случай, когда пространство элементарных событий состоит из конечного числа N точек, вероятность каждой из которых просто равна $1/N$. В этом случае вероятность любого события A равна числу точек, входящих в A, деленному на N. В старой литературе элементарные события называли «случаями», а элементарные события, входящие в A,— «благоприятными случаями» (благоприятными для события A). *Если* все элементарные события равновероятны, то вероятность события равна отношению числа благоприятных случаев к числу

[1]) Буль Джордж (Boole J.; 1815—1864) — английский математик и логик. Он заложил основы математической логики. Его именем названы, например, булевы алгебры.— *Прим. перев.*

всех возможных случаев. К сожалению, этой формулировкой неоднократно злоупотребляли, стараясь «определить» понятие вероятности. Часто утверждали, что в *любом* конечном пространстве элементарных событий вероятности этих событий должны быть равны между собой. Это не так [1]). При однократном бросании неправильной монеты пространство элементарных событий по-прежнему содержит всего две точки (выпадение герба и выпадение решетки), но эти элементарные события могут иметь какие угодно вероятности p и q, лишь бы было $p+q=1$. Новорожденный — либо мальчик, либо девочка, но эти две возможности не равновероятны [по многолетним наблюдениям частота рождения мальчиков несколько больше 0,51; см., например, таблицы в гл. 31 книги Г. Крамера, цитированной на стр. 21.— *Перев.*]. Еще одно опровержение содержит пример 6, б). Полезность пространств с равновероятными элементарными событиями проявляется в основном при изучении азартных игр и в комбинаторном анализе.

§ 8. ЗАДАЧИ

1. Из цифр 1, 2, 3, 4, 5 сначала выбирают одну, а затем из оставшихся четырех — вторую. Допустим, что все 20 возможных исходов равновероятны. Найти вероятность того, что а) в первый раз, б) во второй раз, в) оба раза будет выбрана нечетная цифра.

2. В пространстве элементарных событий примера 2, а) припишем равные вероятности всем 27 точкам. Используя обозначения примера 4, г), проверить формулу (7.4) для двух событий $A_1=S_1$ и $A_2=S_2$. Сколько точек входит в событие S_1S_2?

3. Рассмотрим все 24 возможные перестановки цифр 1, 2, 3, 4 и свяжем с каждой вероятность 1/24. Пусть A_i— событие, состоящее в том, что цифра i оказалась на i-м месте ($i=1, 2, 3, 4$). Проверить формулу (7.4).

4. Монету бросают до тех пор, пока 2 раза подряд она не выпадает одной и той же стороной. Каждому возможному исходу, при котором бросания заканчиваются на n-м шаге, припишем вероятность $1/2^{n-1}$. Описать пространство элементарных событий. Найти вероятности следующих событий: а) опыт окончится до шестого бросания; б) потребуется четное число бросаний.

5. В пространстве элементарных событий примера 5, б) припишем каждой точке из (*), содержащей ровно k букв, вероятность $1/2^k$. (Иначе говоря, aa и bb имеют вероятность 1/4, acb имеет вероятность 1/8 и т. д.) а) Показать, что сумма вероятностей всех точек из (*) равна единице (в силу этого каждая из двух точек (**) имеет вероятность, равную нулю). б) Установить, что с вероятностью 5/14 победителем будет игрок a. Вероятность того, что победит игрок b, будет такой же, но аналогичная вероятность для c равна 2/7.

[1]) Слабость приведенного выше утверждения понимали и ранее; например, еще А. А. Марков писал: «В известных теоретических вопросах равновозможность рассматриваемых событий представляется нашему уму вполне ясно; в других мы условимся, какие именно события считать равновозможными. В практических же вопросах мы можем быть вынуждены считать равновозможными и такие события, равновозможность которых весьма сомнительна.» (Марков А. А. Исчисление вероятностей.— М.: ГИЗ, 1924, с. 2—3.) А. А. Марков (1856—1922) — выдающийся русский математик, специалист по теории чисел, теории вероятностей и математическому анализу. С 1886 г. профессор Петербургского университета, с 1890 г. академик Петербургской Академии Наук.— *Прим. перев.*

в) Вероятность того, что победитель не определится до k-го тура включительно, равна $1/2^{k-1}$; проверить это.

6. Видоизменить пример 5, б), допустив, что партии могут заканчиваться вничью. Описать соответствующее пространство элементарных событий. Как ввести вероятности?

7. В задаче 3 показать, что $A_1A_2A_3 \subset A_4$ и $A_1A_2A_3' \subset A_4'$.

8. Используя обозначения примера 4, г), показать, что а) $S_1S_2D_3=0$; б) $S_1D_2 \subset E_3$; в) $E_3-D_2S_1 \supset S_2D_1$.

9. Бросают две игральные кости. Пусть A — событие, состоящее в том, что сумма очков — нечетное число, а B — событие, состоящее в том, что хотя бы на одной из костей выпала единица. Описать события AB, $A \cup B$, AB'. Найти их вероятности при условии, что все 36 элементарных событий равновероятны.

10. В примере 2, ж) выяснить смысл следующих событий: а) ABC; б) $A-AB$; в) $AB'C$.

11. Проверить, что в примере 2, ж) $AC' \subset B$.

12. *Бридж* (см. примечание к § 1). Пусть N_k, где $k=1, 2, 3, 4$ — событие, состоящее в том, что игрок «Север» получил не менее k тузов. Пусть S_k, E_k и W_k — аналогичные события для игроков «Юг», «Восток» и «Запад». Что можно сказать о числе тузов у игрока «Запад» при осуществлении событий а) W_1'; б) N_2S_2; в) $N_1'S_1'E_1'$; г) W_2-W_3; д) $N_1S_1E_1W_1$; е) N_3W_1; ж) $(N_2 \cup S_2)E_2$?

13. Проверить, что в предыдущей задаче

а) $S_3 \subset S_2$;	г) $N_2S_2 \subset W_1'$;
б) $S_3W_2=0$;	д) $(N_2 \cup S_2)W_3=0$;
в) $N_2S_1E_1W_1=0$;	е) $W_4=N_1'S_1'E_1'$.

14. Проверить следующие соотношения [1])

а) $(A \cup B)'=A'B'$;	д) $(A \cup B)-AB=AB' \cup A'B$;
б) $(A \cup B)-B=A-AB=AB'$;	е) $A' \cup B'=(AB)'$;
в) $AA=A \cup A=A$;	ж) $(A \cup B)C=AC \cup BC$.
г) $(A-AB) \cup B=A \cup B$;	

15. Найти простые выражения для событий

а) $(A \cup B)(A \cup B')$; б) $(A \cup B)(A' \cup B')(A \cup B')$; в) $(A \cup B)(B \cup C)$.

16. Установить, какие из следующих соотношений правильны:

а) $(A \cup B)-C=A \cup (B-C)$;	ж) $(A \cup B)-A=B$;
б) $ABC=AB(C \cup B)$;	з) $AB'C \subset A \cup B$;
в) $A \cup B \cup C=A \cup (B-AB) \cup (C-AC)$;	и) $(A \cup B \cup C)'=A'B'C'$;
г) $A \cup B=(A-AB) \cup B$;	к) $(A \cup B)'C=A'C \cup B'C$;
д) $AB \cup BC \cup CA \supset ABC$;	л) $(A \cup B)'C=A'B'C$;
е) $(AB \cup BC \cup CA) \subset (A \cup B \cup C)$;	м) $(A \cup B)'C=C-C(A \cup B)$.

17. Пусть A, B, C — три произвольно выбранных события. Найти выражения для событий, состоящих в том, что из A, B, C

а) произошло только A;
б) произошли A и B, но C не произошло;
в) все три события произошли;
г) произошло хотя бы одно из этих событий;
д) произошло хотя бы два события;
е) произошло одно и только одно из этих событий;

---

[1]) Заметим, что $(A \cup B)'$ означает событие, противоположное событию $A \cup B$; оно не совпадает с $A' \cup B'$. Аналогично $(AB)'$ не то же самое, что $A'B'$.

ж) произошло два и только два события;
з) ни одно событие не произошло;
и) произошло не более двух событий.

18. Объединение $A \cup B$ двух событий может быть выражено как объединение двух несовместных событий, именно $A \cup B = A \cup (A - AB)$. Выразить аналогичным образом объединение трех событий A, B, C.

19. Используя результат задачи 18, доказать, что

$$\mathbf{P}\{A \cup B \cup C\} = \mathbf{P}\{A\} + \mathbf{P}\{B\} + \mathbf{P}\{C\} - \mathbf{P}\{AB\} - \mathbf{P}\{AC\} - \mathbf{P}\{BC\} + \mathbf{P}\{ABC\}.$$

[Эта формула является частным случаем формулы (1.5) гл. IV.]

ГЛАВА II

ЭЛЕМЕНТЫ КОМБИНАТОРНОГО АНАЛИЗА

В этой главе поясняются основные понятия комбинаторного анализа и развиваются соответствующие вероятностные основания; в последней части главы описываются некоторые простые аналитические приемы. Для дальнейшего изучения книги не требуется глубокого знания комбинаторного анализа, и читатель, специально им не интересующийся, может после беглого просмотра сразу перейти к главе V, продолжающей основную теоретическую линию гл. I. Лучше всего читать отдельные параграфы этой главы в связи с имеющими к ним отношение темами в последующих главах.

При изучении простых азартных игр, процедур случайного выбора, задач о размещении и упорядочении и т. д. мы, как правило, имеем дело с конечными пространствами элементарных событий, в которых всем точкам приписываются равные вероятности. В этом случае, для того чтобы найти вероятность некоторого события A, мы должны разделить число элементарных событий в A («благоприятные случаи») на общее число элементарных событий («возможные случаи»). Подсчет числа случаев облегчается систематическим использованием нескольких правил, к обзору которых мы сейчас перейдем. Простота и экономия мышления могут быть достигнуты постоянным применением нескольких стандартных приемов, и мы будем следовать этому способу действий вместо того, чтобы описывать кратчайший вычислительный метод в каждом конкретном случае [1]).

§ 1. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Пары. *Из m элементов $a_1, \ldots, a_m$ и n элементов $b_1, \ldots, b_n$ можно образовать mn пар (a_j, b_k), содержащих по одному элементу из каждой группы.*

---

[1]) Интересующийся этими вопросами читатель найдет обширный материал по элементарному комбинаторному анализу, обратившись к классической книге У. Уитворта (Whitworth W. A., Choice and chance, 5th ed., London, 1901; New York, Stechert G. E., 1942). Ее дополняет задачник того же автора DCC exercises, переизданный в Нью-Йорке в 1945 г. и содержащий 700 задач с полными решениями. [См. также книгу Сачков В. Н. Введение в комбинаторные методы дискретной математики.— М.: Наука, 1982 и задачник Комбинаторный анализ. Задачи и упражнения/Под ред. К. А. Рыбникова.— М.: Наука, 1982.— *Перев.*]

Доказательство. Составим из этих пар прямоугольную таблицу (наподобие таблицы умножения) с m строками и n столбцами так, чтобы пара (a_j, b_k) стояла на пересечении j-й строки и k-го столбца. Тогда каждая пара появляется один и только один раз, и утверждение становится очевидным. ▶

Примеры. а) *Карты для бриджа* (см. примечание к § 1 гл. I). В качестве множеств элементов берутся четыре масти и тринадцать значений соответственно. Каждая карта определяется своей мастью и значением, и существует $4\cdot 13=52$ таких комбинаций (карт)

б) «*Семипозиционные торшеры*». Так называются торшеры с тремя обычными лампами и подсветкой, которая может устанавливаться на трех уровнях или вообще не использоваться. Каждая из этих четырех возможностей может комбинироваться с включением 0, 1, 2 или 3 ламп. Следовательно, существует $4\cdot 4=16$ возможных комбинаций, из которых одна — назовем ее (0, 0) — означает, что торшер выключен. Остается пятнадцать (а не семь) способов включения торшера. ▶

Комбинации. *Дано n_1 элементов $a_1, \ldots, a_{n_1}$, n_2 элементов $b_1, \ldots, b_{n_2}, \ldots, n_r$ элементов $x_1, \ldots, x_{n_r}$; число возможных комбинаций $(a_{j_1}, b_{j_2}, \ldots, x_{j_r})$, содержащих по одному элементу каждого типа, равно $n_1\cdot n_2\cdot\ldots\cdot n_r$.*

Доказательство. Если $r=2$, то утверждение сводится к первому правилу. Если $r=3$, то рассматриваем пару (a_i, b_j) как элемент нового типа. Существует n_1n_2 таких пар и n_3 элементов c_k. Каждая тройка (a_i, b_j, c_k) в свою очередь является парой, состоящей из (a_i, b_j) и элемента c_k; следовательно, число троек равно $n_1n_2n_3$. Утверждение для любого r получается по индукции. ▶

Многие применения основаны на следующей переформулировке последней теоремы: *при r последовательных выборах (решениях) с ровно n_k возможными исходами на k-м шаге можно получить $n_1\cdot n_2\cdot\ldots\cdot n_r$ различных результатов.*

Примеры. в) *Классификация по многим признакам.* Предположим, что люди классифицируются по полу, семейному положению и профессии. Различные категории играют роль элементов. Если имеется 17 профессий, то всего будет $2\cdot 2\cdot 17=68$ классов.

г) В *сельскохозяйственном эксперименте* проверяют влияние трех различных факторов (таких, как применение удобрений, опрыскивание и температура). Если эти факторы могут иметь r_1, r_2 и r_3 уровней (или концентраций) соответственно, то всего существует $r_1r_2r_3$ комбинаций или способов воздействия.

д) «*Размещение шаров по ящикам*» равнозначно выбору ящика для каждого шара. Для r шаров мы имеем r независимых выборов, и поэтому *r шаров можно разместить в n ящиках n^r различными спо* *собами*. Вспоминая пример гл. I, 2, б), мы видим, что очень многие мыслимые эксперименты с абстрактной точки зрения эквивалентны размещению шаров по ящикам. Например, рассматривая грани игральной кости как «ящики», получим, что при бросании r раз игральной кости имеется 6^r возможных исходов, в 5^r из которых ни разу не появилась единица. Поэтому при предположении, что все исходы равновероятны, событие «в r бросаниях не было единицы» имеет вероятность $(5/6)^r$. Мы можем наивно полагать, что при шести бросаниях «единица должна появиться», однако вероятность этого события равна только $1-(5/6)^6$, т. е. меньше, чем 2/3 (ср. с примером 3, б)).

е) *Вывешивание флагов* [1]). Рассмотрим менее тривиальный пример, а именно предположим, что r флагов различных цветов вывешиваются на n шестах в ряд. Сколькими способами это можно сделать? Мы пренебрегаем, конечно, абсолютным положением флага на шесте и тем обстоятельством, что число флагов на шесте практически ограничено. Мы предполагаем только, что флаги на каждом шесте располагаются в определенном порядке сверху вниз.

Вывешивание можно осуществить последовательным выбором места для каждого из r флагов. Для первого флага выбирается один из n шестов. Этот шест, таким образом, делится на две части, и, следовательно, имеется $n+1$ возможных выборов положения для второго флага. Аналогично показывается, что для третьего флага возможно $n+2$ выбора и т. д. Отсюда следует, что *возможно* $n(n+1)\cdot\ldots\cdot(n+r-1)$ *различных способов вывешивания*. ▶

§ 2. УПОРЯДОЧЕННЫЕ ВЫБОРКИ

Рассмотрим множество или «генеральную совокупность» из n элементов $a_1, a_2, \ldots, a_n$. *Упорядоченной выборкой объема* r, извлеченной из данной генеральной совокупности, называется любое упорядоченное множество $a_{j_1}, a_{j_2}, \ldots, a_{j_r}$ из r его элементов. Для наглядности можно представить себе, что элементы выбираются один за другим. В этом случае возможны две процедуры. Первая — *выбор с возвращением*; здесь каждое извлечение делается из полной генеральной совокупности, так что один и тот же элемент может быть выбран несколько раз. Такие выборки — упорядоченные множества, в которых допускаются повторения. Вторая процедура — *выбор без возвращения*; здесь элемент, выбранный однажды, исключают из генеральной совокупности, так что выборка превращается в упорядоченное множество без повторений. Очевидно, в этом случае объем выборки r не может быть больше объема генеральной совокупности n.

[1]) Finucan H. M., A teaching sequence for ${}^nH_r$, The Math. Gazette, 48 (1964), 440—441.

При выборе с возвращением каждый из r элементов может быть выбран n способами, и поэтому число возможных выборок будет n^r; это получается из последней теоремы при $n_1=n_2=\ldots=n$. При выборе без возвращения мы имеем n возможных исходов для первого элемента, но только $n-1$ для второго, $n-2$ для третьего и т. д.; таким образом, всего существует $n(n-1)\cdot\ldots\cdot(n-r+1)$ выборок. Произведения такого типа появляются очень часто, поэтому удобно ввести обозначение [1])

$$(n)_r=n(n-1)\cdot\ldots\cdot(n-r+1). \tag{2.1}$$

Ясно, что $(n)_r=0$ для таких целых r и n, что $r>n$. Таким образом, мы получили следующую теорему.

Теорема. *Для генеральной совокупности из n элементов и фиксированного объема выборки r существует n^r различных выборок с возвращением и $(n)_r$ выборок без возвращения.*

Отметим частный случай, когда $r=n$. При выборе без возвращения выборка объема n включает всю генеральную совокупность и представляет собой переупорядочеиие (или *перестановку*) ее элементов. Таким образом, n элементов $a_1, \ldots, a_n$ могут быть упорядочены $(n)_n=n\cdot(n-1)\cdot\ldots\cdot 2\cdot 1$ различными способами. Вместо $(n)_n$ мы пишем $n!$, что является более употребительиым обозначением. Мы видим, что наша теорема имеет такое следствие.

Следствие. *Число различных перестановок из n элементов равно*

$$n!=n(n-1)\cdot\ldots\cdot 2\cdot 1. \tag{2.2}$$

Примеры. а) Три человека A, B и C составляют упорядоченную выборку из генеральной совокупности людей. Их дни рождения есть выборка из генеральной совокупности всех календарных дней, их возрасты есть выборка, составленная из трех чисел.

б) Если под «словом из десяти букв» понимать (возможно, бессмысленную) последовательность из десяти букв, то такое слово представляет собой выборку из генеральной совокупности, состоящей из 26 букв [2]). Поскольку повторения разрешаются, существует 26^{10} таких слов. С другой стороны, в типографии буквы существуют не только умозрительно, но также физически в виде литер. Для простоты предположим, что имеется ровно 1000 литер каждой буквы. Чтобы набрать слово, наборщик должен выбрать десять литер, и здесь повторения исключаются. Поэтому слово может быть набрано $(26000)_{10}$ различными способами. Практически это число мало отличается от 26000^{10} и превышает 10^{44}.

в) Мистер и миссис Смит образуют выборку объема два из генеральной совокупности всех людей; в то же время они образуют

[1]) Обозначение $(n)_r$ не общепринято, но мы будем систематически использовать его в этой книге (*даже когда n не является целым числом*).

[2]) Речь идет об английском алфавите, содержащем 26 букв.— *Прим. перев.*

выборку единичного объема из генеральной совокупности всех супружеских пар. Этот пример показывает, что объем выборки определяется только относительно данной генеральной совокупности. Бросание монеты r раз является одним из способов получения выборки объема r из генеральной совокупности, состоящей из двух букв $Г$ и $Р$. Это же упорядоченное множество из r букв $Г$ и $Р$ представляет собой элементарное событие в пространстве, соответствующем бросанию монеты r раз.

г) *Об упорядочении и выборе на практике.* Когда путем выбора исследуется число курящих в некоторой генеральной совокупности людей, может интуитивно показаться, что порядок внутри выборки не имеет значения, и поэтому начинающий склонен считать такие выборки неупорядоченными. Однако выводы при анализе выборки возможны лишь на основе некоторых вероятностных предположений, для чего необходимо иметь подходящую модель для мыслимого эксперимента формирования выборки. Такой эксперимент, очевидно, включает действия, исходы которых можно отличить один от другого, т. е. они как-то помечаются. Для теоретических целей в качестве меток проще всего использовать целые числа, что приводит к упорядочению выборки. На практике могут быть предпочтительнее другие процедуры, но даже указание «третий парень, которого Джоунз интервьюировал во вторник» является меткой. Иначе говоря даже в том случае, когда порядок внутри выборок можно в конечном счете игнорировать, мыслимый эксперимент включает упорядоченные выборки, и мы увидим, что это приводит к надлежащему приписыванию вероятностей. ►

Последовательное извлечение r элементов из генеральной совокупности объема n является экспериментом, возможные исходы которого представляют собой выборки объема r. Их число равно n^r или $(n)_r$ в зависимости от того, как производится выбор: с возвращением или без возвращения. В каждом из этих случаев наш мыслимый эксперимент описывается пространством элементарных событий, в котором каждая отдельная точка означает выборку объема r.

До сих пор мы не говорили о вероятностях, связанных с выборками. Обычно мы будем приписывать всем им равные вероятности и говорить о случайной выборке. Слово «случайный» не определено строго, но применительно к выборкам или выбору оно имеет единственное значение. Использование термина *случайный выбор* предполагает равновероятность всех исходов. Подобным образом, когда мы говорим о *случайной выборке фиксированного объема r, прилагательное «случайная» означает, что все возможные выборки имеют одну и ту же вероятность*, а именно n^{-r} при выборе с возвращением и $1/(n)_r$ при выборе без возвращения, где n — объем генеральной совокупности, из которой производится выбор. Если n велико, а r относительно мало, то отношение $(n)_r/n^r$ близко к единице. Это дает основание ожидать, что для больших генеральных совокупностей и относительно малых выборок оба способа выбора практически эквивалентны (см. задачи 1 и 2 § 11, а также задачу 35 гл. VI, 10).

Мы ввели терминологию, взятую из практики, но ничего не сказали о применимости наших моделей случайного выбора к действительности. Бросание монеты или игральной кости и аналогичные действия можно рассматривать как эксперименты по практическому осуществлению случайного выбора с возвращением, и наши вероятности численно близки к частотам, наблюдаемым в длинной серии экспериментов, хотя полностью симметричных монет или игральных костей не существует. Типичным примером случайного выбора без возвращения является последовательное вытаскивание карт из перетасованной колоды (предполагается, что колода перетасована гораздо тщательнее, чем это делается обычно). При выборе из генеральной совокупности людей статистик сталкивается со значительными и часто непредвиденными трудностями, и горький опыт показывает, что трудно получить даже грубое подобие случайности.

Упражнение. При выборе без возвращения вероятность того, что любой фиксированный элемент генеральной совокупности содержится в случайной выборке объема r, равна

$$1-(n-1)_r/(n)_r=1-(n-r)/n=r/n.$$

При выборе с возвращением вероятность того, что некоторый элемент входит хотя бы в одну выборку, равна $1-(1-1/n)^r$.

§ 3. ПРИМЕРЫ

Примеры этого параграфа представляют собой частные случаи следующей задачи. Из генеральной совокупности, содержащей n элементов, извлекается случайная выборка с возвращением объема r. Найдем вероятность события, состоящего в том, что ни один элемент не появляется дважды, т. е. что наша выборка могла быть также получена выбором без возвращения. Последняя теорема показывает, что всего существует n^r различных выборок, из которых $(n)_r$ удовлетворяют указанному условию. Предполагая, что все исходы имеют одинаковую вероятность, заключаем, что *вероятность отсутствия повторения в нашей выборке равна*

$$p=(n)_r/n^r=n(n-1)\cdot\ldots\cdot(n-r+1)/n^r. \qquad (3.1)$$

Следующие ниже конкретные интерпретации этой формулы демонстрируют некоторые удивительные явления.

а) *Случайно выбранные числа.* Пусть генеральная совокупность состоит из десяти цифр 0, 1, ..., 9. Каждая последовательность из пяти цифр представляет собой выборку объема $r=5$, и мы предположим, что каждая такая упорядоченная последовательность имеет вероятность 10^{-5}. В соответствии с (3.1) *вероятность того, что все пять последовательных случайных цифр различны, равна* $p=(10)_5\cdot10^{-5}=0{,}3024$.

Мы интуитивно ожидаем, что в больших математических таблицах с большим числом десятичных знаков последние пять цифр будут обладать многими свойствами случайных чисел (в обычных логарифмических и многих других таблицах табличные разности примерно постоянны, и поэтому последние цифры изменяются регулярным образом). В качестве эксперимента были выбраны шестнадцатизначные таблицы [1]) и было найдено количество чисел, в которых последние пять цифр различны. В первых двенадцати группах по сто чисел каждая количество чисел с различными последними пятью цифрами составило 30, 27, 30, 34, 26, 32, 37, 36, 26, 31, 36, 32 соответственно. Теория малых выборок показывает, что величина флуктуаций хорошо согласуется с ожидаемыми пределами. Средняя частота равна 0,3142 и довольно близка к теоретической вероятности 0,3024 (см. пример гл. VII, 4, ж)).

Далее рассмотрим число $e=2{,}71828\ldots$. Из первых 800 десятичных знаков [2]) образуем 160 групп по 5 цифр, а группы объединим в 16 «связок» по 10 групп в каждой. Число групп, в которых все пять цифр различны, в последовательных «связках» равно соответственно

$$3,\ 1,\ 3,\ 4,\ 4,\ 1,\ 4,\ 4,\ 4,\ 2,\ 3,\ 1,\ 5,\ 4,\ 6,\ 3.$$

Частоты также колеблются около значения 0,3024, и теория малых выборок подтверждает, что величины флуктуаций не превосходят ожидаемых. Средняя частота нашего события в 160 группах равна $52/160=0{,}325$ и достаточно близка к $p=0{,}3024$.

б) *Если n шаров случайным образом размещаются в n ящиках, то вероятность того, что каждый ящик занят, равна* $n!/n^n$. Это поразительно мало: для $n=7$ она равна только 0,00612... . Отсюда следует, например, что *если в некотором городе каждую неделю происходит семь автомобильных катастроф, то (в предположении, что все возможные распределения по дням недели равновероятны) практически все недели будут содержать дни, в которые произошло не менее двух катастроф, и в среднем только в одной неделе из* 165 *катастрофы будут равномерно распределены по дням.* Этот пример показывает неожиданное свойство чистой случайности. (Все возможные расположения семи шаров в семи ящиках указаны в табл. 1 § 5. Вероятность того, что два или более ящиков останутся пустыми,

---

[1]) Как было указано в предыдущем издании книги, здесь речь идет о следующем издании: Tables of probability function, v. I, Nat. Bureau of Standards, 1941. [Имеется перевод: Таблицы вероятностных функций. — 2-е стереотип. изд.— М.: ВЦ АН СССР, 1970.] — *Прим. перев.*

[2]) По поводу более полных исследований, проведенных с помощью современных ЭВМ, см. Stoneham R. G., A study of 60 000 digits of the transcendental e, Amer. Math. Monthly, 72 (1965), 483—500 и Pathria R. K., A statistical study of the first 10 000 digits of π, Mathematics of Computation, 16 (1962), 188—197. [См. также Shanks D., Wrench J. W., Jr., Calculation of π to 100 000 decimals, Mathematics Computation, 16 (1962), 76—99.— *Перев.*]

составляет около 0,87.) Для $n=6$ вероятность $n!n^{-n}$ равна 0,01543.... Это показывает, сколь маловероятно, что при шести бросаниях правильной игральной кости выпадут *все* грани. (Вероятность того, что определенная грань не выпадет, близка к 1/3; см. пример 1, д).)

в) *Лифт.* Лифт начинает движение с $r=7$ пассажирами и останавливается на $n=10$ этажах. Какова вероятность p того, что никакие два пассажира не выйдут на одном и том же этаже? Для строгой постановки задачи предположим, что все возможные комбинации выхода пассажиров из лифта равновероятны (что является довольно грубым приближением). Тогда

$$p=10^{-7}(10)_7=(10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4)\cdot 10^{-7}=0{,}06048.$$

Если такое событие однажды произошло, то нужно считать это исключительным явлением и можно ставить примерно 300 против 1, что подобное не повторится (ср. с ответом к задаче 43 § 10).

г) ***Дни рождения.*** Дни рождения r человек образуют выборку объема r из совокупности всех дней года. Годы не одинаковы по продолжительности, и мы знаем, что рождаемость в течение года не остается постоянной. Однако в первом приближении можно считать, что в году 365 дней, и рассматривать случайный выбор людей вместо случайного выбора дней рождения.

В этих предположениях, используя равенство (3.1), получаем, что *вероятность того, что все r дней рождения различны, равна*[1])

$$p=\frac{(365)_r}{365^r}=\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdot\ldots\cdot\left(1-\frac{r-1}{365}\right). \tag{3.2}$$

Численные результаты снова удивительны. Так, для $r=23$ человек мы имеем $p<1/2$, т. е. *для 23 человек вероятность того, что по крайней мере у двух из них дни рождения совпадают, превосходит* 1/2.

Формула (3.2) кажется громоздкой, но для p легко получить хорошее численное приближение. Если r мало, можно пренебречь перекрестными произведениями[2]) и получить в первом приближении[3])

$$p\approx 1-\frac{1+2+\ldots+(r-1)}{365}=1-\frac{r(r-1)}{730}. \tag{3.3}$$

Для $r=10$ точное значение $p=0{,}883\ldots$, тогда как (3.3) дает приближение 0,877.

Для бо́льших[4]) r мы получим гораздо более точное приближение,

[1]) См. von Mises R., Über Aufteilungs- und Besetzungs-Wahrscheinlichkeiten, Revue de la Faculté des Sciences de l'Université d'Istanbul, N. S., 4 (1938—1939), 145—163.

[2]) То есть членами вида $(i/365)(j/365)$, $(i/365)(j/365)(k/365)$ и т. д.— *Прим. перев.*

[3]) Знак $\approx$ означает, что равенство только приближенное. Произведения вида (3.2) встречаются часто, и описываемый метод приближения широко применяется.

[4]) Точнее говоря, для r, бо́льших, чем рассматривавшиеся выше, но малых по сравнению с числом дней в году.— *Прим. перев.*

используя логарифмы. Для малых положительных x $\log(1-x)\approx$ $\approx -x$, и, таким образом, из (3.2) имеем

$$\log p \approx -\frac{1+2+\ldots+(r-1)}{365} = -\frac{r(r-1)}{730}. \tag{3.4}$$

Для $r=30$ это приводит к приближенному значению 0,3037, в то время как точное значение $p=0{,}294$. Для $r\leqslant 40$ погрешность в (3.4) не превосходит 0,08 (к этим вопросам мы вернемся в § 7; см. также ответ к задаче 44 § 10).

§ 4. ПОДМНОЖЕСТВА И РАЗБИЕНИЯ

Как и ранее, будем использовать термин *генеральная совокупность объема n* для обозначения множества из n элементов *безотносительно к их порядку*. Две генеральные совокупности считаются различными только тогда, когда одна из них содержит элемент, не содержащийся в другой.

Рассмотрим подмножество объема r из заданной генеральной совокупности, состоящей из n элементов. Произвольная нумерация элементов этого подмножества превращает его в упорядоченную выборку объема r, и обратно, каждая такая выборка может быть получена указанным путем. Так как r элементов можно занумеровать $r!$ различными способами, отсюда следует, что число упорядоченных выборок объема r в $r!$ раз больше, чем число подмножеств объема r. Поэтому число подмножеств объема r равно $(n)_r/r!$. Выражения такого вида известны как *биномиальные коэффициенты* и обычно обозначаются так:

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n(n-1)\cdot\ldots\cdot(n-r+1)}{1\cdot 2\cdot\ldots\cdot(r-1)\cdot r}. \tag{4.1}$$

Таким образом, доказана следующая теорема.

Теорема 1. *Генеральная совокупность из n элементов имеет $\binom{n}{r}$ различных подмножеств объема $r\leqslant n$!*

Иначе говоря, подмножество из r элементов может быть выбрано $\binom{n}{r}$ различными способами. Такое подмножество однозначно определяется $n-r$ элементами, не принадлежащими ему и образующими подмножество объема $n-r$. Отсюда следует, что подмножеств объема $n-r$ существует ровно столько же, сколько и подмножеств объема r, и, следовательно, для $1\leqslant r\leqslant n$ должно быть

$$\binom{n}{r} = \binom{n}{n-r}. \tag{4.2}$$

Для того чтобы непосредственно доказать (4.2), заметим, что биномиальный коэффициент (4.1) можно записать в виде

$$\binom{n}{r}=\frac{n!}{r!\,(n-r)!} \tag{4.3}$$

(для этого достаточно умножить числитель и знаменатель (4.1) на $(n-r)!$). Заметим, что для $r=0$ левая часть (4.2) не определена, а правая определена. Для того чтобы (4.2) было справедливо для всех целых r, таких, что $0\leqslant r\leqslant n$, *мы положим*

$$\binom{n}{0}=1, \quad 0!=1 \tag{4.4}$$

и $(n)_0=1$.

Примеры. а) *Бридж и покер* (см. примечание к § 1 гл. I). Порядок карт игрока несуществен, и, следовательно, имеется $\binom{52}{13}=$ $=635\,013\,559\,600$ различных комбинаций карт у одного игрока при игре в бридж и $\binom{52}{5}=2\,598\,960$ при игре в покер. Найдем вероятность x того, что при игре в покер фиксированный игрок имеет пять различных по значению карт. Значения карт могут быть выбраны $\binom{13}{5}$ способами, и для каждого значения мы свободны в выборе любой из четырех мастей. Отсюда следует, что $x=$ $=4^5\cdot\binom{13}{5}\binom{52}{5}^{-1}$, что приблизительно равно 0,5071. Для бриджа вероятность тринадцати различных значений карт равна $4^{13}\binom{52}{13}^{-1}$, или приблизительно 0,0001057.

б) Каждый из 50 штатов представлен двумя сенаторами. Рассмотрим события, состоящие в том, что в комитете из 50 случайно выбранных сенаторов: 1) представлен данный штат, 2) представлены все штаты.

В первом случае проще найти вероятность q противоположного события, а именно что данный штат *не* представлен. Всего имеется 100 сенаторов и 98 из них не из данного штата. Следовательно,

$$q=\binom{98}{50}\binom{100}{50}^{-1}=\frac{50\cdot 49}{100\cdot 99}=0{,}24747\ldots .$$

Далее, из теоремы § 2 следует, что комитет, содержащий по одному сенатору из каждого штата, может быть образован 2^{50} различными способами. Поэтому вероятность того, что *все* штаты представлены в комитете, равна $p=2^{50}\binom{100}{50}^{-1}$. Используя формулу Стирлинга (см. § 9), можно показать, что $p\approx\sqrt{2\pi}\cdot 5\cdot 2^{-50}\approx 4{,}126\cdot 10^{-14}$.

в) *Задача о размещении.* Еще раз рассмотрим случайное размещение r шаров по n ящикам (т. е. каждое из n^r возможных размещений имеет вероятность n^{-r}). Для того чтобы найти *вероятность* p_k *того, что фиксированный ящик содержит ровно* k *шаров*, заметим, что k шаров можно выбрать $\binom{r}{k}$ способами, а оставшиеся $r-k$ шаров можно разместить в оставшихся $n-1$ ящиках $(n-1)^{r-k}$ способами. Отсюда следует, что

$$p_k=\binom{r}{k}\cdot\frac{1}{n^r}\cdot(n-1)^{r-k}=\binom{r}{k}\cdot\frac{1}{n^k}\cdot\left(1-\frac{1}{n}\right)^{r-k}. \qquad (4.5)$$

Это частный случай так называемого *биномиального распределения*, которое будет изучено в гл. VI. Численные значения будут приведены в табл. 3 гл. IV. ▶

Различие между различимыми и неразличимыми элементами имеет сходство с отношением между подмножеством и соответствующей упорядоченной выборкой. Вычеркивая все индексы в упорядоченной совокупности (или группе) из r элементов $a_1, \ldots, a_r$, получаем совокупность r неразличимых букв. Обратно, произвольным образом занумеровав r неразличимых букв, получим упорядоченное множество $a_1, \ldots, a_r$. Эта процедура дает $r!$ различных наборов при условии, разумеется, что любая перестановка a_i и a_k производит переупорядочение. Следующие примеры показывают, как этот принцип можно распространить на ситуации, когда элементы a_k лишь частично являются неразличимыми.

Примеры. г) *Флаги одного или двух цветов.* В примере 1, е) было показано, что r флагов можно вывесить на n шестах $N=n(n+1)\times\times\ldots\times(n+r-1)$ различными способами. Теперь мы рассмотрим ту же задачу для флагов одного цвета (рассматриваемых как неразличимые). Занумеровав флаги при таком способе вывешивания, получим ровно $r!$ способов вывешивания r различимых флагов, и, *следовательно, r флагов одного цвета можно вывесить $N/r!$ способами.*

Предположим, далее, что p флагов красные (и неразличимые между собой) и q — синие (где $p+q=r$). Легко видеть, что каждый способ вывешивания r занумерованных флагов может быть получен нумерованием красных флагов от 1 до p и синих флагов — от $p+1$ до $p+q$. Отсюда следует, что число различных способов вывешивания теперь равно $N/(p!q!)$.

д) *Последовательности, содержащие два типа элементов.* Рассмотрим число последовательностей длины $p+q$, состоящих из p букв альфа и q букв бета. Занумеровав альфы от 1 до p и беты от $p+1$ до $p+q$, получим упорядоченную последовательность из $p+q$ различимых элементов. Таких последовательностей имеется $(p+q)!$, и ровно $p!q!$ среди них соответствуют одному и тому же порядку альф и бет. Таким образом, *p альф и q бет можно расположить* *равно*

$$\frac{(p+q)!}{p!\,q!}=\binom{p+q}{p}=\binom{p+q}{q}$$

различимыми способами.

Тот же самый результат непосредственно следует из теоремы 1 и того факта, что все расположения p альф и q бет могут быть получены выбором p из $p+q$ возможных мест и помещением на эти места альф.

е) Число кратчайших путей (составленных из сторон клеток шахматной доски), соединяющих нижний левый и верхний правый углы шахматной доски, равно $\binom{16}{8}=12\,870$. ▶

Теорема 2. *Пусть $r_1, \ldots, r_k$ — целые числа, такие, что*

$$r_1+r_2+\ldots+r_k=n, \quad r_i\geqslant 0. \tag{4.6}$$

Число способов, которыми генеральную совокупность из n элементов можно разделить на k упорядоченных частей (разбить на k подмножеств), из которых первая содержит r_1 элементов, вторая r_2 элементов и т. д., равно

$$\frac{n!}{r_1!\,r_2!\cdot\ldots\cdot r_k!}. \tag{4.7}$$

(Числа (4.7) называются *полиномиальными коэффициентами.*)

Отметим, что порядок подмножеств существен в том смысле, что ($r_1=2$, $r_2=3$) и ($r_1=3$, $r_2=2$) представляют собой разные разбиения; однако игнорируется порядок внутри групп. Отметим также, что $0!=1$, так что обращение r_i в нуль не влияет на формулу (4.7). Так как $r_i=0$ допускается, n элементов разбиваются на *k или менее* подмножеств. Случай $r_i>0$ разбиения *ровно* на k подмножеств рассматривается в задаче 7 § 11.

Доказательство. Повторное использование (4.3) показывает, что число (4.7) можно переписать в виде

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdot\ldots\cdot\binom{n-r_1-\ldots-r_{k-2}}{r_{k-1}}. \tag{4.8}$$

С другой стороны, для того, чтобы осуществить требуемое разбиение, мы должны сначала выбрать r_1 элементов из данных n; из оставшихся $n-r_1$ элементов мы выбираем вторую группу размера r_2 и т. д. После образования $(k-1)$-й группы остается $n-r_1-r_2-\ldots-r_{k-1}=r_k$ элементов, образующих последнюю группу. Мы заключаем, что (4.8) на самом деле выражает число способов, которыми может быть выполнена операция разбиения. ▶

Примеры. ж) *Бридж.* При этой игре 52 карты делятся на четыре равные группы, и поэтому число различных раскладов равно $52!\times$

$\times(13!)^{-4}=(5,36\ldots)\cdot 10^{28}$. Найдем теперь вероятность того, что каждый игрок имеет туза. Четыре туза можно упорядочить $4!=24$ способами, и каждый порядок представляет одну возможность получения одного туза каждым игроком. Оставшиеся 48 карт можно распределить $(48!)\cdot(12!)^{-4}$ способами. Следовательно, искомая вероятность равна $24\cdot 48!\cdot(13)^4/52!=0,105\ldots$.

з) *Игральные кости.* В результате бросания двенадцати игральных костей может быть 6^{12} различных исходов, каждому из которых мы приписываем одинаковую вероятность. Событие, состоящее в том, что каждая грань появляется дважды, может осуществиться столькими способами, сколькими двенадцать костей можно разбить на шесть групп по две в каждой. Следовательно, вероятность этого события есть $12!/(2^6\cdot 6^{12})=0,003438\ldots$.

§ 5*). ПРИЛОЖЕНИЕ К ЗАДАЧАМ О РАЗМЕЩЕНИИ

Примеры гл. I, 2 показывают широкую применимость модели случайного размещения r шаров по n ящикам. Во многих ситуациях необходимо считать шары *неразличимыми*. Например, в статистических исследованиях распределения автомобильных катастроф по дням недели или дней рождения по дням года интересуются лишь числом событий, а не ими самими. Далее, бросание r игральных костей эквивалентно размещению r шаров в $n=6$ ящиках. Хотя возможно следить за r отдельными результатами, обычно предпочитают указывать число единиц, двоек и т. д. В таких случаях мы можем по-прежнему предполагать шары занумерованными, но сосредоточить внимание на событиях, не зависящих от нумерации. Такое событие полностью описывается *числами заполнения* $r_1, r_2, \ldots, r_n$, где r_k означает число шаров в k-м ящике. Каждый набор из n целых чисел, удовлетворяющих соотношениям

$$r_1+r_2+\ldots+r_n=r, \quad r_k\geqslant 0, \tag{5.1}$$

описывает возможную комбинацию чисел заполнения. *В случае неразличимых шаров два размещения различимы только тогда, когда соответствующие наборы* $(r_1, \ldots, r_n)$ *не одинаковы.* Докажем теперь следующую лемму.

Лемма. (i) *Число различимых размещений (т. е. число различных решений уравнения* (5.1)) *равно* [1])

$$A_{r,n}=\binom{n+r-1}{r}=\binom{n+r-1}{n-1}. \tag{5.2}$$

(ii) *Число различимых размещений, при которых ни один ящик не остается пустым, равно* $\binom{r-1}{n-1}$.

*) Материал этого параграфа полезен и нагляден, но не будет непосредственно использоваться в дальнейшем.

[1]) Частный случай $r=100$, $n=4$ был использован в примере гл. I, 2, д).

Доказательство. Обозначим шары звездочками и изобразим n ящиков в виде n промежутков между $n+1$ вертикальными черточками. Например, |***|*||||****| используется для обозначения размещения $r=8$ шаров в $n=6$ ящиках так, что эти ящики содержат соответственно 3, 1, 0, 0, 0, 4 шаров. При таком обозначении в начале и в конце обязательно стоят черточки, но остальные $n-1$ черточек и r звездочек могут быть расположены в произвольном порядке. Отсюда вытекает, что число различимых размещений равно числу способов выбора r мест из $n+r-1$, а именно $A_{r,n}$.

Условие, что ни один из ящиков не оказался пустым, означает, что нет двух стоящих рядом черточек. Между r звездочками имеются $r-1$ промежутков, в $n-1$ из которых помещено по черточке; таким образом, мы имеем $\binom{r-1}{n-1}$ выборов, и утверждение доказано. ▶

Примеры. а) Существует $\binom{r+5}{5}$ различимых исходов бросания r неразличимых игральных костей.

б) *Частные производные.* Частные производные порядка r аналитической функции $f(x_1, \ldots, x_n)$ от n переменных не зависят от порядка дифференцирования, а зависят только от того, сколько раз функция дифференцируется по каждому переменному. Таким образом, каждое переменное играет роль ящика, и, следовательно, в этом случае существует $\binom{n+r-1}{r}$ различных частных производных порядка r. Функция трех переменных имеет пятнадцать производных четвертого порядка и 21 производную пятого порядка. ▶

Рассмотрим теперь n фиксированных целых чисел, удовлетворяющих (5.1). Число размещений r шаров по n ящикам, при которых в ящиках содержится соответственно $r_1, \ldots, r_n$ шаров, дается теоремой 2 § 4. Предполагая, что все n^r возможных размещений равновероятны, получаем, что вероятность такого заполнения равна

$$[r!/(r_1!\, r_2! \cdot \ldots \cdot r_n!)]\, n^{-r}. \tag{5.3}$$

Такое распределение вероятностей было использовано во всех упоминавшихся до сих пор приложениях, и считалось само собой разумеющимся, что это использование присуще интуитивному понятию случайности. Никакое альтернативное распределение, основанное на вероятностных или интуитивных соображениях, даже не предлагалось. Поэтому большой методологический интерес представляет тот факт, что *опыт* заставил физиков заменить распределение (5.3) другими, которые вначале противоречили интуиции. Это будет обсуждено в следующем разделе. В физике (5.3) известно как *распределение Максвелла — Больцмана.*

В ряде случаев необходимо пойти дальше и рассматривать сами ящики как неразличимые: этим достигается то, что порядок чисел заполнения становится несущественным. Следующий пример поясняет стандартный метод решения получающихся при этом задач.

Пример. в) *Размещения $r=7$ шаров по $n=7$ ящикам* (ящики можно интерпретировать как дни недели, а шары как телефонные звонки, письма, происшествия и т. д.). Для определенности рассмотрим размещения с числами заполнения 2, 2, 1, 1, 1, 0, 0, *расположенными в произвольном порядке*. Эти семь чисел заполнения порождают разбиение семи ящиков на три подмножества (категории), состоящие соответственно из двух ящиков, содержащих по два шара, трех ящиков, содержащих по одному шару, и двух пустых ящиков. Такое разбиение на три группы объема 2, 3 и 2 может быть выполнено $7!/(2!\cdot 3!\cdot 2!)$ способами. Каждому отдельному сопоставлению чисел заполнения семи ящикам соответствует $7!/(2!\cdot 2!\cdot 1!\cdot 1!\cdot 1!\cdot 0!\cdot 0!)=7!/(2!\cdot 2!)$ различных распределений $r=7$ шаров по семи ящикам. Таким образом, *общее число размещений, таких, что числа заполнения совпадают с числами* 2, 2, 1, 1, 1, 0, 0, *расположенными в определенном порядке, равно*

$$[7!/(2!\,3!\,2!)]\,[7!/(2!\,2!)]. \tag{5.4}$$

Отметим, что этот результат был получен двойным применением формулы (4.7), а именно и к шарам, и к ящикам. Тот же самый результат можно было бы получить и переформулировать многими способами, однако настоящий метод обеспечивает самый простой путь решения разнообразных задач (см. задачи 43—45 § 10). В табл. 1 приведены числа, аналогичные (5.4), и вероятности для всех возможных наборов чисел заполнения в случае $r=n=7$.

Таблица 1

Случайное размещение 7 шаров по 7 ящикам

Числа заполнения	Число размещений равно $7!\times 7!$, деленному на	Вероятность (число размещений, деленное на 7^7)
1, 1, 1, 1, 1, 1, 1	$7!\times 1!$	0,006 120
2, 1, 1, 1, 1, 1, 0	$5!\times 2!$	0,128 518
2, 2, 1, 1, 1, 0, 0	$2!\ 3!2!\times 2!2!$	0,321 295
2, 2, 2, 1, 0, 0, 0	$3!3!\times 2!2!2!$	0,107 098
3, 1, 1, 1, 1, 0, 0	$4!2!\times 3!$	0,107 098
3, 2, 1, 1, 0, 0, 0	$2!3!\times 3!2!$	0,214 197
3, 2, 2, 0, 0, 0, 0	$2!4!\times 3!2!2!$	0,026 775
3, 3, 1, 0, 0, 0, 0	$2!4!\times 3!3!$	0,017 850
4, 1, 1, 1, 0, 0, 0	$3!3!\times 4!$	0,035 699
4, 2, 1, 0, 0, 0, 0	$4!\times 4!2!$	0,026 775
4, 3, 0, 0, 0, 0, 0	$5!\times 4!3!$	0,001 785
5, 1, 1, 0, 0, 0, 0	$2!4!\times 5!$	0,005 355
5, 2, 0, 0, 0, 0, 0	$5!\times 5!2!$	0,001 071
6, 1, 0, 0, 0, 0, 0	$5!\times 6!$	0,000 357
7, 0, 0, 0, 0, 0, 0	$6!\times 7!$	0,000 008

▶

а. Статистика Бозе — Эйнштейна и Ферми — Дирака

Рассмотрим механическую систему, состоящую из r неразличимых частиц. В статистической механике обычно разбивают фазовое пространство на большое число n малых областей или ячеек, так что каждая частица приписывается ровно одной ячейке. В результате состояние всей системы описывается как случайное размещение r частиц по n ячейкам. На первый взгляд кажется, что (во всяком случае при подходящем выборе n ячеек) все n^r размещений будут равновероятны. Если это так, то физики говорят о *статистике Максвелла — Больцмана* (термин «статистика» используется здесь в смысле, специфическом для физики). Делались многочисленные попытки доказать, что физические частицы ведут себя в соответствии со статистикой Максвелла — Больцмана, однако современная теория, вне сомнения, показала, что эта статистика *не применима ни к каким известным частицам*; ни в одном случае все n^r размещений не являются примерно равновероятными. Были введены две различные вероятностные модели, каждая из которых удовлетворительно описывает поведение некоторого класса частиц. Достоинства конкретной модели зависят от успешности ее применения. Ни одна из них не претендует на универсальность, и не исключено, что когда-нибудь для некоторого класса частиц будет введена новая модель.

Вспомним, что здесь мы имели дело только с *неразличимыми* частицами. Мы имеем r частиц и n ячеек. *В статистике Бозе — Эйнштейна рассматриваются только различимые размещения, и каждому из них приписывается вероятность* $1/A_{r,n}$, где $A_{r,n}$ определяется формулой (5.2). В статистической механике показано, что это предположение справедливо для фотонов, атомных ядер и атомов, содержащих четное число элементарных частиц [1]). Для описания других частиц должно быть введено третье возможное распределение вероятностей. *Статистика Ферми — Дирака* основана на следующих предположениях: 1) *в одной ячейке не могут находиться две или более частиц и* 2) *все различимые размещения, удовлетворяющие первому условию, имеют одинаковую вероятность.* Для выполнения первого предприятия необходимо, чтобы $r \leqslant n$. Тогда размещение полностью описывается указанием того, какие из n ячеек содержат частицу, и, так как существует r частиц, соответствующие ячейки могут быть выбраны $\binom{n}{r}$ способами. Следовательно, *в статистике Ферми — Дирака существует* $\binom{n}{r}$ *возможных размещений, каждое из которых имеет вероятность* $\binom{n}{r}^{-1}$.

Эта модель применима к электронам, нейтронам и протонам. Здесь мы имеем поучительный пример невозможности выбора и обоснования вероятностной модели на основании априорных соображений. Действительно, нет оснований говорить, что фотон и протон не подчиняются одним и тем же вероятностным законам. (Существенные

[1]) См. Margenau H., Murphy G. M., The mathematics of physics and chemistry, New York, Van Nostrand, 1943, Ch. 12.

различия между статистиками Максвелла — Больцмана и Бозе — Эйнштейна обсуждаются в задачах 14—19 § 11.)

Подведем итог: *вероятность того, что ячейки с номерами* 1, 2, ..., n *содержат* $r_1, r_2, \ldots, r_n$ *частиц соответственно (где* $r_1+\ldots+r_n=r$), *дается формулой* (5.3) *в случае статистики Максвелла — Больцмана; она равна* $1/A_{r,n}$ *в случае статистики Бозе — Эйнштейна, и при условии, что каждое* r_j *равно* 0 *или* 1, *равна* $\binom{n}{r}^{-1}$ *в случае статистики Ферми — Дирака.*

Примеры. а) Пусть $n=5$, $r=3$. Размещение $|*|-|*|*|-|$ имеет вероятность 6/125, 1/35 или 1/10 в зависимости от того, используется ли статистика Максвелла—Больцмана, Бозе—Эйнштейна или Ферми—Дирака. Смотри также пример гл. I, 6, б).

б) *Опечатки.* Книга содержит n символов (букв), r из которых являются опечатками. Распределение опечаток соответствует распределению r частиц по n ячейкам, причем в каждой ячейке не может содержаться больше одной частицы. Поэтому естественно предположить, что приближенно для *опечаток имеет место статистика Ферми — Дирака.* (См. задачу 38 § 10.)

6. Применения к сериям

Рассмотрим некоторую упорядоченную последовательность, состоящую из элементов двух типов. Отрезок этой последовательности, состоящий из однотипных элементов и ограниченный с обеих сторон элементами другого типа, называется *серией*. Например, последовательность $\alpha\alpha\alpha\beta\alpha\alpha\beta\beta\beta\alpha$ начинается серией альф длиной 3, за которой следуют серии длиной 1, 2, 3, 1 соответственно. Серии из альф и бет чередуются так, что общее число серий всегда равно числу пар стоящих рядом разнотипных элементов плюс единица.

Примеры применений. Теория серий имеет многочисленные применения в статистике, однако главным образом она используется в задачах проверки случайности и однородности.

а) *При проверке случайности* задача состоит в том, чтобы решить, свойственна ли данному наблюдению случайность или в его основе лежат определенные закономерности. В качестве простого примера предположим, что наблюдение[1]) привело к следующему распределению свободных и занятых мест в кафе: *СЗССЗСССЗСССЗСЗС*. Заметим, что никакие два соседних места не заняты. Может ли это быть случайным? Из пяти занятых и одиннадцати свободных мест можно получить не более одиннадцати серий и в данном наблюдении получено именно такое их число. Позже будет показано, что если бы все варианты выбора мест пятью посетителями были равновероятны, то вероятность одиннадцати серий была бы равна 0,0578... . Столь малая вероятность вселяет подозрение, что имеет место тенденция к умышленному разделению. Это подозрение не может быть доказано статистическими методами, но дальнейшее подтверждение этого может быть получено в результате последующих наблюдений. Если кафе часто посещается семьями, то должна быть тенденция занимать соседние места, и это должно привести к относительно малому числу серий. Аналогично подсчет серий мальчиков и девочек, сидящих в классе, может обнаружить перемешивание, большее или меньшее случайного. Неправдоподобные расположения дают ключ к пониманию

[1]) Swed F. S., Eisenhart C., Tables for testing randomness of grouping in a sequence of alternatives, Ann. Math. Statist., 14 (1943), 66—87.

вызывающих их причин: *избыток серий указывает на сильное перемешивание, недостаток — на сильную группировку.* Конечно, эти заключения нельзя считать надежным доказательством, однако существующие эффективные статистические методы позволяют свести к минимуму риск сделать неверные выводы.

Теория серий, как показал Шухарт, оказывается полезной при контроле качества продукции. Изготовленные шайбы могут различаться по толщине. Длинные серии толстых шайб указывают на возможные неполадки в производственном процессе и заставляют устранять причины; таким образом предупреждается появление брака и достигается бо́льшая однородность изготовляемой продукции.

В полевых биологических экспериментах подсчитывают последовательности здоровых и больных растений, и длинные серии последних заставляют подозревать заражение. Метеоролог следит за чередованием сухих и влажных месяцев [1], чтобы найти ключ к разгадке закономерностей установления погоды.

б) Для понимания типичных задач *проверки однородности* предположим, что на двух группах пациентов испытываются два вида лекарств или что мы интересуемся сравнительной эффективностью двух методов (в медицине, сельском хозяйстве или промышленности). Практически мы будем иметь два множества наблюденных значений, скажем $\alpha_1, \alpha_2, \ldots, \alpha_a$ и $\beta_1, \beta_2, \ldots, \beta_b$, соответствующих двум методам или представляющих собой некоторые характеристики (такие, как вес) элементов двух совокупностей. Альфы и беты являются *числами*, которые мы будем предполагать расположенными в порядке возрастания: $\alpha_1 \leqslant \alpha_2 \leqslant \ldots \leqslant \alpha_a$ и $\beta_1 \leqslant \beta_2 \leqslant \ldots \leqslant \beta_b$. Теперь объединим оба множества в одну последовательность, элементы которой расположены в порядке возрастания. Крайним случаем является случай, когда все альфы предшествуют всем бетам, и это может рассматриваться как указание на большую разницу между двумя методами или совокупностями. С другой стороны, если оба метода одинаково эффективны, то альфы и беты должны располагаться в более или менее случайном порядке. А. Вальд и Дж. Вольфовиц [2] показали, что теория серий может быть успешно применена для обнаружения малых систематических различий (иллюстрирующий это пример, рассмотренный, правда, другим методом, будет дан в гл. III,1,б)). ▶

Большое число задач, связанных с сериями, могут быть решены чрезвычайно простым способом. Пусть даны a неразличимых альф и b неразличимых бет; из примера 4,д) мы знаем, что существуют $\binom{a+b}{a}$ различимых упорядочений. Если имеется n_1 серий из альф, то число серий из бет с необходимостью равно одному из чисел $n_1 \pm 1$ или n_1. Получение из a альф n_1 серий эквивалентно размещению их по n_1 ячейкам так, что ни одна из последних не останется пустой. Из последней леммы следует, что это может быть сделано $\binom{a-1}{n_1-1}$ различимыми способами. Отсюда, например, следует, что существует $\binom{a-1}{n_1-1}\binom{b-1}{n_1}$ последовательностей, содержащих n_1 серий из альф и n_1+1 серий из бет (продолжение в задачах 20—25 § 11).

в) В физике теория серий используется для изучения кооперативных явлений. В теории одномерных решеток Изинга энергия зависит от числа соседних элементов разного типа, т. е. от числа серий. ▶

§ 6. ГИПЕРГЕОМЕТРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ

Многие задачи комбинаторики могут быть сведены к следующей модели. В генеральной совокупности из n элементов имеется n_1

[1] Cochran W. G., An extension of Gold's method of examining the apparent persistence of one type of weather, Quarterly Journal of the Royal Meteorological Society, 64, No. 277 (1938), 631—634.

[2] Wald A., Wolfowitz J., On a test whether two samples are from the same population, Ann. Math. Statist., 2(1940), 147—162.

элементов красного цвета и $n_2=n-n_1$ черного. Случайным образом выбирается группа из r элементов. Найдем вероятность q_k того, что так выбранная группа будет содержать ровно k красных элементов. Здесь k может быть любым целым числом между нулем и наименьшим из чисел n_1 и r.

Для того чтобы найти q_k, заметим, что выбранная группа состоит из k красных и $r-k$ черных элементов. Красные элементы могут быть выбраны $\binom{n_1}{k}$ различными способами, а черные $\binom{n-n_1}{r-k}$ способами. Так как любой выбор красных элементов может комбинироваться с любым выбором черных, имеем

$$q_k=\binom{n_1}{k}\binom{n-n_1}{r-k}\binom{n}{r}^{-1}. \tag{6.1}$$

Определенный таким образом набор вероятностей называется *гипергеометрическим распределением* [1]). Используя (4.3), можно переписать (6.1) в виде

$$q_k=\binom{r}{k}\binom{n-r}{n_1-k}\binom{n}{n_1}^{-1}. \tag{6.2}$$

Замечание. Вероятности q_k определены только для k, не превосходящих r или n_1, но, так как при $b>a$ $\binom{a}{b}=0$, из формул (6.1) и (6.2) следует, что $q_k=0$, если либо $k>n_1$, либо $k>r$. Следовательно, определения (6.1) и (6.2) могут использоваться для всех $k\geqslant 0$ при условии, что соотношение $q_k=0$ интерпретируется как невозможность такого выбора.

Примеры. а) *Проверка качества.* При контроле качества продукции выборочной проверке подвергается партия из n изделий. Дефектные изделия в партии играют роль красных элементов. Их число n_1, конечно, не известно. Производится выборка объема r и определяется число k дефектных изделий в ней. Тогда формула (6.1) позволяет нам сделать выводы относительно истинного значения n_1; это типичная задача статистического оценивания, выходящая, однако, за рамки данной книги.

б) В примере 4, б) генеральная совокупность состоит из $n=100$ сенаторов, из которых $n_1=2$ представляют данный штат (являются «красными»). Случайным образом выбирается группа из $r=50$ сенаторов. Она может включать $k=0$, 1 или 2 сенаторов из данного штата. Из (6.2), учитывая (4.4), находим

$$q_0=q_2=(50\cdot 49)/(100\cdot 99)=0{,}24747\ldots,\qquad q_1=50/99=0{,}50505\ldots.$$

Это значение q_0 было получено другим способом в примере 4, б).

[1]) Это название объясняется тем, что производящая функция (см. гл. XI) последовательности $\{q_k\}$ может быть выражена через гипергеометрические функции.

в) *Оценка размера популяций по данным повторного отлова* [1]. Предположим, что из озера вылавливают 1000 рыб, помечают их красной краской и выпускают обратно. При повторном отлове 1000 рыб среди них оказалось 100 помеченных. Какие выводы можно сделать относительно числа рыб в озере? Это типичная задача *статистического оценивания*. Мы зашли бы слишком далеко, если бы стали описывать разнообразные методы, которые может использовать современный статистик, однако мы покажем, как гипергеометрическое распределение дает ключ к решению этой задачи. Мы предполагаем, конечно, что результаты двух отловов можно рассматривать как случайные выборки из совокупности всех рыб в озере (на практике это предположение исключает тот случай, когда два отлова производятся в одном и том же месте и за короткий промежуток времени). Мы предполагаем также, что число рыб в озере не меняется между двумя отловами.

Обобщим задачу, рассмотрев выборки произвольного объема. Пусть n — (неизвестное) число рыб в озере, n_1 — число рыб в первом улове (играющих роль красных элементов), r — число рыб во втором улове, k — число помеченных рыб во втором улове, $q_k(n)$ — вероятность того, что второй улов содержит ровно k помеченных рыб.

При такой постановке задачи $q_k(n)$, очевидно, вычисляется по формуле (6.1). На практике n_1, r и k могут наблюдаться, а n неизвестно. Отметим, что мы рассматриваем n как неизвестное число, которое никоим образом не зависит от случая. Мы знаем, что было поймано n_1+r-k различных рыб и поэтому $n \geqslant n_1+r-k$. Это все, что можно сказать с полной уверенностью. В нашем примере $n_1=r==1000$ и $k=100$; возможно, что в озере обитает только 1900 рыб, однако, отправляясь от этого предположения, мы придем к заключению, что осуществилось событие фантастически малой вероятности. Действительно, при предположении, что имеется всего $n=1900$ рыб, вероятность того, что две выборки объема 1000 каждая полностью исчерпают всю генеральную совокупность, согласно (6.1), равна

$$\binom{1000}{100}\binom{900}{900}\binom{1900}{1000}^{-1} = \frac{(1000!)^2}{100!\,1900!}.$$

Формула Стирлинга (см. § 9) показывает, что эта вероятность является величиной порядка 10^{-430}, и в такой ситуации здравый смысл заставляет отбросить наше предположение как неправдоподобное. Та же причина заставит нас отбросить предположение, что n очень

[1] Приводя этот пример в первом издании, мы не знали, что описанный метод широко используется на практике. Из работ по этому вопросу отметим Bailey N. T. J., On estimating the size of mobile populations from recapture data, Biometrika, 38 (1951), 293—306 и Chapman D. G., Some properties of the hypergeometric distribution with applications to zoological sample censuses, University of California Publications in Statistics, 1 (1951), 131—160.

велико, скажем равно миллиону. Эти соображения побуждают нас искать такое значение n, при котором $q_k(n)$ достигает своего наибольшего значения, поскольку для такого n наше наблюдение имеет максимальную вероятность. Для каждого фиксированного набора наблюденных значений n_1, r, k значение n, при котором $q_k(n)$ максимально, обозначается через $\hat{n}$ и называется *оценкой максимального правдоподобия* для n. Это понятие было введено Фишером. Для нахождения $\hat{n}$ рассмотрим отношение

$$q_k(n)/[q_k(n-1)] = (n-n_1)(n-r)/[(n-n_1-r+k)n]. \qquad (6.3)$$

Простые вычисления показывают, что это отношение больше единицы при $nk<n_1r$ и меньше единицы при $nk>n_1r$. Это означает, что с возрастанием n последовательность $q_k(n)$ сначала возрастает, а затем убывает; она достигает максимума, когда n является наибольшим целым числом, не превосходящим n_1r/k, так что $\hat{n}$ приблизительно равно n_1r/k. В нашем конкретном примере оценка максимального правдоподобия для числа рыб равна $\hat{n}=10\ 000$.

Истинное значение n может быть больше или меньше $\hat{n}$, и можно поставить задачу нахождения пределов, о которых можно с достаточной степенью уверенности сказать, что n находится внутри их. С этой целью проверим предположение, что n меньше, чем 8500. Подставим в (6.1) $n=8500$, $n_1=r=1000$ и найдем вероятность того, что второй улов содержит не более 100 помеченных рыб. Эта вероятность равна $x=q_0+q_1+\ldots+q_{100}$. Непосредственное вычисление x громоздко, однако, используя нормальное приближение (см. гл. VII), мы легко находим, что $x=0{,}04$. Аналогично, если $n=12\ 000$, то вероятность того, что второй улов содержит 100 или более помеченных рыб, примерно равна 0,03. Эти данные подтверждают предположение, что число рыб находится где-то между 8500 и 12 000. Существуют другие способы получения этих выводов и другие методы оценивания, но мы не намерены обсуждать детали. ▶

Из определения вероятностей q_k следует, что

$$q_0+q_1+q_2+\ldots=1.$$

Из формулы (6.2) поэтому вытекает, что для любых положительных целых n, n_1 и r

$$\binom{r}{0}\binom{n-r}{n_1}+\binom{r}{1}\binom{n-r}{n_1-1}+\ldots+\binom{r}{n_1}\binom{n-r}{0}=\binom{n}{n_1}. \qquad (6.4)$$

Это тождество часто бывает полезным. Мы доказали его лишь для положительных целых n и r, но оно остается справедливым без этого ограничения для произвольных положительных или отрицательных n и r (это равенство не имеет смысла, когда n_1 не является положительным целым). (Указания на два способа доказательства даны в задачах 8 и 9 § 12.)

Гипергеометрическое распределение легко обобщить на случай, когда исходная генеральная совокупность объема n содержит несколько классов элементов. Например, пусть генеральная совокупность содержит три класса объемом n_1, n_2 и $n-n_1-n_2$ соответственно. Если извлекается выборка объема r, то вероятность того, что она содержит k_1 элементов первого класса, k_2 элементов второго и $r-k_1-k_2$ элементов третьего, по аналогии с (6.1) равна

$$\binom{n_1}{k_1}\binom{n_2}{k_2}\binom{n-n_1-n_2}{r-k_1-k_2}\binom{n}{r}^{-1}. \tag{6.5}$$

Разумеется, необходимо, чтобы

$$k_1 \leqslant n_1, \quad k_2 \leqslant n_2, \quad r-k_1-k_2 \leqslant n-n_1-n_2.$$

Пример. г) *Бридж.* Совокупность из 52 карт состоит из четырех классов по тринадцати элементов в каждом. Вероятность того, что у какого-либо игрока будет пять пик, четыре червы, три бубны и одна трефа, равна

$$\binom{13}{5}\binom{13}{4}\binom{13}{3}\binom{13}{1}\binom{52}{13}^{-1}. \qquad \blacktriangleright$$

§ 7. ПРИМЕРЫ, СВЯЗАННЫЕ С ВРЕМЕНЕМ ОЖИДАНИЯ

В этом параграфе мы отступим от непосредственного изучения комбинаторного анализа, для того чтобы рассмотреть некоторые пространства элементарных событий нового типа, к которым приводит простая модификация задач о размещении. Рассмотрим еще раз мыслимый эксперимент случайного размещения шаров по n ящикам. Однако на этот раз мы не будем фиксировать заранее число шаров r, а будем помещать шары один за другим до тех пор, пока не возникнет некоторая предписанная ситуация. Две такие возможные ситуации будут обсуждены подробно: (i) *Случайное размещение шаров продолжается до тех пор, пока некоторый шар впервые попадет в уже занятый ящик.* Процесс заканчивается при первом появлении ящика с двумя шарами. (ii) *Мы фиксируем какой-либо ящик (скажем, ящик номер* 1) *и продолжаем размещение шаров все время, пока этот ящик остается пустым.* Процесс заканчивается тогда, когда в заданный ящик впервые попадет какой-то шар.

Несколько интерпретаций этой модели прольют свет на существо проблемы.

Примеры. а) *Дни рождения.* В примере 3, г) с днями рождения число дней в году $n=365$ и дни соответствуют ящикам, а люди — шарам. Наша модель (i) теперь равнозначна следующему: если мы выбираем людей случайно одного за другим, то сколько их нужно взять, чтобы получить пару с одинаковыми днями рождения? Модель (ii) соответствует ожиданию появления в выборке человека с *моим* днем рождения.

б) *Задача о ключе.* Человек хочет открыть свою дверь. У него n ключей, из которых только один подходит к двери. По причине, о которой можно только догадываться, он пробует ключи случайно, так что при каждой попытке каждый ключ имеет вероятность быть выбранным n^{-1} и все возможные исходы, отвечающие данному числу попыток, равновероятны. Это частный случай модели (ii). Интересно сравнить описанный случайный поиск ключа с более систематическим подходом (задача 11 § 10; см. также задачу 5 гл. V, 8).

в) В предыдущем примере мы можем заменить выбор ключей выбором из произвольной совокупности, скажем из набора купонов. Опять мы хотим узнать, когда ожидается первое повторение выбранного элемента или когда некоторый фиксированный элемент будет выбран в первый раз.

г) *Монеты и игральные кости.* В примере гл. 1, 5, а) монета бросается до тех пор, пока не появится герб. Это частный случай модели (ii) с $n=2$. При бросании игральной кости до появления единицы применяется та же модель с $n=6$. (Другие времена ожидания рассматриваются в задачах 21, 22 и 36 § 10 и в задаче 12 § 11.) ►

Мы начнем с более простой для понимания модели (i). Для того, чтобы указать, что первый, второй, ..., r-й шары заняли ящики с номерами $j_1, j_2, \ldots, j_r$ и что процесс закончился на r-м шаге, удобно использовать набор символов $(j_1, j_2, \ldots, j_r)$. Это означает, что j_i — целое число между 1 и n; кроме того, $j_1, \ldots, j_{r-1}$ все различны, а j_r равно одному из них. Каждое упорядоченное множество такого типа представляет собой элементарное событие. Для r возможны только значения $2, 3, \ldots, n+1$, так как повторное попадание в ящик не может произойти до размещения второго шара или после того, как $(n+1)$-й шар занял ящик. Связь настоящей задачи с прежней моделью размещения фиксированного числа шаров по n ящикам приводит нас к приписыванию каждому элементарному событию $(j_1, \ldots, j_r)$, соответствующему ровно r шарам, вероятности n^{-r}. Мы покажем, что такое соглашение допустимо (т. е. что сумма наших вероятностей равна единице) и что это приводит к разумным результатам.

При фиксированном r совокупность всех элементарных событий $(j_1, \ldots, j_r)$ представляет собой *событие, состоящее в том, что процесс закончится на r-м шаге.* В соответствии с (2.1) числа $j_1, \ldots, j_{r-1}$ могут быть выбраны $(n)_{r-1}$ различными способами, а j_r мы выбираем из $r-1$ чисел $j_1, \ldots, j_{r-1}$. Отсюда следует, что *вероятность того, что процесс закончится на r-м шаге, равна*

$$q_r = \frac{(n)_{r-1}\cdot(r-1)}{n^r} = \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-2}{n}\right)\frac{r-1}{n}, \qquad (7.1)$$

причем $q_1=0$ и $q_2=1/n$. Вероятность того, что процесс будет продолжаться после r-го шага, равна $p_r=1-(q_1+q_2+\ldots+q_r)$, откуда

$p_1=1$ и

$$p_r=\frac{(n)_r}{n^r}=\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right), \tag{7.2}$$

что может быть доказано простой индукцией. В частности, $p_{n+1}=0$ и $q_1+\ldots+q_{n+1}=1$, как и предполагалось. Кроме того, при $n=365$ формула (7.2) сводится к (3.2) и в общем случае наша новая модель приводит к тем же количественным результатам, что и предыдущая модель с фиксированным числом шаров.

Модель (ii) отличается от (i) тем, что она связана с *бесконечным пространством элементарных событий*. Последовательности $(j_1, \ldots, j_r)$ подчинены условию, что числа $j_1, \ldots, j_{r-1}$ отличны от фиксированного заранее числа $a\leqslant n$, а $j_r=a$. Кроме того, нет никакой априорной причины, что процесс должен когда-нибудь закончиться. При фиксированном r мы снова припишем каждому элементарному событию $(j_1, \ldots, j_r)$ вероятность n^{-r}. Для каждого из $j_1, \ldots, j_{r-1}$ имеем $n-1$ возможностей, а для j_r выбора совсем нет. Мы получаем поэтому, что *вероятность того, что процесс закончится на r-м шаге, равна*

$$q_r^*=[(n-1)/n]^{r-1}(1/n), \quad r=1,\ 2,\ \ldots . \tag{7.3}$$

Суммируя эту геометрическую прогрессию, найдем, что $q_1^*+q_2^*+\ldots=1$. Так как сумма вероятностей q_r^* равна единице, нет необходимости вводить элементарное событие, соответствующее тому, что никакой шар не попадет в ящик с фиксированным номером a. Для *вероятности*

$$p_r^*=1-(q_1^*+\ldots+q_r^*)$$

того, что процесс будет продолжаться после r-го шага, получаем

$$p_r^*=(1-1/n)^r, \quad r=1,\ 2,\ \ldots, \tag{7.4}$$

как и следовало ожидать.

Медианой распределения $\{p_r\}$ называется такое значение r, для которого $p_1+\ldots+p_{r-1}\leqslant 1/2$, но $p_1+\ldots+p_r>1/2$; это означает, что приблизительно одинаково вероятно, что процесс продолжится после медианы или закончится до нее[1]. (В примере 3, г) о днях рождения медиана $r=23$.) Чтобы вычислить медиану для $\{p_r\}$, перейдем к логарифмам, как это делалось в (3.4). Когда r мало по сравнению с n, мы видим, что $-\log p_r$ близок к $r^2/(2n)$.

[1] В общем случае медианой ряда значений $x_1\leqslant x_2\leqslant\ldots\leqslant x_n$, которые какая-то величина принимает с вероятностями $p_1, p_2, \ldots, p_n$ соответственно, называется значение x с таким индексом k, что $\sum\limits_{i=0}^{k} p_i\geqslant 1/2$ и $\sum\limits_{i=k}^{n} p_i\geqslant 1/2$.— *Прим. перев.*

Отсюда следует, что *медиана* $\{p_r\}$ близка к $\sqrt{n \cdot 2 \cdot \log 2}$, или приближенно к $(6/5)\sqrt{n}$. Интересно, что медиана возрастает как корень квадратный из объема совокупности. В отличие от этого *медиана распределения* $\{p_r^*\}$ *близка к* $n \cdot \log 2$, или $0{,}7n$, и возрастает линейно с ростом n. Вероятность того, что время ожидания в модели (ii) превысит n, равна $(1-n^{-1})^n$ или приблизительно $e^{-1}=0{,}36788\ldots$.

§ 8. БИНОМИАЛЬНЫЕ КОЭФФИЦИЕНТЫ

Мы использовали биномиальные коэффициенты $\binom{n}{r}$ только для целых положительных n, но очень удобно расширить это определение. Число $(x)_r$, введенное формулой (2.1), а именно

$$(x)_r = x(x-1)\ldots(x-r+1), \tag{8.1}$$

определено для всех действительных x, если r является целым положительным числом. При $r=0$ положим $(x)_0=1$. Тогда

$$\binom{x}{r} = \frac{(x)_r}{r!} = \frac{x(x-1)\ldots(x-r+1)}{r!} \tag{8.2}$$

определяет биномиальные коэффициенты для всех значений x и всех положительных целых r. Для $r=0$ положим, как и в (4.4), $\binom{x}{0}=1$ *и* $0!=1$. *Для отрицательных целых r определим*

$$\binom{x}{r}=0, \quad r<0. \tag{8.3}$$

Мы никогда не будем использовать символ $\binom{x}{r}$, *если r не является целым числом.*

Легко проверить, что при таком определении мы имеем, например,

$$\binom{-1}{r}=(-1)^r, \quad \binom{-2}{r}=(-1)^r(r+1). \tag{8.4}$$

Впоследствии будут использоваться три важные свойства биномиальных коэффициентов. Первое: *для любого целого положительного n*

$$\binom{n}{r}=0, \text{ если либо } r>n, \text{ либо } r<0. \tag{8.5}$$

Второе: *для любого числа x и любого целого r*

$$\binom{x}{r-1}+\binom{x}{r}=\binom{x+1}{r}. \tag{8.6}$$

Эти соотношения легко получают из определения. Доказательство третьего соотношения можно найти в курсах математического анализа: *для любого числа a и любого $-1<t<1$ справедлива формула бинома Ньютона*

$$(1+t)^a=1+\binom{a}{1}t+\binom{a}{2}t^2+\binom{a}{3}t^3+\ldots . \qquad (8.7)$$

Если a — целое положительное число, то все слагаемые в правой части, содержащие степени t выше t^a, автоматически обращаются в нуль и формула справедлива при всех t. Если a не является целым положительным числом, правая часть (8.7) представляет собой *бесконечный* ряд.

Используя (8.4), легко получить, что при $a=-1$ разложение (8.7) превращается в *геометрический ряд*

$$1/(1+t)=1-t+t^2-t^3+t^4-+\ldots . \qquad (8.8)$$

Интегрируя (8.8), получаем еще одну формулу, которая окажется полезной впоследствии, а именно *разложение натурального логарифма в ряд Тейлора*

$$\log(1+t)=t-(1/2)t^2+(1/3)t^3-(1/4)t^4+\ldots . \qquad (8.9)$$

Часто используются две другие формы соотношения (8.9) . Заменив t на $-t$, получим

$$\log[1/(1-t)]=t+(1/2)t^2+(1/3)t^3+(1/4)t^4+\ldots . \qquad (8.10)$$

Сложив два последних равенства, имеем

$$(1/2)\log[(1+t)/(1-t)]=t+(1/3)t^3+(1/5)t^5+\ldots . \qquad (8.11)$$

Все эти разложения имеют место только при $-1<t<1$.

Много полезных соотношений, получающихся из (8.7), будет приведено в § 12. Здесь мы укажем только, что при целом $a=n$ и $t=1$ из формулы (8.7) следует, что

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^n. \qquad (8.12)$$

Эта формула допускает простую комбинаторную интерпретацию. Левая часть представляет собой число способов, которыми множество из n элементов можно разделить на два подмножества, если объем первого подмножества может быть любым числом $k=0, 1, \ldots, n$. С другой стороны, такое разделение можно выполнить непосредственно, положив, что каждый элемент может принадлежать либо первой, либо второй группе. (Аналогичные рассуждения показывают, что сумма полиномиальных коэффициентов равна k^n.)

§ 9. ФОРМУЛА СТИРЛИНГА

Важным инструментом аналитической теории вероятностей является классическая формула [1]), известная как

[1]) Stirling J., Methodus differentialis, 1730.

Формула Стирлинга:

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}, \tag{9.1}$$

где символ $\sim$ указывает на то, что отношение двух выражений стремится к единице при $n \to \infty$.

Эта формула неоценима для многих теоретических исследований и может быть также использована для получения хороших численных приближений. Хотя разность левой и правой частей (9.1) неограниченно возрастает, в действительности имеет значение лишь относительная погрешность. Она равномерно убывает, и приближение Стирлинга необыкновенно точно даже для малых n. Действительно, правая часть (9.1) дает 0,9221 для $1!$, 1,919 для $2!$ и 118,019 для $5! = 120$. Относительные погрешности будут 8, 4 и 2% соответственно. Для $10! = 3\,628\,800$ приближение по формуле Стирлинга дает 3 598 600 с погрешностью 0,8%. Для $100!$ получается погрешность только 0,08%.

Вывод формулы Стирлинга. Наша первая задача — получить оценку для

$$\log n! = \log 1 + \log 2 + \ldots + \log n. \tag{9.2}$$

Так как $\log x$ — монотонная функция от x, имеем

$$\int_{k-1}^{k} \log x\, dx < \log k < \int_{k}^{k+1} \log x\, dx. \tag{9.3}$$

Суммируя по $k = 1, \ldots, n$, получаем

$$\int_{0}^{n} \log x\, dx < \log n! < \int_{1}^{n+1} \log x\, dx, \tag{9.4}$$

или

$$n \log n - n < \log n! < (n+1) \log (n+1) - n. \tag{9.5}$$

Это двойное неравенство наводит на мысль сравнить $\log n!$ с некоторой величиной, близкой к среднему арифметическому крайних членов. Простейшей такой величиной является $(n+1/2)\log n - n$, и поэтому будем оценивать разность [1])

$$d_n = \log n! - (n+1/2)\log n + n. \tag{9.6}$$

Заметим, что

$$d_n - d_{n+1} = (n+1/2)\log [(n+1)/n] - 1. \tag{9.7}$$

Но

$$\frac{n+1}{n} = \frac{1+1/(2n+1)}{1-1/(2n+1)}, \tag{9.8}$$

[1]) Проведенное ниже изящное рассуждение и неравенство (9.14) принадлежат Роббинсу (Robbins H. E., Amer. Math. Monthly, 62 (1955), 26—29).

и, используя разложение (8.11), получаем

$$d_n-d_{n+1}=\frac{1}{3\,(2n+1)^2}+\frac{1}{5\,(2n+1)^4}+\dots\,. \tag{9.9}$$

Из сравнения правой части с геометрической прогрессией со знаменателем $(2n+1)^{-2}$ видно, что

$$0<d_n-d_{n+1}<\frac{1}{3\,[(2n+1)^2-1]}=\frac{1}{12n}-\frac{1}{12\,(n+1)}\,. \tag{9.10}$$

Из (9.9) заключаем, что последовательность $\{d_n\}$ убывает, в то время как (9.10) показывает, что последовательность $\{d_n-(12n)^{-1}\}$ возрастает. Отсюда следует, что существует конечный предел

$$C=\lim d_n. \tag{9.11}$$

Но в силу (9.6) соотношение $d_n\to C$ эквивалентно

$$n!\sim e^C\cdot n^{n+1/2}\,e^{-n}. \tag{9.12}$$

Это и есть формула Стирлинга, только константа C еще не определена. В гл. VII, 2 будет доказано, что $e^C=\sqrt{2\pi}$. Доказательство этого факта является элементарным и не зависит от материала гл. IV—VI; оно отложено до гл. VII потому, что оно естественно связано с теоремой о нормальном приближении[1]).

Уточнения. Мы можем получить неравенство в другую сторону, аналогичное (9.10). В самом деле, из (9.9) очевидно, что

$$d_n-d_{n+1}>\frac{1}{3\,(2n+1)^2}>\frac{1}{12n+1}-\frac{1}{12\,(n+1)+1}\,. \tag{9.13}$$

Отсюда следует, что последовательность $\{d_n-(12n+1)^{-1}\}$ убывает. А так как $\{d_n-(12n)^{-1}\}$ возрастает, выполняется двойное неравенство

$$C+1/(12n+1)<d_n<C+1/(12n). \tag{9.14}$$

Заменив в последней формуле d_n его выражением (9.6) и положив (пока без доказательства) $e^C=\sqrt{2\pi}$, получим

$$\sqrt{2\pi}n^{n+1/2}\,e^{-n}\cdot e^{(12n+1)^{-1}}<n!<\sqrt{2\pi}n^{n+1/2}\,e^{-n}\,e^{(12n)^{-1}}. \tag{9.15}$$

Это двойное неравенство дополняет формулу Стирлинга замечательным образом. Отношение крайних членов близко к $1-(12n^2)^{-1}$, и, следовательно, *величина в правой части* (9.15) *превзойдет* $n!$, *но с погрешностью меньшей, чем* $9n^{-2}$ %. В действительности погрешность намного меньше[2]): для $n=2$ правая часть (9.15) дает 2,0007, для $n=5$ получаем 120,01.

---

[1]) Обычно доказательство равенства $e^C=\sqrt{2\pi}$ основывается на формуле Валлиса. Простое прямое доказательство см. в статье автора (Feller W., Amer. Math. Monthly, 1967).

[2]) Основываясь на (9.9), можно показать, что $d_n=C+(12n)^{-1}-(360n^3)^{-1}+\dots$, где многоточие означает слагаемые, убывающие быстрее, чем n^{-4}.

ЗАДАЧИ

Замечание. В § 10 включены стандартные упражнения. Задачи теоретического характера и различные дополнения к тексту содержатся в § 11 и 12.

§ 10. УПРАЖНЕНИЯ И ПРИМЕРЫ

Замечание. В каждом случае предполагается, что все комбинации имеют равные вероятности.

1. Сколько можно образовать различных инициалов, если каждый человек имеет одну фамилию и а) ровно два имени, б) не больше двух имен, в) не больше трех имен?

2. В азбуке Морзе буквы представляются последовательностями тире и точек с возможными повторениями. Сколько букв можно составить из десяти или менее символов [1])?

3. Каждая кость домино помечается двумя числами. Кости симметричны, так что числа в парах не упорядочены. Сколько различных костей можно образовать, используя числа $1, 2, \ldots, n$?

4. Числа $1, 2, \ldots, n$ расположены в случайном порядке. Найти вероятность того, что числа а) 1 и 2, б) 1, 2 и 3 расположены рядом в указанном порядке.

5. Игрок A бросает шесть игральных костей и выигрывает, если выпадет хотя бы одна единица. Игрок B бросает двенадцать игральных костей и выигрывает, если выпадут хотя бы две единицы. У кого больше вероятность выиграть [2])?

Указание. Найти вероятности проигрыша.

6. а) Найти вероятность того, что из трех случайно выбранных цифр ровно 2, 1, 0 будут повторяться. б) Решить ту же задачу для четырех случайно выбранных цифр.

7. Найти вероятность p_r того, что среди r случайно выбранных цифр нет двух равных. Вычислить значение p_{10} по формуле Стирлинга.

8. Чему равна вероятность того, что среди k случайно выбранных цифр а) не встретится 0; б) не встретится 1; в) не встретится ни 0, ни 1; г) не встретится хотя бы одна из двух цифр 0 или 1? Пусть A и B обозначают события а) и б). Выразить остальные события через A и B.

9. Найти вероятность того, что при случайном размещении n шаров по n ящикам ровно один ящик останется пустым.

10. На автомобильной стоянке двенадцать мест расположены в один ряд. Некто заметил, что на стоянке находится восемь автомобилей и что четыре свободных места примыкают друг к другу (образуют одну серию). Является ли такое расположение четырех свободных мест неожиданным (указывающим на отсутствие случайности)?

11. Человеку дают n ключей из которых только один подходит к его двери. Он испытывает их последовательно (выбор без возвращения). Этот процесс может потребовать $1, 2, \ldots, n$ испытаний. Показать, что каждый из этих исходов имеет вероятность n^{-1}.

12. Каждая из n палок разламывается на две части — длинную и короткую. Затем $2n$ обломков объединяются в n пар, каждая из которых образует новую «пал.

[1]) То есть сколько существует последовательностей не более чем из десяти тире и точек? — *Прим. перев.*

[2]) Это перефразировка вопроса, заданного в 1693 г. И. Ньютону знаменитым в свое время Сэмюелем Пипсом. [Пипс С. (Pepys S., 1633—1703) — выдающийся общественный деятель; в 1684 г. был избран президентом Королевского Общества. Знаменит прежде всего своим тайным дневником, частично опубликованным в 1825 г.— *Перев.*] Ньютон ответил, что «легкие вычисления» показывают преимущество A. В подтверждение он позже представил эти вычисления, но не смог убедить Пипса. Короткий документальный отчет см. в статье Schell E. D., Samuel Pepys, Isaac Newton, and probability, The Amer. Statistician, 14 (1960), 27—30, где содержатся ссылки на Private correspondence and miscellaneous papers of Samuel Pepys, London, G. Bell and Sons, 1926.

ку». Найти вероятность того, что а) части будут соединены в первоначальном порядке, б) все длинные части будут соединены с короткими [1]).

13. *Проверка статистической гипотезы.* Некий профессор Корнеллского университета был оштрафован двенадцать раз за незаконную ночную стоянку автомобиля. Все двенадцать штрафов налагались во вторник или в четверг. Найти вероятность этого события. (Имело ли смысл арендовать гараж только на вторники и четверги?)

14. *Продолжение.* Из двенадцати штрафов ни один не был наложен в воскресенье. Свидетельствует ли это о том, что в воскресенье штрафы не налагаются?

15. Ящик содержит девяносто годных и десять дефектных шурупов. Если использовать десять шурупов, какова вероятность того, что ни один из них не окажется дефектным?

16. Из генеральной совокупности, состоящей из пяти символов a, b, c, d, e, производится выборка с возвращением объема 25. Найти вероятность того, что выборка содержит по пяти символов каждого вида. Сравнить результат с таблицами случайных чисел [2]), отождествив цифры 0 и 1 с a, цифры 2 и 3 с b и т. д.

17. Пусть n человек, среди которых находятся A и B, становятся в ряд. Какова вероятность того, что между A и B окажется ровно r людей? Показать, что, если они становятся не в ряд, а в круг, эта вероятность не зависит от r и, следовательно, равна $1/(n-1)$. (При круговом расположении рассматривать только дуги, идущие от A к B в положительном направлении.)

18. Чему равна вероятность того, что два бросания трех игральных костей дадут один и тот же результат, если а) кости различимы, б) кости не различимы?

19. Показать, что более вероятно получить хотя бы одну единицу при бросании четырех игральных костей, чем хотя бы одну пару единиц при 24 бросаниях двух костей. Ответ известен под названием парадокса де Мере [3]).

20. Из генеральной совокупности, содержащей n элементов, извлекается выборка объема r. Найти вероятность того, что ни один из N данных элементов не будет содержаться в выборке, предполагая, что выбор производится а) без возвращения, б) с возвращением. Сравнить численные результаты при двух способах выбора, когда (i) $n=100$, $r=N=3$ и (ii) $n=100$, $r=N=10$.

21. *Распространение слухов.* В городе проживает $n+1$ человек. Один из них, узнав новость, сообщает ее другому, тот — третьему и т. д., причем передача новости осуществляется следующим образом: человек, которому сообщена новость, случайным образом выбирает одного из n жителей и сообщает новость ему, тот поступает точно так же и т. д. Найти вероятность того, что новость будет передана r раз без а) возвращения к человеку, который узнал ее первым, б) повторного сообщения кому-либо. Решить ту же задачу, когда на каждом шаге новость сообщается одним человеком группе из N случайно выбранных людей. (Первая часть задачи является частным случаем при $N=1$.)

22. *Цепь писем.* В генеральной совокупности из $n+1$ людей человек, называемый «прародителем», посылает письма двум случайно выбранным людям, об-

[1]) Когда клетки подвергаются воздействию губительной радиации, некоторые хромосомы (играющие роль наших «палок») разрываются. «Длинными» считаются те части, которые содержат так называемые центромеры. Если соединяются две «длинных» или две «коротких» части, то клетка гибнет. См. Catcheside D. G., The effect of X-ray dosage upon the frequency of induced structural changes in the chromosomes of Drosophila Melanogaster, Journal of Genetics, 36 (1938), 307—320.

[2]) Соответствие порой оказывается удивительно точным; см. Greenwood J. A., Stuart E. E., Review of Dr. Feller's critique, Journal for Parapsychology, 4 (1940), 298—319, в особенности с. 306.

[3]) Неоднократно утверждалось, что эта задача возникла за игорным столом и что в 1654 г. де Мере предложил ее Паскалю. Как полагают, этот случай оказал большое стимулирующее воздействие на развитие теории вероятностей. На самом деле эта задача была поставлена Кардано (1501—1576); см. Ore O., Pascal and the invention of probability theory, Amer. Math. Monthly, 67 (1960), 409—419 и Cardano, the gambling scholar, Princeton, Princeton Univ. Press, 1953.

разующим «первое поколение». Они делают то же самое, и вообще каждый получатель письма посылает письма двум случайно выбранным людям независимо от всего предыдущего. Найти вероятность того, что «поколения» с номерами 1, 2, ..., r не включают в себя «прародителя». Найти медиану распределения, считая n достаточно большим.

23. *Семейная задача.* В некоторой семье четыре сестры моют посуду по очереди. Из четырех разбитых тарелок три разбито младшей сестрой, и поэтому ее называют неуклюжей. Можно ли ее оправдать, приписывая эти неудачи случайности? Обсудить связь со случайным размещением шаров.

24. Найти вероятность того, что а) дни рождения двенадцати человек придутся на двенадцать разных месяцев года (предположить, что все месяцы равновероятны), б) дни рождения шести человек придутся в точности на два месяца.

25. Найти вероятность того, что для данных 30 человек шесть из двенадцати месяцев года содержат по два дня рождения и шесть — по три.

26. В чулане лежат n пар ботинок. Случайно выбираются $2r$ ботинок ($2r<n$). Чему равна вероятность того, что среди них а) не будет ни одной пары, б) будет ровно одна пара, в) ровно две пары?

27. На автомобильной стоянке N мест расположены в один ряд. Прибывший на стоянку автомобиль занимает одно из свободных мест (не с краю). По возвращении его владелец обнаруживает, что ровно r мест еще заняты. Какова вероятность того, что оба соседних места свободны?

28. Группа из $2N$ мальчиков и $2N$ девочек делится на две равные части. Найти вероятность p того, что каждая часть содержит одинаковое число мальчиков и девочек. Оценить p, используя формулу Стирлинга.

29. Доказать, что при игре в бридж вероятность p получения «Западом» ровно k тузов равна вероятности того, что произвольный игрок имеет ровно k тузов. (Это интуитивно ясно. Заметим, однако, что эти две вероятности относятся к двум различным экспериментам, так как во втором случае случайно выбираются тринадцать карт, а в первом случае распределяются все 52 карты.)

30. Вероятность того, что при игре в бридж «Восток» получит m, а «Юг» n пик равна вероятности того, что в двух наборах по тринадцати случайным образом извлеченных из колоды карт первый содержит m и второй n пик.

31. Чему равна вероятность того, что у «Севера» и «Юга» вместе имеется ровно k тузов, где $k=0, 1, 2, 3, 4$?

32. Пусть a, b, c, d — четыре неотрицательных целых числа, таких, что $a+b+c+d=13$. Найти вероятность $p(a, b, c, d)$ того, что при игре в бридж игроки «Север», «Восток», «Юг», «Запад» имеют a, b, c, d пик соответственно. Описать схему размещения красных и черных шаров по ящикам, которая включает эту задачу в качестве частного случая.

33. Используя результат задачи 32, найти вероятность того, что один игрок получит a, второй b, третий c и последний d пик, если а) $a=5$, $b=4$, $c=3$, $d=1$; б) $a=b=c=4$, $d=1$; в) $a=b=4$, $c=3$, $d=2$. Заметим, что эти три случая существенно различаются.

34. Пусть a, b, c, d — целые числа, такие, что $a+b+c+d=13$. Найти вероятность $q(a, b, c, d)$ того, что один из игроков в бридж будет иметь a пик, b черв, c бубен и d треф, и показать, что задача не сводится ни к одному из случайных размещений тринадцати шаров по четырем ящикам. Почему?

35. *Распределение тузов в наборе из r карт.* Вычислить вероятности $p_0(r)$, $p_1(r)$, ..., $p_4(r)$ того, что среди r карт, случайно извлеченных из колоды при игре в бридж, соответственно 0, 1, ..., 4 туза. Проверить, что $p_0(r)=p_4(52-r)$.

36. *Продолжение: время ожидания.* Найти вероятности $f_1(r)$, ..., $f_4(r)$ того, что при последовательном извлечении карт из колоды первый, ..., четвертый туз появится при r-м испытании. *Угадать медианы* времени ожидания для первого, ..., четвертого туза, а затем вычислить их.

37. Найти вероятность того, что каждый из двух наборов карт содержит ровно k тузов, если каждый набор составлен из r карт и выбор производится из а) одной и той же колоды для игры в бридж, б) двух таких колод. Показать, что при $r=13$

вероятность события в случае а) равна вероятности того, что два фиксированных игрока получат ровно по k тузов каждый.

38. *Опечатки.* Каждая страница книги содержит N символов, причем возможны опечатки. Книга содержит $n=500$ страниц и $r=50$ опечаток. Показать, что а) вероятность того, что страницы с номерами $1, 2, \ldots, n$ содержат соответственно $r_1, r_2, \ldots, r_n$ опечаток, равна

$$\binom{N}{r_1}\binom{N}{r_2}\cdots\binom{N}{r_n}\binom{nN}{r}^{-1};$$

б) для больших N эта вероятность может быть приближенно представлена формулой (5.3). Заключить отсюда, что *r опечаток распределены по n страницам приблизительно в соответствии со случайным размещением r шаров по n ящикам.* (*Замечание.* Распределение r опечаток по N имеющимся местам соответствует статистике Ферми — Дирака. Наше утверждение может быть сформулировано как общее предельное свойство статистики Ферми — Дирака; см. разд. а § 5.)

Замечание. Следующие задачи имеют отношение к материалу § 5.

39. Найти число различимых способов размещения r_1 неразличимых предметов одного вида и r_2 неразличимых предметов другого вида по n ящикам.

40. Сколько существует различимых результатов совместного бросания r_1 игральных костей и r_2 монет?

41. Сколькими различимыми способами можно расположить в ряд r_1 белых, r_2 черных и r_3 красных шаров?

42. Найти вероятность того, что при случайном расположении в ряд 52-х карт для игры в бридж никакие два туза не будут находиться рядом.

43. *Лифт.* В примере 3, в) лифт начинает движение с семью пассажирами и останавливается на десяти этажах. Различные распределения пассажиров по этажам символически можно представить записью вида (3, 2, 2), которая интерпретируется как событие, состоящее в том, что три пассажира вышли из лифта на одном этаже, еще два пассажира на другом этаже и последние два — на еще одном этаже. Найти вероятности пятнадцати возможных распределений пассажиров от (7) до (1, 1, 1, 1, 1, 1, 1).

44. *Дни рождения.* Найти вероятности возможных распределений дней рождения 22 человек.

45. Найти вероятность того, что у игрока в покер будет а) флеш ройяль (десятка, валет, дама, король и туз одной масти); б) каре (четыре карты одного значения); в) фул (три карты одного значения и две карты другого значения); г) стрит (пять последовательных по значению карт произвольных мастей); д) тройка (три карты одного значения плюс две карты других значений, различающиеся по значению между собой); е) две двойки (две пары карт одного значения в каждой паре плюс карта отличного от них значения); ж) одна двойка (пара карт одинакового значения плюс три отличные от них по значению и различные по значению между собой карты).

§ 11. ЗАДАЧИ И ДОПОЛНЕНИЯ ТЕОРЕТИЧЕСКОГО ХАРАКТЕРА

1. Генеральная совокупность из n элементов включает в себя np красных и nq черных ($p+q=1$). По схеме выбора с возвращением извлекается случайная выборка объема r. Показать, что вероятность того, что она содержит ровно k красных элементов, равна

$$\binom{r}{k}p^kq^{r-k}. \tag{11.1}$$

2. *Предельная теорема для гипергеометрического распределения.* Если n велико и $n_1/n=p$, то вероятность q_k, задаваемая соотношениями (6.1) и (6.2), близка к (11.1). Точнее,

$$\binom{r}{k}\left(p-\frac{k}{n}\right)^k\left(q-\frac{r-k}{n}\right)^{r-k} < q_k < \binom{r}{k}p^kq^{r-k}\left(1-\frac{r}{n}\right)^{-r}. \tag{11.2}$$

Сравнение этой задачи с предыдущей показывает, *что для больших генеральных совокупностей практически нет разницы между выбором с возвращением и выбором без возвращения.*

3. Из генеральной совокупности, содержащей n элементов, по схеме выбора *без возвращения* извлекается случайная выборка объема r. Вероятность u_r того, что в выборке будут содержаться N данных элементов, равна

$$u_r = \binom{n-N}{r-N}\binom{n}{r}^{-1}. \tag{11.3}$$

(Соответствующей формулой для выбора *с возвращением* является (11.10), но она не может быть получена прямыми рассуждениями. Другой вид записи (11.3) см. в задаче 9 гл. IV, 6.)

4. *Предельный случай.* Если $n \to \infty$ и $r \to \infty$ так, что $r/n \to p$, то $u_r \to p^N$ (см. задачу 13).

Замечание[1]). Задачи 5—13 относятся к классической задаче о размещении (статистика Максвелла—Больцмана); иначе говоря, r частиц размещаются по n ячейкам и каждое из n^r возможных распределений имеет вероятность n^{-r}.

5. Вероятность p_k того, что данная ячейка содержит ровно k частиц, задается биномиальным распределением (4.5). Наиболее вероятным значением k является целое число ν, такое, что $(r-n+1)/n < \nu \leqslant (r+1)/n$. (Иначе говоря, утверждается, что $p_0 < p_1 < \ldots < p_{\nu-1} \leqslant p_\nu > p_{\nu+1} > \ldots > p_r$; ср. с задачей 15.)

6. *Предельное распределение.* Если $n \to \infty$ и $r \to \infty$ так, что среднее число $\lambda = r/n$ частиц на ячейку остается постоянным, то

$$p_k \to e^{-\lambda}\lambda^k/k!. \tag{11.4}$$

(Это — *распределение Пуассона*, обсуждаемое в гл. VI; относительно соответствующей теоремы для статистики Бозе—Эйнштейна см. задачу 16).

7. Пусть $A(r, n)$ — число размещений, при которых *ни одна ячейка не остается пустой*. При помощи комбинаторных рассуждений показать, что

$$A(r, n+1) = \sum_{k=1}^{r} \binom{r}{k} A(r-k, n). \tag{11.5}$$

Вывести отсюда, что

$$A(r, n) = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} (n-\nu)^r. \tag{11.6}$$

Указание. Воспользоваться индукцией: предположить, что (11.6) выполняется, и в соответствии с ним выразить $A(r-k, n)$ в (11.5). Изменить порядок суммирования и использовать формулу бинома для представления $A(r, n+1)$ в виде разности двух простых сумм. Заменить во второй сумме $\nu+1$ новым индексом суммирования и использовать (8.6).

Замечание. Формула (11.6) дает теоретическое решение поставленной задачи, но, очевидно, было бы неблагодарной работой использовать ее для вычисления, скажем, вероятности x того, что в деревне с $r = 1900$ жителями каждый день в году является днем рождения. В гл. IV, 2 мы выведем (11.6) другим способом и по-

---

[1]) Задачи 5—19 играют определенную роль в квантовой статистике, теории светочувствительных материалов, счетчиков Гейгера — Мюллера и т. п. Поэтому некоторые формулы часто обсуждались и заново открывались в физической литературе, зачастую без выявления их классического и совершенно элементарного характера. Возможно, все эти задачи (хотя и в иной постановке) имеются в книге Уитворта, цитированной в начале этой главы.

лучим простую приближенную формулу (показывающую, например, что приблизительно $x=0,135$).

8. Показать, что *число распределений, оставляющих ровно m ячеек пустыми, равно*

$$E_m(r, n)=\binom{n}{m} A(r, n-m)=\binom{n}{m}\sum_{\nu=0}^{n-m}(-1)^{\nu}\binom{n-m}{\nu}(n-m-\nu)^r. \qquad (11.7)$$

9. Не используя предыдущие результаты, показать, что *вероятность*

$$p_m(r, n)=n^{-r}E_m(r, n)$$

того, что имеется ровно m пустых ячеек, удовлетворяет уравнению

$$p_m(r+1, n)=p_m(r, n)(n-m)/n+p_{m+1}(r, n)(m+1)/n. \qquad (11.8)$$

10. Используя результаты задач 7 и 8, показать прямыми вычислениями, что (11.8) выполняется. Показать, что этот метод обеспечивает *новый вывод (индукцией по r)* (11.6).

11. Из задачи 8 получить, что *вероятность $x_m(r, n)$ того, что не менее m ячеек остаются пустыми, равна*

$$\binom{n}{m}\sum_{\nu=0}^{n-m}(-1)^{\nu}\binom{n-m}{\nu}\left(1-\frac{m+\nu}{n}\right)^r\frac{m}{m+\nu}. \qquad (11.9)$$

(При $m\geqslant n$ это выражение равно нулю, как и должно быть.)

Указание. Показать, что $x_m(r, n)-p_m(r, n)=x_{m+1}(r, n)$.

12. *Вероятность того, что каждая из N заданных ячеек занята, равна*

$$u(r, n)=n^{-r}\sum_{k=0}^{r}\binom{r}{k}A(k, N)(n-N)^{r-k}. \qquad (11.10)$$

Вывести отсюда, что

$$u(r, n)=\sum_{\nu=0}^{N}(-1)^{\nu}\binom{N}{\nu}\left(1-\frac{\nu}{n}\right)^r. \qquad (11.11)$$

(Использовать формулу бинома. Для $N=n$ имеем $u(r, n)=n^{-r}A(r, n)$. Заметим, что (11.11) является аналогом (11.3) для *выбора с возвращением*[1]. Другой вывод см. в задаче 8 гл. IV, 6.)

13. *Предельный случай.* При предельном переходе, описанном в задаче 4, получается $u(r, n)\longrightarrow(1-e^{-p})^N$.

Замечание. В задачах 14—19 r и n имеют тот же смысл, что и ранее, но мы предполагаем, что шары неразличимы и все различимые размещения имеют равные вероятности (статистика Бозе — Эйнштейна).

[1]) Заметим, что $u(r, n)$ можно интерпретировать как вероятность того, что время ожидания до момента, когда N-й элемент присоединится к выборке, меньше r. Результат может быть применен к случайному выбору цифр: здесь $u(r, 10)-u(r-1, 10)$ есть вероятность того, что потребуется последовательность из r элементов, для того чтобы она включала все 10 цифр. Этот результат можно использовать в качестве критерия случайности. Р. Гринвуд (Greenwood R. E., Coupon collector's test for random digits, Mathematical Tables and Other Aids to Computation, 9 (1955), 1—5) табулировал это распределение и сравнил его с результатами, которые дают соответствующие времена ожидания для первых 2035 десятичных знаков π и первых 2486 десятичных знаков e. Медиана времени ожидания полного набора всех десяти цифр равна 27. Вероятность того, что это время ожидания превзойдет 50, больше 0,05, а вероятность того, что время ожидания превзойдет 75, приблизительно равна 0,0037.

14. Вероятность того, что заданный ящик содержит ровно k шаров, равна

$$q_k=\binom{n+r-k-2}{r-k}\binom{n+r-1}{r}^{-1}. \tag{11.12}$$

15. Показать, что при $n>2$ наиболее вероятным числом шаров в любом заданном ящике является нуль или, точнее, $q_0>q_1>\dots$ (ср. с задачей 5).

16. *Предельная теорема.* Пусть $n\to\infty$ и $r\to\infty$ так, что среднее число шаров на ящик r/n стремится к λ; тогда

$$q_k\to\lambda^k/(1+\lambda)^{k+1}. \tag{11.13}$$

(Правая часть известна под названием *геометрического распределения.*)

17. Вероятность того, что ровно m ящиков остаются пустыми, равна

$$p_m=\binom{n}{m}\binom{r-1}{n-m-1}\binom{n+r-1}{r}^{-1}. \tag{11.14}$$

18. Вероятность того, что группа из m фиксированных ящиков содержит в сумме ровно j шаров, равна

$$q_j(m)=\binom{m+j-1}{m-1}\binom{n-m+r-j-1}{r-j}\binom{n+r-1}{r}^{-1}. \tag{11.15}$$

19. *Предельное распределение.* При переходе к пределу, как в задаче 4, имеем

$$q_j(m)\to\binom{m+j-1}{m-1}\frac{p^j}{(1+p)^{m+j}}. \tag{11.16}$$

(Правая часть является частным случаем *отрицательного биномиального распределения,* которое будет введено в гл. VI, 8.)

Теоремы о сериях. *В задачах 20—25 мы рассматриваем последовательности из r_1 альф и r_2 бет и предполагаем, что все они имеют одинаковые вероятности (см. пример 4, д)). Эта часть задач связана с разд. 6 § 5.*

20. Вероятность того, что последовательность содержит ровно k серий обоих типов, равна

$$p_{2\nu}=2\binom{r_1-1}{\nu-1}\binom{r_2-1}{\nu-1}\binom{r_1+r_2}{r_1}^{-1} \tag{11.17}$$

при четном k $(k=2\nu)$ и

$$p_{2\nu+1}=\left\{\binom{r_1-1}{\nu}\binom{r_2-1}{\nu-1}+\binom{r_1-1}{\nu-1}\binom{r_2-1}{\nu}\right\}\binom{r_1+r_2}{r_1}^{-1} \tag{11.18}$$

при нечетном k $(k=2\nu+1)$.

21. *Продолжение.* Доказать, что наиболее вероятным числом серий является целое число k, такое, что $2r_1r_2/(r_1+r_2)<k<2r_1r_2/(r_1+r_2)+3$. (*Указание.* Рассмотреть отношения $p_{2\nu+2}/p_{2\nu}$ и $p_{2\nu+1}/p_{2\nu-1}$).

22. Вероятность того, что последовательность начинается с серии альф длиной $\nu\geqslant 0$, равна $(r_1)_\nu r_2/(r_1+r_2)_{\nu+1}$. (*Указание.* Выбрать ν альф и бету, которая должна следовать за ними.) Что вытекает из этой теоремы при $\nu=0$?

23. Вероятность появления ровно k серий из альф равна

$$p_k=\binom{r_1-1}{k-1}\binom{r_2+1}{k}\binom{r_1+r_2}{r_1}^{-1}. \tag{11.19}$$

Указание. Это легко следует из второй части леммы § 5. Кроме того, формулу (11.19) можно вывести из (11.17) и (11.18), но этот вывод более трудоемок.

24. Вероятность того, что n-й альфе предшествуют ровно m бет, равна

$$\binom{r_1+r_2-n-m}{r_2-m}\binom{m+n-1}{m}\binom{r_1+r_2}{r_1}^{-1}. \tag{11.20}$$

25. Вероятность того, что альфы образуют k серий, из которых k_1 имеют длину 1, k_2 длину 2, ..., k_ν длину ν $(k_1+\ldots+k_\nu=k)$, равна

$$\frac{k!}{k_1!\,k_2!\ldots k_\nu!}\binom{r_2+1}{k}\binom{r_1+r_2}{r_1}^{-1}. \tag{11,21}$$

§ 12. ЗАДАЧИ И ТОЖДЕСТВА, СОДЕРЖАЩИЕ БИНОМИАЛЬНЫЕ КОЭФФИЦИЕНТЫ

1. Для целого $n\geqslant 2$

$$\begin{gathered}1-\binom{n}{1}+\binom{n}{2}-+\ldots=0,\\ \binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\ldots=n2^{n-1},\\ \binom{n}{1}-2\binom{n}{2}+3\binom{n}{3}-+\ldots=0,\\ 2\cdot1\binom{n}{2}+3\cdot2\binom{n}{3}+4\cdot3\binom{n}{4}+\ldots=n(n-1)\,2^{n-2}.\end{gathered} \tag{12.1}$$

Указание. Использовать формулу бинома.

2. Доказать, что для целых положительных n и k

$$\binom{n}{0}\binom{n}{k}-\binom{n}{1}\binom{n-1}{k-1}+\binom{n}{2}\binom{n-2}{k-2}\ldots\pm\binom{n}{k}\binom{n-k}{0}=0. \tag{12,2}$$

Более общее соотношение имеет вид [1])

$$\sum\binom{n}{\nu}\binom{n-\nu}{k-\nu}t^\nu=\binom{n}{k}(1+t)^k. \tag{12,3}$$

3. Для любого $a>0$

$$\binom{-a}{k}=(-1)^k\binom{a+k-1}{k}. \tag{12.4}$$

Если a является целым, то это утверждение может быть доказано также повторным дифференцированием геометрической прогрессии $\sum x^k=(1-x)^{-1}$.

4. Доказать, что

$$\begin{gathered}\binom{2n}{n}2^{-2n}=(-1)^n\binom{-1/2}{n},\\ \frac{1}{n}\binom{2n-2}{n-1}2^{-2n+1}=(-1)^{n-1}\binom{1/2}{n}.\end{gathered} \tag{12.5}$$

5. Для целых неотрицательных n и r и для любых действительных a

$$\sum_{\nu=0}^{n}\binom{a-\nu}{r}=\binom{a+1}{r+1}-\binom{a-n}{r+1}. \tag{12.6}$$

Указание. Воспользуйтесь (8.6). Обычно используется частный случай $n=a$.

6. Для произвольного a и целого $n\geqslant 0$

$$\sum_{\nu=0}^{n}(-1)^\nu\binom{a}{\nu}=(-1)^n\binom{a-1}{n}. \tag{12,7}$$

Указание. Воспользуйтесь (8.6).

---

[1]) Читателю следует вспомнить соглашение (8.5): *если* ν пробегает *все* целые числа, то в сумме (12,3) только конечное число слагаемых отлично от нуля.

7. Для целых положительных r и k

$$\sum_{\nu=0}^{r}\binom{\nu+k-1}{k-1}=\binom{r+k}{k}. \tag{12.8}$$

а) Доказать это равенство, используя (8.6). б) Показать, что (12.8) — частный случай (12.7). в) Показать по индукции, что (12.8) приводит к новому доказательству первой части леммы § 5. г) Показать, что (12.8) эквивалентно равенству

$$\sum_{j=0}^{n}\binom{j}{m}=\binom{n+1}{m+1}. \tag{12.8a}$$

8. В § 6 мы отмечали, что члены гипергеометрического распределения должны давать в сумме единицу. Это равнозначно тому, что для любых целых положительных a, b, n

$$\binom{a}{0}\binom{b}{n}+\binom{a}{1}\binom{b}{n-1}+\ldots+\binom{a}{n}\binom{b}{0}=\binom{a+b}{n}. \tag{12.9}$$

Доказать это по индукции. *Указание*. Сначала доказать, что (12.9) выполняется для $a=1$ и любого b.

9. *Продолжение*. Сравнивая коэффициенты при t^n в обеих частях равенства

$$(1+t)^a(1+t)^b=(1+t)^{a+b}, \tag{12.10}$$

доказать, что (12.9) справедливо для любых чисел a, b (и целых n).

10. Используя (12.9), доказать, что

$$\binom{n}{0}^2+\binom{n}{1}^2+\binom{n}{2}^2+\ldots+\binom{n}{n}^2=\binom{2n}{n}. \tag{12.11}$$

11. Используя (12.11), доказать, что

$$\sum_{\nu=0}^{n}\frac{(2n)!}{(\nu!)^2(n-\nu)!^2}=\binom{2n}{n}^2. \tag{12.12}$$

12. Доказать, что для целых $0<a<b$

$$\sum_{k=1}^{a}(-1)^{a-k}\binom{a}{k}\binom{b+k}{b+1}=\binom{b}{a-1}. \tag{12.13}$$

Указание. Используя (12.4), показать, что (12.11) является частным случаем (12.9). Другое доказательство состоит в сравнении коэффициентов при t^{a-1} в тождестве $(1-t)^a(1-t)^{-b-2}=(1-t)^{a-b-2}$.

13. Вывести из (12.9) тождества

$$\binom{a}{k}-\binom{a}{k-1}+-\ldots\mp\binom{a}{1}\pm 1=\binom{a-1}{k} \tag{12.14}$$

и

$$\sum_{\nu}(-1)^{\nu}\binom{a}{\nu}\binom{n-\nu}{r}=\binom{n-a}{n-r}, \tag{12.15}$$

имеющие место, если k, n и r — целые положительные числа. *Указание*. Использовать (12.4).

14. Используя (12.9), доказать, что[1] для произвольных a, b и целого k

$$\sum_{j=0}^{k} \binom{a+k-j-1}{k-j} \binom{b+j-1}{j} = \binom{a+b+k-1}{k}. \qquad (12.16)$$

Указание. Дважды применить (12.4). Другой способ — использовать (12.10), изменив знаки показателей.

Отметим важные частные случаи $b=1, 2$.

15. Обращаясь к задачам § 11, отметим, что (11.12), (11.14), (11.15) и (11.16) определяют вероятности, поэтому в каждом случае их сумма должна равняться единице. Показать, что это вытекает соответственно из (12.8), (12.9), (12.16) и формулы бинома.

16. Из определения $A(r, n)$ в задаче 7 § 11 следует, что $A(r, n)=0$, если $r<n$, и $A(n, n)=n!$. Иначе говоря,

$$\sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} k^r = \begin{cases} 0, & \text{если } r<n, \\ n!, & \text{если } r=n. \end{cases} \qquad (12.17)$$

а) Доказать (12.17) непосредственной редукцией от n к $n-1$. б) Доказать (12.17), рассматривая r-ю производную $(1-e^t)^n$ при $t=0$. в) Обобщить (12.17), исходя из (11.11) вместо (11.6).

17. Доказать по индукции, что если $0 \leqslant N \leqslant n$, то для любого целого $r \geqslant 0$

$$\sum_{\nu=0}^{N} (-1)^{\nu} \binom{N}{\nu} (n-\nu)_r = \binom{n-N}{r-N} r!. \qquad (12.18)$$

(Заметим, что правая часть обращается в нуль, если $r<N$ или $r>n$.) Проверить (12.18), рассматривая r-ю производную $t^{n-N}(t-1)^N$ при $t=1$.

18. Доказать по индукции (используя формулу бинома), что

$$\binom{n}{1}\frac{1}{1} - \binom{n}{2}\frac{1}{2} + \ldots + (-1)^{n-1}\binom{n}{n}\frac{1}{n} = 1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}. \qquad (12.19)$$

Проверить (12.19) интегрированием тождества $\sum\limits_{0}^{n-1} (1-t)^{\nu} = \{1-(1-t)^n\}\, t^{-1}$.

19. Доказать, что для любого целого положительного m

$$(x+y+z)^m = \sum \frac{m!}{a!\,b!\,c!}\, x^a y^b z^c, \qquad (12.20)$$

где суммирование проводится по всем целым неотрицательным a, b, c, таким, что $a+b+c=m$.

20. Показать, что $\Gamma(a+1)=a\Gamma(a)$[2] для любого $a>0$, откуда

$$\binom{-a}{k} = (-1)^k \frac{\Gamma(a+k)}{k!\,\Gamma(a)}. \qquad (12.21)$$

21. Доказать, что для любых положительных целых a и b

$$\frac{(a+1)(a+2)\ \ldots\ (a+n)}{(b+1)(b+2)\ \ldots\ (b+n)} \sim \frac{b!}{a!}\, n^{a-b}. \qquad (12.22)$$

---

[1] Более изящное доказательство см. в задаче 15 гл. IX, 9.

[2] Функция $\Gamma(a)$ (гамма-функция) определена ниже; см. формулу (12.23). — *Прим. перев.*

22. *Гамма-функция* определяется равенством

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt, \tag{12.23}$$

где $x > 0$. Показать, что $\Gamma(x) \sim \sqrt{2\pi}\, e^{-x} x^{x-1/2}$. (Заметим, что если $x = n$ — целое число, то $\Gamma(n) = (n-1)!$.)

23. Пусть a и r — произвольные положительные числа, а n — целое положительное число. Показать, что

$$a(a+r)(a+2r)\ldots(a+nr) \sim Cr^{n+1} n^{n+1/2+a/r}. \tag{12.24}$$

Постоянная C равна $\sqrt{2\pi}/\Gamma(a/r)$.

24. Используя результаты предыдущей задачи, показать, что

$$\frac{a(a+r)(a+2r)\ldots(a+nr)}{b(b+r)(b+2r)\ldots(b+nr)} \sim \frac{\Gamma(b/r)}{\Gamma(a/r)} n^{(a-b)/r}. \tag{12.25}$$

25. Вывести из (8.10) неравенство

$$e^{-t/(1-t)} < 1 - t < e^{-t}, \quad 0 < t < 1. \tag{12.26}$$

ФЛУКТУАЦИИ ПРИ БРОСАНИИ МОНЕТЫ И СЛУЧАЙНЫЕ БЛУЖДАНИЯ

Эта глава стоит в стороне от нашей основной темы, которая будет теперь продолжена лишь в гл. V. Ее материал традиционно служил в качестве первого шага к более общим теориям. Простые методы вскоре приведут нас к результатам большой теоретической и практической важности. Мы столкнемся с теоретическими заключениями, которые окажутся не просто неожиданными, но будут прямо противоречить интуиции и здравому смыслу. Они покажут, что широко распространенные представления о случайных флуктуациях лишены основания и что смысл закона больших чисел часто неправильно истолковывается. Например, в различных приложениях предполагается, что наблюдения за результатами бросаний одной и той же монеты в течение длительного промежутка времени будут давать те же статистические характеристики, что и наблюдение результатов очень большого числа независимых бросаний в данный момент. Это не так. Действительно, мы придем к выводу, что (на распространенном ныне жаргоне) в популяции обычных монет бо́льшую их часть следует считать неправильными. (Эмпирические иллюстрации см. в § 6 и примере 4, б).)

До недавнего времени материал этой главы получали с использованием аналитических методов, так что результаты казались довольно глубокими. Поэтому элементарный метод [1]), используемый ниже, является хорошим примером вновь открытой силы комбинаторных методов. Результаты, справедливые для более широкого класса случайных флуктуаций [2]), будут обсуждаться в томе 2. Все результаты будут получены заново, другими методами. Таким образом, эта глава предназначена для начинающих читателей, которые не спешат приниматься за систематическую теорию, или читателей, интересующихся духом теории вероятностей без желания специа-

*) Эта глава может быть опущена или прочитана параллельно с последующими главами. Ссылки на нее будут сделаны в гл. X (законы больших чисел), XI (времена первого достижения), XIII (рекуррентные события) и XIV (случайные блуждания), но материал главы в дальнейшем непосредственно использоваться не будет.

[1]) Открытие возможности элементарного подхода было основной причиной второго издания этой книги (1957 г.). Настоящий вариант является новым и значительно усовершенствованным, так как в нем удалось избежать различных сложных комбинаторных приемов.

[2]) Смотри примечание к следствию из теоремы 2 § 4.

лизироваться по ней. Для других читателей сравнение методов окажется поучительным и интересным. Таким образом, *настоящая глава может быть прочитана по усмотрению читателя независимо от остальной части книги или параллельно с ней.*

§ 1. ОСНОВНЫЕ ПОНЯТИЯ. ПРИНЦИП ОТРАЖЕНИЯ

С формальной точки зрения мы будем иметь дело с упорядоченными конечными множествами плюс единиц и минус единиц. Рассмотрим $n=p+q$ символов $\varepsilon_1, \ldots, \varepsilon_n$, каждый из которых означает $+1$ или -1; предположим, что имеется p плюс единиц и q минус единиц. Частная сумма $s_k=\varepsilon_1+\ldots+\varepsilon_k$ представляет собой разность между числом плюсов и минусов, находящихся на первых k местах. Тогда

$$s_k-s_{k-1}=\varepsilon_k=\pm 1, \quad s_0=0, \quad s_n=p-q, \tag{1.1}$$

где $k=1, 2, \ldots, n$.

Мы будем применять геометрическую терминологию и использовать ортогональные координаты t, x; для определенности будем предполагать, что ось t горизонтальна, а ось x вертикальна. Последовательность $(\varepsilon_1, \ldots, \varepsilon_n)$ будет изображаться ломаной, в которой k-е звено имеет наклон ε_k, а k-я вершина имеет ординату s_k. Такие линии будут называться путями.

Определение. *Пусть $n>0$ и x — целые числа. Путь $(s_1, s_2, \ldots, s_n)$ из начала координат в точку (n, x) представляет собой ломаную, вершины которой имеют абсциссы $0, 1, \ldots, n$ и ординаты $s_0, s_1, \ldots, s_n$, удовлетворяющие (1.1) с $s_n=x$.*

Мы будем называть n *длиной* пути. Существует 2^n путей длины n. Если среди ε_k имеется p положительных и q отрицательных, то

$$n=p+q, \quad x=p-q. \tag{1.2}$$

Путь из начала координат в произвольную точку (n, x) существует только тогда, когда n и x удовлетворяют (1.2). В этом случае p мест для положительных ε_k могут быть выбраны из $n=p+q$ имеющихся мест

$$N_{n,x}=\binom{p+q}{p}=\binom{p+q}{q} \tag{1.3}$$

различными способами. Для удобства *положим* $N_{n,x}=0$, *если n и x не удовлетворяют* (1.2). При таком соглашении *существуют ровно $N_{n,x}$ различных путей из начала координат в произвольную точку (n, x).*

Прежде чем перейти к главной теме этой главы, а именно к теории случайных блужданий, мы проиллюстрируем возможные приложения нашей схемы.

Примеры. а) *Теорема о баллотировке* (ballot theorem). Следующее интересное утверждение доказано в 1878 г. У. Уитвортом и заново в 1887 г. Ж. Бертраном.

Предположим, что на выборах кандидат P набрал p голосов, а кандидат Q набрал q голосов, причем $p>q$. *Вероятность того, что при последовательном подсчете голосов P все время был впереди Q, равна* $(p-q)/(p+q)$.

Подобные задачи под названием задач о баллотировке привлекли внимание специалистов по комбинаторному анализу. Новый расцвет комбинаторных методов увеличил их популярность, и сейчас это выражается в том, что очень многие важные задачи могут быть переформулированы как варианты некоторой обобщенной задачи о баллотировк: [1]).

Весь протокол голосования может быть изображен в виде пути длины $p+q$, в котором $\varepsilon_k=+1$, если k-й голос подан за P; обратно, каждый путь из начала координат в точку $(p+q,\ p-q)$ можно интерпретировать как протокол голосования с данными итогами p и q. Ясно, что s_k есть число голосов, на которое P опережал Q или отставал от него после учета k-го голоса. Кандидат P лидировал на всем протяжении выборов только в том случае, когда $s_1>0,\ \ldots,\ s_n>0$, т. е. когда все вершины лежат строго выше оси t. (Путь такого типа из O в N_1 изображен на рис. 1.) В теореме о баллотиров-

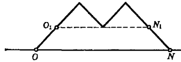

Рис. 1. Положительные пути. Рисунок показывает также, что существует ровно столько же положительных путей из начала координат в точку $(2n,\ 0)$, сколько неотрицательных путей из начала координат в точку $(2n-2,\ 0)$.

ке неявно предполагается, что все возможные пути равновероятны. Тогда ее утверждение сводится к теореме, доказанной в конце этого параграфа как непосредственное следствие леммы об отражении.

б) *Ранговый критерий Гальтона* [2]). Предположим, что некоторая величина (например, высота растения) измеряется у каждого из r подвергнутых воздействию объектов и также у каждого из r контрольных объектов. Обозначим результаты измерений $a_1,\ \ldots,\ a_r$ и $b_1,\ \ldots,\ b_r$ соответственно. Для ясности предположим, что каждая группа упорядочена по убыванию: $a_1>a_2>\ldots$ и $b_1>b_2>\ldots$. (Во

[1]) Историю вопроса и обзор литературы можно найти в статье Barton D. E., Mallows C. L., Some aspects of the random sequence, Ann. Math. Statist., 36 (1965), 236—260. Авторы этой статьи обсуждают также различные приложения. Самое последнее обобщение со многими приложениями к теории очередей принадлежит Л. Такачу.

[2]) Hodges J. L., Biometrica, 42 (1955), 261—262.

избежание тривиальности мы предполагаем, что никакие два результата наблюдения не равны.) Объединим теперь эти две последовательности в одну с числом элементов $n=2r$ и упорядоченную по убыванию. Предшествование всех значений a всем значениям b указывает на весьма высокую эффективность воздействия, тогда как случайная комбинация значений a и b означает полную его бесполезность. Таким образом, об эффективности воздействия можно судить на основании числа различных значений a, предшествующих соответствующим b, т. е. на основании числа индексов k, для которых $a_k > b_k$. Эта идея впервые была использована Ф. Гальтоном в 1876 г. для исследования данных, предоставленных ему Чарльзом Дарвином. В этом случае r было равно 15 и в 13 случаях значения a предшествовали соответствующим значениям b. Не зная реальных вероятностей, Гальтон заключил, что воздействие *было* эффективным. Однако в предположении полной случайности вероятность того, что значения a предшествуют соответствующим b в 13 или более случаях, равна 3/16. Это означает, что в трех из шестнадцати случаев совершенно бесполезное воздействие будет казаться не менее эффективным, чем воздействие, эффективное по Гальтону. Это показывает, что численный анализ может быть полезным дополнением к нашей не совсем надежной интуиции.

Для интерпретации при помощи путей положим $\epsilon_k=+1$ или -1 в зависимости от того, элементом какого типа является k-й член объединенной последовательности: a или b. Полный путь длины $2r$ соединяет начало координат с точкой $(2r, 0)$ на оси t. Событие $a_k > b_k$ осуществляется тогда и только тогда, когда s_{2k-1} содержит по меньшей мере k плюс единиц, т. е. когда $s_{2k-1} > 0$. Это влечет $s_{2k} \geqslant 0$, и поэтому $(2k-1)$-е и $2k$-е звенья лежат выше оси t. Отсюда следует, что неравенство $a_k > b_k$ выполняется ν раз тогда и только тогда, когда 2ν звеньев лежат выше оси t. В § 9 мы докажем неожиданный результат, состоящий в том, что вероятность этого равна $1/(r+1)$ и не зависит от ν. (Соответствующие критерии, основанные на теории серий, см. в разд. 6 гл. II, 5.)

в) *Критерии типа Колмогорова — Смирнова.* Предположим, что мы наблюдаем две популяции одинакового биологического вида (животных или растений), находящиеся в разных местах, или что мы хотим сравнить продукцию двух аналогичных станков. Для определенности рассмотрим только одну поддающуюся измерению характеристику (например, высоту, вес или толщину) и предположим, что для каждой из двух совокупностей мы имеем r наблюдений, скажем $a_1, \ldots, a_r$ и $b_1, \ldots, b_r$. Грубо говоря, вопрос заключается в том, совместимы ли эти данные с гипотезой о том, что обе совокупности статистически идентичны. В этом виде задача несколько неопределенна, но для наших целей нет необходимости обсуждать более точную ее постановку в современной статистической теории. Достаточно сказать, что критерии основаны на сравнении двух эмпирических распределений. Для каждого t обозначим через

$A(t)$ отношение k/r, где k — число индексов i, для которых $a_i \leqslant t$. Так определенная на действительной оси функция называется *эмпирическим распределением* значений a. Эмпирическое распределение B определяется аналогично.

Изящная математическая теория, созданная Н. В. Смирновым [1]), позволяет найти распределение вероятностей максимума отклонения $|A(t)-B(t)|$ и других величин, которые могут быть использованы для проверки указанной гипотезы. Теория довольно сложна, но она была значительно упрощена и сделана более понятной Б. В. Гнеденко, которому пришла удачная мысль связать ее с геометрической теорией путей. Как и в предыдущем примере, свяжем с двумя выборками путь длины $2r$, ведущий из начала координат в точку $(2r, 0)$. Сказать, что две совокупности статистически неразличимы, все равно, что сказать, что все возможные пути равновероятны. Тогда легко видеть, что $|A(t)-B(t)|>\xi$ для некоторого t тогда и только тогда, когда $|s_k|>\xi r$ для некоторого k. Вероятность этого события есть просто вероятность того, что путь длины $2r$, ведущий из начала координат в точку $(0, 2r)$, не ограничен полосой между $\pm\xi r$. Эта вероятность давно известна, потому что она связана с задачей о разорении игрока и физическими задачами диффузии с поглощающими барьерами (см. задачу 3).

Этот пример выходит за рамки настоящей книги, но он показывает, как теорию случайных блужданий можно примеиять к задачам совершенно различной природы.

г) *Игра с бросанием правильной монеты и ее связь со случайными процессами.* Путь длины n можно интерпретировать как результат мыслимого эксперимента, состоящего в n последовательных бросаниях монеты. Если $+1$ приписывается гербам, то s_k равна положительной или отрицательной разности между суммарным числом гербов и решеток после k-го бросания. При классическом описании вводится фиктивный игрок Петр, который в каждом испытании выигрывает или проигрывает определенную денежную единицу. Тогда последовательность $s_1, s_2, \ldots, s_n$ представляет собой последовательные значения его прибыли. Вскоре будет показано, что они являются результатами случайной флуктуации с совершенно неожиданными свойствами.

Образный язык азартных игр не уменьшает важности модели с бросанием монеты. В самом деле, эта модель может служить как первое приближение ко многим более сложным случайным процессам в физике, экономике и теории обучения. Такие величины, как энергия физической частицы, состояние индивидуума и накопленный опыт крысы предполагаются изменяющимися вследствие по-

---

[1]) Смирнов Н. В. Об уклонениях эмпирической кривой распределения.— Матем. сборник, 1939, т. 6 (48), № 1, с. 3—24; см. также Смирнов Н. В. Теория вероятностей и математическая статистика. Избранные труды.— М.: Наука, 1970.— *Прим. перев.*

следовательных случайных воздействий некоторого рода. В целях упрощения первоначального изучения предполагают, что отдельные приращения имеют одинаковую величину и что их знак определяется игрой с бросанием монеты. Более совершенная модель принимает в расчет, что приращения и их вероятности изменяются от испытания к испытанию, но даже простая модель с бросанием монеты приводит к удивительным, даже потрясающим результатам. Эти результаты имеют большое практическое значение, так как показывают, что вопреки распространенным взглядам законы, характеризующие длинные серии отдельных наблюдений, обнаруживают большие отклонения от свойств популяции. Иначе говоря, популярные в наше время психологические тесты привели бы к тому, что в совокупности из «правильных» монет большинство было названо «неправильными».

Отсюда следует, что случайные флуктуации при бросании монеты типичны для более общих случайных процессов накопления. Во всяком случае это приводит к констатации того, что, коль скоро даже простая игра с бросанием монеты приводит к парадоксальным результатам, противоречащим нашей интуиции, последняя не может служить надежным советчиком в более сложных ситуациях. ▶

Столь же удивительно, сколь и приятно, что наиболее важные заключения могут быть получены из приведенной ниже простой леммы.

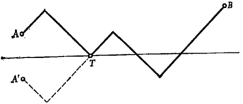

Рис. 2. Принцип отражения.

Пусть $A=(a, \alpha)$ и $B=(b, \beta)$ — точки с целочисленными координатами, лежащие в положительном квадранте: $b>a\geqslant 0$, $\alpha>0$, $\beta>0$. Под отражением A относительно оси t понимается точка $A'=(a, -\alpha)$ (рис. 2). Путь из A в B определяется обычным образом.

Лемма. (Принцип отражения [1]).) *Число путей из A в B, которые касаются оси t или пересекают ее, равно числу всех путей из A' в B.*

[1]) Принцип отражения часто используется в различных формах, однако без геометрической интерпретации он кажется остроумным, но непонятным трюком. В вероятностной литературе этот принцип приписывается Д. Андрэ (1887 г.). Он появляется в связи с разностными уравнениями для случайных блужданий

Доказательство. Рассмотрим путь $(s_a=\alpha,\ s_{a+1},\ \ldots,\ s_b=\beta)$ из A в B, имеющий одну или несколько вершин на оси t. Пусть t — абсцисса первой такой вершины (рис. 2), т. е. выберем t так, что $s_a>0,\ \ldots,\ s_{t-1}>0,\ s_t=0$. Тогда

$$(-s_a,\ -s_{a+1},\ \ldots,\ -s_{t-1},\ s_t=0,\ s_{t+1},\ s_{t+2},\ \ldots,\ s_b)$$

представляет собой путь, ведущий из A' в B и имеющий в качестве первой вершины на оси t точку $T=(t, 0)$. Звенья AT и $A'T$ являются отражениями одно другого, поэтому существует взаимно однозначное соответствие между всеми путями из A' в B и такими путями из A в B, которые имеют вершины на оси t. Это доказывает лемму. ►

В качестве непосредственного следствия мы докажем результат, обсуждавшийся в примере а). Это будет служить отправной точкой для всей теории настоящей главы.

Теорема о баллотировке. *Пусть n и x — целые положительные числа. Существует ровно $(x/n)N_{n,x}$ путей $(s_1,\ \ldots,\ s_n=x)$ из начала координат в точку (n, x), таких, что $s_1>0,\ \ldots,\ s_n>0$.*

Доказательство. Очевидно, существует столько же допустимых путей, сколько существует путей из точки (1,1) в точку (n, x), которые не касаются оси t и не пересекают ее. В силу последней леммы число таких путей равно

$$N_{n-1,\,x-1}-N_{n-1,\,x+1}=\binom{p+q-1}{p-1}-\binom{p+q-1}{p},$$

где p и q определены в (1.2). Простые вычисления показывают, что правая часть равна $N_{n,x}(p-q)/(p+q)$, как и утверждалось. ►

§ 2. СЛУЧАЙНЫЕ БЛУЖДАНИЯ; ОСНОВНЫЕ ПОНЯТИЯ И ОБОЗНАЧЕНИЯ

Игра с бросанием идеальной монеты теперь будет описана при помощи случайных блужданий, которые привлекают своей наглядностью и лучше приспособлены для обобщений. Как показано в предыдущем примере, если путь $(s_1,\ \ldots,\ s_\rho)$ рассматривается как запись результатов ρ последовательных бросаний монеты, то частные суммы $s_1,\ \ldots,\ s_\rho$ представляют собой последовательные значения прибыли. Для геометрического описания удобно предположить, что бросания выполняются через равные интервалы времени, так что n-е испытание осуществляется в момент времени n. Последова-

в гл. XIV, 9, соответствующими некоторым дифференциальным уравнениям в частных производных, для которых принцип отражения является хорошо известным приемом, называемым *методом подобия*. Он обычно приписывается Максвеллу и Кельвину. Об использовании повторных отражений см. задачи 2 и 3.

тельные частные суммы $s_1, \ldots, s_n$ будут изображаться точками на вертикальной оси x; они будут называться положениями «частицы», совершающей случайное блуждание. Заметим, что частица движется единичными шагами вверх или вниз на прямой. Путь представляет собой график такого движения. Например, путь из O в N на рис. 1 означает случайное блуждание, заканчивающееся через шесть шагов возвращением в начало.

Каждый путь длины ρ можно интерпретировать как результат некоторого случайного блуждания; имеется 2^ρ таких путей, и мы приписываем вероятность $2^{-\rho}$ каждому. (Другие распределения вероятностей будут введены в гл. XIV. Для того чтобы отличить рассматриваемое случайное блуждание от других, назовем его *симметричным*.)

Мы завершили определение пространства элементарных событий и вероятностей на нем, но беспокоит зависимость их от числа ρ. Для того чтобы выяснить его значение, рассмотрим событие, состоящее в том, что путь проходит через точку (2, 2). Первые два шага должны быть положительными, и поэтому существует $2^{\rho-2}$ путей с этим свойством. Как и следовало ожидать, вероятность этого события равна 1/4 независимо от значения ρ. В более общем случае для любого $k \leqslant \rho$ существует $2^{\rho-k}$ путей с фиксированным поведением на первых k шагах. Отсюда следует, что *событие, определенное первыми $k \leqslant \rho$ шагами, имеет вероятность, не зависящую от* ρ. Поэтому на практике число ρ не играет роли, если оно достаточно велико. Иначе говоря, любой путь длиной n может рассматриваться как начальная часть очень длинного пути, и нет необходимости определять длину последнего. И с интуитивной, и с формальной точки зрения наиболее удобно рассматривать неограниченные последовательности испытаний, но это привело бы к использованию несчетных пространств элементарных событий. Поэтому в дальнейшем будем считать, что длины ρ путей, составляющих пространство элементарных событий, больше числа шагов, фигурирующего в наших формулах. За исключением этого, мы будем с радостью позволять себе забывать о ρ.

Для того чтобы согласовать символику с той, которая будет использоваться в общей теории, мы обозначим отдельные перемещения через $\mathbf{X}_1, \mathbf{X}_2, \ldots$, а положения частицы через $\mathbf{S}_1, \mathbf{S}_2, \ldots$. Таким образом,

$$\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n, \quad \mathbf{S}_0 = 0. \tag{2.1}$$

Для каждого конкретного пути можно указать соответствующие значения $\mathbf{X}_1, \mathbf{X}_2, \ldots$, т. е. $\mathbf{X}_k$ являются функциями пути[1]). Например, для пути на рис. 1, очевидно, $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_4 = 1$, а $\mathbf{X}_3 = \mathbf{X}_5 = \mathbf{X}_6 = -1$.

[1]) По терминологии, которая будет введена в гл. IX, $\mathbf{X}_n$ представляют собой случайные величины.

Как правило, мы будем описывать все события заданием подходящих условий для сумм $\mathbf{S}_k$. Таким образом, событие «в момент времени n частица находится в точке r» будет обозначаться так: $\{\mathbf{S}_n = r\}$. Его вероятность будем обозначать через $p_{n,r}$. (Это событие мы будем называть попаданием[1]) в точку r в момент времени n.) Число $N_{n,r}$ путей из начала координат в точку (n, r) дается соотношением (1.3), и, следовательно,

$$p_{n,r} = \mathbf{P}\{\mathbf{S}_n = r\} = \binom{n}{(n+r)/2} 2^{-n}, \tag{2.2}$$

где подразумевается, что биномиальный коэффициент равен нулю, когда $(n+r)/2$ не является целым числом, лежащим между 0 и n.

Если $\mathbf{S}_k = 0$, то *возвращение в начало* происходит в момент времени k. Здесь k обязательно четно и для $k = 2\nu$ вероятность возвращения в начало равна $p_{2\nu,0}$. Так как эта вероятность часто встречается в дальнейшем, введем для нее специальное обозначение $u_{2\nu}$. Таким образом,

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}. \tag{2.3}$$

Когда биномиальный коэффициент выражен через факториалы, формула Стирлинга (9.1) гл. II непосредственно дает

$$u_{2\nu} \sim 1/\sqrt{\pi\nu}, \tag{2.4}$$

где знак $\sim$ указывает, что отношение левой и правой части стремится к 1 при $\nu \to \infty$; правая часть служит прекрасным приближением[2]) для $u_{2\nu}$ даже для небольших значений ν.

Среди возвращений в начало особый интерес представляет *первое возвращение*. Первое возвращение происходит в момент времени 2ν, если

$$\mathbf{S}_1 \neq 0, \ \ldots, \ \mathbf{S}_{2\nu-1} \neq 0, \ \text{но} \ \mathbf{S}_{2\nu} = 0. \tag{2.5}$$

Вероятность этого события будет обозначаться $f_{2\nu}$. По определению $f_0 = 0$.

Связь между вероятностями f_{2n} и u_{2n} представляет большой интерес. Попадание в начало в момент времени $2n$ может быть либо первым возвращением, либо первое возвращение произошло в момент времени $2k < 2n$ и далее за время $2n - 2k$ вновь произойдет возвращение в начало. Вероятность последнего события равна $f_{2k}u_{2n-2k}$, так как имеется $2^{2k}f_{2k}$ путей длины $2k$, оканчивающихся с первым возвращением, и $2^{2n-2k}u_{2n-2k}$ путей из

[1]) В оригинале visit. — *Прим. перев.*

[2]) Для истинного значения $u_{10} = 0{,}2461$ мы получаем приближение 0,2523; для $u_{20} = 0{,}1762$ — приближение 0,1784. Относительная погрешность убывает, грубо говоря, обратно пропорционально ν.

точки $(2k, 0)$ в точку $(2n, 0)$. Отсюда следует, что

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \dots + f_{2n} u_0, \quad n \geqslant 1. \tag{2.6}$$

(См. задачу 5.)

Нормальное приближение. Формула (2.2) не пригодна для прямого вычисления вероятностей попадания S_n в заданные границы. Для этих целей служит приближенная формула, которая представляет собой частный случай центральной предельной теоремы и будет доказана в гл. VII, 2[1]).

Вероятность того, что $a < S_n < b$, получается суммированием вероятностей $p_{n,r}$ по всем r, лежащим между a и b. Для ее определения достаточно знать вероятности выполнения всех неравенств вида $S_n > a$. Такие вероятности могут быть оценены благодаря тому факту, что для всех x при $n \to \infty$

$$\mathbf{P}\{S_n > x\sqrt{n}\} \to 1 - \mathfrak{N}(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-t^2/2}\, dt, \tag{2.7}$$

где $\mathfrak{N}$ — нормальная функция распределения, определенная в гл. VII, 1. На данном этапе ее свойства не представляют для нас особого интереса. Существование предела означает, что при больших n отношения $S_n/\sqrt{n}$ имеют приблизительно одни и те же вероятности, и, таким образом, одно и то же приближение может быть использовано одновременно для всех больших n.

Хорошее представление о вероятностях (2.7) дает табл. 1. Более подробной и точной является табл. 1 гл. VII.

Таблица 1

x	0,5	1,0	1,5	2,0	2,5	3,0
$\mathbf{P}(S_n > x\sqrt{n})$	0,309	0,159	0,067	0,023	0,006	0,001

§ 3. ОСНОВНАЯ ЛЕММА

Как мы видели, вероятность возвращения в начало в момент времени 2ν равна величине $u_{2\nu}$, задаваемой равенством (2.3). По мере развития теории случайных блужданий неожиданно выяснилось, что эта вероятность входит почти во все формулы. Одну из причин этого раскрывает следующая простая лемма, не представляющая большого самостоятельного интереса, но позволяющая доказать более глубокие теоремы следующего параграфа.

Лемма 1[2]). *Вероятность того, что до момента $2n$ включительно не произойдет ни одного возвращения в начало, равна ве-*

[1]) Частный случай, который нам здесь потребуется, исследуется в гл. VII, 2 *отдельно* (без ссылки на общее биномиальное распределение). Доказательство простое и может быть помещено в данном месте.

[2]) Эта лемма очевидна, если использовать вид производящей функции $\sum f_{2k} s^{2k}$ (см. формулу (3.6) гл. XI), и должна быть отмечена за ее познавательное значение. Ее важность была установлена недавно. Геометрическое доказательство см. в задаче 7.

роятности того, что возвращение в начало произойдет в момент времени $2n$, т. е.

$$\mathbf{P}\{\mathbf{S}_1 \neq 0, \ \ldots, \ \mathbf{S}_{2n} \neq 0\} = \mathbf{P}\{\mathbf{S}_{2n} = 0\} = u_{2n}. \tag{3.1}$$

Здесь, конечно, $n > 0$. При осуществлении события в левой части все $\mathbf{S}_j$ либо положительны, либо отрицательны. Оба этих события являются одинаково вероятными, поэтому мы можем переписать (3.1) в виде

$$\mathbf{P}\{\mathbf{S}_1 > 0, \ \ldots, \ \mathbf{S}_{2n} > 0\} = (1/2)\, u_{2n}. \tag{3.2}$$

Доказательство. Рассматривая все возможные значения $\mathbf{S}_{2n}$, очевидно, имеем

$$\mathbf{P}\{\mathbf{S}_1 > 0, \ \ldots, \ \mathbf{S}_{2n} > 0\} = \sum_{r=1}^{\infty} \mathbf{P}\{\mathbf{S}_1 > 0, \ \ldots, \ \mathbf{S}_{2n-1} > 0, \ \mathbf{S}_{2n} = 2r\} \tag{3.3}$$

(где все слагаемые при $r > n$ равны нулю). По теореме о баллотировке число путей, удовлетворяющих условию, указанному в правой части, равно $N_{2n-1,\,2r-1} - N_{2n-1,\,2r+1}$, и, следовательно, r-е слагаемое суммы равно

$$(1/2)\,(p_{2n-1,\,2r-1} - p_{2n-1,\,2r+1}).$$

Отрицательная часть r-го слагаемого и положительная часть $(r+1)$-го слагаемого взаимно уничтожаются, поэтому в результате сумма (3.3) сводится к $(1/2)\,p_{2n-1,\,1}$. Легко проверить, что $p_{2n-1,\,1} = u_{2n}$, чем завершается доказательство. ▶

Лемма может быть переформулирована различными способами, например

$$\mathbf{P}\{\mathbf{S}_1 \geqslant 0, \ \ldots, \ \mathbf{S}_{2n} \geqslant 0\} = u_{2n}. \tag{3.4}$$

В самом деле, путь длиной $2n$, все вершины которого расположены выше оси t, проходит через точку $(1, 1)$. Взяв эту точку в качестве нового начала координат, мы получим путь длины $2n-1$, все вершины которого лежат выше новой оси t или на этой оси. Отсюда следует, что

$$\mathbf{P}\{\mathbf{S}_1 > 0, \ \ldots, \ \mathbf{S}_{2n} > 0\} = (1/2)\,\mathbf{P}\{\mathbf{S}_1 \geqslant 0, \ \ldots, \ \mathbf{S}_{2n-1} \geqslant 0\}. \tag{3.5}$$

Но $\mathbf{S}_{2n-1}$ — нечетное число, и следовательно, из неравенства $\mathbf{S}_{2n-1} \geqslant 0$ вытекает также, что $\mathbf{S}_{2n} \geqslant 0$. Вероятность в правой части (3.5) равна вероятности в левой части (3.4), и тем самым (3.4) доказано. (См. задачу 8.)

Лемма 1 непосредственно приводит к явному выражению для распределения вероятностей первого возвращения в начало. Сказать «первое возвращение произошло в момент времени $2n$» все равно что сказать «условие $\mathbf{S}_1 \neq 0, \ \ldots, \ \mathbf{S}_{2k} \neq 0$ выполняется для $k = n-1$, но не выполняется для $k=n$». Используя (3.1), отсюда получаем

$$f_{2n}=u_{2n-2}-u_{2n}, \quad n=1,\ 2,\ \ldots\ . \tag{3.6}$$

После простых преобразований приходим к выражению

$$f_{2n}=[1/(2n-1)]\,u_{2n}. \tag{3.7}$$

Таким образом, нами доказана следующая лемма.

Лемма 2. *Вероятность того, что первое возвращение в начало произошло в момент времени* $2n$, *дается соотношением* (3.6) *или* (3.7).

Из (3.6) следует, что $f_2+f_4+\ldots=1$. Применительно к бросанию монеты это означает, что если игра продолжается достаточно долго, то в конечном счете уравнивание шансов становится практически несомненным. Этого можно было ожидать и из интуитивных соображений, удивление вызывает лишь число необходимых для этого испытаний. Например, вероятность того, что уравнивания не было в течение 100 испытаний, приблизительно равна 0,08.

§ 4. ПОСЛЕДНЕЕ ПОПАДАНИЕ И ПРОДОЛЖИТЕЛЬНЫЕ ЛИДИРОВАНИЯ

Мы готовы теперь к строгому анализу природы флуктуаций при случайных блужданиях. Результаты поразительны. Согласно широко распространенному убеждению, так называемый «закон средних» должен гарантировать, что при длительном бросании монеты каждый из игроков будет в выигрыше примерно половину времени и лидерство будет нередко переходить от одного игрока к другому. Представим себе, что имеются записи огромного числа результатов игр с бросанием идеальной монеты, причем каждая из игр содержит ровно $2n$ испытаний. Выберем случайным образом одну из них и найдем момент времени, когда в последний раз счет в игре был ничейным (иначе говоря, номер последнего испытания, в результате которого суммарные числа появившихся гербов и решеток стали равными). Это число четно, и мы обозначим его через $2k$ (так что $0\leqslant k\leqslant n$). Частая смена лидерства должна означать, что k будет сравнительно близко к n, но это не так. Действительно, приведенная ниже теорема обнаруживает удивительный факт: распределение k симметрично в том смысле, что любое значение k имеет в точности ту же вероятность, что и $n-k$. Эта симметрия означает, в частности, что неравенства $k>n/2$ и $k<n/2$ равновероятны [1]). *С вероятностью* $1/2$ *уравнение счета не будет иметь места во второй половине игры*

[1]) Симметрия распределения для k была обнаружена эмпирически на ЭВМ и доказана теоретически без знания точного распределения (4.1). См. Blackwell D., Deuel P., Freedman D., Ann. Math. Statist., 35 (1964), 1344.

независимо от ее длительности. Более того, вероятности *максимальны* вблизи крайних точек; наиболее вероятными значениями k являются 0 и n. Эти результаты показывают, что интуиция ведет к ошибочной вероятностной картине случайных флуктуаций. Поясним сказанное численными примерами.

Примеры. а) Предположим, что очень много игр с бросанием монеты проводится одновременно, каждая со скоростью одно бросание в секунду, днем и ночью, в течение года. В среднем в одной из десяти игр последнее уравнение счета произойдет в первые 9 дней, и лидерство не изменится в течение последующих 356 дней. В одном из двадцати случаев последнее уравнивание будет иметь место в первые $2^{1}/_{4}$ дня, и в одном из ста случаев оно осуществится в течение первых 2 часов 10 минут.

б) Предположим, что во время учебного эксперимента, продолжающегося в течение года, один ребенок постоянно был отстающим, за исключением, быть может, первой недели. Другой ребенок постоянно был впереди, за исключением, возможно, последней недели. Можно ли этих двух детей считать одинаково способными? Далее, пусть группа из 11 детей подвергается такому же учебному эксперименту, основанному не на способностях, а исключительно на случае. Один из одиннадцати будет лидером все время, кроме одной недели, другой — отстающим все время, кроме одной недели.

Точные вероятности для возможных значений k дает следующая теорема.

Теорема 1. (Закон арксинуса для последних попаданий.) *Вероятность того, что до момента времени $2n$ включительно последнее попадание в начальное состояние произойдет в момент $2k$, равна*

$$\alpha_{2k,\,2n} = u_{2k}u_{2n-2k}, \quad k=0,\ 1,\ \ldots,\ n. \tag{4.1}$$

Доказательство. Мы интересуемся путями, удовлетворяющими условиями $S_{2k}=0$ и $S_{2k+1}\neq 0,\ \ldots,\ S_{2n}\neq 0$. Первые $2k$ вершин можно выбрать $2^{2k}u_{2k}$ различными способами. Взяв точку $(2k,\ 0)$ в качестве нового начала координат и используя (3.1), мы видим, что остальные $2n-2k$ вершин можно выбрать $2^{2n-2k}u_{2n-2k}$ способами. Разделив на 2^{2n}, получим (4.1). ▶

Из теоремы следует, что числа (4.1) в сумме дают единицу. Распределение вероятностей, которое приписывает точке $2k$ вес $\alpha_{2k,\,2n}$, будет называться *дискретным распределением арксинуса порядка n* по той причине, что функция, обратная синусу, дает отличное численное приближение к нему. Это распределение симметрично в том смысле, что $\alpha_{2k,\,2n}=\alpha_{2n-2k,\,2n}$. При $n=2$ имеется три значения 3/8, 2/8, 3/8; для случая $n=10$ см. табл. 2. Наименьшее значение всегда имеет центральный член.

Таблица 2

Дискретное распределение арксинуса порядка 10

	$k=0$ $k=10$	$k=1$ $k=9$	$k=2$ $k=8$	$k=3$ $k=7$	$k=4$ $k=6$	$k=5$
$\alpha_{2k,\,20}$	0,1762	0,0927	0,0736	0,0655	0,0617	0,0606

Особенности распределений арксинуса хорошо видны из графика функции

$$f(x) = 1/(\pi\sqrt{x(1-x)}), \quad 0 < x < 1. \tag{4.2}$$

При помощи формулы Стирлинга можно показать, что u_{2n} близко к $1/\sqrt{\pi n}$, если n не слишком мало. Это дает следующее приближение:

$$\alpha_{2k,\,2n} \approx (1/n)\, f(x_k), \quad \textit{где}\ x_k = k/n, \tag{4.3}$$

погрешность которого незначительна, за исключением тех случаев когда k очень близко к 0 или к n. Величина, стоящая в правой части

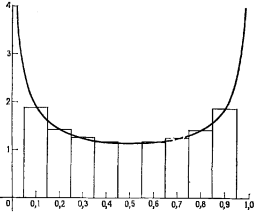

Рис. 3. График функции $f(x) = 1/[\pi\sqrt{x(1-x)}]$. Рисунок поясняет приближение (4.3).

(4.3), равна площади прямоугольника высотой $f(x_k)$, основанием которого является интервал длины $1/n$ с центром в x_k (см. рис. 3). Поэтому для $0<p<q<1$ и больших n сумма вероятностей $\alpha_{2k,\,2n}$ при $pn<k<qn$ приблизительно равна площади области, лежащей

ниже графика f в пределах интервала $p<x<q$. Это остается справедливым и для $p=0$, $q=1$, потому что полная площадь области, лежащей ниже графика, равна единице, что выполнено также и для суммы всех $\alpha_{2k, 2n}$. К счастью, интеграл от правой части (4.2) можно найти в явном виде, и мы получаем, что *для фиксированного* $0<x<1$ *и достаточно большого* n выполняется приближенное равенство

$$\sum_{k<xn} \alpha_{2k,\,2n} \approx (2/\pi) \arcsin \sqrt{x}. \tag{4.4}$$

Заметим, что правая часть не зависит от n; это означает, что табл. 3 подходит для всех распределений арксинуса высокого порядка. (На самом деле приближения оказываются довольно хорошими даже для относительно малых значений n.)

Таблица 3

Непрерывное распределение арксинуса

$A(x) = (2/\pi) \arcsin \sqrt{x}$

x	$A(x)$	x	$A(x)$	x	$A(x)$
0,00	0,000	0,20	0,295	0,40	0,436
0,01	0,064	0,21	0,303	0,41	0,442
0,02	0,090	0,22	0,311	0,42	0,449
0,03	0,111	0,23	0,318	0,43	0,455
0,04	0,128	0,24	0,326	0,44	0,462
0,05	0,144	0,25	0,333	0,45	0,468
0,06	0,158	0,26	0,341	0,46	0,474
0,07	0,171	0,27	0,348	0,47	0,481
0,08	0,183	0,28	0,355	0,48	0,487
0,09	0,194	0,29	0,362	0,49	0,494
				0,50	0,500
0,10	0,205	0,30	0,369		
0,11	0,215	0,31	0,376		
0,12	0,225	0,32	0,383		
0,13	0,235	0,33	0,390		
0,14	0,244	0,34	0,396		
0,15	0,253	0,35	0,403		
0,16	0,262	0,36	0,410		
0,17	0,271	0,37	0,416		
0,18	0,279	0,38	0,423		
0,19	0,287	0,39	0,429		

Для $x > 1/2$ использовать равенство $A(1-x) = 1 - A(x)$.

Мы видели, что вопреки общераспространенным суждениям вполне вероятно, что в длительной игре с бросанием монеты один из игроков практически остается все время в выигрыше, а другой — в проигрыше. Следующая теорема объясняет то же самое явление путем анализа доли времени, проводимой частицей на положительной стороне. Интуитивно может показаться, что наиболее вероятен случай, когда эта доля будет близка к 1/2, но верно обратное: значение, ближайшее к 1/2, является наименее вероятным, тогда как крайние точки $k=0$ и $k=n$ имеют наибольшую вероятность. Анализ облегчается тем благоприятным обстоятельством, что в теореме вновь фигурирует дискретное распределение арксинуса (4.1), которое дважды встретится нам в § 8.

Теорема 2. (Закон дискретного арксинуса для времени пребывания.) *Вероятность того, что в интервале времени от* 0 *до* $2n$ *частица* $2k$ *единиц времени проведет на положительной стороне и* $2n-2k$ *единиц времени на отрицательной, равна* $\alpha_{2k,2n}$.

(Полное время, проведенное на положительной стороне, обязательно четно.)

Следствие[1]). *Если* $0<x<1$, *то вероятность того, что не более* xn *единиц времени будет проведено на положительной стороне и не менее* $(1-x)n$ *на отрицательной, стремится к* $(2/\pi)\arcsin\sqrt{x}$ *при* $n\to\infty$.

Примеры. в) Вероятность того, что за двадцать бросаний монеты лидерство ни разу не перейдет от одного игрока к другому, приблизительно равна 0,352. Вероятность того, что более удачливый игрок будет лидировать в течение 16 или более единиц времени, примерно равна 0,685. (При $x=4/5$ следствие дает приближенное значение 0,590.) Вероятность того, что каждый игрок лидировал на протяжении 10 единиц времени, равна всего лишь 0,06 [2]).

г) Пусть n велико. С вероятностью 0,20 частица проведет около 97,6% времени на той стороне, куда она попала в самом начале. В одном из десяти случаев частица проведет на этой стороне 99,4% времени.

---

[1]) Поль Леви (Paul Lévy, Sur certains processus stochastiques homogènes, Compositia Mathematica, 7 (1939), 283—339) установил закон арксинуса для броуновского движения и указал на связь с игрой с бросанием монеты. Общий предельный закон арксинуса для числа положительных частных сумм в последовательности взаимно независимых случайных величин был доказан П. Эрдёшем и М. Кацем (Erdős P., Kac M., On the number of positive sums of independent random variables, Bull. Amer. Math. Soc., 53 (1947), 1011—1020). Широкая область применения предельного закона арксинуса казалась в то время непостижимой. Целая теория была создана заново, когда Э. Спарре Андерсен сделал удивительное открытие, заключающееся в том, что многие аспекты теории флуктуаций сумм независимых случайных величин имеют исключительно комбинаторную природу (см. Mathematica Scandinavica, 1(1953), 263—285; 2 (1954), 195—223). Первоначальные доказательства были очень сложными, но они дали новые методы исследования и сейчас сильно упрощены. Теорему 2 впервые доказали Чжун Кайлай и Феллер довольно сложными методами (см. гл. XII, 5—6 первого издания этой книги); теорема 1 — новая.

[2]) Указанные в этом примере вероятности заимствованы из таблицы, имевшейся в предыдущих изданиях книги и исключенной автором из настоящего издания. На с. 101 воспроизводится эта таблица, в которую включены данные о распределении лидерства при 20 бросаниях монеты,

д) В примере а) монета бросается со скоростью одно бросание в секунду в течение 365 дней. Следующая таблица показывает время t_p, такое, что с фиксированной вероятностью p менее удачливый игрок будет лидировать в течение всего времени, не превосходящего t_p.

p	t_p	p	t_p
0,9	153,95 дня	0,3	19,89 дня
0,8	126,10 »	0,2	8,93 »
0,7	99,65 »	0,1	2,24 »
0,6	75,23 »	0,05	13,5 часа
0,5	53,45 »	0,02	2,16 »
0,4	34,85 »	0,01	32,4 минуты

►

Доказательство теоремы 2. Рассмотрим пути фиксированной длины $2n$ и обозначим через $b_{2k,\,2n}$ вероятность того, что ровно $2k$ звеньев лежат выше оси t. Мы должны доказать, что

$$b_{2k,\,2\nu}=\alpha_{2k,\,2\nu}. \tag{4.5}$$

Из (3.4) следует, что $b_{2\nu,\,2\nu}=u_{2\nu}$, и в силу симметрии имеем также $b_{0,2\nu}=u_{2\nu}$. Поэтому достаточно доказать (4.5) для $1\leqslant k\leqslant \nu-1$.

Предположим, что ровно $2k$ из $2n$ единиц времени проведено на положительной стороне и $1\leqslant k\leqslant \nu-1$. В этом случае первое возвращение в начало должно осуществиться в некоторый момент времени $2r<2n$, и имеются две возможности. Во-первых, $2r$ единиц времени до первого возвращения могут быть проведены на положительной стороне. В этом случае $r\leqslant k\leqslant n-1$ и отрезок пути после вершины $(2r,\,0)$ имеет ровно $2k-2r$ звеньев над осью. Очевидно, число таких путей равно

$$(1/2)\cdot 2^{2r}f_{2r}\cdot 2^{2n-2r}\,b_{2k-2r,\,2n-2r}.$$

Во-вторых, $2r$ единиц времени до первого возвращения могут быть проведены на отрицательной стороне. В этом случае отрезок пути после вершины $(2r,\,0)$ имеет ровно $2k$ звеньев над осью, откуда

	$k=0$ $k=20$	$k=2$ $k=18$	$k=4$ $k=16$	$k=6$ $k=14$	$k=8$ $k=12$	$k=10$
$p_{k,20}$	0,1762	0,0927	0,0736	0,0655	0,0617	0,0606
$P_{k,20}$	0,3524	0,5379	0,6851	0,8160	0,9394	1

Здесь $p_{k,20}=u_ku_{20}$ — вероятность того, что k отрезков пути лежат на положительной стороне (т. е. «Петр лидировал после k из 20 испытаний»), а $P_{k,20}$ — вероятность того, что один из игроков лидировал не менее чем после k испытаний, другой же — не более чем после $20-k$ испытаний. — *Прим. перев.*

$n-r \geqslant k$. Число таких путей равно

$$(1/2)\cdot 2^{2r} f_{2r}\cdot 2^{2n-2r}\, b_{2k,\, 2n-2r}.$$

Следовательно, при $1 \leqslant k \leqslant n-1$

$$b_{2k,\,2n} = (1/2)\sum_{r=1}^{k} f_{2r} b_{2k-2r,\,2n-2r} + (1/2)\sum_{r=1}^{n-k} f_{2r} b_{2k,\,2n-2r}. \qquad (4.6)$$

Применим теперь индукцию. Соотношение (4.5), очевидно, справедливо при $\nu=1$; предположим, что оно справедливо при $\nu \leqslant n-1$. Тогда (4.6) сведется к

$$b_{2k,\,2n} = (1/2)\, u_{2n-2k}\sum_{r=1}^{k} f_{2r} u_{2k-2r} + (1/2)\, u_{2k}\sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r}. \qquad (4.7)$$

С учетом (2.6) получаем, что первая сумма равна u_{2k}, тогда как вторая равна u_{2n-2k}. Следовательно, (4.5) справедливо при $\nu=n$. ►

(Удивительный результат, связанный с законом арксинуса, содержится в задаче 4 гл. XIV, 9.)

§ 5*). ПЕРЕМЕНЫ ЗНАКА

При теоретическом исследовании случайных флуктуаций мы столкнулись со многими парадоксами. Например, можно было бы наивно ожидать, что в длительной игре с бросанием монеты наблюдаемое число смен лидерства возрастает примерно пропорционально продолжительности игры. В игре, которая по времени длится вдвое больше, Петр будет лидировать вдвое чаще. Это интуитивное рассуждение ошибочно. Мы покажем, что число смен лидерства в n испытаниях в определенном смысле возрастает лишь как $\sqrt{n}$: в $100n$ испытаниях можно ожидать лишь в 10 раз больше перемен лидерства, чем в n испытаниях. Это еще раз доказывает, что время ожидания между последовательными уравниваниями является, как правило, фантастически долгим.

Вернемся к терминологии теории случайного блуждания. Будем говорить, что в момент времени n произошла *перемена знака*, если $\mathbf{S}_{n-1}$ и $\mathbf{S}_{n+1}$ имеют противоположные знаки, т. е. если путь пересекает ось. В этом случае $\mathbf{S}_n=0$, и, следовательно, n — четное (положительное) целое число.

Теорема 1 [1]. *Вероятность* $\xi_{r,\,2n+1}$ *того, что до момента времени* $2n+1$ *произойдет ровно* r *перемен знака, равна* $2p_{2n+1,\,2r+1}$;

*) Материал этого параграфа не будет явно использоваться в дальнейшем.

[1] По поводу аналогичных теорем для числа возвращений в начало см. задачи 9 и 10. По поводу другого доказательства см. задачу 11.

иначе говоря,

$$\xi_{r,\,2n+1} = 2\mathbf{P}\{\mathbf{S}_{2n+1} = 2r+1\}, \quad r = 0,\ 1,\ \ldots\ . \tag{5.1}$$

Доказательство. Мы начнем с того, что переформулируем теорему в более удобном виде. Если первый шаг ведет в точку (1, 1), возьмем эту точку в качестве начала новой системы координат. Тогда пересечению горизонтальной оси в старой системе соответствует пересечение прямой, лежащей непосредственно под новой осью, т. е. пересечение уровня -1. Аналогичная процедура применима в том случае, когда $\mathbf{S}_1 = -1$, и, таким образом, видно, что теорема полностью эквивалентна следующему *утверждению*: вероятность того, что до момента времени $2n$ уровень -1 будет пересечен ровно r раз, равна $2p_{2n+1,\,2r+1}$.

Рассмотрим сначала случай $r=0$. Уровень -1 не будет пересечен тогда и только тогда, когда путь не коснется уровня -2 (и не пересечет его). В этом случае $\mathbf{S}_{2n}$ — неотрицательное четное целое число. Для $k \geqslant 0$ из основной леммы об отражении § 1 заключаем, что число путей из (0, 0) в $(2n,\ 2k)$, которые касаются уровня -2, равно числу путей в $(2n,\ 2k+4)$. Вероятность попасть в точку $(2n,\ 2k)$, не коснувшись уровня -2, равна, следовательно, $p_{2n,\,2k} - p_{2n,\,2k+4}$. Вероятность того, что путь не касается уровня -2, равна сумме этих величин при $k=0,\ 1,\ 2,\ \ldots$. Внутренние члены взаимно уничтожаются, и мы находим, что наша вероятность равна $p_{2n,\,0} + p_{2n,\,2}$. Это доказывает утверждение при $r=0$, потому что

$$p_{2n+1,\,1} = (1/2)\,(p_{2n,\,0} + p_{2n,\,2}). \tag{5.2}$$

Это очевидным образом следует из того факта, что каждый путь, проходящий через точку $(2n+1,\ 1)$, проходит либо через $(2n,\ 0)$, либо через $(2n,\ 2)$.

Далее, пусть $r=1$. Путь, пересекающий уровень -1 в момент времени $2\nu-1$, может быть разделен на участок от (0, 0) до $(2\nu,\ -2)$ и путь длины $2n-2\nu$, выходящий из $(2\nu,\ -2)$. К последнему участку применим результат при $r=0$, поменяв местами плюс и минус. Таким образом, число путей длины $2n-2\nu$, выходящих из точки $(2\nu,\ -2)$ и не пересекающих уровень -1, равно числу путей из $(2\nu,\ -2)$ в $(2n+1,\ -3)$. Но каждый такой путь комбинируется с начальным участком от (0, 0) до $(2n+1,\ -3)$. Отсюда следует, что число путей длины $2n$, пересекающих уровень -1 ровно один раз, равно числу путей из начала координат в точку $(2n+1,\ -3)$, т. е. $2^{2n+1}p_{2n+1,\,3}$. Это доказывает утверждение для $r=1$.

Утверждение для произвольного r доказывается теперь по индукции, причем рассуждение второй части доказательства используется без изменения. (Доказательство проведено для частного случая $r=1$ только по той причине, что в общем случае потребуются громоздкие обозначения.) ▶

Неожиданным следствием теоремы является то, что *вероятность* $\xi_{r,n}$ *r перемен знака в n испытаниях убывает с ростом r*:

$$\xi_{0,n} \geqslant \xi_{1,n} > \xi_{2,n} > \dots \tag{5.3}$$

Это означает, что независимо от числа бросаний событие, состоящее в том, что лидерство ни разу не переменится, более вероятно, чем любое заранее фиксированное число перемен.

Примеры. а) Вероятности x_r для ровно r перемен знака в 99 испытаниях имеют следующие значения:

r	x_r	r	x_r
0	0,1592	7	0,0517
1	0,1529	8	0,0375
2	0,1412	9	0,0260
3	0,1252	10	0,0174
4	0,1066	11	0,0111
5	0,0873	12	0,0068
6	0,0686	13	0,0040

б) Вероятность того, что в 10 000 испытаниях не будет ни одной перемены знака, приблизительно равна 0,0160. Вероятности x_r для ровно r перемен убывают очень медленно; при $r=10$, 20, 30 соответствующие значения x_r равны 0,0156, 0,0146, 0,0130. Вероятность того, что в 10 000 испытаниях произойдет не более 10 смен лидерства, приблизительно равна 0,174; иначе говоря, почти в каждой шестой такой серии испытаний будет не более 10 перемен лидерства. ▶

Приятным свойством равенства (5.1) является то, что оно дает возможность применить нормальное приближение, рассмотренное в § 2. Предположим, что n велико и x—фиксированное положительное число. Вероятность того, что до момента времени n произойдет менее $x\sqrt{n}$ перемен знака, практически совпадает с $2\mathbf{P}\{S_n < 2x\sqrt{n}\}$, а последняя вероятность, согласно (2.7), стремится к $\mathfrak{N}(2x)-1/2$ при $n \to \infty$. Таким образом, мы имеем следующую теорему.

Теорема 2. (Нормальное приближение.) *Вероятность того, что до момента времени n произойдет менее* $x\sqrt{n}$ *перемен знака, стремится к* $2\mathfrak{N}(2x)-1$ *при* $n \to \infty$.

Отсюда следует, что *медиана* числа перемен знака примерно равна $0,337\sqrt{n}$; это означает, что при достаточно больших n вероятности того, что число перемен знака будет меньше $0,337\sqrt{n}$ или больше него, приблизительно равны. С вероятностью 1/10 произойдет менее $0,0628\sqrt{n}$ перемен знака и т. д. [1])

---

[1]) Такое приближение дает 1/10 для вероятности не более 6 уравниваний в 10 000 испытаниях. Это заниженная оценка: истинное значение составляет около 0,112.

§ 6. РЕЗУЛЬТАТ ЭКСПЕРИМЕНТА

На рис. 4 приведен результат машинного эксперимента — моделирования на ЭВМ 10 000 бросаний монеты; эти же данные содержатся в таблице примера гл. I, 6, в). Верхняя часть графика содержит результаты первых 550 испытаний; следующие две части изображают полную запись результатов 10 000 испытаний, причем масштаб в горизонтальном направлении здесь изменен в отношении 1 : 10. Масштаб в вертикальном направлении одинаков для всех графиков.

Когда большинство людей смотрят на график, их поражают длины интервалов между последовательными пересечениями оси. Фактически же график изображает довольно умеренный случай — был выбран самый умеренный из трех имеющихся графиков. Более поразительный пример получается, если на этот же самый график посмотреть в *обратном* направлении, т. е. изменить на противоположный порядок, в котором в действительности производились эти 10 000 испытаний (см. § 8). Теоретически серия, изображенная на графике, и обратная серия, получаемая в обратном направлении, одинаково законны в качестве представителей идеального случайного блуждания. Обратное случайное блуждание имеет следующие характеристики. Выходя из начала координат,

путь находится на

отрицательной стороне			*положительной стороне*		
первые	7804	*шага*	*следующие*	8	*шагов*
следующие	2	*шага*	*следующие*	54	*шага*
следующие	30	*шагов*	*следующие*	2	*шага*
следующие	48	*шагов*	*следующие*	6	*шагов*
следующие	2046	*шагов*			
Всего	9930	*шагов*	*Всего*	70	*шагов*
Доля времени:	0,993		*Доля времени*:	0,007	

Это *выглядит* абсурдно, однако вероятность того, что при 10 000 бросаниях идеальной монеты одна из сторон будет лидировать на протяжении более чем 9930 испытаний, а вторая менее чем 70 превосходит 1/10. Иначе говоря, в среднем *результаты одной из десяти последовательностей испытаний будут выглядеть более абсурдно, чем только что описанные*. Для сравнения заметим, что вероятность равновесия, лучшего, чем на графике, равна всего лишь 0,072.

График, приведенный на рис. 4, имеет 78 перемен знака и 64 других возвращений в начало. Обратная серия имеет 8 перемен знака и 6 других возвращений в начало. Опрос мнения специалистов обнаруживает, что даже опытные статистики ожидают значительно больше чем 78 перемен знака в 10 000 испытаний, и никто не считает всерьез возможным только 8 перемен знака. На самом деле вероятность не более 8 перемен знака превышает 0,14, тогда как вероят-

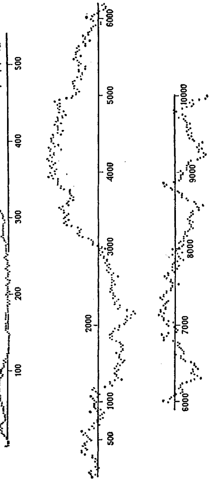

Рис. 4. Результаты 10 000 бросаний идеальной монеты (описанные в § 6).

ность более 78 перемен знака приблизительно равна 0,12. Что касается числа перемен знака в этих двух сериях, то они равноправны, и теоретически ни одно из них не должно вызывать удивления. Если они кажутся неожиданными, то причиной тому наша достойная осуждения интуиция и наши сомнительные ссылки на таинственный «закон средних».

§ 7. МАКСИМУМЫ И ПЕРВЫЕ ДОСТИЖЕНИЯ

Большинство наших результатов выводится из основной леммы 1 § 3, которая, в свою очередь, является простым следствием принципа отражения. Сосредоточим теперь наше внимание на других интересных следствиях этого принципа.

Вместо путей, проходящих выше оси t, рассмотрим пути, которые лежат ниже прямой $x=r$, т. е. пути, удовлетворяющие условию

$$\mathbf{S}_0 < r, \quad \mathbf{S}_1 < r, \ \ldots, \ \mathbf{S}_n < r. \tag{7.1}$$

В этом случае мы говорим, что *максимум* пути $< r$. (Максимум неотрицателен, потому что $\mathbf{S}_0 = 0$.) Пусть $A=(n, k)$ — точка с ординатой $k \leqslant r$. Путь из точки 0 в точку A касается прямой $x=r$ или пересекает ее, если нарушается условие (7.1). Согласно принципу отражения, число таких путей равно числу путей из начала координат в точку $A'=(n, 2r-k)$, которая является отражением точки A относительно прямой $x=r$. Это доказывает следующую лемму.

Лемма 1. *Пусть $k \leqslant r$. Вероятность того, что путь длины n ведет в точку $A=(n, k)$ и имеет максимум $\geqslant r$, равна $p_{n,\,2r-k} = \mathbf{P}\{\mathbf{S}_n = 2r-k\}$.*

Вероятность того, что максимум равен r, равна разности $p_{n,\,2r-k} - p_{n,\,2r+2-k}$. Суммируя по всем $k \leqslant r$, получаем вероятность того, что произвольный путь длины n имеет максимум, в точности равный r. После сокращений сумма приводится к виду $p_{n,\,r} + p_{n,\,r+1}$. Если n и r имеют одинаковую четность, то $p_{n,\,r+1}=0$, в противном случае $p_{n,\,r}=0$. Таким образом, справедлива следующая теорема.

Теорема 1. *Вероятность того, что максимум пути длины n равен $r \geqslant 0$, совпадает с положительным членом пары $p_{n,\,r}$ и $p_{n,\,r+1}$.*

Для $r=0$ и четных моментов времени это утверждение сводится к равенству

$$\mathbf{P}\{\mathbf{S}_1 \leqslant 0, \ \mathbf{S}_2 \leqslant 0, \ \ldots, \ \mathbf{S}_{2n} \leqslant 0\} = u_{2n}, \tag{7.2}$$

эквивалентному соотношению (3.4), которое представляет собой один из вариантов основной леммы. Следовательно, теорема 1 обобщает эту лемму.

Введем теперь понятие, играющее важную роль в общей теории случайных процессов. Момент времени n называется моментом *первого достижения точки* $r>0$, если

$$S_1 < r, \ \ldots, \ S_{n-1} < r, \quad S_n = r. \tag{7.3}$$

В настоящем контексте было бы предпочтительнее говорить о первом попадании, однако термин «первое достижение», берущий начало из физической литературы [1]), более употребителен; кроме того, термин «попадание» не применим к непрерывным процессам.

Очевидно, путь, удовлетворяющий (7.3), должен проходить через точку $(n-1, r-1)$, и его максимум до момента $n-1$ должен быть равен $r-1$. Мы видели, что вероятность такого события равна $p_{n-1, r-1} - p_{n-1, r+1}$, и, таким образом, справедлива следующая теорема.

Теорема 2. *Вероятность* $\varphi_{r, n}$ *того, что первое достижение* r *произошло в момент времени* n, *равна*

$$\varphi_{r, n} = (1/2)\left[p_{n-1, r-1} - p_{n-1, r+1}\right]. \tag{7.4}$$

Простые вычисления показывают, что

$$\varphi_{r, n} = \frac{r}{n}\binom{n}{(n+r)/2} 2^{-n} \tag{7.5}$$

(как всегда, биномиальный коэффициент полагается равным нулю, если $(n+r)/2$ не является целым). По поводу другого вывода формулы см. п. 6 § 8.

Распределение (7.5) наиболее интересно, когда r велико. Для того чтобы получить вероятность того, что первое достижение r произойдет до момента времени N, нужно просуммировать $\varphi_{r, n}$ по всем $n \leqslant N$. Из нормального приближения (2.7) следует, что только те члены дают значимый вклад в сумму, для которых r^2/n не слишком велико и не слишком близко к 0. Для таких членов оценки из гл. VII, 2 дают приближение

$$\varphi_{r, n} \sim \sqrt{2/\pi}\,(r/\sqrt{n^3})\, e^{-r^2/(2n)}. \tag{7.6}$$

При суммировании нужно иметь в виду, что n должно иметь ту же четность, что и r. Сумма является интегральной суммой Римана для интеграла в (7.7), и мы приходим к следующему утверждению.

Теорема 3. (Предельная теорема для первых достижений.) *Для фиксированного t вероятность того, что первое достижение r произойдет до момента времени tr^2, при $r\to\infty$ стремится к* [2]).

---

[1]) В оригинале first passage; в литературе по теории вероятностей этому обычно соответствует термин «первое достижение», в физической литературе — термин «первое прохождение». В дальнейшем будут использоваться оба термина (в зависимости от контекста).— *Прим. перев.*

[2]) Формула (7.7) определяет так называемое положительное устойчивое распределение с показателем 1/2. По поводу обобщений теоремы 3 см. задачу 14 гл. XIV, 9.

$$\sqrt{2/\pi}\int\limits_{1/\sqrt{T}}^{\infty} e^{-s^2/2}\,ds = 2\left[1-\mathfrak{R}\left(1/\sqrt{T}\right)\right], \tag{7.7}$$

где $\mathfrak{R}$ — *нормальное распределение, определенное в гл.* VII, 1.

Отсюда следует, что, грубо говоря, время ожидания первого достижения r возрастает как *квадрат* r: вероятность того, что первое достижение произойдет после момента времени $(9/4)r^2$, близка к 1/2.

Распределение времен первого достижения непосредственно приводит к распределению момента времени, когда частица в r-й раз возвращается в начало.

Теорема 4. *Вероятность того, что r-е возвращение в начало произойдет в момент времени n, равна величине* $\varphi_{r,\,n-r}$ *из* (7.5).

Это означает, что r-е возвращение в момент времени n имеет такую же вероятность, как первое достижение r в момент времени $n-r$.

Доказательство [1]). Рассмотрим путь из начала координат в точку $(n, 0)$, все звенья которого лежат ниже оси и ровно $r-1$ внутренних вершин находятся на оси. Для простоты будем называть такие пути характерными [2]). (На рис. 5 изображен такой путь с $n=20$ и $r=5$.) Характерный путь состоит из r участков, крайние точки которых лежат на оси, и для каждого характерного пути мы можем

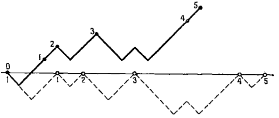

Рис. 5. Первые достижения и возвращения в начало.

построить 2^r различных путей, приписывая вершинам каждого отдельного участка определенный знак (т. е. отражая участки относительно оси). Действуя так, мы получим все пути, заканчивающиеся r-м возвращением, и, следовательно, имеется ровно в 2^r раз больше путей, заканчивающихся r-м возвращением в момент времени n, чем характерных путей. Поэтому теорема может быть переформулирована следующим образом: имеется столько характерных

---

[1]) По поводу доказательства методом производящих функций см. формулу (3.17) гл. XI.

[2]) В оригинале — representative.— *Прим. перев.*

путей длины n, сколько существует путей длины $n-r$, оканчивающихся первым достижением r. Это имеет место потому, что если в характерном пути вычеркнуть r звеньев, у которых левые крайние точки лежат на оси, то получится путь длины $n-r$, заканчивающийся первым достижением r. Эта процедура обратима: для того чтобы получить исходный характерный путь, нужно, исходя из начала координат, вставить r звеньев с отрицательным наклоном и $r-1$ вершин, которые соответствуют первым достижениям $1, 2, \ldots, r-1$ (см. рис. 5). ►

Отсюда следует, что предельную теорему для первого достижения можно применить для нахождения вероятности r-го возвращения при $r\to\infty$: *вероятность того, что r-е возвращение в начало произойдет до момента времени tr^2, стремится к величине* (7.7).

Этот результат обнаруживает другое неожиданное свойство флуктуаций при случайных блужданиях. Случайное блуждание в известном смысле начинается сначала каждый раз, когда частица возвращается в начало. Время до r-го возвращения есть, таким образом, сумма времен ожидания, которые можно интерпретировать как «результаты измерения одной и той же физической величины в одинаковых условиях». Считается, что среднее из r таких наблюдений должно сходиться к «истинному значению». Однако в данном случае сумма является величиной такого же порядка, как r^2, и поэтому *среднее увеличивается примерно пропорционально r*. Более глубокий анализ показывает, что одно из r времен ожидания является величиной того же порядка, что и вся сумма, а именно r^2. На практике такое явление часто приписывается «ошибке эксперимента» или отбрасывается как «постороннее». Трудно заметить то, что не ожидаешь увидеть.

§ 8. ДВОЙСТВЕННОСТЬ. ПОЛОЖЕНИЕ МАКСИМУМА

Каждому пути соответствует конечная последовательность плюс единиц и минус единиц; рассматривая их в обратном порядке, получаем новый путь. Геометрически новый путь получается поворотом данного пути на 180 градусов вокруг его правой крайней точки и выбором последней в качестве начала новой системы координат. Каждому множеству путей таким образом ставится в соответствие новое множество такой же мощности. Если шаги исходного случайного блуждания суть $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, то шаги нового случайного блуждания определяются следующим образом:

$$\mathbf{X}_1^* = \mathbf{X}_n, \ \ldots, \ \mathbf{X}_n^* = \mathbf{X}_1. \tag{8.1}$$

Вершины нового случайного блуждания определяются частными суммами

$$\mathbf{S}_k^* = \mathbf{X}_1^* + \ldots + \mathbf{X}_k^* = \mathbf{S}_n - \mathbf{S}_{n-k} \tag{8.2}$$

(откуда $S_0^* = 0$ и $S_n^* = S_n$). Мы будем называть такое случайное блуждание *двойственным*. Каждому событию, определенному для исходного случайного блуждания, соответствует событие равной вероятности в двойственном случайном блуждании, и, таким образом, почти каждое вероятностное соотношение имеет двойственное. Этот простой метод получения новых соотношений является более полезным, чем это может показаться на первый взгляд. В полной мере его мощь будет показана только в томе 2 в связи с общими случайными блужданиями и теорией очередей, но даже в данном контексте мы сможем без особых усилий получить некоторые новые интересные результаты.

Для того чтобы показать это, рассмотрим несколько пар двойственных событий, выделяя в каждом случае заслуживающий внимания аспект. В следующих примерах n считается заданным и для краткости крайняя правая точка пути (n, S_n) будет называться *конечной точкой*. Удобно начать с известных событий в двойственном случайном блуждании.

а) *Времена первого достижения.* Из (8.2) следует, что события, определенные соответственно посредством

$$S_j^* > 0, \quad j = 1, 2, \ldots, n, \tag{8.3}$$

и

$$S_n > S_j, \quad j = 0, 1, \ldots, n-1, \tag{8.4}$$

двойственны по отношению друг к другу. Второе из них означает, что конечная точка не достигалась до момента времени n. Из (3.2) мы знаем, что первое событие имеет вероятность $(1/2)\,u_{2\nu}$, когда $n = 2\nu > 0$ четное; для $n = 2\nu + 1$ вероятность та же самая, потому что из $S_{2\nu}^* > 0$ следует $S_{2\nu+1}^* > 0$. Таким образом, *вероятность того, что первое достижение уровня конечной точки произойдет в момент времени n, равна* $(1/2)\,u_{2\nu}$, где $\nu = (1/2)\,n$ или $\nu = (1/2)\,(n-1)$. (Это, очевидно, справедливо также для $n = 1$, но неверно для $n = 0$.) Принцип двойственности привел нас к интересному результату, который нелегко получить непосредственно.

б) *Продолжение.* В предыдущем утверждении конечная точка заранее не фиксировалась. Задание точки r в качестве точки первого достижения означает добавление к (8.4) условия $S_n = r$. Двойственное событие состоит из путей, идущих из начала координат в точку (n, r), у которых все промежуточные вершины расположены выше оси. Число таких путей находится непосредственно из леммы об отражении [с $A = (1, 1)$ и $B = (n, r)$], и мы получаем, таким образом, новое доказательство (7.4).

в) *Максимум в конечной точке.* Если в (8.3) и (8.4) строгие неравенства заменить нестрогими, то получится новая пара двойственных событий. Второе событие осуществляется каждый раз, когда S_n максимально, даже если этот максимум уже достигался в некоторый предыдущий момент времени[1]). Обращаясь к (3.4), видим, что *вероятность этого события равна* $u_{2\nu}$, где $\nu=(1/2)n$ или $\nu=(1/2)(n+1)$. Заслуживает внимания, что эти вероятности равны удвоенным вероятностям, найденным в п. а.

г) Событие, заключающееся в том, что было k возвращений в начало, двойственно событию, заключающемуся в том, что до момента времени n было k попаданий в конечную точку. Сходное утверждение использовалось при изучении перемен знака. (Относительно вероятностей см. § 5, а также задачи 9 и 10).

д) *Закон арксинуса для первого попадания в конечную точку.* Рассмотрим случайно выбранный путь длины $n=2\nu$. В п. а мы видели, что с вероятностью $(1/2)\,u_{2\nu}$ значение $S_{2\nu}$ положительно и таково, что ни один из членов последовательности $S_0, S_1, \ldots, S_{2\nu-1}$ не равен $S_{2\nu}$. Это же справедливо для отрицательных $S_{2\nu}$, и, следовательно, вероятность того, что значение $S_{2\nu}$ не будет достигнуто до момента времени 2ν, равна $u_{2\nu}$; эта же величина определяет вероятность события $S_{2\nu}=0$, в котором конечное значение достигается уже в момент времени 0. Рассмотрим теперь более общее событие, состоящее в том, что первое попадание в конечную точку имеет место в момент времени $2k$ (иначе говоря, мы требуем, чтобы $S_{2k}=S_{2\nu}$, но $S_j \neq S_{2\nu}$ при $j<2k$). Это событие двойственно событию, состоящему в том, что последнее попадание в начало имело место в момент времени $2k$, а в § 4 мы видели, что такие попадания подчиняются дискретному распределению арксинуса. Мы, таким образом, получили неожиданный результат: *с вероятностью* $\alpha_{2k,\,2\nu}=u_{2k}u_{2\nu-2k}$ *первое попадание в конечную точку* $S_{2\nu}$ *имеет место в момент времени* $2\nu-2k$ $(k=0, 1, \ldots, \nu)$. Отсюда следует, в частности, что моменты времени $2k$ и $2\nu-2k$ равновероятны. Кроме того, очень ранние и очень поздние первые попадания гораздо более вероятны, чем первые попадания в другие моменты времени.

е) *Закон арксинуса для положения максимума.* В качестве последнего примера применения принципа двойственности мы покажем, что результаты, полученные в п. а и в, позволяют непосредственно определить распределение вероятностей для моментов, в которые последовательность $S_0, S_1, \ldots, S_n$ достигает максимального значения. К сожалению, это максимальное значение может достигаться неоднократно, и, таким образом, мы должны делать различие между первым и последним максимумом. Результаты, впрочем, аналогичны.

Пусть для простоты n четно: $n=2\nu$. *Первый* максимум имеет место в момент времени k, если

$$S_0 < S_k, \ldots, \quad S_{k-1} < S_k, \tag{8.5а}$$
$$S_{k+1} \leqslant S_k, \ldots, \quad S_{2\nu} \leqslant S_k. \tag{8.5б}$$

[1]) По терминологии гл. 12 тома 2 мы рассматриваем *слабые лестничные точки* в отличие от *строгих* лестничных точек, которые определяются условиями п. а.

Запишем k в виде $k=2\rho$ или $k=2\rho+1$. Согласно п. а, вероятность события (8.5а) равна $(1/2)\,u_{2\rho}$, исключая случай, когда $\rho=0$. Событие (8.5б) зависит только от участка пути после момента времени k, и его вероятность, очевидно, равна вероятности того, что у пути длины $2\nu-k$ все вершины лежат ниже оси t или на ней. В п. в было показано, что эта вероятность равна $u_{2\nu-2\rho}$. Таким образом, *если* $0<k<2\nu$, *то вероятность того, что в последовательности* $S_0, \ldots, S_{2\nu}$ *первый максимум имеет место в момент времени* $k=2\rho$ *или* $k=2\rho+1$, *равна* $(1/2)\,u_{2\rho}\,u_{2\nu-2\rho}$. *При* $k=0$ *и* $k=2\nu$ *эти вероятности равны* $u_{2\nu}$ *и* $(1/2)\,u_{2\nu}$ *соответственно.*

(Для *последнего* максимума вероятности для моментов времени 0 и 2ν меняются местами; остальные вероятности остаются неизменными при условии, что k записано в виде $k=2\rho$ или $k=2\rho-1$.)

Мы видим, что при подходящем соединении четных и нечетных индексов положение максимума описывается дискретным распределением арксинуса. При игре с бросанием монеты вопреки интуиции гораздо вероятнее, что максимальный суммарный выигрыш будет в самом начале (или в самом конце) игры, чем в ее середине.

§ 9. ТЕОРЕМА О РАВНОРАСПРЕДЕЛЕННОСТИ

Мы закончим эту главу доказательством теоремы, упоминавшейся в примере 1, б) в связи с ранговым критерием Гальтона. Она поучительна тем, что показывает, как, казалось бы, безобидные изменения в условиях могут изменить характер результата.

В § 4 было показано, что число звеньев, лежащих выше оси t, описывается дискретным распределением арксинуса. Мы рассмотрим теперь ту же задачу, но сосредоточим наше внимание на путях, ведущих из начала координат в точку на оси t. Результат является неожиданным сам по себе, а также потому, что в корне отличается от закона арксинуса.

Теорема. *Число путей длины* $2n$, *таких, что* $S_{2n}=0$ *и ровно* $2k$ *звеньев лежат над осью* t, *не зависит от* k *и равно* $2^{2n}u_{2n}/(n+1)=2^{2n+1}f_{2n+2}$. *(Здесь* $k=0, 1, \ldots, n$.)

Доказательство. Рассмотрим отдельно случаи $k=0$ и $k=n$. Число путей в $(2n, 0)$, все звенья которых лежат выше оси t, равно числу путей из $(1, 1)$ в $(2n, 0)$, которые не касаются прямой, проходящей непосредственно под осью t. В силу принципа отражения это число равно

$$\binom{2n-1}{n}-\binom{2n-1}{n+1}=\frac{1}{n+1}\binom{2n}{n}. \tag{9.1}$$

Это доказывает утверждение при $k=n$ и в силу симметрии также при $k=0$.

При $1 \leqslant k \leqslant n-1$ воспользуемся индукцией. Теорему легко проверить для случая $n=1$; предположим, что она справедлива для всех путей, длина которых меньше $2n$. Обозначим через $2r$ момент первого возвращения. Имеются две возможности. Если участок пути до момента времени $2r$ расположен в положительной полуплоскости, то $1 \leqslant r \leqslant k$ и второй участок имеет ровно $2k-2r$ звеньев над осью. По предположению индукции путь, удовлетворяющий этим условиям, может быть выбран

$$2^{2r-1} f_{2r} \cdot \frac{2^{2n-2r}}{n-r+1} u_{2n-2r} = \frac{2^{2n-2}}{r(n-r+1)} u_{2r-2} u_{2n-2r} \qquad (9.2)$$

различными способами. С другой стороны, если участок до первого возвращения в начало находится в отрицательной полуплоскости, то конечный участок длины $2n-2r$ содержит ровно $2k$ положительных звеньев и, следовательно, в этом случае $n-r \geqslant k$. Для фиксированного r число путей, удовлетворяющих этим условиям, также равно (9.2). Таким образом, число путей обоих типов получается суммированием (9.2) по $1 \leqslant r \leqslant k$ и $1 \leqslant r \leqslant n-k$ соответственно. Во второй сумме заменим индекс суммирования r на $\rho = n+1-r$. Тогда ρ меняется от $k+1$ до n и слагаемые имеют вид (9.2) с заменой r на ρ. Отсюда следует, что число путей, у которых k звеньев лежат в положительной полуплоскости, получается суммированием (9.2) по $1 \leqslant r \leqslant n$. Поскольку k не входит в (9.2), сумма не зависит от k, что и утверждалось. Поскольку общее число путей равно $2^{2n} u_{2n}$, отсюда находим (делением на $n+1$) число путей каждой категории. (Относительно непосредственного вычисления см. задачу 13.)

Имеет место аналогичная теорема для положения максимума (см. задачу 14).

§ 10. ЗАДАЧИ

1. а) Если $a > 0$ и $b > 0$, то число путей $(s_1, s_2, \ldots, s_n)$, таких, что $s_1 > -b, \ldots, s_{n-1} > -b$, $s_n = a$, равно $N_{n,a} - N_{n,a+2b}$.

б) Если $b > a > 0$, то имеется $N_{n,a} - N_{n,2b-a}$ путей, удовлетворяющих условиям $s_1 < b, \ldots, s_{n-1} < b$, $s_n = a$.

2. Пусть $a > c > 0$ и $b > 0$. Число путей, которые касаются прямой $x=a$ и затем ведут в точку (n, c), не касаясь прямой $x=-b$, равно $N_{n,2a-c} - N_{n,2a+2b+c}$. (Заметим, что сюда относятся пути, касающиеся прямой $x=-b$, *до того* как они коснутся прямой $x=a$.)

3. *Повторные отражения.* Пусть a и b положительны и $-b < c < a$. Число путей, ведущих в точку (n, c) и не имеющих общих точек с прямыми $x=-b$ и $x=a$, равно сумме

$$\sum \left(N_{n,2k(a+b)+c} - N_{n,2k(a+b)+2a-c}\right);$$

эта сумма берется по всем целым k от $-\infty$ до ∞, но лишь конечное число слагаемых отлично от нуля.

Указание. Применить и развить метод предыдущей задачи.

Замечание. Эта задача связана с так называемой *задачей о разорении*, которая возникает при исследовании азартной игры, когда два игрока имеют начальные капиталы a и b, так что игра заканчивается, если суммарный выигрыш достигает значения a или $-b$. По поводу связи со статистическими критериями см. пример 1,в).

(Метод повторного отражения будет вновь применен в задаче 17 гл. XIV, 9 и в связи с теорией диффузии в томе 2, гл. X, 5).

4. Используя лемму 1 § 3, показать (без вычислений), что

$$u_0u_{2n}+u_2u_{2n-2}+\ldots+u_{2n}u_0=1.$$

5. Показать, что

$$u_{2n}=(-1)^n\binom{-1/2}{n}, \quad f_{2n}=(-1)^{n-1}\binom{1/2}{n}.$$

Вывести тождество предыдущей задачи так же, как (2.6) выводится из формулы (12.9) гл. II.

6. Доказать геометрически, что существует ровно столько же путей, которые заканчиваются в точке $(2n+2, 0)$ и все внутренние вершины которых лежат строго выше оси, сколько существует путей, которые заканчиваются в точке $(2n, 0)$ и все вершины которых лежат выше оси или на ней. Таким образом,

$$\mathbf{P}\{S_1\geqslant 0, \ \ldots, \ S_{2n-1}\geqslant 0, \ S_{2n}=0\}=2f_{2n+2}.$$

Указание. См. рис. 1.

7. Доказать лемму 1 § 3 геометрически, показав, что взаимно однозначное соответствие между двумя классами путей устанавливается следующим построением. Обозначим *самую левую* точку минимума заданного пути, ведущего в точку $(2n, 0)$, через $M=(k, m)$. Отразим участок, ведущий из начала координат в точку M, относительно вертикальной прямой $t=k$ и передвинем отраженный участок так, чтобы его начальная точка совпала с точкой $(2n, 0)$. Если M принять за начало координат, то в новой системе координат новый путь ведет из начала в точку $(2n, 2m)$ и все его вершины лежат выше оси или на ней. (Это построение принадлежит Э. Нелсону.)

8. Доказать формулу (3.5) непосредственно, рассмотрев пути, не имеющие общих точек с прямой $x=-1$.

9. Вероятность того, что *до* момента времени $2n$ произошло ровно r возвращений в начало, равна вероятности того, что возвращение имело место в момент времени $2n$ и ему предшествовало по меньшей мере r возвращений. *Указание.* Использовать лемму 1 § 3.

10. *Продолжение.* Обозначим через $z_{r,2n}$ вероятность того, что до момента времени $2n$ включительно произошло ровно r возвращений в начало. Используя предыдущую задачу, показать, что $z_{r,2n}=\rho_{r,2n}+\rho_{r+1,2n}+\ldots$, где $\rho_{r,2n}$ — вероятность того, что r-е возвращение произошло в момент времени $2n$. Применив теорему 4 § 7, доказать, что

$$z_{r,2n}=\frac{1}{2^{2n-r}}\cdot\binom{2n-r}{n}.$$

11. *Другой вывод формулы для вероятности числа перемен знака.* Показать, что

$$\xi_{r,2n-1}=(1/2)\sum_{k=1}^{n-1}f_{2k}\left[\xi_{r-1,2n-1-2k}+\xi_{r,2n-1-2k}\right].$$

Применяя индукцию, предположим, что (5.1) имеет место для всех моментов времени, предшествующих $2n-1$. Показать, что это равенство сводится к

$$\xi_{r,2n-1}=2\sum_{1}^{n-1}f_{2k}\rho_{2n-2k,2r},$$

а это вероятность того, что точка $(2n, 2r)$ будет достигнута после возвращения в начало. Рассмотрев первый шаг и применив теорему о баллотировке, показать, что (5,1) справедливо.

12. Вероятность того, что $S_{2n}=0$ и наибольшее из $S_1, \ldots, S_{2n-1}$ равно k, есть $P\{S_{2n}=2k\}-P\{S_{2n}=2k+2\}$. Доказать это, используя отражение.

13. При доказательстве теоремы § 9 было показано, что

$$\sum_{r=1}^{n} \{1/[r(n-r+1)]\} u_{2r-2}u_{2n-2r}=[1/(n+1)]\, u_{2n}.$$

Показать, что это соотношение эквивалентно (2.6). *Указание.* Разложить дробь.

14. Рассмотрим путь длиной $2n$, такой, что $S_{2n}=0$. Расположим звенья в круговом порядке, отождествив 0 и $2n$, так что первое и последнее звенья окажутся соседними. Применив циклическую перестановку, рассмотреть этот путь в новой системе координат с началом в точке (k, S_k). Показать, что такая перестановка сохраняет максимум, но сдвигает его на k шагов. Доказать, что когда применены все $2n$ циклических перестановок, число случаев, в которых максимум достигается в момент времени r, не зависит от r.

Рассмотрим теперь случайный выбор пути, такого, что $S_{2n}=0$, и найдем место максимума, если последний является единственным; если существует несколько точек максимума, то выберем одну из них случайным образом. В результате получим число между 0 и $2n-1$. Показать, что все возможные значения равновероятны.

В этой главе мы будем иметь дело с событиями, которые определяются посредством некоторых других событий $A_1, A_2, \ldots, A_N$. Например, при игре в бридж событие A «по крайней мере один игрок имеет полную масть» является объединением четырех событий A_k «игрок с номером k имеет полную масть» ($k=1, 2, 3, 4$). Из событий A_k одно, два или более могут осуществиться одновременно, и за счет этого частичного наложения вероятность события A не равна сумме четырех вероятностей $\mathbf{P}\{A_k\}$. Мы покажем, как для заданного множества событий $A_1, \ldots, A_N$ вычислить вероятности того, что 0, 1, 2, 3, ... из них осуществились [1]).

§ 1. ОБЪЕДИНЕНИЕ СОБЫТИЙ

Если A_1 и A_2 — два события, то $A_1 \cup A_2$ обозначает событие, состоящее в том, что осуществились либо A_1, либо A_2, либо они оба. Из гл. I мы знаем (формула (7.4)), что

$$\mathbf{P}\{A\}=\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}-\mathbf{P}\{A_1A_2\}. \tag{1.1}$$

Мы хотим обобщить эту формулу на случай N событий $A_1, A_2, \ldots, A_N$, т. е. вычислить вероятность того, что осуществится хотя бы одно из событий A_k. Символически это событие записывается в виде

$$A = A_1 \cup A_2 \cup \ldots \cup A_N.$$

Для нашей цели недостаточно знать вероятности отдельных событий A_k, нам должна быть дана полная информация о всевозможных пересечениях. Это означает, что для каждой пары (i, j), каждой тройки (i, j, k) и т. д. мы должны знать вероятности одновременного осуществления A_i и A_j, A_i, A_j и A_k и т. д. Для удобства будем обозначать эти вероятности буквой p с соответствующими индексами. Таким образом,

$$p_i=\mathbf{P}\{A_i\}, \quad p_{ij}=\mathbf{P}\{A_iA_j\}, \quad p_{ijk}=\mathbf{P}\{A_iA_jA_k\}, \quad \ldots . \tag{1.2}$$

*) Материал этой главы непосредственно использоваться в дальнейшем не будет. Очень важна только первая теорема.

[1]) По поводу дополнительной информации см. Fréchet M., Les probabilités associées a un système d'événements compatibles et dépendants, Actualités scientifiques et industrielles, № 859 (1940); № 942 (1943), Paris.

Порядок индексов несуществен, однако в целях однозначности мы всегда будем *писать индексы в порядке возрастания*; так, мы пишем $p_{3,7,11}$, а не $p_{7,3,11}$. Два индекса не совпадают. Для суммы всех p с r индексами будем использовать обозначение S_r, т. е. положим

$$S_1=\sum p_i,\quad S_2=\sum p_{ij},\quad S_3=\sum p_{ijk},\ \ldots . \tag{1.3}$$

Здесь $i<j<k<\ldots\leqslant N$, так что в суммах каждая комбинация встречается один и только один раз; следовательно, S_r содержит $\binom{N}{r}$ слагаемых. Последняя сумма S_N сводится к единственному $p_{1,2,3,\ldots,N}$, которое является вероятностью одновременного осуществления всех N событий. При $N=2$ имеется только две суммы S_1 и S_2 и (1.1) можно записать в виде

$$\mathbf{P}\{A\}=S_1-S_2. \tag{1.4}$$

Обобщение на произвольное число N событий дает следующая теорема.

Теорема. *Вероятность P_1 осуществления хотя бы одного из событий $A_1, A_2, \ldots, A_N$ равна*

$$P_1=S_1-S_2+S_3-S_4+-\ldots\pm S_N. \tag{1.5}$$

Доказательство. Докажем (1.5) так называемым методом включения и исключения (см. задачу 26). Для того чтобы вычислить P_1, мы должны сложить вероятности всех элементарных событий, которые содержатся хотя бы в одном из A_i, но каждое элементарное событие должно быть взято только один раз. Действуя последовательно, мы сначала возьмем все элементарные события, которые содержатся только в одном A_i, затем те, которые содержатся ровно в двух из A_i и т. д., и наконец элементарные события (если они существуют), содержащиеся во всех A_i. Пусть E — любое элементарное событие, содержащееся ровно в n из N событий A_i. Не ограничивая общности, мы можем занумеровать события так, что E содержится в $A_1, A_2, \ldots, A_n$, но не содержится в $A_{n+1}, A_{n+2}, \ldots, A_N$. Тогда $\mathbf{P}\{E\}$ является слагаемым в тех $p_i, p_{ij}, p_{ijk}, \ldots$, у которых индексы изменяются от 1 до n. Следовательно, $\mathbf{P}\{E\}$ входит n раз в S_1, $\binom{n}{2}$ раз — в S_2 и т. д. В итоге, выразив правую часть (1.5) через вероятности элементарных событий, мы убедимся, что $\mathbf{P}\{E\}$ входит в P_1 с коэффициентом

$$n-\binom{n}{2}+\binom{n}{3}-+\ldots\pm\binom{n}{n}. \tag{1.6}$$

Остается показать, что это число равно 1. Это непосредственно следует из сравнения (1.6) с биномиальным разложением $(1-1)^n$ (см. формулу (8.7) гл. II). Последнее начинается с 1, и слагаемые из

(1.6) содержатся в нем с противоположными знаками, следовательно, для любого $n \geqslant 1$ выражение (1.6) равно 1. ►

Примеры. а) Пусть при игре в бридж A_i есть событие «игрок с номером i имеет полную масть». Тогда $p_i = 4\binom{52}{13}^{-1}$; событие, состоящее в том, что два игрока i и j имеют полную масть, осуществляется в $4 \cdot 3$ случаях и имеет вероятность $p_{ij} = 12\binom{52}{13}^{-1}\binom{39}{13}^{-1}$; аналогично находим

$$p_{ijk} = 24\binom{52}{13}^{-1}\binom{39}{13}^{-1}\binom{26}{13}^{-1}.$$

Наконец, $p_{1,2,3,4} = p_{1,2,3}$, так как если три игрока имеют полную масть, то ее имеет и четвертый. Вероятность того, что *хотя бы один* игрок имеет полную масть, равна поэтому $P_1 = 4p_1 - 6p_{1,2} + 4p_{1,2,3} - p_{1,2,3,4}$. Используя формулу Стирлинга, находим, что приблизительно $P_1 = (1/4)\,10^{-10}$. В этом частном случае P_1 очень близко к сумме вероятностей A_i, но это скорее исключение, чем правило.

б) *Пары (совпадения).* Следующая задача, известная в различных вариантах и приводящая к неожиданному ответу, была предложена Монтмором (1708). Она была обобщена Лапласом и многими другими авторами.

Две одинаковые случайным образом перетасованные колоды, содержащие по N различных карт каждая, сравниваются друг с другом. Если карта находится на одном и том же месте в обеих колодах, то мы говорим, что имеет место *пара (совпадение)*. Совпадения могут быть на любом из N мест и на нескольких местах одновременно. Этот эксперимент может быть описан и в шуточной форме. Например, две колоды карт можно заменить множеством писем и предназначенных для них конвертов, и своенравный секретарь запечатывает письма в конверты случайным образом. Или можно представить себе, что шляпы в гардеробе перепутаны и распределяются между посетителями случайным образом. Совпадение имеет место, если посетитель получает свою собственную шляпу. Поучительно попытаться угадать, как вероятность совпадения зависит от N: как вероятность получения своей шляпы хотя бы одним из 8 гостей соотносится с соответствующей вероятностью в собрании из 10 000 человек? Кажется удивительным, что эта вероятность практически не зависит от N и близка к 2/3. (Более серьезные приложения см. в задачах 10 и 11.)

Вероятности наличия ровно 0, 1, 2, 3, ... совпадений будут вычислены в § 4, а здесь мы найдем только вероятность P_1 хотя бы одного совпадения. Для удобства занумеруем карты числами 1, 2, ..., N в том порядке, в котором они расположены в одной из колод, и предположим, что каждая перестановка карт во второй колоде имеет вероятность $1/N!$. Пусть событие A_k означает, что имеет место совпадение на k-м месте. Это означает, что карта с номером k находится на k-м месте, а оставшиеся $N-1$ карт могут быть расположены в произвольном порядке. Очевидно, $p_k=(N-1)!/N!=1/N$. Аналогично для любой комбинации i, j имеем $p_{ij}=(N-2)!/N!=1/[N(N-1)]$ и т. д. Сумма S_r содержит $\binom{N}{r}$ слагаемых, каждое из которых равно $(N-r)!/N!$. Следовательно, $S_r=1/r!$, и в силу (1.5) искомая вероятность равна

$$P_1=1-1/2!+1/3!-+\ldots\pm 1/N!. \tag{1.7}$$

Заметим, что $1-P_1$ представляет собой $N+1$ первых членов разложения

$$e^{-1}=1-1+1/2!-1/3!+1/4!-+\ldots, \tag{1.8}$$

и, следовательно, приблизительно

$$P_1\approx 1-e^{-1}=0{,}63212\ldots . \tag{1.9}$$

Точность приближения видна из следующей таблицы истинных значений P_1:

$N=$	3	4	5	6	7
$P_1=$	0,66667	0,62500	0,63333	0,63196	0,63214.

▶

§ 2. ПРИЛОЖЕНИЕ К КЛАССИЧЕСКОЙ ЗАДАЧЕ О РАЗМЕЩЕНИИ

Вернемся теперь к задаче о случайном размещении r шаров по n ящикам, предполагая, что каждое размещение имеет вероятность n^{-r}. Найдем вероятность $p_m(r,n)$ того, что ровно m ящиков остаются пустыми [1].

Пусть событие A_k заключается в том, что ящик с номером k пуст ($k=1, 2, \ldots, n$). При выполнении этого события все r шаров размещаются в остальных $n-1$ ящиках, и это может быть сделано $(n-1)^r$ различными способами. Аналогично существует $(n-2)^r$ размещений, при которых два фиксированных ящика остаются пустыми и т. д. Следовательно,

$$p_i=(1-1/n)^r,\quad p_{ij}=(1-2/n)^r,\quad p_{ijk}=(1-3/n)^r,\ \ldots, \tag{2.1}$$

и поэтому для любого $\nu\leqslant n$

$$S_\nu=\binom{n}{\nu}\left(1-\frac{\nu}{n}\right)^r. \tag{2.2}$$

Вероятность того, что хотя бы один ящик пуст, задается соотношением (1.5), и поэтому *вероятность того, что все ящики за-*

[1]) Эта вероятность была получена совершенно другим способом в задаче 8 гл. II, 11. Сравните также с заключительным замечанием § 3.

няты, равна

$$p_0(r, n) = 1 - S_1 + S_2 - + \ldots = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} \left(1 - \frac{\nu}{n}\right)^r. \quad (2.3)$$

Рассмотрим теперь размещение, при котором ровно m ящиков остаются пустыми. Эти m ящиков могут быть выбраны $\binom{m}{n}$ способами. Тогда r шаров распределяются между остальными $n-m$ ящиками так, что каждый из них занят; число таких распределений равно $(n-m)^r p_0(r, n-m)$. Разделив на n^r, найдем *вероятность того, что ровно m ящиков остаются пустыми:*

$$p_m(r, n) = \binom{n}{m} \left(1 - \frac{m}{n}\right)^r p_0(r, n-m) =$$

$$= \binom{n}{m} \sum_{\nu=0}^{n-m} (-1)^\nu \binom{n-m}{\nu} \left(1 - \frac{m+\nu}{n}\right)^r. \quad (2.4)$$

Мы уже использовали модель случайного выбора r цифр для иллюстрации случайного размещения r предметов по $n=10$ ящикам. Пустым ящикам соответствуют в этом случае отсутствующие цифры: если m ящиков пусты, то в данной последовательности встречаются $10-m$ различных цифр. Численная иллюстрация приводится в табл. 1.

Таблица 1

Вероятности $p_m(r, 10)$, полученные по формуле (2.4)

m	$r=10$	$r=18$	m	$r=10$	$r=18$
0	0,000 363	0,134 673	5	0,128 596	0,000 876
1	0,016 330	0,385 289	6	0,017 189	0,000 014
2	0,136 080	0,342 987	7	0,000 672	0,000 000
3	0,355 622	0,119 425	8	0,000 005	0,000 000
4	0,345 144	0,016 736	9	0,000 000	0,000 000

$p_m(r, 10)$ — вероятность того, что ровно m из цифр 0, 1, ..., 9 *не* появятся в последовательности из r случайных цифр.

Ясно, что непосредственное вычисление по (2.4) возможно в случае относительно малых n и r. С другой стороны, задача о размещении представляет собой интерес, когда n велико. Каковы шансы, что при размещении 10 000 шаров по 1000 ящикам хотя бы один ящик останется пустым? Каковы шансы, что для группы из 2000 человек есть хотя бы один день в году, не являющийся чьим-либо днем рождения? К счастью, на подобные вопросы можно ответить с помощью замечательно простой приближенной формулы, погрешность которой стремится к нулю при $n \to \infty$. Это приближение и приводящие к нему рассуждения типичны для многих *предельных теорем* теории вероятностей.

Таким образом, наша цель заключается в нахождении предельного вида (2.4) при $n \to \infty$ и $r \to \infty$. Соотношение между r и n в принципе может быть произвольным, но если среднее число r/n шаров на ящик чрезмерно велико, то мы не можем ожидать наличия пустых ящиков; в этом случае $p_0(r, n)$ близка к единице и все $p_m(r, n)$ при $m \geqslant 1$ малы. С другой стороны, если r/n стремится к нулю, то практически все ящики должны быть пустыми, и в этом случае $p_m(r, n) \to 0$ для любого фиксированного m. Поэтому действительный интерес представляет только промежуточный случай.

Мы начнем с оценки величины S_ν в (2.2). Так как

$$(n-\nu)^\nu < (n)_\nu < n^\nu,$$

имеем

$$n^\nu (1-\nu/n)^{\nu+r} < \nu! S_\nu < n^\nu (1-\nu/n)^r. \tag{2.5}$$

Из разложения (8.10) гл. II следует, что при $0 < t < 1$ значение $-\log(1-t)$ лежит между t и $t/(1-t)$. Поэтому

$$\{ne^{-(\nu+r)/(n-\nu)}\}^\nu < \nu!\, S_\nu < \{ne^{-r/n}\}^\nu. \tag{2.6}$$

Положим для краткости

$$ne^{-r/n} = \lambda \tag{2.7}$$

и предположим, что r и n возрастают таким образом, что λ остается внутри конечного интервала $0 < a < \lambda < b$. Для каждого фиксированного ν отношение крайних членов в (2.6) стремится к единице, и поэтому

$$0 \leqslant \lambda^\nu/\nu! - S_\nu \to 0. \tag{2.8}$$

Это соотношение очевидно справедливо, когда $\lambda \to 0$, и, следовательно, (2.8) остается справедливым и тогда, когда r и n возрастают таким образом, что λ остается ограниченным. Но

$$e^{-\lambda} - p_0(r, n) = \sum_{\nu=0}^{\infty} (-1)^\nu \{\lambda^\nu/\nu! - S_\nu\} \tag{2.9}$$

и из (2.8) следует, что правая часть стремится к нулю. Кроме того, множитель при $p_0(r, n-m)$ в (2.4) равен S_m, и поэтому для каждого фиксированного m имеем

$$p_m(r, n) - e^{-\lambda}\lambda^m/m! \to 0. \tag{2.10}$$

Тем самым доказана следующая теорема.

Таблица 2

Пуассоновское приближение (2.11) для вероятностей наличия ровно m пустых ящиков при случайном распределении r шаров по $n = 1000$ ящикам

r	λ	$p(m:\lambda)$											
		$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	$m=11$
5000	6,74	0,0012	0,0080	0,0269	0,0604	0,1017	0,1371	0,1540	0,1482	0,1249	0,0935	0,0630	0,0386
5500	4,09	0,0167	0,0685	0,1400	0,1909	0,1951	0,1596	0,1088	0,0636	0,0325	0,0148	0,0060	0,0023
6000	2,48	0,0838	0,2077	0,2575	0,2128	0,1320	0,0655	0,0271	0,0096	0,0030	0,0008	0,0002	
6500	1,50	0,2231	0,3347	0,2510	0,1255	0,0471	0,0141	0,0035	0,0008	0,0001			
7000	0,91	0,4027	0,3661	0,1666	0,0506	0,0115	0,0021	0,0003					
7500	0,55	0,5777	0,3163	0,0873	0,0162	0,0023	0,0003						
8000	0,34	0,7126	0,2406	0,0414	0,0049	0,0004							
8500	0,20	0,8187	0,1637	0,0164	0,0011	0,0001							
9000	0,12	0,8869	0,1064	0,0064	0,0003								

Теорема[1]). *Если n и r стремятся к бесконечности, так что $\lambda = ne^{-r/n}$ остается ограниченным, то для каждого фиксированного m выполняется соотношение* (2.10).

Аппроксимирующее выражение

$$p(m;\ \lambda) = e^{-\lambda}\lambda^m/m! \tag{2.11}$$

определяет так называемое *распределение Пуассона*, которое имеет большое значение и описывает разнообразные явления; оно будет изучаться в гл. VI.

На практике можно использовать $p(m;\ \lambda)$ в качестве приближения, когда n велико. Для небольших значений n необходима оценка погрешности, однако мы заниматься этим не будем.

Примеры. а) В табл. 2 приведены приближенные значения вероятностей того, что m ящиков окажутся пустыми при общем числе ящиков, равном 1000, и числе шаров, меняющемся от 5000 до 9000. При $r=5000$ медиана числа пустых ящиков равна шести: наличие семи или более пустых ящиков имеет приблизительно такую же вероятность, как шести и менее. Даже в случае 9000 шаров в 1000 ящиках мы имеем около одного шанса из девяти найти пустой ящик.

б) В статистике дней рождения (пример гл. II, 3, г)) $n=365$, а r равно числу людей. При $r=1900$ приблизительно $\lambda=2$. *Вероятности $P_{[m]}$ того, что в поселке с* 1900 *жителями m дней в году не являются днями рождения, приблизительно таковы:*

$$P_{[0]}=0{,}135,\quad P_{[1]}=0{,}271,\quad P_{[2]}=0{,}271,\quad P_{[3]}=0{,}180,$$
$$P_{[4]}=0{,}090,\quad P_{[5]}=0{,}036,\quad P_{[6]}=0{,}012,\quad P_{[7]}=0{,}003.$$

в) Когда $n\log n+an$ шаров размещаются в n ящиках и n велико, вероятность того, что все ящики заняты, равна $1-e^{-a}$. ▶

Вместо пустых ящиков можно рассмотреть ящики, содержащие ровно k шаров. Здесь с небольшими изменениями применимы рассуждения, использованные выше в частном случае $k=0$. Как показал Мизес, вероятность того, что ровно m ящиков содержат по k шаров, также аппроксимируется распределением Пуассона (2.11), но в этом случае λ определяется следующим образом:

$$\lambda = n\,(e^{-r/n}/k!)\,(r/n)^k. \tag{2.12}$$

§ 3. ОСУЩЕСТВЛЕНИЕ m ИЗ N СОБЫТИЙ

Теорема § 1 может быть усилена следующим образом.

Теорема. *Для любого целого m, такого, что $1\leqslant m\leqslant N$, вероятность $P_{[m]}$ того, что одновременно осуществляется ровно m из N*

[1]) Принадлежит (с другим доказательством) Мизесу (von Mises R., Über Aufteilungs- und Besetzungswahrscheinlichkeiten, Revue de la Faculté des Sciences de l'Université d'Istanbul, N. S., 4 (1939), 145—163).

событий $A_1, \ldots, A_N$, равна

$$P_{[m]} = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} - + \ldots \pm \binom{N}{m} S_N. \quad (3.1)$$

Замечание. Согласно (1.5), вероятность $P_{[0]}$ того, что не осуществилось ни одно из событий A_j, равна

$$P_{[0]} = 1 - P_1 = 1 - S_1 + S_2 - S_3 \pm \ldots \mp S_N. \quad (3.2)$$

Это показывает, что (3.1) дает правильное значение также и для $m=0$, если мы положим $S_0 = 1$.

Доказательство. Будем рассуждать так же, как при доказательстве (1.5). Пусть E — произвольное элементарное событие; предположим, что оно содержится ровно в n из N событий A_j. Тогда $\mathbf{P}\{E\}$ дает вклад в P_m лишь тогда, когда $n=m$. Для того чтобы узнать, каким образом $\mathbf{P}\{E\}$ входит в правую часть (3.1), заметим, что $\mathbf{P}\{E\}$ входит в суммы $S_1, S_2, \ldots, S_n$, но не входит в суммы $S_{n+1}, \ldots, S_N$. Отсюда следует, что $\mathbf{P}\{E\}$ не дает вклада в правую часть (3.1), если $n < m$. Если $n=m$, то $\mathbf{P}\{E\}$ входит в одно и только в одно слагаемое суммы S_m. Для завершения доказательства теоремы остается показать, что при $n > m$ вклады $\mathbf{P}\{E\}$ в члены правой части $S_m, S_{m+1}, \ldots, S_n$ взаимно уничтожаются. Действительно, $\mathbf{P}\{E\}$ входит в S_k с коэффициентом $\binom{n}{k}$, равным числу наборов по k событий из n событий, содержащих E. Таким образом, при $n > m$ полный вклад $\mathbf{P}\{E\}$ в правую часть (3.1) равен

$$\binom{n}{m} - \binom{m+1}{m}\binom{n}{m+1} + \binom{m+2}{m}\binom{n}{m+2} - + \ldots. \quad (3.3)$$

Если выразить биномиальные коэффициенты через факториалы, то легко видеть, что это выражение сводится к следующему:

$$\binom{n}{m}\left\{\binom{n-m}{0} - \binom{n-m}{1} + - \ldots \pm \binom{n-m}{n-m}\right\}. \quad (3.4)$$

В скобках имеем биномиальное разложение для $(1-1)^{n-m}$, так что (3.3) обращается в нуль, что и утверждалось. ▶

Читателю предлагается проверить, что подстановка выражения (2.2) в (3.1) *непосредственно* приводит к (2.4).

§ 4. ПРИЛОЖЕНИЕ К ЗАДАЧАМ О СОВПАДЕНИЯХ И К ЗАДАЧЕ ОБ УГАДЫВАНИИ

В примере 1, б) мы рассматривали совпадения карт в двух колодах и нашли, что $S_k = 1/k!$. Подставив это выражение в (3.1), получим следующий результат.

При случайном сопоставлении двух одинаковых колод, состоящих из N различных карт, вероятности $P_{[m]}$ получить ровно m совпадений равны

$$
\begin{aligned}
P_{[0]}&=1-1+\frac{1}{2!}-\frac{1}{3!}+-\ldots\pm\frac{1}{(N-2)!}\mp\frac{1}{(N-1)!}\pm\frac{1}{N!},\\
P_{[1]}&=1-1+\frac{1}{2!}-\frac{1}{3!}+-\ldots\pm\frac{1}{(N-2)!}\mp\frac{1}{(N-1)!},\\
P_{[2]}&=\frac{1}{2!}\left\{1-1+\frac{1}{2!}-\frac{1}{3!}+-\ldots\pm\frac{1}{(N-3)!}\mp\frac{1}{(N-2)!}\right\},\\
P_{[3]}&=\frac{1}{3!}\left\{1-1+\frac{1}{2!}-\frac{1}{3!}+-\ldots\pm\frac{1}{(N-3)!}\right\},\qquad (4.1)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
P_{[N-2]}&=\frac{1}{(N-2)!}\left\{1-1+\frac{1}{2!}\right\},\\
P_{[N-1]}&=\frac{1}{(N-1)!}\{1-1\}=0,\quad P_{[N]}=\frac{1}{N!}.
\end{aligned}
$$

Последнее соотношение очевидно. Обращение в нуль $P_{[N-1]}$ выражает тот факт, что невозможно получить $N-1$ совпадение без того, чтобы все N карт в каждой колоде были расположены в одинаковом порядке.

Выражения в фигурных скобках в правых частях (4.1) являются первыми членами разложения e^{-1}. *Поэтому для больших N мы имеем приближенное равенство*

$$P_{[m]}\approx e^{-1}/m!. \qquad (4.2)$$

Ниже в табл. 3 колонки, озаглавленные $P_{[m]}$, дают точные значения $P_{[m]}$ при $N=3,\ 4,\ 5,\ 6,\ 10$. В последней колонке приведены предельные значения

$$p_m=e^{-1}/m!. \qquad (4.3)$$

Числа p_m дают довольно хорошее приближение для $P_{[m]}$ даже при небольших значениях N.

Для чисел p_m, определенных формулой (4.3), мы имеем

$$\sum p_k=e^{-1}\left(1+1+\frac{1}{2!}+\frac{1}{3!}+\ldots\right)=e^{-1}e=1.$$

Следовательно, p_k могут трактоваться как вероятности. Отметим, что (4.3) является частным случаем *распределения Пуассона* (2.11) с $\lambda=1$.

Пример. *Проверка способности угадывать.* При дегустации вин, в физических экспериментах и т. д. проверяемого просят назвать неизвестный ему порядок расположения N предметов, скажем карт. Действительная способность угадывать с его стороны будет проявляться как отклонение от случайности. Для оценки степени этой способности мы должны уметь находить вероятности случайных удач. Угадывающий может придерживаться разных систем, среди которых мы упомянем три крайние возможности. (i) Угадывающий выбирает одну карту и называет все время ее. При такой системе он будет иметь одно и только одно правильное угадывание в каждой серии; случайные флуктуации исключены. (ii) Угадывающий называет каждую карту один раз, так что каждая серия из N угадываний соответствует перетасовке колоды. Если проверяемый не обладает способностью угадывать, то можно применить формулы (4.1). (iii) Третья возможность заключается в том, что каждое из N угадываний производится совершенно независимо от других. Здесь имеется N^N возможных комбинаций. Несомненно, что каждый человек имеет свои привычки и склонен называть определенные комбинации чаще других, однако в первом приближении можно считать все N^N комбинаций равновероятными. Так как m верных и $N-m$ неверных угадываний могут быть произведены $\binom{N}{m}(N-1)^{N-m}$ различными способами, вероятность ровно m верных угадываний равна

$$b_m = \binom{N}{m} \frac{(N-1)^{N-m}}{N^N}. \tag{4.4}$$

(Это частный случай биномиального распределения, которое было получено в примере гл. II, 4, в).)

Таблица 3

Вероятности m правильных угадываний в колоде из N различных карт

m	$N=3$		$N=4$		$N=5$		$N=6$		$N=10$		p_m
	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	
0	0,333	0,296	0,375	0,316	0,367	0,328	0,368	0,335	0,36788	0,34868	0,367879
1	0,500	0,444	0,333	0,422	0,375	0,410	0,367	0,402	0,36788	0,38742	0,367879
2	...	0,222	0,250	0,211	0,167	0,205	0,187	0,201	0,18394	0,19371	0,183940
3	0,167	0,037	...	0,047	0,083	0,051	0,056	0,053	0,06131	0,05740	0,061313
4			0,042	0,004	...	0,006	0,021	0,008	0,01534	0,01116	0,015328
5					0,008	0,000	...	0,001	0,00306	0,00149	0,003066
6							0,001	0,000	0,00052	0,00014	0,000511
7									0,00007	0,00001	0,000073
8									0,00001	.....	0,000009
9									.....	.....	0,000001
10									.....	.....	0,000000

$P_{[m]}$ заданы формулой (4.1), b_m — формулой (4.4). В последнем столбце приведены предельные вероятности (4.3) (распределение Пуассона).

В табл. 3 приведены вероятности успеха, когда угадывание производится по системе (ii) или (iii). Чтобы оценить достоинства этих двух методов, необходимо найти средние значения и величины случайных флуктуаций. Оказывается, что среднее число правильных угадываний одинаково при всех системах; случайные отклонения несколько больше при системе (ii), чем при (iii). Просмотр табл.3 показывает, что практически разница не очень велика. ▶

§ 5. РАЗЛИЧНЫЕ ДОПОЛНЕНИЯ

а. Осуществление по меньшей мере m событий

В обозначениях § 3 *вероятность P_m того, что одновременно осуществляются m или более из событий $A_1, \ldots, A_N$, равна*

$$P_m = P_{[m]} + P_{[m+1]} + \ldots + P_{[N]}. \tag{5.1}$$

Для того чтобы выразить P_m через S_k, проще всего действовать методом индукции, начиная с выражения (1.5) для P_1 и используя рекуррентное соотношение $P_{m+1} = P_m - P_{[m]}$. При $m \geqslant 1$ получаем

$$P_m = S_m - \binom{m}{m-1} S_{m+1} + \binom{m+1}{m-1} S_{m+2} - \binom{m+2}{m-1} S_{m+3} + \ldots \pm \binom{N-1}{m-1} S_N. \tag{5.2}$$

Соотношение (5.2) можно вывести также и непосредственно, используя рассуждения, которые привели к (3.1).

б. Прочие тождества

Коэффициенты S_ν можно выразить через $P_{[k]}$ или P_k следующим образом:

$$S_\nu = \sum_{k=\nu}^{N} \binom{k}{\nu} P_{[k]} \tag{5.3}$$

и

$$S_\nu = \sum_{k=\nu}^{N} \binom{k-1}{\nu-1} P_k. \tag{5.4}$$

Идея доказательства. При заданных значениях $P_{[m]}$ соотношения (3.1) можно рассматривать как линейные уравнения относительно неизвестных S_ν, и мы должны доказать, что (5.3) представляет собой их единственное решение. Если подставить (5.3) в выражение (3.1) для $P_{[m]}$, то коэффициент при $P_{[k]}$ $(m \leqslant k \leqslant N)$ в правой части будет равен

$$\sum_{\nu=m}^{k} (-1)^{\nu-m} \binom{\nu}{m} \binom{k}{\nu} = \binom{k}{m} \sum_{\nu=m}^{k} (-1)^{\nu-m} \binom{k-m}{\nu-m}. \tag{5.5}$$

При $k = m$ это выражение равно 1. При $k > m$ сумма представляет собой биномиальное разложение $(1-1)^{k-m}$ и поэтому равна нулю. Следовательно, подстановка (5.3) сводит (3.1) к тождеству $P_{[m]} = P_{[m]}$. Единственность решения (3.1) следует из того, что каждое последующее уравнение содержит только одно новое неизвестное по сравнению с предыдущим, так что S_ν могут быть вычислены последовательно. Справедливость (5.4) может быть доказана аналогичным образом.

в. Неравенства Бонферрони

Ряд неравенств как для $P_{[m]}$, так и для P_m может быть получен следующим образом. *Если либо в формуле* (3.1), *либо в* (5.2) *оставлены лишь слагаемые, содержащие* $S_m, S_{m+1}, \ldots, S_{m+r-1}$, *а слагаемые, содержащие* $S_{m+r}, S_{m+r+1}, \ldots, S_N$, *отброшены, то погрешность* (*т. е. истинное значение минус приближенное значение*) *имеет знак первого отброшенного члена* [*а именно*, $(-1)^r$] *и не превосходит его по абсолютной величине*. Таким образом, при $r=1$ и $r=2$

$$S_m-(m+1)\,S_{m+1} \leqslant P_{[m]} \leqslant S_m \tag{5.6}$$

и

$$S_m-mS_{m+1} \leqslant P_m \leqslant S_m. \tag{5.7}$$

Идея доказательства. Тождество (3.1) показывает, что утверждение (5.6) эквивалентно неравенству

$$\sum_{\nu=t}^{N}(-1)^{\nu-t}\binom{\nu}{m}S_\nu \geqslant 0 \tag{5.8}$$

для любого t. Для того чтобы представить левую часть в виде линейной комбинации $P_{[k]}$, воспользуемся соотношением (5.3). При $t \leqslant k \leqslant N$ коэффициент при $P_{[k]}$ равен

$$\sum_{\nu=t}^{k}(-1)^{\nu-t}\binom{\nu}{m}\binom{k}{\nu}=\binom{k}{m}\sum_{\nu=t}^{k}(-1)^{\nu-t}\binom{k-m}{\nu-m}.$$

Последняя сумма равна $\binom{k-m-1}{t-m-1}$ и, следовательно, положительна (задача 13 гл. II, 12).

Дальнейшие неравенства можно найти в монографии Фреше, цитированной в начале главы.

§ 6. ЗАДАЧИ

Замечание. Предполагается, что в каждом случае все комбинации равновероятны.

1. В чулане находится десять пар ботинок. Случайно выбираются четыре ботинка. Найти вероятность того, что среди выбранных четырех ботинок есть по крайней мере одна пара.

2. Бросаются пять игральных костей. Найти вероятность того, что по меньшей мере на трех из них выпадут одинаковые грани (проверить методами гл. II, 5).

3. Найти вероятность того, что при пяти бросаниях монеты герб выпадет по меньшей мере три раза подряд.

4. Решить задачу 3 для случая по меньшей мере пяти последовательных выпадений герба при десяти бросаниях.

5. Решить задачи 3 и 4 для выпадения единицы, если бросается игральная кость вместо монеты.

6. Две кости бросаются r раз. Найти вероятность p_r того, что каждая из шести комбинаций (1, 1), ..., (6, 6) появится по меньшей мере один раз.

7. *Четверки у игрока в бридж.* При игре в бридж среди тринадцати карт одного игрока могут содержаться 0, 1, 2 или 3 четверки карт одного значения. Вычислить соответствующие вероятности.

8. *Выбор с возвращением.* Из генеральной совокупности n людей извлекается выборка объема r. Найти вероятность u_r того, что в этой выборке будут содержаться N заданных людей. (Это задача 12 гл. II, 11.)

9. *Выбор без возвращения.* Ответить на вопрос задачи 8 в этом случае и показать, что $u_r \longrightarrow p^N$. (Это задача 3 гл. II, 11, но настоящий метод приводит к другому по форме результату. Доказать тождественность результатов.)

10. Доказать, что в общем разложении детерминанта порядка N число слагаемых, содержащих один или более диагональных элементов, равно $N!P_1$, где P_1 определяется соотношением (1.7).

11. Доказать, что число способов, которыми 8 ладей могут быть расставлены на шахматной доске так, что ни одна из них не бьет другую и ни одна не стоит на белой диагонали, равно $8!(1-P_1)$, где P_1 определяется соотношением (1.7) при $N=8$.

12. *Задача о выборке (о собирании купонов).* Колода карт состоит из s одинаковых наборов, содержащих каждый n карт, занумерованных числами 1, 2, ..., n. Из колоды извлекается без возвращения случайная выборка объема $r \geqslant n$. Вычислить вероятность u_r того, что каждое число представлено в выборке. (Применительно к колоде карт для игры в бридж при $s=4$, $n=13$ получается вероятность того, что среди r выбранных карт содержатся все 13 значений, а при $s=13$, $n=4$ — вероятность того, что в выборке представлены все четыре масти.)

13. *Продолжение.* Показать, что при $s \longrightarrow \infty$ $u_r \longrightarrow p_0(r, n)$, где последняя величина определена формулой (2.3). Это означает, что в пределе наш выбор становится случайным выбором с возвращением из совокупности чисел 1, 2, ..., n.

14. *Продолжение.* Из результатов задачи 12 вывести, что при $r < n$

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (ns-ks)_r = 0,$$

а при $r=n$

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (ns-ks)_n = s^n n!.$$

Чтобы проверить это, вычислить r-ю производную при $x=0$ функции

$$[1/(1-x)^{ns-r+1}]\,[1-(1-x)^s]^n.$$

15. В задаче 12 о выборке найти вероятность того, что потребуется ровно r извлечений для получения выборки, содержащей все числа. Перейти к пределу при $s \to \infty$.

16. В ядре клетки содержатся N хромосом, между любыми двумя из которых возможен обмен частями. Предположив, что произошло r обменов (которые могут осуществиться $\binom{N}{2}^r$ различными способами), найти вероятность того, что в них участвовало ровно m хромосом [1]).

17. Найти вероятность того, что среди 5 случайно выбранных карт из колоды для игры в покер отсутствует ровно k мастей.

[1]) О случае $N=6$ см. Catcheside D. G., Lea D. E., Thoday J. M., Types of chromosome structural change introduced by the irradiation of tradescantia microspores, Journal of Genetics, 47 (1945—1946), 113—149.

18. Найти вероятность того, что набор из тринадцати карт из колоды для игры в бридж содержит ровно k пар «туз — король» одной масти.

19. *Кратные совпадения*. Две одинаковые колоды из N различных карт каждая сравниваются одновременно с такой же третьей (контрольной) колодой. Найти вероятность u_m того, что будет ровно m двойных совпадений. Показать, что $u_0 \to 1$ при $N \to \infty$ (откуда вытекает, что $u_m \to 0$ при $m \geqslant 1$).

20. *Сложные совпадения*. Процедура сравнения из предыдущей задачи модифицируется следующим образом. Из $2N$ карт случайно выбираются N, и только эти N карт сравниваются с контрольной колодой. Найти вероятность отсутствия совпадений. Доказать, что она стремится к $1/e$ при $N \to \infty$.

21. *Сложные совпадения*. Решить задачу 20, если вместо двух используется r колод.

22. В классической задаче о размещении вероятность $P_{[m]}(k)$ того, что ровно в m ящиках окажется по k шаров, равна

$$P_{[m]}(k) = \frac{(-1)^m \, n! \, r!}{m! \, n^r} \sum_j (-1)^j \frac{(n-j)^{r-jk}}{(j-m)! \, (n-j)! \, (r-jk)! \, (k!)^j},$$

где суммирование распространяется на те $j \geqslant m$, для которых $j \leqslant n$ и $kj \leqslant r$.

23. Доказать последнее утверждение § 2 в случае $k=1$.

24. Используя (3.1), найти вероятность того, что в случае статистики Бозе — Эйнштейна ровно m ячеек останутся пустыми.

25. Проверить, что формула, полученная в задаче 24, тождественна формуле (11.14) гл. II.

26. Доказать (1,5) индукцией по N.

ГЛАВА V

УСЛОВНАЯ ВЕРОЯТНОСТЬ. СТОХАСТИЧЕСКАЯ НЕЗАВИСИМОСТЬ

С этой главы мы продолжим систематическое изложение основ теории вероятностей.

§ 1. УСЛОВНАЯ ВЕРОЯТНОСТЬ

Понятие условной вероятности является основным инструментом теории вероятностей, и удручает тот факт, что его крайняя простота отчасти затемняется чрезвычайно сложной терминологией. Следующие ниже рассуждения естественным путем ведут к формальному определению.

Предварительные примеры

Предположим, что в генеральной совокупности из N человек N_A страдают дальтонизмом и N_H являются женщинами. Обозначим через A и H соответственно события, состоящие в том, что случайно выбранный человек страдает дальтонизмом или что он является женщиной. Тогда (см. определение случайного выбора, гл. II, 2)

$$\mathbf{P}\{A\}=N_A/N, \quad \mathbf{P}\{H\}=N_H/N. \tag{1.1}$$

Мы можем теперь сосредоточить внимание на подмножестве, состоящем из женщин. Вероятность того, что лицо, случайно выбранное из этого подмножества, страдает дальтонизмом, равна N_{HA}/N_H, где N_{HA}—число женщин-дальтоников. Мы не приходим здесь к новому понятию, но нуждаемся в новом обозначении, указывающем на то, какое именно подмножество исследуется. Наиболее широко распространено обозначение $\mathbf{P}\{A\mid H\}$, которое может читаться как «вероятность события A (дальтонизм) при условии события H (случайно выбранное лицо является женщиной)». В символической записи:

$$\mathbf{P}\{A\mid H\}=N_{AH}/N_H=\mathbf{P}\{AH\}/\mathbf{P}\{H\}. \tag{1.2}$$

Разумеется, каждое подмножество может рассматриваться как отдельная генеральная совокупность; мы употребляем термин подмножество только для удобства, для того чтобы подсознательно помнить, что мы имеем более широкую генеральную совокупность.

Страховую компанию может интересовать частота повреждений, вызванных молнией и приносящих фиксированный ущерб (событие A). Вероятно, эта компания имеет различные категории застрахованных объектов: индустриальные, городские, сельские и т. д. Изучение отдельно ущерба, нанесенного индустриальным объектам, означает исследование события A лишь в связи с событием H — «ущерб причин индустриальному объекту». Формула (1.2) вновь применима очевидным образом. Заметим, однако, что для страховой компании, специализирующейся на индустриальных объектах, множество H совпадает со всем пространством элементарных событий и $\mathbf{P}\{A|H\}$ сводится к $\mathbf{P}\{A\}$.

Наконец, рассмотрим игрока «Север» при игре в бридж. После сдачи карт он знает свои карты и интересуется только распределением остальных 39 карт. Это побуждает рассматривать совокупность всех возможных распределений оставшихся 39 карт как новое пространство элементарных событий, однако совершенно очевидно, что удобнее рассматривать их в связи с данным распределением 13 карт игрока «Север» (событие H) и говорить о событии A (скажем, «Юг» имеет двух тузов) при условии события H. Формула (1.2) вновь применима. ▶

По аналогии с (1.2) введем формальное определение.

Определение. *Пусть H — событие положительной вероятности. Для любого события A будем писать*

$$\mathbf{P}\{A|H\}=\mathbf{P}\{AH\}/\mathbf{P}\{H\}. \tag{1.3}$$

Так определенная величина будет называться условной вероятностью A при условии H (или при заданном H).

Если все элементарные события имеют равные вероятности, то $\mathbf{P}\{A|H\}$ равно отношению N_{AH}/N_H числа элементарных событий, общих для A и H, к числу элементарных событий в H.

Условные вероятности остаются неопределенными, когда событие H имеет нулевую вероятность. Эта ситуация не существенна в дискретном случае, но важна в общей теории.

Хотя обозначение $\mathbf{P}\{A|H\}$ само по себе является практичным, его словесная расшифровка этим свойством не обладает и используется реже. Так, в нашем вводном примере мы рассматривали вероятность того, что женщина страдает дальтонизмом, не говоря «условная вероятность того, что случайно выбранное лицо страдает дальтонизмом, при условии, что оно является женщиной». Часто слова «при условии H» заменяют словами «если известно, что H осуществилось». Короче говоря, наши формулы и символы недвусмысленны, но словесные выражения часто неформальны и требуют четкого истолкования.

Для стилистической ясности вероятности на исходном пространстве элементарных событий будут иногда называться *безусловными* (*абсолютными*) *вероятностями* в отличие от условных вероятностей. Строго говоря, прилагательное «безусловная» излишне и может быть опущено.

Рассмотрение условных вероятностей различных событий при одном и том же данном событии H равносильно выбору H в качестве нового пространства элементарных событий с вероятностями, пропорциональными первоначальным. Коэффициент пропорциональности $\mathbf{P}\{H\}$ необходим для того, чтобы сделать вероятность нового пространства элементарных событий равной единице. Из сказанного следует, что *все основные теоремы о вероятностях остаются справедливыми для условных вероятностей, взятых относительно некоторого фиксированного события* H. Например, основное соотношение для вероятности осуществления одного из событий A или B или их обоих принимает вид

$$\mathbf{P}\{A \cup B \mid H\} = \mathbf{P}\{A \mid H\} + \mathbf{P}\{B \mid H\} - \mathbf{P}\{AB \mid H\}. \qquad (1.4)$$

Аналогично переносятся на случай условных вероятностей все теоремы гл. IV, касающиеся вероятностей осуществления m из N событий, однако они нам не понадобятся.

Формула (1.3) часто используется в следующем виде:

$$\mathbf{P}\{AH\} = \mathbf{P}\{A \mid H\} \cdot \mathbf{P}\{H\}. \qquad (1.5)$$

Это так называемая теорема об умножении вероятностей. Для того чтобы обобщить ее на случай трех событий A, B, C, положим сначала $H=BC$ и затем дважды применим (1.5); получим

$$\mathbf{P}\{ABC\} = \mathbf{P}\{A \mid BC\} \cdot \mathbf{P}\{B \mid C\} \cdot \mathbf{P}\{C\}. \qquad (1.6)$$

Аналогично проводится обобщение на случай любого числа событий.

В заключение приведем простую, но часто используемую формулу. Пусть $H_1, \ldots, H_n$ — множество взаимно исключающих событий, таких, что одно из них непременно произойдет (т. е. объединение $H_1, \ldots, H_n$ дает все пространство элементарных событий). Тогда любое событие A может осуществиться только одновременно с некоторым H_j, или символически

$$A = AH_1 \cup AH_2 \cup \ldots \cup AH_n. \qquad (1.7)$$

Поскольку AH_j попарно несовместны, их вероятности складываются. Применяя (1.5) к $H=H_j$ и суммируя, получаем

$$\mathbf{P}\{A\} = \sum \mathbf{P}\{A \mid H_j\} \cdot \mathbf{P}\{H_j\}. \qquad (1.8)$$

Эта формула полезна потому, что оценка условной вероятности $\mathbf{P}\{A \mid H_j\}$ часто оказывается проще, чем прямое нахождение $\mathbf{P}\{A\}$.

Примеры. а) *Выбор без возвращения.* Из генеральной совокупности, содержащей n элементов $1, 2, \ldots, n$, извлекается упорядоченная выборка. Пусть i и j — два различных элемента. Какова вероятность того, что вторым извлеченным элементом будет

j (событие A), если первым извлеченным элементом был i (событие H)? Ясно, что $\mathbf{P}\{AH\}=1/[n(n-1)]$ и $\mathbf{P}\{A|H\}=1/(n-1)$. Это выражает тот факт, что второй выбор относится к генеральной совокупности, состоящей из $n-1$ элементов, каждый из которых может быть выбран с одинаковой вероятностью. Действительно, большинство естественных *определений* случайного выбора таковы: «*Каковы бы ни были первые r выбранных элементов, на $(r+1)$-м шаге для каждого из оставшихся $n-r$ элементов вероятность быть выбранным равна* $1/(n-r)$». Это определение эквивалентно данному в гл. II, но мы не вводили его ранее, поскольку оно опирается на понятие условной вероятности.

б) Четыре шара последовательно размещаются в четырех ящиках, причем все 4^4 комбинаций равновероятны. Какова вероятность того, что одна из ячеек будет содержать ровно три шара (событие A), если известно, что первые два шара оказались в разных ячейках (событие H)? При данном H событие A может осуществляться двумя способами, и, таким образом, $\mathbf{P}\{A|H\}=2\cdot 4^{-2}=1/8$. (Легко непосредственно проверить, что события H и AH содержат $12\times\times 4^2$ и $12\cdot 2$ точек соответственно.)

в) *Распределение по признаку пола*. Рассмотрим семьи с двумя детьми. Обозначим буквами *м* и *д* соответственно мальчика и девочку, и на первом месте будем указывать старшего ребенка. Имеем четыре возможности: *мм*, *мд*, *дм*, *дд*. Это четыре элементарных события, и мы припишем каждому из них вероятность 1/4. Какова вероятность того, что оба ребенка мальчики (событие A), если известно, что в семье уже есть мальчик (событие H)? Событие AH означает *мм*, а H означает *мм* или *мд* или *дм*. Поэтому $\mathbf{P}\{A|H\}=1/3$; примерно для каждой третьей семьи, для которой выполнено событие H, следует ожидать также выполнения события A. Интересно, что большинство людей ожидают ответ 1/2. Это правильный ответ на другой вопрос, а именно: мальчик выбран случайным образом и оказалось, что он из семьи с двумя детьми; какова вероятность того, что второй ребенок мальчик? Различие может быть объяснено эмпирически. В нашей первой задаче мы имели дело с множеством семей, во второй — с множеством мужчин. В последней задаче каждая семья с двумя мальчиками представлена дважды, и этим объясняется различие в результатах.

г) *Неоднородные совокупности*. Предположим, что генеральная совокупность людей состоит из подмножеств или классов H_1, H_2, ... Это могут быть расы, возрастные группы, профессии и т. д. Пусть p_j — вероятность того, что случайно выбранный из генеральной совокупности индивидуум принадлежит H_j. Фраза «q_j есть вероятность того, что индивидуум из H_j является левшой» служит для более короткого выражения следующего факта: «q_j есть условная вероятность события A (является левшой) при условии, что индивидуум принадлежит H_j». Вероятность того, что случайным образом выбранный индивидуум является левшой, равна p_1q_1+

$+p_2q_2+p_3q_3+\ldots$, что представляет собой частный случай (1.8). Если известно, что индивидуум левша, то условная вероятность того, что он принадлежит подмножеству H_j, равна

$$\mathbf{P}\{H_j|A\}=p_jq_j/(p_1q_1+p_2q_2+\ldots). \tag{1.9}$$

2. ВЕРОЯТНОСТИ, ОПРЕДЕЛЯЕМЫЕ ЧЕРЕЗ УСЛОВНЫЕ ВЕРОЯТНОСТИ. УРНОВЫЕ МОДЕЛИ

В предыдущем параграфе нам были заданы вероятности элементарных событий, и мы нашли некоторые условные вероятности. В приложениях многие эксперименты описываются указанием определенных условных вероятностей (хотя прилагательное «условный» обычно опускается). Теоретически это означает, что вероятности на пространстве элементарных событий должны быть найдены из заданных условных вероятностей. Как уже было отмечено выше (пример 1, а)), выбор без возвращения лучше всего определить следующим образом: если известен результат первых r испытаний, то на $(r+1)$-м шаге любой оставшийся элемент может быть выбран с одинаковой вероятностью. Аналогично в примере 1, г) неоднородная популяция полностью описывается указанием абсолютных вероятностей p_j каждого класса и условных вероятностей q_j, характеризующих левшей в каждом классе. Еще несколько примеров разъяснят общую схему лучше, чем прямое описание.

Примеры. а) Возвратимся к примеру гл. I, 5, б), в котором в игре принимают участие три игрока a, b и c. В таблице (*) гл. I,5 описано пространство элементарных событий, но мы еще не определили на нем вероятности. Предположим теперь, что игра такова, что в каждой партии оба игрока имеют вероятность выигрыша 1/2. Это определение не содержит слов «условная вероятность», однако неявно ее использует. Оно означает, что если игрок a участвует в r-й партии (событие H), то вероятность его выигрыша в этой партии равна 1/2. Из (1.5) следует, что вероятность того, что a выиграет в первой и второй партиях, равна 1/4 (символически $\mathbf{P}\{aa\}=1/4$). Повторное применение (1.5) показывает, что $\mathbf{P}\{acc\}=1/8$, $\mathbf{P}\{acbb\}=1/16$ и т. д., т. е. элементарное событие в схеме (*), включающее r букв, имеет вероятность 2^{-r}. Такие же вероятности использованы в задаче 5 гл. I,8, однако теперь их описание более наглядно. (Продолжение см. в задаче 14.)

б) *Семьи.* Мы хотим истолковать следующее утверждение: «вероятность того, что в семье имеется ровно k детей, равна p_k (где $\sum p_k=1$). При любом фиксированном числе детей в семье все возможные их распределения по признаку пола равновероятны». Пусть буква *м* означает мальчика, а *д* — девочку. Пространство элементарных событий состоит из точек 0 (нет детей), *м*, *д*, *мм*, *мд*, *дм*, *дд*, *ммм*,... . Второе из наших предположений более строго формулируется так: если известно, что в семье ровно n

детей, то каждое из 2^n возможных распределений их по признаку пола имеет условную вероятность 2^{-n}. Вероятность условия равна p_n, и из (1.5) находим, что безусловная вероятность любой последовательности из n букв *м* и *д* равна $p_n \cdot 2^{-n}$.

Заметим, что это пример *генеральной совокупности, разделенной на классы:* семьи, имеющие j детей, образуют класс H_j. В качестве упражнения рассмотрим событие A «в семье есть мальчики, но нет девочек». Его вероятность, очевидно, равна $\mathbf{P}\{A\}=p_1 \cdot 2^{-1}+p_2 \times 2^{-2}+\ldots$, что является частным случаем (1.8). Условием H_j в этом случае является событие «в семье j детей». Зададим теперь вопрос: чему равна (условная) вероятность того, что в семье один ребенок, если известно, что в ней нет девочек?» Здесь условием является A. Пусть H — событие «имеется только один ребенок». Тогда AH означает «имеется один ребенок, не являющийся девочкой» и

$$\mathbf{P}\{H \mid A\}=\frac{\mathbf{P}\{AH\}}{\mathbf{P}\{A\}}=\frac{p_1 2^{-1}}{p_1 2^{-1}+p_2 2^{-2}+p_3 2^{-3}+\ldots}; \tag{2.1}$$

это частный случай (1.9).

в) *Урновые модели эффекта последействия.* Рассмотрим для определенности завод, на котором возможны аварии. Каждую аварию можно рассматривать как результат случайности: из урны, содержащей красные и черные шары, через постоянные интервалы времени случайно извлекается один шар, причем красный шар означает аварию. Если вероятность аварии постоянна во времени, то состав урны остается неизменным. Более реальной, однако, является ситуация, когда каждая авария приводит к эффекту последействия, в силу которого вероятность новой аварии либо увеличивается, либо уменьшается. Это соответствует урне, содержимое которой меняется по определенным правилам, зависящим от исходов последовательных извлечений. Легко придумать много таких правил, отвечающих различным ситуациям, однако мы ограничимся обсуждением одной следующей модели [1]).

Урновая модель. Урна содержит b черных и r красных шаров. Случайно извлекается шар. Он возвращается обратно, и, кроме того, добавляется с шаров одного с ним цвета и d шаров другого цвета. Производится новое случайное извлечение из урны (теперь содержащей $r+b+c+d$) шаров, и описанная процедура повторяется. Здесь c и d — произвольные целые числа. Они могут быть выбраны отрицательными, однако в этом случае процесс может окончиться

---

[1]) Идея использовать урновые схемы для описания эффекта последействия (заражения) была, по-видимому, впервые предложена Пойа. Его схема (впервые введенная в работе Eggenberger F., Polya G., Über die Statistik verketteter Vorgänge, Zeitschrift für Angewandte Mathematik und Mechanik, 3 (1923), 279—289) послужила прототипом для многих других моделей, которым посвящена обширная литература. Модель, описанная в тексте, и три ее частных варианта были предложены Фридманом (Friedman B., A simple urn model, Communications on Pure and Applied Mathematics, 2 (1949), 59—70).

после конечного числа извлечений из-за отсутствия шаров. В частности, выбирая $c=-1$ и $d=0$, имеем модель *случайного выбора без возвращения*, который оканчивается через $r+b$ шагов.

Чтобы придать нашему образному описанию точный математический смысл, заметим, что оно определяет условные вероятности, из которых могут быть вычислены некоторые основные вероятности. Точку пространства элементарных событий, соответствующую n извлечениям, можно представить как последовательность n букв $Ч$ и $К$. Событие «первым извлечен черный шар» (т. е. множество всех последовательностей, начинающихся с $Ч$) имеет вероятность $b/(b+r)$. Если первый шар черный, то вероятность (условная) того, что при втором извлечении будет черный шар, равна

$$(b+c)/(b+r+c+d).$$

Вероятность (безусловная) последовательности «черный, черный» (т. е. множество элементарных событий, начинающихся с $ЧЧ$) равна, следовательно, по (1.5)

$$[b/(b+r)][(b+c)/(b+r+c+d)]. \tag{2.2}$$

Вероятность последовательности «черный, черный, черный» равна (2.2), умноженному на $(b+2c)/(b+r+2c+2d)$ и т. д. Таким способом могут быть вычислены вероятности всех элементарных событий. Легко проверить по индукции, что сумма вероятностей всех элементарных событий равна единице, как и должно быть.

Точные выражения для вероятностей получить нелегко, за исключением следующего, самого важного и лучше всего изученного частного случая.

Урновая схема Пойа, характеризующаяся значениями $d=0$, $c>0$. В этом случае после каждого извлечения число шаров цвета выбранного шара увеличивается, тогда как число шаров другого цвета не меняется. Это приводит к тому, что извлечение шара определенного цвета увеличивает вероятность выбора шара того же цвета при следующем извлечении, и мы получаем грубую модель такого явления как *заражение*, при котором каждое заболевание увеличивает вероятность дальнейших заболеваний. Вероятность того, что при $n=n_1+n_2$ извлечениях в первых n_1 были черные шары, а в остальных n_2 красные, равна

$$\frac{b(b+c)(b+2c)\ldots(b+n_1c-c)\cdot r(r+c)\ldots(r+n_2c-c)}{(b+r)(b+r+c)(b+r+2c)\ldots(b+r+nc-c)}. \tag{2.3}$$

Рассмотрим далее любой другой порядок выбора n_1 черных и n_2 красных шаров. Вычисляя вероятность того, что n извлечений приведут к такому порядку цветов, мы получим те же множители, что в (2.3), но расположенные в другом порядке. Отсюда следует, что все возможные последовательности n_1 черных и n_2 красных шаров имеют одинаковые вероятности. Аналитическая простота (и, следовательно, легкость применения) урновой схемы Пойа следует главным образом из этого характеристического свойства. Чтобы найти вероятность $p_{n_1, n}$ того, что при n извлечениях было n_1 черных и n_2 красных шаров в произвольном порядке, мы должны умножить (2.3) на $\binom{n}{n_1}$, а именно на число возможных упорядочений. Использование обобщенных биномиальных коэффициентов позволяет записать эту вероятность в любой из следующих форм:

$$p_{n_1, n}=\frac{\binom{n_1-1+b/c}{n_1}\binom{n_2-1+r/c}{n_2}}{\binom{n-1+(b+r)/c}{n}}=\frac{\binom{-b/c}{n_1}\binom{-r/c}{n_2}}{\binom{-(b+r)/c}{n}}. \qquad (2.4)$$

(Обсуждение схемы Пойа продолжается в задачах 18—24; см. также задачи 9 и 10 гл. XVII, 10.)

В дополнение к схеме Пойа рассмотрим еще одну урновую модель, представляющую интерес, а именно модель Эренфестов.

Модель Эренфестов [1]) *теплообмена между двумя изолированными телами.* В первоначальном описании модели, используемом физиками, рассматриваются два сосуда I и II и k находящихся в них частиц. Случайным образом выбирается частица и переносится в другой сосуд. Эта процедура повторяется. Каково распределение частиц после n шагов? Для того чтобы свести эту задачу к урновым моделям, достаточно назвать частицы в сосуде I красными шарами, а в сосуде II — черными шарами. Тогда после каждого извлечения выбранный шар заменяется шаром другого цвета, т. е. мы имеем $c=-1$, $d=1$. Ясно, что в этом случае процесс может продолжаться сколь угодно долго (если не останется красных шаров, то автоматически извлекается черный и заменяется на красный). (Мы обсудим модель Эренфестов с другой точки зрения в примере гл. XV, 2,д).)

Частный случай $c=0$, $d>0$ был предложен Фридманом в качестве модели *службы безопасности.* Каждый раз, когда происходит авария (т. е. извлекается красный шар), служба безопасности усиливает контроль; если же аварии не происходит, то контроль ослабевает и вероятность аварии увеличивается.

г) *Урновые модели для неоднородных совокупностей. Ложное заражение.* Развивая предыдущий пример, предположим, что каждый человек подвержен несчастным случаям; произойдет или не произойдет этот несчастный случай, определяется случайным извлечением шара из урны. Однако на этот раз будем предполагать, что нет никакого эффекта последействия, так что содержимое урны остается неизменным в течение всего процесса. Зато вероятность несчастного случая или предрасположенность к нему меняется от человека к человеку или от профессии к профессии, и мы предполагаем, что каждому человеку (или каждой профессии) соответствует своя урна.

---

[1]) Ehrenfest P., Ehrenfest T., Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem, Physikalische Zeitschrift, 8(1907), 311—314. Математическую теорию см. Kac M., Random walk and the theory of Brownian motion, Amer. Math. Monthly, 54 (1947), 369—391.

Для того чтобы не усложнять без необходимости изложение, предположим, что имеется всего два типа людей (две профессии) и их числа во всей совокупности относятся как 1 : 5. Рассмотрим теперь урну I, содержащую r_1 красных и b_1 черных шаров, и урну II, содержащую r_2 красных и b_2 черных шаров. Эксперимент «случайным образом выбираем человека и наблюдаем, сколько несчастных случаев произошло с ним в течение n единиц времени» можно интерпретировать следующим образом: *бросается игральная кость; если выпадает единица, то выбираем урну* I, *в противном случае выбираем урну* II. *В каждом случае производится* n *случайных извлечений с возвращением из выбранной урны*. Наш эксперимент описывает ситуацию, возникающую при заключении договора между страховой компанией и новым клиентом.

Используя (1.8), находим, что вероятность извлечения первым красного шара равна

$$\mathbf{P}\{K\} = (1/6)\,[r_1/(b_1+r_1)] + (5/6)\,[r_2/(b_2+r_2)], \tag{2.5}$$

а вероятность последовательности «красный, красный» составляет

$$\mathbf{P}\{KK\} = (1/6)\,[r_1/(b_1+r_1)]^2 + (5/6)\,[r_2/(b_2+r_2)]^2. \tag{2.6}$$

Наша модель не приводит к новым математическим задачам, однако она имеет интересную особенность, приводящую в приложениях к неожиданным выводам. Предположим, что страховая компания регистрирует, что в течение первого года с новым клиентом произошел несчастный случай, и интересуется вероятностью несчастного случая с ним в следующем году. Иначе говоря, какова (условная) вероятность последовательности «красный, красный», если при первом испытании извлечен красный шар. Очевидно, она равна отношению $\mathbf{P}\{KK\}/\mathbf{P}\{K\}$ и *отличается* от $\mathbf{P}\{K\}$. Для иллюстрации предположим, что

$$r_1/(b_1+r_1)=0{,}6 \quad \text{и} \quad r_2/(b_2+r_2)=0{,}06.$$

Вероятность извлечения красного шара в любом испытании равна 0,15, но если при первом испытании был извлечен красный шар, вероятность того, что красный шар будет извлечен при следующем, равна 0,42. Заметим, что наша модель предполагает *отсутствие последействия* для всей совокупности, однако несчастный случай со случайно выбранным человеком увеличивает вероятность того, что с ним же произойдет второй несчастный случай. Здесь мы имеем просто эффект выбора: несчастный случай не оказывает действительного влияния на будущее, а только указывает, что выбранный человек имеет большую предрасположенность к несчастным случаям. Поэтому продолжающиеся наблюдения позволяют уточнить наши прогнозы, даже если в действительности прошлое вообще не влияет на будущее.

В статистической литературе стало привычкой использовать термин *заражение* вместо эффект последействия. *Кажущийся* эффект последействия выбора сначала неправильно интерпретировался как эффект действительного заражения, и поэтому статистики теперь говорят о заражении (или о вероятностных распределениях при заражении) в неясном и вводящем в заблуждение смысле. Возьмем, например, натуралиста, который ловит насекомых в поле. Если после периода неудач он находит насекомое, то вполне вероятно, что он находится вблизи выводка, и в этом случае можно ожидать дальнейших успехов. Иначе говоря, на практике каждый успех увеличивает вероятность дальнейших успехов, но отметим еще раз, что это только эффект увеличения информации, обеспечиваемый выбором. Никакого эффекта последействия здесь нет, и статистики вводят в заблуждение, когда говорят о заражении.

д) Следующий пример широко известен и поучителен, но несколько искусствен. Представим себе совокупность из $N+1$ урн, в каждой из которых содержится N шаров; урна с номером k содержит k красных и $N-k$ белых шаров ($k=0, 1, 2, \ldots, N$). Случайным образом выбирается урна и из нее n раз извлекаются шары, причем выбранный шар каждый раз возвращается обратно. Предположим, что все n выбранных шаров оказались красными (событие A). Найдем (условную) вероятность того, что следующим будет извлечен также красный шар (событие B). Если сначала была выбрана урна с номером k, то вероятность того, что из нее были последовательно извлечены n красных шаров, равна $(k/N)^n$. Следовательно, в силу (1.8)

$$\mathbf{P}\{A\}=(1^n+2^n+\ldots+N^n)/[N^n(N+1)]. \tag{2.7}$$

Событие AB означает, что при $n+1$ извлечениях были красные шары, и поэтому

$$\mathbf{P}\{AB\}=\mathbf{P}\{B\}=(1^{n+1}+2^{n+1}+\ldots+N^{n+1})/[N^{n+1}(N+1)]. \tag{2.8}$$

Искомая вероятность равна $\mathbf{P}\{B|A\}=\mathbf{P}\{B\}/\mathbf{P}\{A\}$.

При больших значениях N числитель в (2.7) относительно мало отличается от площади между осью x и графиком x^n в интервале $(0, N)$. Тогда приближенно имеем

$$\mathbf{P}\{A\}\approx\frac{1}{N^n(N+1)}\int_0^N x^n\,dx=\frac{N}{N+1}\cdot\frac{1}{n+1}\approx\frac{1}{n+1}. \tag{2.9}$$

Аналогичные вычисления, примененные к (2.8), показывают, что для больших N приблизительно

$$\mathbf{P}\{B|A\}\approx(n+1)/(n+2). \tag{2.10}$$

Этот результат можно грубо интерпретировать следующим образом: если все возможные варианты содержимого урны равновероятны и при n испытаниях появлялись красные шары, то вероятность появления красного шара при следующем испытании равна $(n+1)/(n+2)$. Это так называемый закон следования Лапласа (1812).

До развития современной теории понятие равновероятности часто использовалось как синоним для «отсутствия предварительных знаний». Сам Лаплас иллюстрировал применение (2.10) вычислением вероятности того, что Солнце взойдет на следующий день, если оно всходило ежедневно в течение 5000 лет или $n = 1\,826\,213$ дней. Говорят, что Лаплас был готов поставить 1 826 214 против 1 за то, что Солнце не изменит своего поведения; в наше время следовало бы увеличить ставку, так как регулярное движение Солнца наблюдалось в течение еще одного столетия. Чтобы понять, что имел в виду Лаплас и истолковать его выводы, потребовалось бы целое историческое исследование. Его последователи, однако, использовали аналогичные рассуждения повсеместно и рекомендовали физикам и инженерам применять такие методы в случаях, в которых полученные формулы не имеют никакого реального смысла. Мы должны отвергнуть этот метод, даже если допустим, что наша вселенная случайно выбрана из некоторого множества вселенных, в котором все мыслимые возможности равновероятны. Действительно, он претендует на оценку шансов восхода Солнца завтра на основе предположения о том, что оно всходило в прошлом. Однако предполагаемый восход Солнца 5 февраля 3123 г. до н. э. ничуть не более достоверен, чем восход Солнца завтра. У нас одинаковые основания верить в оба этих события. ▶

Замечание о формуле Бейеса. В (1.9) и (2.2) мы нашли некоторые условные вероятности, исходя непосредственно из определения. Начинающему рекомендуется поступать так всегда и не запоминать формулу (2.12), которую мы сейчас выведем. Она воспроизводит в общем виде то, что мы делали в частных случаях, и является просто другой формой записи (1.3). В примерах мы имели группу взаимно исключающих и исчерпывающих все пространство событий $H_1, H_2, \ldots$, т. е. таких, что каждое элементарное событие принадлежит одному и только одному из H_j, и нас интересовала следующая вероятность:

$$\mathbf{P}\{H_k \mid A\} = \mathbf{P}\{AH_k\}/\mathbf{P}\{A\}. \tag{2.11}$$

Если в (2.11) подставить (1.5) и (1.8), то эта формула принимает вид

$$\mathbf{P}\{H_k \mid A\} = \mathbf{P}\{A \mid H_k\}\,\mathbf{P}\{H_k\} \Big/ \sum_j \mathbf{P}\{A \mid H_j\}\,\mathbf{P}\{H_j\}. \tag{2.12}$$

Если события H_k назвать гипотезами, то (2.12) будет «формулой Бейеса для вероятностей гипотез». С математической точки зрения (2.12) представляет собой специальный вид записи (1.3) и ничего больше. Эта формула полезна во многих статистических приложениях, подобных описанным в примерах б) и г), где мы ее и использовали. К сожалению, формула Бейеса была дискредитирована метафизическими приложениями вроде описанного в примере д). На практике подобные рассуждения могут привести к опасным последствиям. Инженер по контролю качества имеет дело с одной конкретной машиной, а не с бесконечным множеством машин, из которого случайно выбрана одна. Между тем ему советуют пользоваться формулой Бейеса на том основании, что она логически приемлема и соответствует нашей манере мышления. Такие доводы использовал Платон для доказательства существования Атлантиды, а некоторые философы используют для доказательства нелепости механики Ньютона. В нашем случае эти доводы не учитывают того обстоятельства, что инженер желает достичь определенного успеха и что он поступит лучше, оценивая и доводя до минимума источники различного рода ошибок в своих предвидениях и догадках. Современные методы статистических проверок и оценок менее наглядны, но более реалистичны. Их можно не только оправдать рассуждениями, но и применять.

§ 3. СТОХАСТИЧЕСКАЯ НЕЗАВИСИМОСТЬ

В примерах, приведенных выше, условная вероятность $\mathbf{P}\{A|H\}$, как правило, не равнялась безусловной вероятности $\mathbf{P}\{A\}$. Грубо говоря, информация о том, произошло или нет событие H, меняет наше отношение к возможности наступления события A. Только в том случае, когда $\mathbf{P}\{A|H\}=\mathbf{P}\{A\}$, эта информация не позволяет сделать никаких выводов об A. В этом случае мы будем говорить, что A стохастически не зависит от H. Из формулы (1.5) следует, что условие $\mathbf{P}\{A|H\}=\mathbf{P}\{A\}$ может быть записано в виде

$$\mathbf{P}\{AH\}=\mathbf{P}\{A\}\cdot\mathbf{P}\{H\}. \tag{3.1}$$

Это равенство симметрично относительно A и H и показывает, что если A стохастически не зависит от H, то и H стохастически не зависит от A. Таким образом, лучше отправляться от следующего симметричного определения.

Определение 1. *Два события A и H называются стохастически независимыми (или, короче, независимыми), если имеет место равенство* (3.1).

Это определение распространяется и на случай $\mathbf{P}\{H\}=0$, в котором $\mathbf{P}\{A|H\}$ не определена. Термин *статистическая независимость* является синонимом термина *стохастическая независимость.*

На практике обычно бывает ясно, что некоторые события должны быть независимыми, иначе вероятностная модель будет абсурдной. Однако, как покажут приводимые ниже примеры, существуют ситуации, в которых стохастическая независимость может быть установлена только путем вычислений.

Примеры. а) Из колоды игральных карт случайным образом извлекается одна карта. Из соображений симметрии мы ожидаем, что события «пика» и «туз» независимы. И действительно, их вероятности равны 1/4 и 1/13, а вероятность их одновременного осуществления равна 1/52.

б) Бросаются две правильные игральные кости. События «единица на первой кости» и «четное число очков на второй» независимы, так как вероятность их одновременного осуществления $3/36=1/12$ равна произведению их вероятностей, а именно 1/6 и 1/2.

в) При случайной перестановке четырех букв (a, b, c, d) события «a предшествует b» и «c предшествует d» независимы. Это интуитивно ясно и легко проверяется.

г) *Распределение по признаку пола.* Вернемся к примеру 1,в), но теперь рассмотрим семьи с тремя детьми. Предположим, что каждая из восьми возможностей *ммм*, *ммд*, . . . , *ддд* имеет вероятность 1/8. Пусть H — событие «в семье имеются дети обоих полов» и A — событие «имеется не более одной девочки». Тогда $\mathbf{P}\{H\}=6/8$ и $\mathbf{P}\{A\}=4/8$. Одновременное осуществление A и H означает одну из возможностей *ммд*, *мдм*, *дмм*, и, таким образом, $\mathbf{P}\{AH\}=3/8=\mathbf{P}\{A\}\cdot\mathbf{P}\{H\}$. Таким образом, для семей с тремя детьми эти два события независимы. Отметим, что это не так в случае семей с двумя или четырьмя детьми. Это показывает, что не всегда очевидно, имеем мы независимость или нет. ►

Если осуществилось событие H, то противоположное событие H' не осуществилось, и наоборот. Из стохастической независимости следует, что из осуществления H нельзя сделать никаких выводов об A; поэтому стохастическая независимость A и H будет означать то же самое, что независимость A и H' (и в силу симметрии событий A' и H, а также A' и H'). Это утверждение легко проверить, используя соотношение $\mathbf{P}\{H'\}=1-\mathbf{P}\{H\}$. Действительно, если имеет место (3.1), то (поскольку $AH'=A-AH$)

$$\mathbf{P}\{AH'\}=\mathbf{P}\{A\}-\mathbf{P}\{AH\}=\mathbf{P}\{A\}-\mathbf{P}\{A\}\cdot\mathbf{P}\{H\}=\mathbf{P}\{A\}\cdot\mathbf{P}\{H'\}, \tag{3.2}$$

как и ожидалось.

Предположим теперь, что три события A, B и попарно независимы, т. е.

$$\begin{aligned}\mathbf{P}\{AB\}&=\mathbf{P}\{A\}\cdot\mathbf{P}\{B\},\\ \mathbf{P}\{AC\}&=\mathbf{P}\{A\}\cdot\mathbf{P}\{C\},\\ \mathbf{P}\{BC\}&=\mathbf{P}\{B\}\cdot\mathbf{P}\{C\}.\end{aligned} \tag{3.3}$$

Можно подумать, что из этих трех соотношений будет следовать равенство

$$\mathbf{P}\{ABC\}=\mathbf{P}\{A\}\,\mathbf{P}\{B\}\,\mathbf{P}\{C\},$$

иначе говоря, что из попарной независимости трех событий следует независимость двух событий AB и C. Это почти всегда так, однако в принципе возможно, что (3.3) имеет место, но в то же время

$$\mathbf{P}\{ABC\}=0.$$

Фактически такие явления настолько редки, что их возможность не была замечена до тех пор, пока С. Бернштейн не построил соответствующий пример. Этот пример является искусственным, однако его существование побуждает искать естественные примеры.

Пример. д) Рассмотрим шесть перестановок букв a, b, c, а также три тройки (a, a, a), (b, b, b) и (c, c, c). Возьмем эти девять троек в качестве точек пространства элементарных событий и припишем каждой вероятность 1/9. Обозначим через A_k событие, состоящее в том, что на k-м месте находится буква a. Очевидно, каждое из этих трех событий имеет вероятность 1/3, в то время как

$$\mathbf{P}\{A_1A_2\}=\mathbf{P}\{A_1A_3\}=\mathbf{P}\{A_2A_3\}=1/9.$$

Три события являются, таким образом, *попарно независимыми*, но *не* взаимно независимыми, потому что $\mathbf{P}\{A_1A_2A_3\}=1/9$. (Осуществление A_1 и A_2 влечет осуществление A_3, и, значит, A_3 не является независимым от A_1A_2.)

Мы получим дополнительные примеры, рассматривая также события B_k и C_k, состоящие соответственно в том, что на k-м месте находятся буквы b и c. Всего имеем теперь девять событий, вероятность каждого из которых равна 1/3. Очевидно, $\mathbf{P}\{A_1B_2\}=1/9$, и вообще *каждые два события с разными индексами независимы*. С другой стороны, буквы, стоящие на первых двух местах, однозначно определяют букву, стоящую на третьем месте, и, таким образом, C_3 не является независимым от каждого из девяти событий $A_1A_2, \ldots, C_1C_2$, относящихся к буквам, стоящим на первых двух местах [1]). Мы вернемся к этому примеру в конце § 1 гл. IX. Дополнительный пример приводится в задаче 26. ►

Желательно сохранить термин стохастическая независимость для случая, когда имеет место не только (3.3), но также

$$\mathbf{P}\{ABC\}=\mathbf{P}\{A\}\mathbf{P}\{B\}\mathbf{P}\{C\}. \tag{3.4}$$

Это равенство обеспечивает независимость событий A и BC, а также B и AC, и C и AB. Кроме того, можно доказать, что $A\cup B$ и C также независимы. Действительно, из основного соотношения (7.4) гл. I имеем

$$\mathbf{P}\{(A\cup B)C\}=\mathbf{P}\{AC\}+\mathbf{P}\{BC\}-\mathbf{P}\{ABC\}. \tag{3.5}$$

Применяя вновь (3.3) и (3.4) к правой части, можем вынести множитель $\mathbf{P}\{C\}$. Второй множитель есть $\mathbf{P}\{A\}+\mathbf{P}\{B\}-\mathbf{P}\{AB\}=\mathbf{P}\{A\cup B\}$ и, таким образом,

$$\mathbf{P}\{(A\cup B)C\}=\mathbf{P}\{(A\cup B)\}\mathbf{P}\{C\}. \tag{3.6}$$

Поэтому представляется правдоподобным предположение, что условий (3.3) и (3.4) в совокупности достаточно для того, чтобы избежать затруднений: каждое событие, которое можно выразить через A и B, не будет зависеть от C.

Определение 2. *События $A_1, A_2, \ldots, A_n$ называются взаимно независимыми, если для любых комбинаций* $1\leqslant i<j<k<\ldots\leqslant n$

---

[1]) Эти рассуждения обобщаются на совокупности из r символов, $r>3$. Пространство элементарных событий в этом случае содержит $r!+r$ точек, а именно $r!$ перестановок символов $a_1, \ldots, a_r$ и r последовательностей вида $(a_j, a_j, \ldots, a_j)$. Каждой перестановке припишем вероятность $1/[r^2(r-2)!]$, а каждой последовательности вида $(a_j, a_j, \ldots, a_j)$ вероятность $1/r^2$. Если событие A_k означает, что a_1 находится на k-м месте, то A_k попарно независимы, но никакие три из этих событий не являются взаимно независимыми.

справедливы правила умножения

$$
\begin{gathered}
\mathbf{P}\{A_iA_j\}=\mathbf{P}\{A_i\}\mathbf{P}\{A_j\},\\
\mathbf{P}\{A_iA_jA_k\}=\mathbf{P}\{A_i\}\mathbf{P}\{A_j\}\mathbf{P}\{A_k\},\\
\cdots\cdots\cdots\cdots\cdots\cdots\\
\cdots\cdots\cdots\cdots\cdots\cdots\\
\mathbf{P}\{A_1A_2\ldots A_n\}=\mathbf{P}\{A_1\}\mathbf{P}\{A_2\}\ldots\mathbf{P}\{A_n\}.
\end{gathered}
\tag{3.7}
$$

В первой строке стоят $\binom{n}{2}$ равенств, во второй — $\binom{n}{3}$ и т.д. Следовательно, имеем

$$\binom{n}{2}+\binom{n}{3}+\ldots+\binom{n}{n}=(1+1)^n-\binom{n}{1}-\binom{n}{0}=2^n-n-1$$

условий, которые должны быть выполнены. С другой стороны, $\binom{n}{2}$ условий в первой строке достаточно для того, чтобы обеспечить *попарную независимость*. Вся система (3.7) выглядит как сложное множество условий, однако скоро станет ясно, что ее справедливость обычно очевидна и не нуждается в проверке. Применяя индукцию, легко увидеть [начиная с $n=2$ и (3.2)], что *в определении 2 систему* (3.7) *можно заменить системой из 2^n равенств, получаемой из последнего равенства* (3.7) *заменой произвольного числа событий A_j их дополнениями A_j'*.

§ 4. ПРОИЗВЕДЕНИЕ ПРОСТРАНСТВ. НЕЗАВИСИМЫЕ ИСПЫТАНИЯ

Мы теперь окончательно готовы к тому, чтобы ввести математический аналог эмпирических процедур, обычно называемых непрерывным экспериментированием, повторяющимся наблюдением, объединением двух выборок, комбинированием двух экспериментов и рассмотрением их как частей единого целого и т. д. Понятие независимых испытаний соответствует интуитивному понятию «экспериментов, повторяемых в одинаковых условиях». Это понятие является основным для теории вероятностей и придает бóльшую строгость примерам, рассмотренным выше.

Сначала нам понадобится понятие, не являющееся специфичным для теории вероятностей. *Прямым произведением* двух множеств A и B называется множество всех упорядоченных пар (a, b) из их элементов. Мы будем обозначать [1]) его через (A, B). Определение тривиально распространяется на тройки (A, B, C), четверки (A, B, C, D) и даже на бесконечные последовательности.

[1]) Другим общеупотребительным обозначением является $A\times B$. Термины «прямое произведение» и «декартово произведение» — синонимы.

Понятие прямого произведения настолько естественно, что мы его уже неявно использовали. Например, мыслимый эксперимент, состоящий в трехкратном бросании монеты, описывается пространством элементарных событий, содержащим восемь точек, а именно троек, которые могут быть образованы из двух букв Γ и P. Это позволяет говорить, что пространство элементарных событий является прямым произведением трех пространств, каждое из которых состоит из двух точек (элементов) Γ и P. Вообще, когда мы говорим о двух последовательных испытаниях, мы имеем в виду пространство $\mathfrak{S}$, точки которого представляют собой пары возможных исходов, и поэтому $\mathfrak{S}$ является прямым произведением двух пространств элементарных событий, соответствующих отдельным испытаниям. Два данных мысленных эксперимента с пространствами элементарных событий $\mathfrak{A}$ и $\mathfrak{B}$ можно рассматривать одновременно или последовательно. Это позволяет рассматривать пары возможных исходов, т. е. ввести прямое произведение $(\mathfrak{A}, \mathfrak{B})$ в качестве нового пространства элементарных событий. Возникает вопрос: как должны быть определены вероятности на новом пространстве элементарных событий? Ответ зависит от обстоятельств, но до рассмотрения этого вопроса мы обратимся к двум примерам, которые помогут понять основные идеи и объяснят применяемую терминологию.

Примеры. а) *Декартовы пространства.* Когда точки плоскости представлены парами действительных чисел (x, y), плоскость становится прямым произведением двух осей. (Тот факт, что геометрию на плоскости можно изучать без использования координат, показывает, что одно и то же пространство можно рассматривать с различных точек зрения.) Трехмерное пространство с точками (x, y, z) можно рассматривать либо как прямое произведение трех осей, либо как прямое произведение плоскости x, y и оси z.

На плоскости множество точек, удовлетворяющее двум условиям $0<x<1$ и $0<y<1$, является прямым произведением двух единичных интервалов. Заметим, однако, что такое описание невозможно для произвольных множеств, таких, как треугольники и эллипсы. Наконец, заметим, что в пространстве x, y, z множество, определенное теми же самыми неравенствами, является бесконечным цилиндром с квадратным основанием. Вообще в трехмерном пространстве каждое множество, в определение которого входят только координаты x и y, можно рассматривать как цилиндр с образующими, параллельными оси z.

б) *Алфавиты и слова.* Пусть A состоит из 26 букв английского алфавита. Тогда прямое произведение (A, A, A) является множеством всевозможных троек букв или, как мы будем говорить, всех трехбуквенных «слов». Так принято в теории связи и в теории кодирования, но неестественно рассматривать слова фиксированной длины, Действительно, сообщение произвольной длины может рассматриваться как «слово», если к алфавиту добавить новый символ для разделения (пробел). Поэтому нет необходимости вводить какие-либо ограничения на длину слов: любое конечное сообщение может рассматриваться как начало потенциально неограниченного сообщения, так же как написанное слово потенциально является первым в некоторых сериях. Попутно отметим, что в теории связи используются произвольные коды и под ее влиянием стало общеупотребительным называть произвольные символы буквами некоторого алфавита. В этом смысле исходы n повторных испытаний называют «сообщением» или «словом» длины n. ▶

Если $\mathfrak{S}$ — произвольное пространство элементарных событий с точками $E_1, E_2, \ldots$, то n-кратное прямое произведение $(\mathfrak{S}, \mathfrak{S}, \ldots \ldots, \mathfrak{S})$ пространства $\mathfrak{S}$ самого на себя называют пространством элементарных событий для последовательности n испытаний, соответствующих $\mathfrak{S}$. Удобно описывать его точки в общем виде символами, такими, как $(x_1, \ldots, x_n)$, где каждое x_i означает некоторую точку из $\mathfrak{S}$. По аналогии с примером а) x_i обычно называют *координатами*. Термины множество и событие могут использоваться как синонимы. *Событие, которое зависит только от исхода первых двух испытаний*, обычно называют множеством, зависящим только от первых двух координат[1]).

Как уже было указано выше, все эти понятия и обозначения переносятся на случай бесконечных последовательностей. Формально это не вызывает трудностей; например, десятичная последовательность 3,1415... представляет число π как точку в пространстве, являющемся бесконечным произведением, только вместо n-я координата говорят n-я десятичная цифра. *Бесконечные произведения пространств очень естественны для теории вероятностей.* Нежелательно фиксировать число бросаний монеты или продолжительность случайного блуждания. Теория становится более гибкой и простой, если мы допустим потенциально неограниченные последовательности испытаний и сосредоточим внимание на событиях, зависящих только от первых нескольких испытаний. Этот более простой и удобный подход, к сожалению, требует применения аппарата теории меры. Задачей настоящего тома является ознакомление с основными идеями теории вероятностей, не осложненное техническими трудностями. По этой причине мы ограничиваемся дискретными пространствами элементарных событий и должны удовлетвориться изучением конечного числа испытаний. Это означает, что, расплачиваясь за простоту, мы вынуждены иметь дело с неопределенными или изменяющимися пространствами элементарных собы-

---

1) То есть если точка $(x_1, x_2, \ldots)$ принадлежит этому множеству, то ему принадлежат все точки $(x_1', x_2', \ldots)$, у которых $x_1' = x_1$, $x_2' = x_2$. По аналогии с примером а) множества, зависящие только от некоторых координат (в любом числе), называют *цилиндрическими*.

тий. Такой вариант теоретически неудовлетворителен, но в известной мере полезен практически.

Обратимся теперь к вопросу о задании вероятностей на произведении пространств. Различные урновые модели § 2 могут быть описаны при помощи повторных испытаний, и мы видели, что вероятности в различных схемах можно определить посредством условных вероятностей. Могут быть рассмотрены различные виды зависимости между последовательными испытаниями, но наиболее важно понятие независимых испытаний или, более общо, независимых экспериментов.

Для определенности рассмотрим два пространства элементарных событий $\mathfrak{A}$ и $\mathfrak{B}$ с точками $\alpha_1, \alpha_2, \ldots$ и $\beta_1, \beta_2, \ldots$, имеющими вероятности $p_1, p_2, \ldots$, и $q_1, q_2, \ldots$ соответственно. Мы интерпретируем произведение пространств $(\mathfrak{A}, \mathfrak{B})$ как пространство элементарных событий, описывающих последовательность двух экспериментов, соответствующих $\mathfrak{A}$ и $\mathfrak{B}$. Говоря, что эти два эксперимента независимы, мы имеем в виду, что два события «первый исход есть α_i» и «второй исход есть β_k» стохастически независимы. Однако дело обстоит так только в том случае, когда вероятности на $(\mathfrak{A}, \mathfrak{B})$ определены по правилу

$$\mathbf{P}\{(\alpha_i, \beta_k)\} = p_i q_k. \tag{4.1}$$

Такое задание вероятностей правомерно[1]), потому что в сумме они дают единицу. Действительно, суммирование по всем элементарным событиям приводит к двойной сумме $\sum\sum p_i q_k$, которая является произведением сумм $\sum p_i$ и $\sum q_k$.

Условимся теперь, что *слова «два независимых эксперимента» означают прямое произведение двух пространств элементарных событий с вероятностями, определенными по правилу умножения* (4.1). *Это соглашение распространяется на понятие n последовательных независимых экспериментов.*

Мы говорим о повторных независимых испытаниях, если пространства элементарных событий, соответствующих каждому испытанию (и вероятности на них), идентичны.

Это соглашение дает нам право, например, для краткости говорить «n независимых бросаний монеты» вместо «пространство элементарных событий, состоящее из 2^n точек, каждой из которых приписывается вероятность 2^{-n}».

Заслуживает упоминания следующее интуитивно очевидное свойство независимых экспериментов. Пусть A — событие из $\mathfrak{A}$, содержащее элементарные события $\alpha_{s_1}, \alpha_{s_2}, \ldots$, и пусть B — событие из $\mathfrak{B}$, содержащее элементарные события $\beta_{t_1}, \beta_{t_2}, \ldots$. Тогда (A, B) есть событие из $(\mathfrak{A}, \mathfrak{B})$, которое состоит из всех пар $(\alpha_{s_j}, \beta_{t_k})$,

[1]) Аналогично определяемые меры встречаются и вне теории вероятностей и называются *произведением мер*.

и, очевидно,

$$\mathbf{P}\{(A,\ B)\} = \sum\sum p_{s_i} q_{t_k} = \left(\sum p_{s_i}\right)\left(\sum q_{t_k}\right) = \mathbf{P}\{A\}\,\mathbf{P}\{B\}. \quad (4.2)$$

Таким образом, правило умножения распространяется на произвольные события в пространствах-сомножителях. Эти рассуждения применимы к n независимым экспериментам и показывают, что *если система из n событий $A_1, \ldots, A_n$ такова, что A_k зависит только от k-го эксперимента, то события $A_1, \ldots, A_n$ взаимно независимы.*

Теория независимых экспериментов является самой простой аналитически и наиболее изученной частью теории вероятностей. Поэтому желательно, если это возможно, интерпретировать сложные эксперименты как последовательность более простых независимых экспериментов. Следующие примеры показывают случаи, когда такая процедура возможна.

Примеры. в) *Перестановки.* Мы рассматривали $n!$ перестановок элементов $a_1, a_2, \ldots, a_n$ как элементарные события и приписывали каждому из них вероятность $1/n!$. *То же самое пространство элементарных событий* можно рассматривать как представляющее $n-1$ последовательных независимых экспериментов следующим образом. Запишем сначала a_1. Первый эксперимент состоит в постановке a_2 перед a_1 или после него. После того как это сделано, имеется три места для a_3, и второй эксперимент состоит в его выборе, устанавливающем относительный порядок a_1, a_2, и a_3. Вообще, когда $a_1, \ldots, a_k$ расположены в некотором порядке, производим эксперимент с номером k, который состоит в выборе одного из $k+1$ мест для a_{k+1}. Иначе говоря, мы имеем последовательность из $n-1$ экспериментов, k-й из которых имеет $k+1$ различных исходов (элементарных событий), имеющих вероятности $1/(k+1)$. Эксперименты независимы, т. е. вероятности перемножаются. Каждая перестановка из n элементов имеет вероятность $(1/2)\cdot(1/3)\ldots(1/n)$, что совпадает с первоначальным определением.

г) *Выбор без возвращения.* Пусть $(a_1, \ldots, a_n)$ — генеральная совокупность. При выборе без возвращения при каждом испытании удаляется один элемент. После k испытаний остается $n-k$ элементов и следующий выбор можно описать заданием номера ν места, занимаемого выбранным элементом ($\nu = 1, 2, \ldots, n-k$). Таким образом, выборка объема r без возвращения превращается в последовательность из r экспериментов, первый из которых имеет n возможных исходов, второй — $n-1$, третий — $n-2$ и т. д. Мы припишем равные вероятности всем исходам отдельного эксперимента и постулируем независимость r экспериментов. Это равносильно приписыванию вероятности $1/(n)_r$ каждому элементарному событию, что соответствует определению случайного выбора. Заметим, что при $n=100$ и $r=3$ выборка (a_{13}, a_{40}, a_{81}) означает выбор чисел 13, 39, 79 соответственно: в третьем эксперименте был выбран семьдесят девятый элемент из сокращенной совокупности из $n-2$ элементов. (При первоначальной нумерации исходы третьего эксперимента зависели бы от исходов первых двух.) Мы видим, что понятие повторных независимых испытаний позволяет изучать выбор как последовательность независимых операций. ▶

§ 5*). ПРИЛОЖЕНИЯ К ГЕНЕТИКЕ

Теория наследственности, ведущая начало от Г. Менделя (1822—1884), дает поучительную иллюстрацию применимости простых вероятностных моделей. Мы ограничимся рассмотрением наиболее

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

простых задач. Биологические явления будут по необходимости описываться в весьма упрощенном виде, и мы сосредоточим внимание на таких фактах, которые относятся к построению математической модели.

Наследование признаков зависит от специальных носителей, называемых *генами*. Все клетки тела, исключая половые клетки (гаметы), несут один и тот же набор генов. Надо отметить, что каждый ген в клетке представлен парой. Читатель может представить их себе как совокупность огромного числа шариков на коротких кусках нитей-хромосом. Хромосомы также входят в каждую клетку попарно, и парные гены занимают одинаковое положение в парных хромосомах. В простейшем случае каждый ген отдельной пары может находиться в одной из двух форм (аллелей) A или a. Тогда можно образовать три различные пары, и в соответствии с этим организм имеет один из трех *генотипов* AA, Aa и aa (между парами Aa и aA различия нет). Например, горох несет пару генов, таких, что ген A обусловливает красную окраску цветка, а a — белую окраску. Три генотипа соответствуют в этом случае красному, розовому и белому цветкам.

Каждая пара генов определяет один наследственный признак, но большинство наблюдаемых свойств организма зависит от нескольких факторов. Для некоторых характеристик (таких, как цвет глаз или свойство быть левшой) преобладающим является влияние одной пары генов, и наблюдаемые в этом случае эффекты подчиняются законам Менделя. Другие характеристики, например рост, определяются совместным действием очень большого числа генов (см. пример гл. X, 5,в)). Здесь мы будем рассматривать генотипы и наследование лишь для одной пары генов; в этом случае возможны три генотипа AA, Aa и aa. Часто имеет место случай, когда для каждой пары генов существует N различных форм $A_1, \ldots, A_N$ и, следовательно, имеется $N(N+1)/2$ генотипов A_1A_1, $A_1A_2, \ldots, A_NA_N$. С очевидными изменениями вся теория применима и в этом случае (см. задачу 27). Последующие вычисления применимы также к случаю, когда ген A является *доминантным*, а ген a — *рецессивным*. Это означает, что особи Aa имеют те же наблюдаемые свойства, что и AA, так что влияние гена a проявляется лишь у особей aa. В природе встречаются все степени частичного доминирования. Типичными примерами частично рецессивных свойств является голубой цвет глаз, свойство быть левшой и т. д.

Половые клетки (гаметы) образуются в результате расщепления и содержат только по *одному* гену каждой пары. Поэтому организмы чистых генотипов AA или aa (или гомозиготы) производят гаметы только одного вида, в то время как организмы Aa (гибриды или гетерозиготы) производят в равном количестве гаметы A и a. Новый организм развивается из двух родительских гамет, от которых он и получает свои гены. Поэтому в каждой паре генов один ген получен от матери, а другой — от отца; это позволяет установить, какому из предков в каждом поколении, даже очень отдаленном, принадлежит любой ген.

Генотип потомка зависит от случайного процесса. При любых обстоятельствах каждый родительский ген может передаться с вероятностью 1/2, и последовательные испытания независимы. Иначе говоря, мы представляем генотипы n потомков как результат n независимых испытаний, каждое из которых эквивалентно бросанию пары монет. Например, при скрещивании особей $Aa \times Aa$ могут появиться генотипы AA, Aa и aa с вероятностями, равными соответственно 1/4, 1/2, 1/4. Скрещивание $AA \times aa$ может привести к появлению лишь особей Aa и т. п.

Рассматривая популяцию в целом, мы представляем образование родительской пары как результат второго случайного процесса. Будем исследовать лишь так называемое *случайное скрещивание*, которое определяется следующим условием: если случайно выбрать r особей в первом поколении, то их родители образуют случайную выборку с возвращением объема r из множества всех возможных родительских пар. Иначе говоря, каждого потомка можно рассматривать как результат случайного выбора родителей, и все выборы взаимно независимы. Случайное скрещивание является идеализированной моделью условий, которые преобладают во многих встречающихся в природе популяциях и в полевых экспериментах. Однако, если один угол поля засеян красным горохом, а другой белым, то пары одного типа будут скрещиваться чаще, чем это допускается гипотезой о случайности. Выбор с предпочтением (например, если блондины предпочитают блондинок) также нарушает условия случайного скрещивания. Самоопыляющиеся растения и искусственное оплодотворение представляют собой полную противоположность случайному скрещиванию. Некоторые системы со специальными законами скрещивания будут в дальнейшем анализироваться математически, но все же основное внимание будет сосредоточено на модели случайного скрещивания.

Генотип потомка есть результат четырех независимых испытаний. Генотипы двух родителей можно выбрать $3 \cdot 3$ способами, их гены — $2 \cdot 2$ способами. К счастью, однако, можно объединить эти два выбора и описать процесс как двойной выбор следующим образом: отцовский и материнский гены выбираются случайно и независимо один от другого из множества всех генов, которые несут мужские или женские особи родительской популяции.

Допустим, что три генотипа AA, Aa и aa встречаются среди мужских и женских особей в одной и той же пропорции $u:2v:w$. Предположим, что $u+2v+w=1$, и назовем u, $2v$ и w *частотами генотипов*. Положим

$$p=u+v, \quad q=v+w. \tag{5.1}$$

Ясно, что число генов A относится к числу генов a как $p:q$, и, так как $p+q=1$, мы будем называть p и q *частотами генов* A и a.

При каждом из двух случайных выборов ген A появляется с вероятностью p, и в силу независимости этих выборов вероятность того, что потомок имеет генотип AA, равна p^2. Генотип Aa может появиться двумя способами, и вероятность этого события равна $2pq$. Таким образом, если выполняется гипотеза о случайном скрещивании, то *потомок имеет один из генотипов AA, Aa или aa с вероятностями*

$$u_1=p^2,\quad 2v_1=2pq,\quad w_1=q^2 \tag{5.2}$$

соответственно.

Примеры. а) Все родители типа Aa (гетерозиготы). В этом случае $u=w=0$, $2v=1$ и $p=q=1/2$. б) Родительские особи AA и aa смешаны в равном отношении; тогда $u=w=1/2$, $v=0$ и $p=q=1/2$. в) Наконец, $u=w=1/4$, $2v=1/2$; снова $p=q=1/2$. Во всех трех случаях распределение генотипов потомков одно и то же: $u_1=1/4$, $2v_1=1/2$, $w_1=1/4$. ▶

Для лучшего понимания сущности соотношений (5.2) фиксируем частоты генов p и q $(p+q=1)$ и рассмотрим все системы, для которых частоты генотипов u, $2v$ и w удовлетворяют условиям $u+v=p$, $v+w=q$. Во всех системах распределение генотипов в первом поколении потомков дается той же формулой (5.2). Выделим частную систему, для которой

$$u=p^2,\quad 2v=2pq,\quad w=q^2. \tag{5.3}$$

Рассмотрим теперь популяцию, в которой — как в примере в) — частоты трех генотипов u, v, w задаются соотношениями (5.3). В силу (5.2) эти частоты остаются неизменными для следующего поколения. Поэтому распределение генотипов вида (5.3) можно назвать *стационарным* или равновесным. Каждому отношению $p:q$ отвечает стационарное распределение.

В больших популяциях действительно наблюдаемые частоты трех генотипов должны быть близки к теоретическим вероятностям, задаваемым соотношениями (5.2) [1]. Достопримечательно, что это распределение является стационарным при любых отношениях $u:2v:w$ в родительском поколении. Иначе говоря, если наблюдаемые частоты в точности совпадают с вычисленными вероятностями, то распределение генотипов уже в первом поколении потомков становится стационарным и повторяется без изменений в последующих поколениях. На практике возможны отклонения, но для больших популяций мы можем сказать: *каков бы ни был состав родительской*

[1] Без этого наша вероятностная модель оказалась бы неприменимой. Данное утверждение можно уточнить, опираясь на закон больших чисел и центральную предельную теорему, которые позволяют оценить эффект случайных флуктуаций.

популяции, при гипотезе о случайном скрещивании уже в первом поколении получается приблизительно стационарное распределение генотипов с неизменными частотами генов. Начиная со второго поколения не существует тенденций к систематическим изменениям; система становится генетически устойчивой уже в первом поколении. Этот факт впервые отметил Г. Харди [1]), который тем самым преодолел некоторые трудности, связанные с применением законов Менделя. Отсюда, в частности, вытекает, что при случайном скрещивании частоты трех генотипов должны установиться в отношении $p^2 : 2pq : q^2$. Это обстоятельство в свою очередь может быть использовано для проверки гипотезы о случайном скрещивании.

Харди также отмечал, что следует сделать ударение на слове «приблизительно». Даже для стационарного распределения мы должны ожидать от поколения к поколению небольших изменений, которые приводят к следующей картине. Какова бы ни была родительская популяция, после случайного скрещивания мы уже в *первом* поколении приходим к стационарному распределению (5.3). В дальнейшем не существует тенденций к систематическим изменениям, однако частоты генов p и q подвержены случайным колебаниям от поколения к поколению и генетическая структура популяций медленно, но меняется. Не существует никаких сил, стремящихся восстановить начальные частоты. Наоборот, наша упрощенная модель приводит к заключению (см. пример гл. XV, 2, и)), что в ограниченных популяциях в конечном счете один из двух генов исчезает, так что в конечном счете популяция будет состоять лишь из чистых генотипов AA или aa. В природе этот процесс происходит не всегда, так как мутации, селекция и другие эффекты приводят к образованию новых генов.

Часто считают, что из теорем Харди вытекает строгая устойчивость в течение неограниченного времени. Обычной ошибкой является убеждение в том, что закон больших чисел действует как наделенная памятью сила, стремящаяся вернуть систему к исходному состоянию; представления такого рода породили много неправильных заключений. Отметим, что закон Харди неприменим к распределению двух пар генов (например, обусловливающих голубой цвет глаз и свойство быть левшой) с девятью генотипами $AABB$, $AABb$, . . . , $aabb$. Здесь все еще имеется тенденция к установлению стационарного распределения, но это распределение не достигается в первом поколении (см. задачу 31).

---

[1]) Hardy G. H., Mendelian proportions in a mixed population, Letter to the Editor, Science, N. S., 28 (1908), 49—50. Используя терминологию гл. IX и XV, мы можем описать ситуацию следующим образом. Частоты трех генотипов в n-м поколении являются случайными величинами, математические ожидания которых определяются по формуле (5.2) и не зависят от n. Действительные значения частот меняются от поколения к поколению и образуют марковский случайный процесс.

§ 6 *). ПРИЗНАКИ, СЦЕПЛЕННЫЕ С ПОЛОМ

В начале предыдущего параграфа отмечалось, что гены являются составными частями хромосом, которые содержатся в клетках парами и передаются как единое целое, т. е. так, что все гены одной и той же хромосомы постоянно соединены вместе [1]). Таким образом, наша схема генной передачи наследственности применима также к хромосомам, рассматриваемым как единицы наследственности. Пол определяется парой хромосом; для женщин это XX, для мужчин XY. Материнский организм всегда передает X-хромосому, и пол потомка определяется хромосомой, полученной от отца. Таким образом, мужские и женские гаметы производятся в равном количестве. Разница в частотах рождения мальчиков и девочек объясняется тем, что они имеют неодинаковые шансы выжить в предродовой период.

Мы уже говорили, что как гены, так и хромосомы встречаются парами, однако имеются исключения, связанные с тем, что гены, расположенные в X-хромосоме, не имеют парных им генов в Y-хромосоме. Женщины имеют две X-хромосомы и, следовательно, двойной набор генов этих хромосом, но мужчины имеют лишь одинарный набор таких генов. Типичными генами, сцепленными с полом, являются гены, обусловливающие дальтонизм и гемофилию. По отношению к ним женщины могут иметь три генотипа AA, Aa и aa, а мужчины, которые обладают лишь *одним* геном — два генотипа A и a. Заметим, что сын всегда наследует отцовскую Y-хромосому, так что сцепленные с полом признаки не могут передаваться от отца к сыну. Однако они могут передаваться от отца к дочери, а от нее — к внуку.

Применим к этой ситуации методы предыдущего параграфа. Вновь предположим, что скрещивание происходит случайным образом, и будем считать, что частоты генотипов AA, Aa и aa среди особей *женского пола* равны соответственно u, $2v$ и w. Как и ранее, положим $p=u+v$, $q=v+w$. Частоты *мужских* генотипов A и a обозначим через p' и q' $(p'+q'=1)$. Тогда p и p' будут частотами генов A среди особей мужского и женского пола соответственно. Вероятности потомку женского пола иметь генотипы AA, Aa, aa будем обозначать u_1, $2v_1$, w_1 соответственно; аналогичные вероятности для мужских генотипов A и a обозначим p_1', q_1'. Потомок мужского пола получает свою X-хромосому от матери, и, следовательно,

$$p_1'=p, \quad q_1'=q. \tag{6.1}$$

---

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

[1]) Эта картина несколько осложняется явлениями разрывов и рекомбинаций хромосом (см. задачу 12 гл. II, 10).

Для трех женских генотипов, повторяя рассуждения § 5, получаем

$$u_1 = pp', \quad 2v_1 = pq' + qp', \quad w_1 = qq', \tag{6.2}$$

откуда

$$p_1 = u_1 + v_1 = (1/2)(p + p'), \quad q_1 = v_1 + w_1 = (1/2)(q + q'). \tag{6.3}$$

Это означает, что среди потомков мужского пола гены A и a встречаются приблизительно с теми же частотами p и q, что и в материнской популяции; соответствующие частоты для потомков женского пола равны примерно p_1 и q_1 или среднему арифметическому частот генов A и a в отцовской и материнской популяции. Мы наблюдаем здесь тенденцию к сближению частот генов. Действительно, из (6.1) и (6.3) следует, что

$$p_1' - p_1 = (1/2)(p - p'), \quad q_1' - q_1 = (1/2)(q - q'), \tag{6.4}$$

и, таким образом, при случайном скрещивании разница в частотах генов A и a среди мужчин и женщин уже в первом поколении сокращается примерно в два раза. Однако эта разница полностью не исчезает, и существует тенденция к дальнейшему сближению. В отличие от ситуации, описываемой законом Харди, распределение генотипов в первом поколении не является стационарным. Однако от поколения к поколению можно проследить за систематической компонентой, пренебрегая случайными флуктуациями и отождествляя теоретические вероятности (6.2) и (6.3) с реальными частотами в первом поколении особей женского пола [1]). Аналогично для второго поколения получим

$$\begin{aligned} p_2 &= (1/2)(p_1 + p_1') = (3/4)\,p + (1/4)\,p', \\ q_2 &= (1/2)(q_1 + q_1') = (3/4)\,q + (1/4)\,q', \end{aligned} \tag{6.5}$$

и, конечно, $p_2' = p_1$ и $q_2' = q_1$. Продолжая рассуждать подобным образом, получим общие выражения для вероятностей p_n и q_n, характеризующих распределение генов A и a среди особей женского пола в n-м поколении. Положим

$$\alpha = (1/3)(2p + p'), \quad \beta = (1/3)(2q + q'); \tag{6.6}$$

тогда

$$\begin{aligned} p_n &= (p_{n-1} + p_{n-1}')/2 = \alpha + (-1)^n (p - p')/(3 \cdot 2^n), \\ q_n &= (q_{n-1} + q_{n-1}')/2 = \beta + (-1)^n (q - q')/(3 \cdot 2^n) \end{aligned} \tag{6.7}$$

и $p_n' = p_{n-1}$, $q_n' = q_{n-1}$. Следовательно,

$$p_n \to \alpha, \quad p_n' \to \alpha, \quad q_n \to \beta, \quad q_n' \to \beta. \tag{6.8}$$

---

[1]) Используя терминологию, введенную в примечании на с. 154, мы можем интерпретировать p_n и q_n как математические ожидания частот генов в n-м поколении особей женского пола. При такой интерпретации формулы для p_n и q_n становятся уже не приближенными, а точными.

Частоты генотипов в популяции особей женского пола, согласно (6.2), равны

$$u_n = p_{n-1}p'_{n-1}, \quad 2v_n = p_{n-1}q'_{n-1} + q_{n-1}p'_{n-1}, \quad w_n = q_{n-1}q'_{n-1}. \quad (6.9)$$

Отсюда следует, что

$$u_n \to \alpha^2, \quad 2v_n \to 2\alpha\beta, \quad w_n \to \beta^2. \quad (6.10)$$

(Отметим, что $\alpha+\beta=1$).

Эти формулы показывают, что вся система от поколения к поколению стремится к устойчивому состоянию, в котором мужские генотипы A и a появляются с частотами α и β, а женские генотипы AA, Aa, aa с частотами α^2, $2\alpha\beta$, β^2 соответственно. На практике приближенное равенство может быть использовано уже после третьего или четвертого поколения. Разумеется, на эту зависимость могут накладываться небольшие случайные флуктуации, но в целом она правильно отражает основную тенденцию.

Главный вывод из всего сказанного состоит в следующем: можно ожидать, что при случайном скрещивании генотипы A и a среди особей мужского пола и генотипы AA, Aa и aa среди особей женского пола должны встречаться приблизительно с частотами α, β, α^2, $2\alpha\beta$, β^2 соответственно, причем $\alpha+\beta=1$.

Приложение. Многие гены, сцепленные с полом, например ген дальтонизма, являются *рецессивными* и вызывают дефекты. Пусть a — такой ген. Тогда дефект имеют все мужчины a и женщины aa. Женщины типа Aa дефекта не имеют, но могут передавать его своим потомкам. Поэтому мы можем ожидать, что *рецессивный сцепленный с полом дефект, который встречается у мужчин с частотой* α, *у женщин будет встречаться с частотой* α^2. Если один мужчина из ста — дальтоник, то женщина с этим недостатком встречается одна на 10 000.

§ 7*). СЕЛЕКЦИЯ

В качестве типичного примера влияния селекции мы исследуем случай, когда особи aa не могут размножаться. Это может случиться, если гены a являются рецессивными летальными, так что появившиеся на свет особи aa сразу же умирают. Кроме того, этого можно добиться искусственным изменением условий размножения или законом, запрещающим особям aa участвовать в скрещивании.

Предположим, что гипотеза о случайном скрещивании применима к совокупности всех особей AA и Aa, но особи aa не участвуют в скрещивании. Пусть частоты генотипов AA, Aa и aa в *полной* популяции равны u, $2v$ и w соответственно. Аналогичные час-

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

тоты в популяции *родителей* равны

$$u^*=u/(1-w), \quad 2v^*=2v/(1-w), \quad w^*=0. \tag{7.1}$$

В дальнейшем можно рассуждать, как в § 5, используя, однако, вместо величин u, $2v$, w величины (7.1). Тогда соотношения (5.1) заменяются соотношениями

$$p=(u+v)/(1-w), \quad q=v/(1-w). \tag{7.2}$$

Вероятности трех генотипов в первом поколении снова определяются формулами (5.2), т. е. $u_1=p^2$, $2v_1=2pq$ и $w_1=q^2$.

Как и ранее, для того чтобы проследить за систематическими изменениями от поколения к поколению, мы должны заменить u, v, w на u_1, v_1, w_1 и таким образом получим вероятности u_2, v_2, w_2 генотипов во втором поколении потомков и т. д. В общем случае, исходя из (7.2), имеем

$$p_n=(u_n+v_n)/(1-w_n), \quad q_n=v_n/(1-w_n) \tag{7.3}$$

и

$$u_{n+1}=p_n^2, \quad 2v_{n+1}=2p_nq_n, \quad w_{n+1}=q_n^2. \tag{7.4}$$

Сопоставляя (7.3) и (7.4), найдем, что

$$p_{n+1}=(u_{n+1}+v_{n+1})/(1-w_{n+1})=p_n/(1-q_n^2)=1/(1+q_n) \tag{7.5}$$

и аналогично

$$q_{n+1}=v_{n+1}/(1-w_{n+1})=q_n/(1+q_n). \tag{7.6}$$

Используя (7.6), нетрудно найти q_n в явном виде. Действительно,

$$q_{n+1}^{-1}=1+q_n^{-1}, \tag{7.7}$$

и отсюда последовательно

$$\begin{gathered} q_1^{-1}=1+q^{-1}, \quad q_2^{-1}=2+q^{-1}, \\ q_3^{-1}=3+q^{-1}, \quad \ldots, \quad q_n^{-1}=n+q^{-1}, \end{gathered} \tag{7.8}$$

или

$$q_n=q/(1+nq), \quad w_{n+1}=[q/(1+qn)]^2. \tag{7.9}$$

Мы видим, что генотип, не способный к воспроизведению (или нежелательный), постепенно исчезает, но процесс этот крайне медленный. При $q=0{,}1$ требуется десять поколений, чтобы наполовину сократить частоту генов a; при этом частота особей, имеющих генотип aa, изменяется с 1 до 0,25%. (Если ген a сцеплен с полом, то процесс вымирания происходит значительно быстрее, как показано в задаче 29; обобщенной схеме селекции посвящена задача 30 [1]).)

[1]) Дальнейший анализ разнообразных евгенических эффектов (которые часто отличны от того, что думают горячие сторонники законов о стерилизации) можно найти в книге Dahlberg G., Mathematical methods for population genetics, New York and Basel, 1948.

§ 8. ЗАДАЧИ

1. Бросаются три игральные кости. Какова вероятность того, что на одной из них выпадет единица, если на всех трех костях выпали разные грани?

2. Известно, что при бросании десяти игральных костей выпала хотя она единица. Какова вероятность p того, что выпало две или более единиц?

3. *Бридж*. При сдаче карт для игры в бридж игрок «Запад» не получил ни одного туза. Какова вероятность того, что его партнер а) не имеет тузов, б) имеет не менее двух тузов? Проверить результат, не используя формулу условной вероятности.

4. *Бридж*. Игроки «Север» и «Юг» имеют вместе десять козырей (козыри — карты определенной масти). а) Найти вероятность, что все три остальных козыря находятся у одного игрока (т. е. либо игрок «Восток», либо игрок «Запад» не имеют козырей). б) Известно, что среди трех оставшихся козырей имеется король. Найти вероятность того, что он «не защищен» (т. е. один игрок имеет короля, а другой — два остальных козыря).

5. Решить задачу о ключах (пример гл. II, 7, б)) при помощи условных вероятностей, используя метод примера 2, а).

6. На заводе, изготовляющем болты, на долю машин A, B и C приходится соответственно 25, 35 и 40% всех изделий. В их продукции брак составляет соответственно 5, 4 и 2%. Случайно выбранный из продукции болт оказался дефектным. Каковы вероятности того, что он был изготовлен на машине A? на машине B? на машине C?

7. Предположим, что 5 мужин из 100 и 25 женщин из 10 000 являются дальтониками. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность того, что это мужчина? (Считать, что мужчин и женщин одинаковое число.)

8. Семь шаров случайным образом распределяются по семи ящикам. Известно, что ровно два ящика остались пустыми. Показать, что вероятность (условная) того, что в одном из ящиков окажется три шара, равна 1/4. Проверить это численно, используя табл. 1 из гл. II, 5.

9. Игральная кость бросается до тех пор, пока не выпадет единица. Предполагая, что при первом испытании единица не выпала, найти вероятность того, что потребуется не менее трех бросаний.

10. *Продолжение*. Допустим, что число испытаний n четно. Какова вероятность того, что $n=2$?

11. Пусть[1]) вероятность p_n того, что в семье ровно n детей, равна αp^n при $n \geq 1$, и $p_0 = 1-\alpha p\,(1+p+p^2+\ldots)$. Предположим, что все комбинации полов n детей равновероятны. Показать, что при $k \geq 1$ вероятность того, что в семье ровно k мальчиков, равна $2\alpha p^k/(2-p)^{k+1}$.

12. *Продолжение*. Пусть известно, что в семье есть по меньшей мере один мальчик. Какова вероятность того, что в ней не менее двух мальчиков?

13. Игральная кость A имеет четыре красных и две белых грани, а кость B — две красных и четыре белых. *Один раз* бросается монета. Если выпал герб, то все время бросается только кость A, если решетка — только кость B. а) Показать, что вероятность получить красную грань в любом испытании равна 1/2. б) Известно, что первые два бросания кости дали красные грани. Найти вероятность того, что третье бросание также даст красную грань. в) Первые n испытаний дали красные грани. Какова вероятность того, что бросалась кость A? г) Какой урновой модели эквивалентна эта игра?

14. Пусть в примере 2, а) x_n — условная вероятность того, что выигравший n-ю партию выиграет игру в целом при условии, что игра не заканчивается n-й партией; пусть y_n и z_n — вероятности выиграть игру в целом для проигравшего и свободного в n-й партии игроков соответственно, а) Пока-

[1]) Согласно Лотка, американская семейная статистика удовлетворяет этой гипотезе с $p=0{,}7358$. См. Lotka A. J. Théorie analytique des associations biologiques, v, 2, Actualités scientifiques et industrielles, No, 780 (1939), Paris, Hermann et Cie,

зать, что

$$x_n = 1/2 + (1/2)\, y_{n+1}, \qquad y_n = (1/2)\, z_{n+1}, \qquad z_n = (1/2)\, x_{n+1}. \tag{*}$$

б) Непосредственно рассматривая пространство элементарных событий, показать, что в действительности $x_n = x$, $y_n = y$, $z_n = z$ не зависят от n. в) Доказать, что вероятность того, что игрок a выиграет в целом, равна 5/14 (в соответствии с задачей 5 гл. 1, 8). г) Показать, что $x_n = 4/7$, $y_n = 1/7$, $z_n = 2/7$ есть единственное ограниченное решение системы (*).

15. Пусть события $A_1, A_2, \ldots, A_n$ независимы и $\mathbf{P}\{A_k\} = p_k$. Найти вероятность p того, что ни одно из этих событий не осуществится.

16. *Продолжение.* Показать, что всегда $p \leqslant \exp\left(-\sum p_k\right)$.

17. *Продолжение.* Из неравенства Бонферрони (5.7) гл. IV вывести, что вероятность одновременного осуществления k или более событий из $A_1, \ldots, A_n$ меньше чем $(p_1 + \ldots + p_n)^k/k!$.

18. *К урновой схеме Пойа, пример* 2, в). Какова вероятность того, что первый шар был черным, если второй шар оказался черным?

19. *К урновой схеме Пойа, пример* 2, в). Используя индукцию, показать, что вероятность извлечения черного шара при любом испытании равна $b/(b+r)$.

20. *Продолжение.* Доказать по индукции следующее утверждение: для любых $m < n$ вероятности того, что m-е и n-е испытания дадут комбинации (черный, черный) или (черный, красный), равны соответственно

$$\frac{b\,(b+c)}{(b+r)\,(b+r+c)}, \qquad \frac{br}{(b+r)\,(b+r+c)}.$$

Обобщить на случай более чем двух испытаний.

21. *Симметрия по времени в схеме Пойа.* Пусть каждая из букв A и B обозначает красный или черный цвет (так что AB может быть любой из четырех комбинаций). Показать, что вероятность цвета A при n-м испытании при условии, что m-е испытание дало цвет B, равна вероятности цвета A при m-м испытании при условии, что n-е испытание дало цвет B.

22. Пусть в схеме Пойа $p_k(n)$ — вероятность извлечения k черных шаров в первых n испытаниях. Доказать рекуррентное соотношение

$$p_k(n+1) = p_k(n)\,\frac{r + (n-k)\,c}{b + r + nc} + p_{k-1}(n)\,\frac{b + (k-1)\,c}{b - r + nc},$$

где $p_{-1}(n)$ принимается равным нулю. Использовать это соотношение для нового доказательства формулы (2.5).

23. *Распределение Пойа.* В (2.4) положим

$$b/(b+r) = p, \qquad r/(b+r) = q, \qquad c/(b+r) = \gamma. \tag{8.1}$$

Показать, что выражение

$$p_{n_1 n} = \binom{-p/\gamma}{n_1}\binom{-q/\gamma}{n_2}\binom{-1/\gamma}{n}^{-1} \tag{8.2}$$

имеет смысл для произвольных (не обязательно рациональных) постоянных $p > 0$, $q > 0$, $\gamma > 0$, таких, что $p + q = 1$. Проверить, что $p_{n_1 n} > 0$ и что

$$\sum_{\nu=0}^{n} p_{\nu n} = 1.$$

Иначе говоря, (8.2) определяет распределение вероятностей на множестве целых чисел $0, 1, \ldots, n$. Это распределение называется распределением Пойа.

24. *Предельная форма распределения Пойа.* Если $n \to \infty$, $p \to 0$, $\gamma \to 0$ так, что $np \to \lambda$, $n\gamma \to \rho^{-1}$, то при фиксированном n_1

$$p_{n_1 n} \to \binom{\lambda\rho + n_1 - 1}{n_1}\left(\frac{\rho}{1+\rho}\right)^{\lambda\rho}\left(\frac{1}{1+\rho}\right)^{n_1}.$$

Проверить это и доказать, что при фиксированных λ и ρ члены в правой части этого соотношения дают в сумме единицу. (Правая часть определяет так называемое *отрицательное биномиальное распределение;* см. гл. VI, 8 и задачу 37 гл. VI, 9).

25. Записать формулу (11.8) гл. II через условные вероятности.

26. *Попарно независимые события, не являющиеся взаимно независимыми.* Бросаются две игральные кости. Рассмотрим три события: A — «на первой кости выпало нечетное число очков», B — «на второй кости выпало нечетное число очков» и C — «сумма очков на обеих костях нечетна (на одной кости четное число очков, на другой — нечетное)». Если все 36 элементарных событий равновероятны, то любые два из событий A, B и C независимы. Вероятность каждого равна 1/2. Однако все три события не могут осуществиться одновременно.

Приложения к биологии

27. Обобщить результаты § 5 на случай, когда каждый ген представлен k формами $A_1, A_2, \ldots, A_k$, так что имеются $k(k+1)/2$ генотипов вместо трех (множественные аллели).

28. *Братско-сестринское скрещивание.* Из некоторой популяции, в которой генотипы AA, Aa, aa встречаются с частотами u, $2v$ и w соответственно, случайным образом выбирается пара родителей. С их потомством повторяется та же процедура. Найти вероятность того, что оба родителя в первом, втором, третьем поколениях потомков имеют генотип AA (см. примеры гл. XV,2,к) и гл. XVI,4,б)).

29. *Селекция.* Пусть a — рецессивный сцепленный с полом ген. Предположим, что процесс селекции делает невозможным участие в скрещивании мужских особей, обладающих геном a. Показать, что если частоты генотипов AA, Aa, aa среди женских особей начального поколения равны u, $2v$, w, то для женских особей в первом поколении потомков $u_1=u+v$, $2v_1=v+w$, $w_1=0$, и, следовательно, $p_1=p+(1/2)\,q$, $q_1=(1/2)q$. Короче говоря, частота гена a среди женских особей уменьшается вдвое.

30. Задачу о селекции § 7 можно обобщить, предполагая, что гибнет лишь некоторая часть $\lambda\,(0<\lambda\leqslant 1)$ особей aa. Показать, что

$$p=(u+v)/(1-\lambda w),\quad q=[v+(1-\lambda)w]/(1-\lambda w).$$

В этом более общем случае вместо (7.3) можно записать

$$p_{n+1}=p_n/(1-\lambda q_n^2),\quad q_{n+1}=(1-\lambda q_n)/(1-\lambda q_n^2)\,q_n.$$

(Общее решение этих уравнений, по-видимому, неизвестно.)

31. Рассмотрим одновременно две пары генов с возможными формами (A, a) и (B, b). Любая особь передает своим прямым потомкам один ген из каждой пары, и мы предположим, что каждая из четырех возможных комбинаций имеет вероятность 1/4 (это предположение справедливо лишь в том случае, когда разные гены принадлежат разным хромосомам, иначе имеет место строгая зависимость). Существует девять генотипов, и мы обозначим их частоты в популяции родителей через U_{AABB}, U_{aaBB}, U_{AAbb}, U_{aabb}, $2U_{AaBB}$, $2U_{Aabb}$, $2U_{AABb}$, $2U_{aaBb}$, $4U_{AaBb}$. Положим

$$\begin{aligned}
p_{AB}&=U_{AABB}+U_{AABb}+U_{AaBB}+U_{AaBb},\\
p_{Ab}&=U_{AAbb}+U_{Aabb}+U_{AABb}+U_{AaBb},\\
p_{aB}&=U_{aaBB}+U_{aaBb}+U_{AaBB}+U_{AaBb},\\
p_{ab}&=U_{aabb}+U_{Aabb}+U_{aaBb}+U_{AaBb}.
\end{aligned}$$

Найти частоты, характеризующие распределение генотипов в первом поколении потомков. Показать, что

$$p_{AB}^{(1)}=p_{AB}-\delta,\qquad p_{Ab}^{(1)}=p_{Ab}+\delta,$$
$$p_{aB}^{(1)}=p_{aB}+\delta,\qquad p_{ab}^{(1)}=p_{ab}-\delta,$$

где $2\delta=p_{AB}p_{ab}-p_{Ab}p_{aB}$. Стационарное распределение определяется системой равенств

$$p_{AB}-2\delta=p_{Ab}+2\delta \text{ и т. д.}$$

(Отметим, что закон Харди *неприменим*; распределение меняется от поколения к поколению.)

32. Предположим, что частоты генотипов в популяции равны $u=p^2$, $2v=2pq$, $w=q^2$. Известно, что некто имеет генотип Aa. Тогда вероятность того, что генотип его брата тоже Aa, равна $(1+pq)/2$.

Замечание. Последующие задачи очень близки. Они связаны с понятием степени родства[1]). Каждая задача продолжает предыдущую. Предположение о случайном скрещивании и обозначения § 5 сохраняются. Мы имеем здесь дело с частным случаем цепей Маркова (см. гл. XV). Использование матриц упрощает запись.

33. Занумеруем генотипы AA, Aa и aa числами 1, 2, 3 соответственно. Пусть p_{ik} (i, $k=1$, 2, 3) — условная вероятность того, что особь имеет генотип k при условии, что одна из родительских особей (мужская или женская) имеет генотип i. Найти все девять вероятностей p_{ik}, предполагая, что другая родительская особь имеет один из генотипов 1, 2, 3 с вероятностями p^2, $2pq$ и q^2 соответственно.

34. Показать, что p_{ik} является в то же время условной вероятностью того, что родительская особь имеет генотип k при условии, что заданный потомок первого поколения имеет генотип i.

35. Доказать, что условная вероятность того, что внук (дед) имеет генотип k при условии, что дед (внук) имеет генотип i, определяется соотношением

$$p_{ik}^{(2)}=p_{i1}p_{1k}+p_{i2}p_{2k}+p_{i3}p_{3k}.$$

(Матрица $(p_{ik}^{(2)})$ является квадратом матрицы (p_{ik}).)

36[2]). Показать, что $p_{ik}^{(2)}$ является также условной вероятностью того, что человек имеет генотип k при условии, что его единокровный (или единоутробный) брат имеет генотип i.

37. Показать, что условная вероятность того, что некто имеет генотип k при условии, что определенный его прадед (правнук) имеет генотип i, дается формулой

$$p_{ik}^{(3)}=p_{i1}^{(2)}p_{1k}+p_{i2}^{(2)}p_{2k}+p_{i3}^{(2)}p_{3k}=p_{i1}p_{1k}^{(2)}+p_{i2}p_{2k}^{(2)}+p_{i3}p_{3k}^{(2)}.$$

(Матрица $(p_{ik}^{(3)})$ — куб матрицы (p_{ik}). Это придает точный смысл понятию степени родства.)

38. Рассмотрим более общий случай, а именно определим вероятности $p_{ik}^{(n)}$ того, что потомок в n-м поколении имеет генотип k, если определенный предок имеет генотип i. Доказать по индукции, что $p_{ik}^{(n)}$ являются элементами следующей матрицы:

$$\begin{pmatrix} p^2+pq/2^{n-1} & 2pq+q\,(q-p)/2^{n-1} & q^2-q^2/2^{n-1} \\ p^2+p\,(q-p)/2^n & 2pq+(1-4pq)/2^n & q^2+q\,(p-q)/2^n \\ p^2-p^2/2^{n-1} & 2pq+p\,(p-q)/2^{n-1} & q^2+pq/2^{n-1} \end{pmatrix}.$$

(Отсюда следует, что влияние предка убывает от поколения к поколению вдвое.)

39. Рассмотреть задачу 36, заменив единокровного (или единоутробного) брата *родным братом*. Показать, что соответствующая матрица имеет вид

$$\begin{pmatrix} (1/4)\,(1+p)^2 & (1/2)\,q\,(1+p) & (1/4)\,q^2 \\ (1/4)\,p\,(1+p) & (1/4)\,(1+pq) & (1/4)\,q(1+q) \\ (1/4)\,p^2 & (1/2)\,p\,(1+q) & (1/4)\,(1+q)^2 \end{pmatrix}.$$

40. Показать, что степень родства между дядей и племянником такая же, как между внуком и дедом.

[1]) Относительно этого понятия см. задачи 33, 35 и 37.— *Прим. перев.*

[2]) Первое издание книги содержало ошибку: слово «брат» (два общих родителя) использовалось там, где подразумевался *единокровный или единоутробный* брат (один общий родитель). Это было указано в статье Li C. C., Sacks L., Biometrica, 40 (1954), 347—360.

БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И РАСПРЕДЕЛЕНИЕ ПУАССОНА

§ 1. ИСПЫТАНИЯ БЕРНУЛЛИ[1])

Повторные независимые испытания называются испытаниями Бернулли, если каждое испытание имеет только два возможных исхода и вероятности исходов остаются неизменными для всех испытаний. Обычно эти две вероятности обозначают через p и q, и исход с вероятностью p называют «*успехом*» $У$, а второй — «*неудачей*» $Н$. Ясно, что p и q должны быть неотрицательными и

$$p+q=1. \tag{1.1}$$

Пространство элементарных событий каждого отдельного испытания состоит из двух точек $У$ и $Н$. Пространство элементарных событий n испытаний Бернулли содержит 2^n точек или последовательностей из n символов $У$ и $Н$; каждая точка представляет собой один возможный исход составного испытания. Поскольку испытания независимы, вероятности перемножаются. Иначе говоря, *вероятность любой конкретной последовательности есть произведение, полученное при замене символов $У$ и $Н$ на p и q соответственно.* Таким образом, $\mathbf{P}\{(УУНУН\ \ldots\ ННУ)\}=ppqpq\ldots qqp$.

Примеры. Наиболее известным примером испытаний Бернулли являются последовательные бросания правильной, или симметричной, монеты; здесь $p=q=1/2$. Если монета несимметрична, то мы по-прежнему считаем последовательные бросания независимыми и тем самым получаем модель испытаний Бернулли, в которой вероятность успеха может быть произвольной. Повторные случайные извлечения из урны с одним и тем же набором шаров представляют собой испытания Бернулли. Испытания Бернулли возникают и при более сложных экспериментах, если мы не будем различать несколько возможных исходов, а опишем каждый результат как A или не-A. Следовательно, в случае бросаний правильной кости различие между появлением единицы ($У$) и появлением не единицы ($Н$) приводит к испытаниям Бернулли с $p=1/6$, в то время как различие между выпадением четного и нечетного числа очков приводит к испытаниям Бернулли с $p=1/2$. Если кость несимметрична, то последовательные бросания все же образуют испытания Бер-

[1]) Яков Бернулли (1654—1705). Его основная работа Ars conjectandi была опубликована в 1713 г.

нулли, но соответствующие вероятности p изменяются. Десятка, валет, дама, король и туз одной масти в покере или две единицы при бросании кости могут представлять собой успех; называя все остальные исходы неудачей, получаем испытания Бернулли с $p = 1/649\ 740$ и $p = 1/36$ соответственно. Упрощения такого типа обычно используются в статистических приложениях. Например, при массовом производстве прокладок они могут различаться по толщине, однако при проверке классифицируются как годные ($У$) или бракованные ($Н$) в зависимости от того, находится их толщина в заданных границах или нет. ▶

Схема испытаний Бернулли — это теоретическая модель, и только опыт может показать, подходит ли она для описания конкретных наблюдений. Предположение о том, что последовательные бросания монеты соответствуют схеме Бернулли, подтверждается экспериментально. Философ К. Марбе[1]) разделяет мнение несведующих людей, считающих, что после семнадцати последовательных выпадений герба появление решетки становится более вероятным. Это убеждение возникает не из-за несовершенства реальных монет, а из-за того, что природа наделяется памятью, или — в нашей терминологии — отрицается стохастическая независимость последовательных испытаний. Теория Марбе не может быть опровергнута логически, но отвергается потому, что она не подтверждается эмпирически.

При проведении выборок, при промышленном контроле качества и т. д. схема испытаний Бернулли представляет собой идеальную модель, хотя никогда полностью не соответствует действительности. Так, в приведенном выше примере о производстве прокладок существует много причин, по которым выпуск продукции не укладывается в схему Бернулли. Машины меняются, поэтому вероятности не остаются постоянными; в работе машин есть некоторое постоянство и, следовательно, появление длинных серий отклонений похожего типа более вероятно, чем если бы исходы были действительно независимыми. Однако с точки зрения контроля качества желательно, чтобы этот процесс соответствовал схеме Бернулли, и важно то, что в определенных границах этого можно добиться. Тогда цель текущего контроля — обнаружить на ранней стадии значительные отклонения от идеальной схемы, указывающие на предстоящие неприятности.

§ 2. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Часто нас интересует не порядок появления успехов в последовательности n испытаний Бернулли, а их общее число. Число успехов может быть равно $0, 1, \ldots, n$, и первая задача заключается в на-

[1]) Marbe K., Die Gleichförmigkeit in der Welt, München, 1916. Теория Марбе получила широкое распространение; ее наиболее известным противником был Мизес.

хождении соответствующих вероятностей. Событие «n испытаний закончились k успехами и $n-k$ неудачами» содержит столько элементарных событий, сколько существует способов размещения k букв на n местах. Иначе говоря, это событие содержит $\binom{n}{k}$ точек и по определению каждая точка имеет вероятность $p^k q^{n-k}$. Тем самым доказана следующая теорема.

Теорема. *Пусть $b(k;\ n,\ p)$ — вероятность того, что n испытаний Бернулли с вероятностями успеха p и неудачи $q=1-p$ закончились k успехами и $n-k$ неудачами. Тогда*

$$b(k;\ n,\ p)=\binom{n}{k}p^k q^{n-k}. \tag{2.1}$$

В частности, q^n есть вероятность того, что успехов не будет, а вероятность того, что будет хотя бы один успех, равна $1-q^n$. ►

Будем рассматривать p как постоянную и обозначим число успехов в n испытаниях через S_n; тогда $b(k;\ n,\ p)=\mathbf{P}\{S_n=k\}$. Согласно общей терминологии, S_n есть *случайная величина*, а функция (2.1) является «распределением» этой случайной величины; будем называть это распределение *биномиальным*. Слово биномиальное отражает тот факт, что (2.1) представляет собой k-й член биномиального разложения $(q+p)^n$. Из этого замечания вытекает, что

$$b(0;\ n,\ p)+b(1;\ n,\ p)+\ldots+b(n;\ n,\ p)=(q+p)^n=1,$$

а это и требуется по определению вероятности. Биномиальное распределение табулировано [1]).

Примеры. а) *Данные Уэлдона о бросании костей.* Предположим, что эксперимент состоит в бросании 12 костей и под «успехом» для каждой кости понимается выпадение шестерки или пятерки. Для правильной кости вероятность успеха равна $p=1/3$ и число успехов должно подчиняться биномиальному распределению $b(k;\ 12,\ 1/3)$. В табл. 1 приведены эти вероятности с соответствующими средними частотами, вычисленными по результатам 26 306 проведенных экспериментов. Согласование кажется хорошим, но на самом деле для столь большого числа опытов оно оказывается очень плохим. Обычно статистики судят о степени согласования с помощью критерия χ^2. Согласно этому критерию, для правильной кости столь большие отклонения, как наблюденные, осуществляются в среднем один раз в 10 000 экспериментах. Следовательно, естест-

[1]) Для $n\leqslant 50$ см. Tables of the binomial probability distribution, Applied Mathematics Series National Bureau of Standards, 6 (1950). Для $50\leqslant n\leqslant 100$ см. Romig H. C., 50—100 Binomial tables, New York, John Wiley and Sons, 1953. Для более широкого интервала значений n см. Tables of the cumulative binomial probability distribution, Harvard Computation Laboratory, 1955; Tables of the cumulative binomial probabilities, Ordnance Corps, ORDP 20-11, 1952. [См. также Большев Л. Н., Смирнов Н. В, Таблицы математической статистики.— 2-е изд.— М.: Наука, 1983.— *Перев.*]

венно предположить, что кости были несимметричны. Полученным наблюдениям соответствовала бы асимметрия, при которой вероятность успеха равна $p=0{,}3377$[1]).

Таблица 1

Данные Уэлдона о бросании костей

k	$b(k;\ 12,\ 1/3)$	Наблюденная частота	$b(k;\ 12,\ 0{,}3377)$
0	0,007 707	0,007 033	0,007 123
1	0,046 244	0,043 678	0,043 584
2	0,127 171	0,124 116	0,122 225
3	0,211 952	0,208 127	0,207 736
4	0,238 446	0,232 418	0,238 324
5	0,190 757	0,197 445	0,194 429
6	0,111 275	0,116 589	0,115 660
7	0,047 689	0,050 597	0,050 549
8	0,014 903	0,015 320	0,016 109
9	0,003 312	0,003 991	0,003 650
10	0,000 497	0,000 532	0,000 558
11	0,000 045	0,000 152	0,000 052
12	0,000 002	0,000 000	0,000 002

б) В § 4 гл. IV нам встречалось биномиальное распределение в задаче об угадывании карт, и в столбцах b_m табл. 3 были приведены члены этого распределения для $n=3$, 4, 5, 6, 10 и $p=n^{-1}$. В задаче о размещении (пример гл. II, 4, в)) мы пришли к другому частному случаю биномиального распределения с $p=n^{-1}$.

в) Сколько испытаний с $p=0{,}01$ нужно провести, чтобы вероятность хотя бы одного успеха была не меньше 1/2? Здесь мы ищем наименьшее целое число n, для которого $1-(0{,}99)^n\geqslant 1/2$, или $-n\log(0{,}99)\geqslant\log 2$; следовательно $n\geqslant 70$.

г) *Задача о снабжении энергией.* Предположим, что $n=10$ рабочих должны время от времени использовать электрическую энергию, и нас интересует ожидаемая общая нагрузка. При грубом приближении считаем, что в течение любого заданного промежутка времени каждому рабочему с одной и той же вероятностью p требуется единица энергии. Если рабочие действуют независимо друг от друга, то вероятность того, что одновременно ровно k рабочим потребуется энергия, равна $b(k;\ n,\ p)$. Если каждый рабочий использует электроэнергию в среднем 12 минут в час, то следует положить $p=1/5$. Тогда вероятность того, что одновременно не менее семи рабочим потребуется электроэнергия, равна $b(7;\ 10,\ 0{,}2)+\ldots$

[1]) Fischer R. A., Statistical methods for research workers, Edinburgh — London, 1932. [Имеется перевод: Фишер Р. Э. Статистические методы для исследователей.— М.: Статиздат, 1958.]

$\ldots + b(10;\ 10,\ 0{,}2) = 0{,}0008643584$. Иначе говоря, если снабжение рассчитано на шесть единиц энергии, то вероятность перегрузки равна 0,00086, т. е. перегрузка ожидается в среднем в течение одной минуты из 1157 мин., или приблизительно одной минуты из 20 часов. Вероятность того, что не менее восьми рабочим одновременно потребуется энергия, равна всего лишь 0,0000779264, или приблизительно в одиннадцать раз меньше.

д) *Проверка сывороток или вакцин* [1]. Предположим, что нормальная частота заболеваний определенной болезнью среди крупного рогатого скота составляет 25%. Для проверки новой вакцины n здоровым животным делается прививка. Как оценить результат эксперимента? Для абсолютно недейственной вакцины вероятность иметь ровно k здоровых животных среди n подвергшихся прививке можно считать равной $b(k;\ n,\ 0{,}75)$. Для $k=n=10$ эта вероятность составляет приблизительно 0,056, а для $k=n=12$ всего лишь 0,032. Таким образом, отсутствие заболеваний среди десяти или двенадцати испытуемых животных можно рассматривать как подтверждение, хотя и не окончательное, эффективности вакцины. Заметим, что вероятность иметь не более одного больного животного из семнадцати, не получивших прививки, приблизительно равна 0,0501. Следовательно, одно заболевание среди семнадцати животных есть *более сильное свидетельство* в пользу вакцины, чем отсутствие заболеваний в партии из десяти животных. Для $n=23$ вероятность не более двух заболеваний приблизительно равна 0,0492, и поэтому два заболевания среди двадцати трех животных лучше свидетельствует в пользу вакцины, чем одно среди семнадцати или отсутствие заболеваний среди десяти.

е) *Другой статистический критерий.* Предположим, что n человек измеряют свое кровяное давление после приема определенного лекарства и до него и получают соответственно результаты $x_1, \ldots, x_n$ и $x_1', \ldots, x_n'$. Будем говорить, что i-е испытание закончилось успехом, если $x_i < x_i'$, и неудачей, если $x_i > x_i'$. (Для простоты можем предположить, что все измерения дали *разные* результаты.) Если бы лекарство не было эффективным, то наши наблюдения соответствовали бы n испытаниям Бернулли с $p=1/2$, и поэтому большее число успехов следует рассматривать как доказательство эффективности лекарства. ▶

§ 3. МАКСИМАЛЬНАЯ ВЕРОЯТНОСТЬ И «ХВОСТЫ»

Из (2.1) видно, что

$$\frac{b(k;\ n,\ p)}{b(k-1;\ n,\ p)} = \frac{(n-k+1)\,p}{kq} = 1 + \frac{(n+1)\,p-k}{kq}. \tag{3.1}$$

Поэтому вероятность $b(k;\ n,\ p)$ больше предыдущей для $k<(n+1)p$ и меньше для $k>(n+1)p$. Если $(n+1)p=m$ — целое число, то

[1] Sukhatme P. V., Panse V. G., Size of experiments for testing sera or vaccines, Indian Journal of Veterinary Science and Animal Husbandry, 13 (1943), 75—82.

$b(m;\ n,\ p)=b(m-1;\ n,\ p)$. Существует только одно целое число m, такое, что

$$(n+1)\,p-1<m\leqslant(n+1)\,p, \tag{3.2}$$

и мы получаем следующую теорему.

Теорема. *При изменении k от 0 до n члены $b(k;\ n,\ p)$ сначала монотонно возрастают, затем монотонно убывают, достигая наибольшего значения при $k=m$; если же $m=(n+1)p$, то наибольшее значение достигается дважды:* $b(m-1;\ n,\ p)=b(m;\ n,\ p)$.

Будем называть $b(m;\ n,\ p)$ *максимальной* вероятностью. Часто m называют «наиболее вероятным числом успехов», но следует иметь в виду, что для больших значений n *все* члены $b(k;\ n,\ p)$ малы. При 100 бросаниях симметричной монеты наиболее вероятное число появлений герба равно 50, но вероятность этого события менее 0,08. В следующей главе мы увидим, что $b(m;\ n,\ p)$ приближенно совпадает с $1/\sqrt{2\pi npq}$.

Обычно больший интерес представляет не вероятность появления ровно r успехов, а вероятность появления по меньшей мере r успехов, т. е.

$$\mathbf{P}\{S_n\geqslant r\}=\sum_{\nu=0}^{\infty}b(r+\nu;\ n,\ p). \tag{3.3}$$

(Этот ряд лишь формально является бесконечным, так как при $\nu>n-r$ члены ряда равны нулю.) Сейчас мы получим верхнюю оценку этой вероятности, которая останется полезной и после того, как в следующей главе будут найдены более точные оценки. Предположим, что $r>np$. В силу (3.1) очевидно, что члены ряда (3.3) убывают быстрее членов геометрической прогрессии со знаменателем $1-(r-np)/rq$, и поэтому

$$\mathbf{P}\{S_n\geqslant r\}\leqslant b(r;\ n,\ p)\,rq/(r-np). \tag{3.4}$$

С другой стороны, существуют более $r-np$ целых k, таких, что $m\leqslant k\leqslant r$. Сумма соответствующих членов биномиального распределения меньше единицы, и каждый из них не меньше $b(r;\ n,\ p)$. Следовательно, $b(r;\ n,\ p)$ не больше $(r-np)^{-1}$, и поэтому

$$\mathbf{P}\{S_n\geqslant r\}\leqslant rq/(r-np)^2, \text{ если } r>np. \tag{3.5}$$

Такие же рассуждения применимы и при оценке левого «хвоста», однако проводить вычисления не нужно. Действительно, событие, заключающееся в том, что наступило не больше r успехов, эквивалентно тому, что по меньшей мере $n-r$ испытаний закончились неудачей. Применяя эквивалент (3.5) для неудач, получаем

$$\mathbf{P}\{S_n\leqslant r\}\leqslant(n-r)p/(np-r)^2, \text{ если } r<np. \tag{3.6}$$

В следующем параграфе мы продемонстрируем пользу этих неравенств при оценке вероятности больших отклонений от наиболее вероятного значения m.

§ 4. ЗАКОН БОЛЬШИХ ЧИСЕЛ

Мы несколько раз отмечали, что наше *интуитивное представление о вероятности* основывается на следующем предположении. Если в n одинаковых испытаниях событие A осуществилось ν раз и если n очень велико, то отношение ν/n должно быть близко к вероятности p события A. Ясно, что формальная математическая теория может никогда не обращаться к действительной жизни, но она должна хотя бы содержать теоретическую модель явления, которое пытается объяснить. В соответствии с этим нам нужно, чтобы неясное вводное замечание стало точной теоремой. С этой целью под «одинаковыми испытаниями» будем понимать «испытания Бернулли» с вероятностью успеха p. Если $\mathbf{S}_n$ — число успехов в n испытаниях, то $\mathbf{S}_n/n$ есть средняя доля успехов, которая должна быть близка к p. Теперь легко придать этому утверждению точный смысл. Рассмотрим, например, вероятность того, что $\mathbf{S}_n/n$ превосходит $p+\varepsilon$, где $\varepsilon>0$ — произвольно малое, но фиксированное число. Эта вероятность совпадает с $\mathbf{P}\{\mathbf{S}_n>n(p+\varepsilon)\}$, и в силу (3.5) она не больше чем $1/(n\varepsilon^2)$. Следовательно, при увеличении n имеем

$$\mathbf{P}\{\mathbf{S}_n>n(p+\varepsilon)\}\to 0.$$

Аналогично доказывается, что $\mathbf{P}\{\mathbf{S}_n<n(p-\varepsilon)\}\to 0$, и поэтому

$$\mathbf{P}\{|\mathbf{S}_n/n-p|<\varepsilon\}\to 1. \tag{4.1}$$

Словесно это выражается так: вероятность того, что средняя доля успехов отличается от p больше чем на произвольное наперед заданное ε, стремится к нулю при увеличении n. Это одна из форм *закона больших чисел*, которая служит основой нашего интуитивного представления о вероятности как мере относительных частот. Для практических приложений эту формулировку необходимо дополнить более точной оценкой вероятности в левой части (4.1); такую оценку позволяет получить нормальное приближение для биномиального распределения (см. типичный пример гл. VII,4,з)). В действительности (4.1) является простым следствием этого приближения (см. задачу 12 гл. VII,7).

Утверждение (4.1) является классическим законом больших чисел. Этот закон представляет весьма ограниченный интерес, и в дальнейшем будет доказан более точный и более полезный *усиленный закон больших чисел* (см. гл. VIII, 4).

Предостережение. Стало обычным делать из закона больших чисел выводы, которые определенно из него не следуют. Если Петр и Павел бросают симметричную монету 10 000 раз, то принято считать, что Петр будет впереди приблизительно половину времени. *Но это неверно.* При большом числе *различных* игр с бросанием монеты есть основания ожидать, что в любой *фиксированный* момент число появлений герба больше приблизительно в половине всех случаев. Однако почти наверняка победивший игрок был впереди практически в течение всей игры. Итак, в противоположность широко распространенному мнению временнóе среднее для любой отдельно взятой игры никак не связано со средним для совокупности различных игр в любой данный момент. Для более глубокого изучения других неожиданных и парадоксальных свойств случайных флуктуаций отсылаем читателя к гл. III, в частности к обсуждению закона арксинуса.

§ 5. ПУАССОНОВСКОЕ ПРИБЛИЖЕНИЕ [1]

Во многих приложениях мы имеем дело с испытаниями Бернулли, в которых n относительно велико и p относительно мало, а произведение

$$\lambda = np \tag{5.1}$$

и не мало, и не велико. В таких случаях удобно использовать для $b(k;\ n,\ p)$ предложенное Пуассоном приближение, вывод которого мы начинаем. Для $k=0$ имеем

$$b(0;\ n,\ p) = (1-p)^n = (1-\lambda/n)^n. \tag{5.2}$$

Переходя к логарифмам и используя разложение Тейлора (8.10) гл. II, находим

$$\log b(0;\ n,\ p) = n\log(1-\lambda/n) = -\lambda - \lambda^2/(2n) - \ldots, \tag{5.3}$$

так что при больших n

$$b(0;\ n,\ p) \approx e^{-\lambda}, \tag{5.4}$$

где знак $\approx$ означает приближенное равенство (в данном случае с точностью до членов порядка n^{-1}). Далее, из (3.1) видно, что для произвольного фиксированного k и достаточно больших n имеем

$$\frac{b(k;\ n,\ p)}{b(k-1;\ n,\ p)} = \frac{\lambda-(k-1)\,p}{kq} \approx \frac{\lambda}{k}. \tag{5.5}$$

Отсюда последовательно заключаем, что

$$b(1;\ n,\ p) \approx \lambda\cdot b(0;\ n,\ p) \approx \lambda e^{-\lambda},$$
$$b(2;\ n,\ p) \approx (1/2)\,\lambda\cdot b(1;\ n,\ p) \approx (1/2)\,\lambda^2 e^{-\lambda},$$

и в общем случае по индукции получаем

$$b(k;\ n,\ p) \approx (\lambda^k/k!)\,e^{-\lambda}. \tag{5.6}$$

Это и есть классическое *пуассоновское приближение для биномиального распределения* [2]. В силу большой важности этого приближе-

[1] Симон Д. Пуассон (1781—1840). Его книга Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités появилась в 1837 г.

[2] О точности приближения см. задачи 33 и 34.

ния введем обозначение

$$p(k;\ \lambda)=e^{-\lambda}\lambda^{k}/k!. \tag{5.7}$$

В этих обозначениях $p(k;\ \lambda)$ будет приближением для $b(k;\ n,\ \lambda/n)$ при достаточно больших n.

Примеры. а) В табл. 3 гл. IV,4 приведены пуассоновские вероятности (5.7) с $\lambda=1$ и для сравнения биномиальные распределения с $p=1/n$ и $n=3$, 4, 5, 6, 10. Можно видеть, что, несмотря на малые значения n, приближение удивительно хорошее.

б) *Эмпирическая иллюстрация.* Появление пары (7, 7) среди 100 пар случайных цифр должно подчиняться биномиальному распределению с $n=100$ и $p=0{,}01$. В табл. 2 приведены числа N_k действительных появлений этой пары в 100 группах по 100 пар случайных цифр в каждой [1]). Отношения $N_k/100$ сравниваются как с теоретическими биномиальными вероятностями, так и с соответствующими пуассоновскими приближениями. Полученные частоты хорошо согласуются с теоретическими вероятностями. (По критерию χ^2 случайные флуктуации приблизительно в 75 из 100 аналогичных случаев давали бы бóльшие отклонения наблюдаемых частот от теоретических вероятностей.)

Таблица 2

Пример пуассоновского приближения

k	$b(k;\ 100,\ 0{,}01)$	$p(k;\ 1)$	N_k
0	0,366 032	0,367 879	41
1	0,369 730	0,367 879	34
2	0,184 865	0,183 940	16
3	0,060 999	0,061 313	8
4	0,014 942	0,015 328	0
5	0,002 898	0,003 066	1
6	0,000 463	0,000 511	0
7	0,000 063	0,000 073	0
8	0,000 007	0,000 009	0
9	0,000 001	0,000 001	0

Первые три столбца иллюстрируют пуассоновское приближение для биномиального распределения. В последнем столбце приведены числа групп по 100 пар случайных цифр, в которых комбинация (7, 7) встречается ровно k раз.

в) *Дни рождения.* Какова вероятность p_k того, что в группе из 500 человек ровно k родились 1 января? Если эти 500 человек выбраны случайно, то можно применить схему из 500 испытаний Бернулли с вероятностью успеха $p=1/365$. Для пуассоновского приближения положим $\lambda=500/365=1{,}3699$. . . .

[1]) Kendall M. G., Smith B., Tables of random sampling numbers, Tracts for Computers, 24, Cambridge, 1940.

Теоретические биномиальные вероятности и их пуассоновские приближения приведены ниже:

k	0	1	2	3	4	5	6
Биномиальные вероятности	0,2537	0,3484	0,2388	0,1089	0,0372	0,0101	0,0023
Пуассоновские вероятности	0,2541	0,3481	0,2385	0,1089	0,0373	0,0102	0,0023

г) *Дефектные изделия.* Предположим, что при производстве шурупов используется статистический контроль качества и поэтому законно применение схемы испытаний Бернулли. Если вероятность того, что шуруп окажется дефектным, равна $p=0{,}015$, то вероятность того, что коробка со 100 шурупами не содержит брака, равна $(0{,}985)^{100}=0{,}22061$. Соответствующее пуассоновское приближение дает $e^{-1{,}5}=0{,}22313\ldots$, что достаточно точно для большинства практических целей. Теперь поставим вопрос: сколько шурупов должно быть в коробке, чтобы вероятность обнаружить в ней по крайней мере 100 недефектных шурупов была не меньше 0,8? Если $100+x$ — искомое число шурупов, то x — небольшое целое. Чтобы применить пуассоновское приближение для $n=100+x$ испытаний, нужно было бы положить $\lambda=np$, но np приблизительно равно $100p=1{,}5$. Теперь задача свелась к нахождению наименьшего целого числа x, для которого

$$e^{-1{,}5}\{1+1{,}5/1+\ldots+(1{,}5)^x/x!\}\geqslant 0{,}8. \tag{5.8}$$

По таблицам[1]) находим, что для $x=1$ левая часть приближенно равна 0,56, а при $x=2$ она равна 0,809. Итак, согласно пуассоновскому приближению, необходимо иметь 102 шурупа. В действительности вероятность найти по меньшей мере 100 недефектных шурупов в коробке со 102 шурупами равна 0,8022... .

д) *Столетние старики.* Каждый отдельный человек в момент рождения имеет мало шансов прожить 100 лет, но в большом обществе число ежегодных рождений велико. Из-за войн, эпидемий и т. п. продолжительности жизни различных людей не являются стохастически независимыми, однако в первом приближении можно n рождений сравнить с n испытаниями Бернулли, в которых успехом является смерть после 100 прожитых лет. В устойчивом обществе, где ни его размеры, ни уровень смертности существенно не изменяются, естественно ожидать, что частота тех лет, когда умирает

---

[1]) Molina E. C., Poisson's exponential binomial limit, New York, Van Nostrand, 1942. [В этих таблицах даны значения $p(k;\lambda)$ и $p(k;\lambda)+p(k+1;\lambda)+\ldots$ при k, меняющемся от 0 до 100.] [См. также книгу Л. Н. Большева и Н. В. Смирнова, указанную в примечании на с. 165.— *Перев.*]

рают ровно k столетних стариков, приближенно равна $p(k;\ \lambda)$, где λ зависит от размера общества и от состояния здоровья его членов. Данные по Швейцарии подтверждают этот вывод [1]).

e) *Опечатки, изюминки и т. п.* Если при наборе книги существует постоянная вероятность того, что любая буква будет набрана неправильно, и если условия набора остаются неизменными, то мы имеем столько испытаний Бернулли, сколько букв в книге. Тогда частота страниц, содержащих ровно k опечаток, будет приближенно равна $p(k;\ \lambda)$, где λ — характеристика наборщика. Возможная усталость наборщика, трудные места текста и т. п. увеличивают вероятность ошибки и могут приводить к скоплениям опечаток. Таким образом, формулу Пуассона можно использовать для обнаружения существенных отклонений от нормы или от требований статистического контроля. Аналогичные рассуждения применимы во многих случаях. Например, если в тесто кладут много изюма, то после перемешивания следует ожидать, что доля булочек, содержащих ровно k изюминок, будет приближенно равна $p(k;\ \lambda)$, где λ — плотность изюма в тесте. ▶

§ 6. РАСПРЕДЕЛЕНИЕ ПУАССОНА

В предыдущем параграфе пуассоновские вероятности (5.7) фигурировали лишь как удобные приближения для биномиального распределения в случае больших n и малых p. В связи с задачами о совпадениях и о размещениях в гл. IV изучались различные вероятностные распределения, которые в пределе также приводили к пуассоновским выражениям $p(k;\ \lambda)$. Здесь мы сталкиваемся с частным случаем того замечательного факта, что существует несколько распределений большой универсальности, которые появляются в самых разнообразных задачах. Тремя основными распределениями, встречающимися во многих задачах теории вероятностей, являются биномиальное распределение, нормальное распределение (которое будет введено в следующей главе) и *распределение Пуассона*

$$p(k;\ \lambda)=e^{-\lambda}\lambda^k/k!, \qquad (6.1)$$

которое мы рассмотрим теперь более подробно.

Заметим сначала, что при сложении равенств (6.1) для $k=0$, 1, 2,. . . в правой части получается ряд Тейлора для e^{λ}, умноженный на $e^{-\lambda}$. Следовательно, для любого фиксированного λ сумма $p(k;\ \lambda)$ равна единице, и поэтому можно представить себе идеальный эксперимент, в котором $p(k;\ \lambda)$ является вероятностью ровно k успехов. Сейчас мы покажем, почему многие физические опыты и статистические наблюдения действительно приводят к такой интер-

---

[1]) Gumbel E. J., Les centenaires, Aktuárske Vedy, Prague, 7 (1937), 1—8.

претации (6.1). Общность и важность применений распределения (6.1) иллюстрируются примерами следующего параграфа. Истинная природа распределения Пуассона станет ясной только в связи с теорией стохастических процессов (см. иные подходы в гл. XII,2 и гл. XVII,2).

Рассмотрим последовательность происходящих с течением времени случайных событий, таких, как радиоактивный распад или вызовы на телефонной станции. Каждое событие представляется точкой на оси времени, и нас интересует случайное распределение этих точек. Существует много различных типов таких распределений, но они изучаются с помощью непрерывных распределений вероятностей, исследование которых откладывается до тома 2. Здесь мы лишь покажем, что простейшие физические предположения приводят к $p(k;\lambda)$ как к вероятности иметь ровно k точек (событий) внутри фиксированного интервала определенной длины. Наши методы неизбежно будут грубыми, но мы вернемся к этой задаче в гл. XII и XVIII, используя более точные методы.

Физические предположения, которые мы хотим выразить математически, заключаются в том, что условия эксперимента остаются неизменными с течением времени и что неперекрывающиеся интервалы времени стохастически независимы в том смысле, что информация о числе событий в одном интервале ничего не говорит об их числе в другом. Непрерывные распределения вероятностей дают возможность выразить эти предположения непосредственно, однако, будучи ограничены дискретными вероятностями, мы должны использовать приближенную конечную модель и затем переходить к пределу.

Представим себе единичный интервал времени, разделенный на n подынтервалов длиной $1/n$. Заданную конечную совокупность точек в этом интервале можно рассматривать как результат случайного процесса, при котором каждый подынтервал имеет одну и ту же вероятность p_n того, что в нем содержится одна или несколько точек совокупности. Подынтервал является либо занятым, либо пустым, и из предположения о независимости неперекрывающихся интервалов времени вытекает, что мы имеем дело с испытаниями Бернулли: вероятность получить ровно k занятых подынтервалов равна $b(k;\,n,\,p_n)$. Полагая теперь $n\to\infty$, мы делаем эту дискретную модель все более и более точной. При этом вероятность того, что весь интервал вообще не содержит точек совокупности, должна стремиться к конечному пределу. Но весь интервал не содержит точек тогда и только тогда, когда ни один подынтервал не занят, а вероятность последнего события равна $(1-p_n)^n$. Переходя к логарифмам, мы видим, что эта величина стремится к пределу одновременно с np_n. Случай $np_n\to\infty$ невозможен, так как он означал бы, что в сколь угодно малый интервал попадают бесконечно много точек совокупности. Поэтому в нашей модели неизбежно существует число λ, такое, что $np_n\to\lambda$. В этом случае вероятность иметь ровно

k занятых подынтервалов стремится к $p(k;\ \lambda)$, и, поскольку мы рассматриваем различные точки, число занятых ячеек в пределе соответствует числу точек совокупности, лежащих в нашем единичном интервале времени [1]).

В приложениях единичный интервал времени необходимо заменить интервалом произвольной длины t. Если опять разделить его на подынтервалы длины $1/n$, то вероятности p_n остаются неизменными, но число подынтервалов равно целому числу, ближайшему к nt. Переход к пределу будет таким же, только λ заменяется на λt. Это приводит нас к интерпретации величины

$$p(k;\ \lambda t) = e^{-\lambda t} (\lambda t)^k / k! \tag{6.2}$$

как вероятности иметь ровно k точек в фиксированном интервале длины t. В частности, вероятность того, что в интервале длины t не будет ни одной точки, равна

$$p(0;\ \lambda t) = e^{-\lambda t}, \tag{6.3}$$

и, следовательно, вероятность иметь одну или несколько точек равна $1 - e^{-\lambda t}$.

Параметр λ — физическая постоянная, которая определяет плотность точек на оси t. Чем больше λ, тем меньше вероятность (6.3) того, что точек нет. Предположим, что некоторый физический эксперимент повторяется большое число N раз и что каждый раз подсчитывается число событий в интервале фиксированной длины t. Пусть N_k — количество экспериментов, в которых наблюдается ровно k событий. Тогда

$$N_0 + N_1 + N_2 + \ldots = N. \tag{6.4}$$

Общее число точек, наблюдаемых в N экспериментах, равно

$$N_1 + 2N_2 + 3N_3 + \ldots = T, \tag{6.5}$$

и T/N — среднее. При больших N мы ожидаем, что

$$N_k \approx Np(k;\ \lambda t) \tag{6.6}$$

(это лежит в основе всех применений понятия вероятности и будет обосновано и уточнено с помощью закона больших чисел в гл. X). Подставляя (6.6) в (6.5), находим

$$\begin{aligned} T \approx N\{p(1;\ \lambda t) + 2p(2;\ \lambda t) + 3p(3;\ \lambda t) + \ldots\} = \\ = Ne^{-\lambda t}\lambda t\{1 + \lambda t/1 + (\lambda t)^2/2! + \ldots\} = N\lambda t, \end{aligned} \tag{6.7}$$

[1]) Можно представить себе и другие возможности. Наша модель может быть полезной при изучении автомобильных катастроф, но она неприменима, если подсчитывается не число самих катастроф, а число разбитых автомобилей. Это объясняется тем, что в некоторых катастрофах участвует более одного автомобиля, и поэтому необходимо рассматривать случаи, когда все точки различны, когда есть две совпадающие точки, три совпадающие точки и т. д. В пределе мы приходим к обобщенному распределению Пуассона (см. гл. XII,2). С точки зрения более общих процессов можно сказать, что мы считаем только число скачков, но не рассматриваем их величину.

и поэтому

$$\lambda t \approx T/N. \tag{6.8}$$

Это соотношение дает нам метод оценки λ по наблюдениям и способ сравнения выводов теории с результатами опытов. Примеры из следующего параграфа проиллюстрируют этот подход.

Пространственные распределения

Мы рассмотрели распределение случайных событий или точек на оси t, но те же самые рассуждения применимы к распределениям точек на плоскости или в пространстве. Вместо интервалов длины t будут области площади или объема t. Основное предположение состоит в том, что вероятность иметь k точек в любой определенной области зависит не от ее формы, а лишь от ее площади или объема. Кроме того, сохраним те же предположения, что и раньше: 1) при малых t вероятность иметь более одной точки в области объема t мала по сравнению с t; 2) неперекрывающиеся области взаимно независимы. Для нахождения вероятности того, что в области объема t содержится ровно k случайных точек, разобьем эту область на n подобластей и в качестве приближения для искомой вероятности возьмем вероятность появления k успехов в n испытаниях. Это означает, что мы пренебрегаем возможностью найти более одной точки в одной и той же подобласти, однако из нашего предположения 1) вытекает, что допускаемая погрешность стремится к нулю при $n \to \infty$. В пределе снова получается распределение Пуассона (6.2). Звезды в космосе, изюминки в кексе, семена сорняков среди семян злака, дефекты в материалах, выводки животных в поле распределены согласно закону Пуассона; см. примеры 7,б) и 7,д).

§ 7. НАБЛЮДЕНИЯ, СООТВЕТСТВУЮЩИЕ РАСПРЕДЕЛЕНИЮ ПУАССОНА[1]

а) *Радиоактивный распад.* Радиоактивное вещество испускает α-частицы; число частиц, достигающих заданной части пространства в течение времени t, дает хорошо известный пример случайных событий, подчиняющихся закону Пуассона. Конечно, вещество непрерывно разрушается и через длительный промежуток времени плотность потока α-частиц уменьшается. Однако для радия необходимы годы, прежде чем такое уменьшение станет заметным; для относительно коротких периодов времени условия можно считать неизменными, и мы имеем идеальную реализацию предположений, которые приводят к распределению Пуассона.

[1]) Распределение Пуассона стало известно под названием закона малых чисел или редких событий. Эти неправильные названия затруднили понимание основной роли распределения Пуассона. Следующие примеры покажут, сколь обманчивы такие «синонимы» распределения Пуассона.

В одном известном эксперименте[1]) радиоактивное вещество наблюдали в течение $N=2608$ интервалов времени по 7,5 секунд каждый; для каждого интервала регистрировалось число частиц, достигших счетчика. В табл. 3 приведены числа N_k интервалов, в течение которых наблюдались ровно k частиц. Общее число частиц равно $T=\sum kN_k=10\,094$, среднее — $T/N=3{,}870$. Теоретические значения $Np(k;\ 3{,}870)$ кажутся достаточно близкими к наблюденным числам N_k. Чтобы судить о степени этой близости, нужно оценить возможные величины случайных флуктуаций. Статистики судят о степени соответствия с помощью критерия χ^2. Согласно этому критерию, при идеальных условиях следует ожидать приблизительно в 17 из 100 аналогичных случаев отклонения более значительные, чем в табл. 3.

Таблица 3

Пример «а». Радиоактивный распад

k	N_k	$Np(k;\ 3{,}870)$	k	N_k	$Np(k;\ 3{,}870)$
0	57	54,399	5	408	393,515
1	203	210,523	6	273	253,817
2	383	407,361	7	139	140,325
3	525	525,496	8	45	67,882
4	532	508,418	9	27	29,189
			$\geqslant 10$	16	17,075
			Всего	2608	2608,000

б) *Падения самолетов-снарядов в Лондоне*. В качестве примера распределения случайных точек по поверхности рассмотрим статистику падений самолетов-снарядов в южной части Лондона во время второй мировой войны. Всю область разделим на $N=576$ небольших участков, каждый площадью $t=1/4$ кв. км. В табл. 4 приведены числа N_k участков ровно с k падениями[2]). Общее число падений равно $T=\sum kN_k=537$, среднее — $\lambda t=T/N=0{,}9323, \ldots$. Совпадение с распределением Пуассона поразительное; согласно критерию χ^2, при идеальных условиях примерно 88% аналогичных наблюдений дали бы худшее соответствие. Интересно отметить, что большинство населения верило в тенденцию точек падения скапливаться в нескольких местах. Если бы это было верно, то

[1]) Rutherford E., Chadwick J., Ellis C., Radiations from radioactive substances, Cambridge, 1920, 172. Табл. 3 и оценка по критерию χ^2 взяты из книги Kramér H., Mathematical methods of statistics, Uppsala and Princeton, 1945, p. 436. [Имеется перевод: Крамер Г. Математические методы статистики.— 2-е изд.— М.: Мир, 1975, с. 472.]

[2]) Данные заимствованы из статьи Clarke R. D., An application of the Poisson distribution, Journal of the Institute of Actuaries, 72 (1946), 481.

следовало бы ожидать бо́льшую долю участков без разрывов либо с бо́льшим числом разрывов и меньшую долю участков промежуточного класса. Табл. 4 показывает, что точки падения были совершенно случайными, все участки равноправными; здесь мы имеем поучительную иллюстрацию того установленного факта, что неискушенному человеку случайность представляется регулярностью или стремлением к скоплению.

Таблица 4

Пример «б». Падения самолетов-снарядов в Лондоне

k	0	1	2	3	4	5 и более
N_k	229	211	93	35	7	1
$Np\ (k;\ 0{,}9323)$	226,74	211,39	98,54	30,62	7,14	1,57

в) *Перестройка хромосом в клетках.* Облучение рентгеновскими лучами вызывает в органических клетках определенные процессы, которые мы называем перестройкой хромосом. Во время облучения вероятность такой перестройки остается постоянной и, согласно теории, числа N_k клеток, содержащих ровно k изменений, должны подчиняться распределению Пуассона. Теория способна также предсказать зависимость параметра λ от интенсивности облучения, температуры и т. д., но мы не будем вдаваться в эти детали. В табл. 5 приведены результаты одиннадцати различных серий экспериментов [1]). Они сгруппированы по степени их соответствия теории. В последнем столбце указаны примерные проценты случаев, в которых при идеальных условиях (по критерию χ^2) случайные флуктуации дали бы худшее согласование с распределением Пуассона. Согласие между теорией и наблюдениями поразительное.

г) *Соединения с неправильным номером.* В табл. 6 показана статистика телефонных соединений с неправильным номером [2]). Всего наблюдалось $N=267$ номеров; N_k обозначает, сколько номеров имели ровно k неправильных соединений. Распределение Пуассона $p(k;\ 8{,}74)$ опять подходит удивительно хорошо. (По критерию χ^2 отклонения близки к среднему значению.) В статье Торндайка читатель найдет сведения о других телефонных статистиках, подчиняющихся закону Пуассона. Иногда (например, в случае группового абонентской линии, вызовов из группы телефонов-автоматов и т. д.) существует очевидная взаимосвязь между событиями, и тогда распределение Пуассона не подходит.

---

[1]) Catcheside D. G., Lea D. E., Thoday J. M., Types of chromosome structural change induced by the irradiation of Tradescantia microspores, Journal of Genetics, 47 (1945—46), 113—136. Наша таблица совпадает с табл. IX этой статьи, мы лишь подсчитали заново χ^2-уровни с одной степенью свободы.

[2]) Данные заимствованы из работы Thorndike F., Applications of Poisson's probability summation, The Bell System Technical Journal, 5 (1926), 604—624. В этой статье содержится графический анализ 32 различных статистик.

Таблица 5

Пример «в». Перестройка хромосом, вызванная рентгеновским облучением

Номер опыта		Клетки с k изменениями				Общее число N	χ^2-уровни в процентах
		0	1	2	3		
1	Наблюденное N_k $Np(k;\ 0{,}35508)$	753 752,3	266 267,1	49 47,4	5 6,2	1073	95
2	Наблюденное N_k $Np(k;\ 0{,}45601)$	434 432,3	195 197,1	44 44,9	9 7,7	682	85
3	Наблюденное N_k $Np(k;\ 0{,}27717)$	280 278,9	75 77,3	12 10,7	1 1,1	368	65
4	Наблюденное N_k $Np(k;\ 0{,}11808)$	2278 2280,2	273 269,2	15 15,9	0 0,7	2566	65
5	Наблюденное N_k $Np(k;\ 0{,}25296)$	593 589,4	143 149,1	20 18,8	3 1,7	759	45
6	Наблюденное N_k $Np(k;\ 0{,}21059)$	639 642,4	141 135,3	13 14,2	0 1,1	793	45
7	Наблюденное N_k $Np(k;\ 0{,}28631)$	359 362,0	109 103,6	13 14,9	1 1,5	482	40
8	Наблюденное N_k $Np(k;\ 0{,}33572)$	493 498,2	176 167,3	26 28,1	2 3,4	697	35
9	Наблюденное N_k $Np(k;\ 0{,}39867)$	793 804,8	339 320,8	62 64,0	5 9,4	1199	20
10	Наблюденное N_k $Np(k;\ 0{,}40544)$	579 588,7	254 238,7	47 48,4	3 7,2	883	20
11	Наблюденное N_k $Np(k;\ 0{,}49339)$	444 461,6	252 227,7	59 56,2	1 10,5	756	5

Таблица 6

Пример «г». Соединения с неправильным номером

k	N_k	$Np(k; 8{,}74)$	k	N_k	$Np(k; 8{,}74)$
0—2	1	2,05	11	20	24,34
3	5	4,76	12	18	17,72
4	11	10,39	13	12	11,92
5	14	18,16	14	7	7,44
6	22	26,45	15	6	4,33
7	43	33,03	$\geqslant 16$	2	4,65
8	31	36,09			
9	40	35,04		267	267,00
10	35	30,63			

д) *Подсчет бактерий и клеток крови.* На рис. 1 воспроизведена фотография чашки Петри с колониями бактерий, которые под микроскопом видны как темные пятнышки. Чашка разделена на маленькие квадраты. В табл. 7 приведены наблюденные в восьми

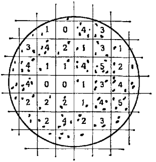

Рис. 1. Бактерии в чашке Петри.

Таблица 7

Пример «д». Число бактерий

k	0	1	2	3	4	5	6	7	χ^2-уровень
Наблюденное N_k	5	19	26	26	21	13	8		97
Пуассоновское	6,1	18,0	26,7	26,4	19,6	11,7	9,5		
Наблюденное N_k	26	40	38	17	7				66
Пуассоновское	27,5	42,2	32,5	16,7	9,1				
Наблюденное N_k	59	86	49	30	20				26
Пуассоновское	55,6	82,2	60,8	30,0	15,4				
Наблюденное N_k	83	134	135	101	40	16	7		63
Пуассоновское	75,0	144,5	139,4	89,7	43,3	16,7	7,4		
Наблюденное N_k	8	16	18	15	9	7			97
Пуассоновское	6,8	16,2	19,2	15,1	9,0	6,7			
Наблюденное N_k	7	11	11	11	7	8			53
Пуассоновское	3,9	10,4	13,7	12,0	7,9	7,1			
Наблюденное N_k	3	7	14	21	20	19	7	9	85
Пуассоновское	2,1	8,2	15,8	20,2	19,5	15	9,6	9,6	
Наблюденное N_k	60	80	45	16	9				78
Пуассоновское	62,6	75,8	45,8	18,5	7,3				

Последнее значение N_k в каждой строке учитывает квадраты с бо́льшим числом пятнышек и поэтому в соответствующем столбце следовало бы указать «k или более».

экспериментах с восемью различными видами бактерий числа квадратов, содержащих ровно k пятнышек [1]). Здесь мы имеем пример важного практического применения распределения Пуассона к распределениям случайных точек по поверхности. ▶

§ 8. ВРЕМЯ ОЖИДАНИЯ. ОТРИЦАТЕЛЬНОЕ БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Рассмотрим последовательность n испытаний Бернулли. Нас интересует, сколько испытаний предшествуют r-му успеху, где r — фиксированное положительное целое. Общее число успехов в n испытаниях может быть, конечно, меньше r, но вероятность того, что r-й успех произойдет при ν-м испытании, где $\nu \leqslant n$, очевидно, не зависит от n и зависит только от ν, r и p. Поскольку необходимо, чтобы $\nu \geqslant r$, удобно писать $\nu = k+r$. *Вероятность того, что r-й успех произойдет при $(r+k)$-м испытании, где $k=0, 1, \ldots$, будем обозначать через $f(k; r, p)$.* Она равна *вероятности иметь ровно k неудач перед r-м успехом.* Такое событие происходит тогда и только тогда, когда среди $r+k-1$ испытаний есть ровно k неудач, а следующее, или $(r+k)$-е, испытание заканчивается успехом; соответствующие вероятности равны $\binom{r+k-1}{k} p^{r-1}q^k$ и p; следовательно,

$$f(k; r, p) = \binom{r+k-1}{k} p^r q^k. \tag{8.1}$$

Переписывая биномиальный коэффициент в соответствии с формулой (12.4) гл. II, находим эквивалентную форму записи

$$f(k; r, p) = \binom{-r}{k} p^r (-q)^k, \quad k = 0, 1, 2, \ldots . \tag{8.2}$$

Предположим теперь, что *испытания Бернулли проводятся до тех пор, пока не наступит r успехов.* Типичное элементарное событие представляется в виде последовательности, содержащей случайное число k букв H и ровно r букв $У$ и заканчивающейся буквой $У$; вероятность такого события есть, по определению, $p^r q^k$. Мы должны, однако, поставить вопрос, возможно ли, что испытания никогда не закончатся, т. е. существует ли бесконечная последовательность испытаний с числом успехов, меньшим чем r. Так как $\sum_{k=0}^{\infty} f(k; r, p)$ есть вероятность того, что r-й успех произойдет после конечного числа испытаний, то возможностью существования бесконечной последовательности с числом успехов, меньшим чем r,

[1]) Таблица заимствована из книги Neyman J., Lectures and conferences on mathematical statistics (mimeographed), Dept. of Agriculture, Washington, 1938.

можно пренебречь тогда и только тогда, когда

$$\sum_{k=0}^{\infty} f(k;\ r,\ p) = 1. \tag{8.3}$$

Но это так в силу следующего свойства биномиального ряда:

$$\sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k = (1-q)^{-r} = p^{-r}. \tag{8.4}$$

Умножая (8.4) на p^r, получаем (8.3).

В нашей задаче о времени ожидания r обязательно является положительным целым, но величина, определяемая либо (8.1), либо (8.2), неотрицательна и (8.3) справедливо для любого положительного r. *Для произвольных фиксированных* $r>0$ *и* $0<p<1$ *последовательность* $\{f(k;\ r,\ p)\}$ *называется отрицательным биномиальным распределением.* Оно встречается во многих приложениях (и мы уже сталкивались с ним в задаче 24 гл. V,8 как с предельной формой распределения Пойа). Для положительных целых r последовательность $\{f(k;\ r,\ p)\}$ можно интерпретировать как *вероятностное распределение времени ожидания r-го успеха*; в такой форме оно также называется *распределением Паскаля.* При $r=1$ оно сводится к *геометрическому распределению* $\{pq^k\}$.

Примеры. а) *Задача Банаха о спичечных коробках* [1]). Некий математик всегда носит в правом и левом карманах по коробке спичек. Когда ему нужна спичка, он наугад выбирает один из карманов. Последовательные выборы образуют, таким образом, испытания Бернулли с $p=1/2$. Предположим, что в начальный момент каждая коробка содержала ровно N спичек, и рассмотрим момент, когда математик *впервые* вытащит пустую коробку. В этот момент другая коробка может содержать 0, 1, 2, . . ., N спичек; соответствующие вероятности обозначим через u_r. Договоримся считать «успехом» выбор коробки из левого кармана. Из левого кармана будет вынута пустая коробка, а в правом в это время будет ровно r спичек тогда и только тогда, когда $(N+1)$-му успеху предшествуют $N-r$ неудач. Вероятность этого события равна $f(N-r;\ N+1,\ 1/2)$. Точно так же можно рассуждать и о правом кармане и, следовательно, искомая вероятность равна

$$u_r = 2f(N-r;\ N+1,\ 1/2) = \binom{2N-r}{N} 2^{-2N+r}. \tag{8.5}$$

Численные значения в случае $N=50$ даны в табл. 8. (См. также задачи 21 и 22 ниже и задачу 11 гл. IX,9.)

---

[1]) Этот пример возник из шутливого замечания о привычках Банаха (сделанного Г. Штейнгаузом) во врученном ему адресе. Пример стал неожиданно популярным в литературе, и по этой причине его название сохранено. Приписывать Банаху авторство этой задачи, конечно, было бы ошибкой.

Таблица 8

Вероятности (8.5) в задаче о спичечных коробках

r	u_r	U_r	r	u_r	U_r
0	0,079 589	0,079 589	15	0,023 171	0,917 941
1	0,079 589	0,159 178	16	0,019 081	0,937 022
2	0,078 785	0,237 963	17	0,015 447	0,952 469
3	0,077 177	0,315 140	18	0,012 283	0,964 752
4	0,074 790	0,389 931	19	0,009 587	0,974 338
5	0,071 674	0,461 605	20	0,007 338	0,981 676
6	0,067 902	0,529 506	21	0,005 504	0,987 180
7	0,063 568	0,593 073	22	0,004 041	0,991 220
8	0,058 783	0,651 855	23	0,002 901	0,994 121
9	0,053 671	0,705 527	24	0,002 034	0,996 155
10	0,048 363	0,753 890	25	0,001 392	0,997 547
11	0,042 989	0,796 879	26	0,000 928	0,998 475
12	0,037 676	0,834 555	27	0,000 602	0,999 077
13	0,032 538	0,867 094	28	0,000 379	0,999 456
14	0,027 676	0,894 770	29	0,000 232	0,999 688

u_r есть вероятность того, что в момент, когда впервые одна из коробок окажется пустой, другая содержит ровно r спичек (предполагается, что вначале каждая коробка содержала 50 спичек). $U_r=u_0+u_1+\ldots+u_r$ означает вероятность того, что вторая коробка содержит не более r спичек.

б) *Обобщение: настольный теннис.* Суть рассмотренной выше задачи становится яснее, если разным коробкам приписать разные вероятности. Для разнообразия проиллюстрируем этот вариант иначе. Предположим, что Петр и Павел играют в игру, которую можно рассматривать как последовательность испытаний Бернулли с вероятностями p и q, характеризующими мастерство игроков. В обычном настольном теннисе побеждает игрок, первым набравший 21 очко, т. е. победивший при 21 подаче. Для сравнения с предыдущим примером рассмотрим более общую ситуацию, когда для общей победы нужна победа при $2\nu+1$ подаче. Тогда в игре будет самое меньшее $2\nu+1$ и самое большее $4\nu+1$ подач (испытаний). Обозначим через a_r вероятность общей победы Петра при $4\nu+1-r$ испытаниях. Это событие происходит тогда и только тогда, когда в первых $4\nu-r$ испытаниях Петр набрал 2ν побед и, следовательно, победил в $(2\nu+1)$-м испытании. Таким образом,

$$a_r=\binom{4\nu-r}{2\nu}p^{2\nu+1}q^{2\nu-r}. \tag{8.6}$$

В нашей игре $a_0+\ldots+a_{2N}$ есть вероятность победы Петра. Вероятность того, что игра закончится ровно после $4\nu+1-r$ испытаний, равна a_r+b_r, где b_r определяется формулой (8.6), в которой p и q меняются местами.

Если положить $2\nu=N$ и $p=q=1/2$, то вероятности a_r+b_r сводятся к вероятностям u_r из предыдущего примера. ►

§ 9. ПОЛИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Биномиальное распределение нетрудно обобщить на случай последовательных независимых испытаний, каждое из которых заканчивается одним из нескольких возможных исходов. Обозначим возможные исходы каждого испытания через $E_1,\ldots,E_r$ и предположим, что вероятность осуществления исхода E_i для каждого испытания равна p_i $(i=1,\ldots,r)$. При $r=2$ получаем испытания Бернулли; в общем случае числа p_i должны удовлетворять только следующему условию:

$$p_1+\ldots+p_r=1, \quad p_i\geqslant 0. \tag{9.1}$$

Результатом n испытаний будет, например, последовательность вида $E_3E_1E_2\ldots$. *Вероятность того, что в n испытаниях E_1 произойдет k_1 раз, E_2 произойдет k_2 раз и т. д., равна*

$$[n!/(k_1!\,k_2!\ldots k_r!)]\,p_1^{k_1}p_2^{k_2}p_3^{k_3}\ldots p_r^{k_r}, \tag{9.2}$$

где k_i — произвольные неотрицательные целые числа, удовлетворяющие очевидному условию

$$k_1+k_2+\ldots+k_r=n. \tag{9.3}$$

При $r=2$ (9.2) сводится к биномиальному распределению с $p_1=p$, $p_2=q$, $k_1=k$, $k_2=n-k$. Доказательство в общем случае опирается на формулу (4.7) гл. II и проводится так же, как и для биномиального распределения.

Выражение (9.2) задает *полиномиальное распределение*, полиномиальное потому, что выражение (9.2) совпадает с общим членом *разложения полинома* $(p_1+\ldots+p_r)^n$. Это распределение применяется к *выбору с возвращением*, когда выбираемые объекты разбиты более чем на два класса (например, по профессиям).

Примеры. а) Какова вероятность получить каждую грань дважды при бросании двенадцати костей? Здесь $E_1,\ldots,E_6$ соответствуют шести граням, все k_i равны 2 и все p_i равны 1/6. Следовательно, ответ будет $12!2^{-6}\,6^{-12}=0{,}0034\ldots$.

б) *Выбор.* Пусть генеральная совокупность из N элементов разбита на классы $E_1,\ldots,E_r$ объема $Np_1,\ldots, Np_r$ соответственно. Полиномиальное распределение дает вероятности возможных результатов случайного выбора с возвращением объема n из этой совокупности, в) *Сложные*[1]) *испытания Бернулли*. Две последовательности испытаний Бернулли с вероятностями успеха и неудачи соответственно p_1, q_1 и p_2, q_2 можно рассматривать как один сложный эксперимент с четырьмя возможными исходами в каждом испытании, а именно комбинациями $(У, У)$, $(У, Н)$, $(Н, У)$, $(Н, Н)$. Предположение о независимости двух исходных последовательностей приводит к тому, что вероятности четырех исходов равны соответственно p_1p_2, p_1q_2, q_1p_2, q_1q_2. Если четыре целых числа k_1, k_2, k_3, k_4 в сумме дают n, то вероятность того, что в n испытаниях $УУ$ появится k_1 раз, $УН$ появится k_2 раз и т. д., равна

$$[n!/(k_1!\,k_2!\,k_3!\,k_4!)]\,p_1^{k_1+k_2}q_1^{k_3+k_4}p_2^{k_1+k_3}q_2^{k_2+k_4}. \tag{9.4}$$

Частный случай рассмотренной схемы возникает при *выборочном контроле*. Изделие признается стандартным или бракованным с вероятностями соответственно p и q. Оно может быть проверено или не проверено с вероятностями соответственно p' и q'. Решение о проверке изделия принимается без информации о его качестве, так что испытания независимы (см. задачи 25 и 26 ниже и задачу 12 гл. IX, 9).

§ 10. ЗАДАЧИ

1. Предполагая все комбинации полов детей равновероятными, определить, какую долю семей с шестью детьми будут составлять семьи с тремя мальчиками и тремя девочками.

2. Игрок в бридж при трех последовательных сдачах не получил туза. Есть ли у него основания жаловаться на невезение?

3. Сколько случайных цифр нужно взять, чтобы вероятность появления среди них цифры 7 была не меньше 9/10?

4. Сколько раз нужно сдать колоду карт для игры в бридж (сдачи независимы), чтобы вероятность того, что фиксированный игрок по меньшей мере один раз получит четыре туза, была не меньше 1/2? Решить задачу и в том случае, когда игрок заранее не фиксируется.

5. Пусть вероятность попадания в цель равна 1/5 и производится десять независимых выстрелов. Какова вероятность попадания в цель по меньшей мере дважды?

6. Найти в задаче 5 условную вероятность попадания в цель по меньшей мере дважды, если известно, что по крайней мере одно попадание произошло.

7. Найти вероятность того, что среди 13 карт, наугад выбранных из колоды в 52 карты, имеются ровно две карты красной масти. Сравнить ее с соответствующей вероятностью для испытаний Бернулли с $p=1/2$.

8. Какова вероятность того, что дни рождения шести людей приходятся на два месяца, оставляя ровно десять месяцев свободными? Предполагается независимость и равновероятность всех месяцев.

9. Найти вероятность того, что при бросании 6 правильных костей единица выпадает а) по меньшей мере один раз, б) ровно один раз, в) ровно два раза. Сравнить эти вероятности с их пуассоновскими приближениями.

10. Известно, что левши составляют в среднем 1%. Оценить вероятность того, что среди 200 людей окажется по меньшей мере четверо левшей.

11. Книга в 500 страниц содержит 500 опечаток. Оценить вероятность того, что на заданной странице не менее трех опечаток.

---

[1]) В оригинале multiple.— *Прим. перев.*

12. Дальтоники составляют 1% лиц в некоторой группе людей. Как велика должна быть случайная выборка (с возвращением), чтобы вероятность иметь в ней одного дальтоника была не меньше 0,95?

13. В предыдущей задаче найти вероятность того, что в выборке из 100 человек а) нет дальтоников, б) дальтоников два или больше.

14. Оценить среднее число изюминок, которое должно быть в одной булочке, чтобы не более одной булочки из ста было без изюма.

15. Вероятность иметь на одной руке в покере десятку, валета, даму, короля и туза одной масти равна $p=1/649\,740$. Сколь велико должно быть n, чтобы вероятность того, что ни разу при n сдачах не окажется такого набора карт, меньше $1/e \approx 1/3$? (*Замечание*. Решение не требует никаких вычислений.)

16. Книга в n страниц содержит в среднем λ опечаток на страницу. Оценить вероятность того, что хотя бы на одной странице будет более k опечаток.

17. Предположим, что существует два типа звезд (или изюминок в кексе, или дефектов материала). Заданный участок содержит j звезд первого типа с вероятностью $p(j;\, a)$ и k звезд второго типа с вероятностью $p(k;\, b)$; эти два события предполагаются независимыми. Доказать, что тот же участок содержит всего n звезд с вероятностью $p(n;\, a+b)$. (Сформулировать результат и предположения задачи в абстрактных терминах.)

18. *Задача об уличном движении*. Поток транспорта через пешеходный переход таков, что вероятность проезда машины в течение любой заданной секунды постоянна и равна p; известно также, что нет связи между проездом машин в разное время. Рассматривая секунды как неделимые единицы времени, приходим к схеме испытаний Бернулли. Предположим, что пешеход может перейти улицу, если ни одна машина не будет проезжать в течение следующих трех секунд. Найти вероятность того, что он должен ждать ровно $k=0, 1, 2, 3, 4$ секунд. (Соответствующие общие формулы не очевидны и будут выведены при изложении теории серий успехов в гл. XIII,7.)

19. Двое бросают симметричную монету n раз каждый. Найти вероятность того, что у них выпадет одинаковое число гербов.

20. Для последовательности испытаний Бернулли с вероятностью успеха p найти вероятность того, что a успехов произойдут прежде, чем b неудач. (*Замечание*. Результат определяется самое большее за $a+b-1$ испытаний. Эта задача имеет значение для классической теории игр в связи с вопросом о том, как делить ставку, если игра прервана в момент, когда одному игроку не хватает до победы a очков, а второму b.)

21. В задаче Банаха о спичечных коробках (пример 8,а)) найти вероятность того, что в момент, когда первая коробка оказалась пустой (но не была вынута пустой), во второй содержится ровно r спичек (где $r=1, 2, \ldots, N$).

22. *Продолжение*. Используя предыдущий результат, найти вероятность x того, что отсутствие спичек было обнаружено впервые не в той коробке, которая опустела первой. Показать, что полученное выражение сводится к виду $x=\binom{2N}{N}2^{-2N-1}$ или приближенно $(1/2)\,(N\pi)^{-1/2}$.

23. Корректуры некоторой книги независимо читали два корректора, которые нашли соответственно k_1 и k_2 ошибок; k_{12} ошибок нашли оба. Дать приемлемую оценку неизвестного числа n ошибок в корректурах. (Предположить, что проверка корректур соответствует схеме испытаний Бернулли, в которой проверяющие обнаруживают ошибку с вероятностями соответственно p_1 и p_2. Использовать закон больших чисел.) *Замечание*. В этой задаче в простых терминах описывается экспериментальный прием, использованный Резерфордом для подсчета сцинтилляций.

24. Чтобы оценить численность популяций животных путем их частичного отлова [1]), ловушки расставляются последовательно r раз. Пусть каждое животное

[1]) Moran P. A. P., A mathematical theory of animal trapping, Biometrika, 38 (1951), 307—311.

попадает в ловушку с вероятностью q; предположим, что первоначально было n животных и что изменение ситуации между двумя последовательными отловами заключается в изменении численности популяции (так как пойманные животные исключаются из рассмотрения). Найти вероятность того, что r последовательных отловов дадут соответственно $n_1, n_2, \ldots, n_r$ пойманных животных.

25. *Сложные испытания Бернулли.* В примере 9,в) найти условную вероятности p и q соответственно событий $(У, Н)$ и $(Н, У)$, предполагая, что одна из этих комбинаций осуществилась. Показать, что $p>1/2$ в случае $p_1>p_2$ и $p<1/2$ в случае $p_2>p_1$.

26. *Продолжение* [1]. Показать, что если в n парах испытаний ровно m пар привели к одной из комбинаций $(У, Н)$ или $(Н, У)$, то вероятность появления $(У, Н)$ ровно k раз равна $b(k;\ m,\ p)$.

27. *Смесь биномиального и пуассоновского распределений.* Пусть некоторое насекомое откладывает r яиц с вероятностью $p(r;\ \lambda)$, а вероятность развития насекомого из яйца равна p. Предполагая взаимную независимость развития яиц, показать, что вероятность появления k новых насекомых дается распределением Пуассона с параметром λp. *Замечание.* Другой пример аналогичной ситуации: вероятность разрыва k хромосом равна $p(k;\ \lambda)$, а вероятность воссоединения фрагментов равна p (другие примеры такого рода см. пример гл. IX,1,г) и в гл. XII,1).

---

28. Доказать следующую теорему [2]: максимальная вероятность полиномиального распределения (9.2) удовлетворяет неравенствам

$$np_i-1 < k_i \leqslant (n+r-1)\,p_i, \qquad i=1,\ 2,\ \ldots,\ r. \tag{10.1}$$

Указание. Сначала доказать, что вероятность является максимальной тогда и только тогда, когда $(p_i k_j \leqslant p_j (k_i+1)$ для каждой пары $(i,\ j)$. Сложить эти неравенства для всех j и для всех $i \neq j$.

29. Доказать, что вероятности $p(k;\ \lambda)$ распределения Пуассона достигают максимума, когда k есть наибольшее целое, не превосходящее λ.

Замечание. В задачах 30—34 используется пуассоновское приближение для биномиального распределения. Предполагается, что $\lambda=np$.

30. Показать, что отношение $a_k=b(k;\ n,\ p,)/p(k;\ \lambda)$ при изменении k от 0 до ∞ сначала растет, а затем убывает, достигая максимума, когда k есть наибольшее целое, не превосходящее $\lambda+1$.

31. Доказать, что при увеличении k вероятности $b(k;\ n,\ p)$ сначала меньше, затем больше и потом опять меньше, чем $p(k;\ \lambda)$.

32. Доказать, что если $n\to\infty$ и $p\to 0$ так, что величина $np=\lambda$ остается постоянной, то

$$b(k;\ n,\ p)\to p(k;\ \lambda)$$

равномерно по k.

33. Показать, что

$$\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} \geqslant b\,(k;\ n,\ p) \geqslant \frac{\lambda^k}{k!}\left(1-\frac{k}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}. \tag{10.2}$$

[1] Wald A., Sequential tests of statistical hypotheses, Ann. Math. Statist., 16 (1945), 166. Вальд использует приведенные выше результаты, чтобы создать практический метод сравнения двух эмпирически полученных последовательностей испытаний (например, результаты работы двух машин) для выбора той, в которой вероятность успеха больше. Он сводит задачу к более простой задаче нахождения того, будет ли в последовательности испытаний Бернулли частота успеха значительно отличаться от 1/2.

[2] В первом издании утверждалось лишь, что $|k_i-np_i|\leqslant r$. Настоящее уточнение и его изящное доказательство принадлежат П. Морану.

34. Вывести из (10.2) неравенства

$$p(k;\lambda)\,e^{k\lambda/n} > b(k;n,p) > p(k;\lambda)\,e^{-k^2/(n-k)-\lambda^2/(n-\lambda)}. \tag{10.3}$$

Указание. Использовать (12.26) из гл. II.

Замечание. Хотя неравенства (10.2) являются очень грубыми, соотношения (10.3) дают превосходные оценки погрешности приближения. При помощи вычислений, аналогичных выполненным в гл. II, 9, неравенства (10.3) нетрудно уточнить. Между прочим, из результата задачи 30 очевидно вытекает, что экспоненту в левой части (10.3) можно заменить величиной $m\lambda/n$, не превышающей $(p+n^{-1})\lambda$.

Другие предельные теоремы

35. *Биномиальное приближение для гипергеометрического распределения.* Генеральная совокупность из N элементов состоит из красных и черных элементов, число которых относится как $p:q$ (где $p+q=1$). Берется выборка без возвращения объема n. Вероятность того, что она содержит ровно k красных элементов, дается гипергеометрическим распределением, введенным в гл. II, 6. Доказать, что эта вероятность стремится к $b(k;n,p)$ при $N\to\infty$.

36. Предположим, что в условиях предыдущей задачи p мало, n велико, а $\lambda=np$ и не мало, и не велико. Тогда гипергеометрическое распределение можно аппроксимировать распределением Пуассона $p(k;\lambda)$. Проверить это непосредственно, не используя пуассоновское приближение для биномиального распределения.

37. Пусть в *отрицательном биномиальном распределении* $\{f(k;r,p)\}$ из § 8 $q\to0$ и $r\to\infty$ таким образом, что произведение $rq=\lambda$ постоянно. Показать, что

$$f(k;r,p)\to p(k;\lambda).$$

(*Замечание.* Тем самым получаем предельную теорему для распределения Пойа; см. задачу 24 гл. V, 8.)

38. *Многомерное распределение Пуассона.* Доказать, что когда n велико, а $np_j=\lambda_j$ при $j=1,\ldots,r-1$ и не мало, и не велико, полиномиальное распределение (9.2) можно аппроксимировать распределением

$$e^{-(\lambda_1+\ldots+\lambda_{r-1})}\,\lambda_1^{k_1}\lambda_2^{k_2}\ldots\lambda_{r-1}^{k_{r-1}}/(k_1!\,k_2!\ldots k_{r-1}!).$$

Доказать также, что члены этого распределения в сумме дают единицу. (Заметим, что в задаче 17 мы имели дело с двумерным распределением Пуассона.)

39. а) Вывести (3.6) непосредственно из (3.5), используя очевидное соотношение

$$b(k;n,p)=b(n-k;n,q).$$

б) Вывести биномиальное распределение как с помощью общей формулы суммирования (3.1) гл. IV, так и по индукции.

40. Доказать, что $\sum kb(k;n,p)=np$ и $\sum k^2b(k;n,p)=n^2p^2+npq$.

41. Доказать, что $\sum k^2p(k;\lambda)=\lambda^2+\lambda$.

42. Проверить тождество

$$\sum_{\nu=0}^{k} b(\nu;n_1,p)\,b(k-\nu;n_2,p)=b(k;n_1+n_2,p) \tag{10.4}$$

и пояснить его вероятностный смысл. *Указание.* Использовать равенство (6,4) гл. II.

Замечание. Соотношение (10.4) представляет собой частный случай формулы свертки, которая будет введена в гл. XI; другой пример дает (10.5).

43. Проверить тождество

$$\sum_{\nu=0}^{k} p(\nu;\ \lambda_1)\ p(k-\nu;\ \lambda_2) = p(k;\ \lambda_1+\lambda_2). \qquad (10.5)$$

44. Пусть

$$B(k;\ n,\ p) = \sum_{\nu=0}^{k} b(\nu;\ n,\ p) \qquad (10.6)$$

есть вероятность появления не более k успехов в n испытаниях. Тогда

$$\begin{gathered} B(k;\ n+1,\ p) = B(k;\ n,\ p) - pb(k;\ n,\ p), \\ B(k+1;\ n+1,\ p) = B(k;\ n,\ p) + qb(k+1;\ n,\ p). \end{gathered} \qquad (10.7)$$

Проверить эти равенства, а) исходя из определения, б) аналитически.

45. Доказать, что в тех же обозначениях[1])

$$B(k;\ n,\ p) = (n-k)\binom{n}{k}\int_0^q t^{n-k-1}(1-t)^k dt \qquad (10.8)$$

и

$$1 - B(k;\ n,\ p) = n\binom{n-1}{k}\int_0^p t^k (1-t)^{n-k-1}\, dt. \qquad (10.9)$$

Указание. Проинтегрировать по частям или продифференцировать обе части по p. Вывести одну формулу из другой.

46. Доказать, что

$$p(0;\ \lambda) + \ldots + p(n;\ \lambda) = (1/n!)\int_\lambda^\infty e^{-x} x^n\, dx. \qquad (10.10)$$

---

[1]) Интеграл в (10.9) представляет собой *неполную бета-функцию*. Таблица для $1-B(k;\ n,\ p)$ с точностью до 7 знаков после запятой для k и n, не превосходящих 50, и $p=0{,}01$, 0,02 0,03, ... даны в работе Pearson K., Tables of the incomplete beta function, London, Biometrica Office. 1934. [Имеется перевод: Пирсон К. Таблицы неполной бета-функции.— М.: ВЦ АН СССР, 1974.]

ГЛАВА VII

НОРМАЛЬНОЕ ПРИБЛИЖЕНИЕ ДЛЯ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Нормальное приближение для биномиального распределения имеет важное теоретическое и практическое значение. Оно сыграло большую роль в развитии теории вероятностей, так как привело к первой предельной теореме. С современной точки зрения эта теорема является лишь частным случаем *центральной предельной теоремы*, к которой мы обратимся в гл. X, но полное исследование которой должно быть отложено до тома 2.

Частный случай $p=1/2$ рассматривался в гл. III при выводе предельных теорем о первом достижении, о числе перемен знака и т. д. Этот частный случай особенно прост и отдельно исследуется в § 2.

§ 1. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Во избежание дальнейших перерывов в изложении основных идей сделаем здесь отступление и введем две чрезвычайно важные функции.

Определение. *Функция, определяемая формулой*

$$\mathfrak{n}(x) = (1/\sqrt{2\pi})\, e^{-x^2/2}, \tag{1.1}$$

называется плотностью нормального распределения; ее интеграл

$$\mathfrak{N}(x) = (1/\sqrt{2\pi}) \int\limits_{-\infty}^{x} e^{-y^2/2}\, dy \tag{1.2}$$

называется нормальной функцией распределения.

График $\mathfrak{n}(x)$ — симметричная колоколообразная кривая, изображенная на рис. 1. Заметим, что вдоль осей x и y мы используем разный масштаб. Максимум $\mathfrak{n}(x)$ равен $1/\sqrt{2\pi} \approx 0{,}399$, так что в обычной декартовой системе график $\mathfrak{n}(x)$ был бы более пологим. (Обозначения $\mathfrak{n}$ и $\mathfrak{N}$ нестандартны. В первых двух изданиях были более традиционные обозначения φ и Φ, но в томе 2 для последовательности изложения необходимо, чтобы эти символы использовались в других целях.)

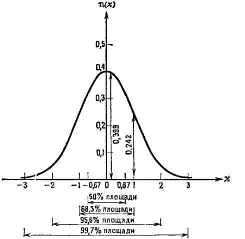

Рис. 1. Плотность нормального распределения $\mathfrak{n}$.

Лемма 1. *Площадь, ограниченная графиком функции* $\mathfrak{n}(x)$ *и осью* x, *равна* 1, *т. е.*

$$\int_{-\infty}^{\infty} \mathfrak{n}(x)\,dx = 1. \tag{1.3}$$

Доказательство. Имеем

$$\left\{ \int_{-\infty}^{\infty} \mathfrak{n}(x)\,dx \right\}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathfrak{n}(x)\,\mathfrak{n}(y)\,dx\,dy = \\ = [1/(2\pi)] \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy. \tag{1.4}$$

Переходя в этом двойном интеграле к полярным координатам, получаем

$$[1/(2\pi)]\int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2/2} r\,dr = \int_0^{\infty} e^{-r^2/2} r\,dr = e^{-r^2/2}\Big|_0^{\infty} = 1, \tag{1.5}$$

что и доказывает лемму. ▶

Из определения и леммы вытекает, что $\mathfrak{N}(x)$ *монотонно возрастает от* 0 *до* 1. Ее график (рис. 2) представляет собой Sобразную кривую, причем

$$\mathfrak{N}(-x) = 1 - \mathfrak{N}(x). \tag{1.6}$$

В табл. 1 приведены значения[1] $\mathfrak{N}(x)$ для положительных x; значения $\mathfrak{N}(-x)$ получаются при помощи (1.6).

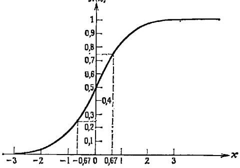

Рис. 2. Нормальная функция распределения $\mathfrak{N}$.

Для многих целей удобно иметь простую оценку «хвоста» $1-\mathfrak{N}(x)$ при больших x. Такую оценку дает следующая лемма.

Лемма 2[2]. *При* $x \to \infty$

$$1 - \mathfrak{N}(x) \sim x^{-1}\mathfrak{n}(x); \tag{1.7}$$

точнее, для всех $x > 0$ *справедливо двойное неравенство*

$$(x^{-1} - x^{-3})\,\mathfrak{n}(x) < 1 - \mathfrak{N}(x) < x^{-1}\mathfrak{n}(x). \tag{1.8}$$

(См. задачу 1.)

Доказательство. Очевидно,

$$[1 - 3x^{-4}]\,\mathfrak{n}(x) < \mathfrak{n}(x) < [1 + x^{-2}]\,\mathfrak{n}(x). \tag{1.9}$$

Члены этого неравенства представляют собой производные членов из (1.8), взятые с обратным знаком, поэтому (1.8) вытекает из (1.9) после интегрирования по полупрямой $[x, \infty)$.

---

[1]) Более подробные таблицы см. в Tables of Probability Functions, v. 2, Nat. Bureau of Standards, New York, 1942. [Имеется перевод: Таблицы вероятностных функций, Том II. — М.: ВЦ АН СССР, 1959.] Значения функций $\mathfrak{n}(x)$ и $\mathfrak{N}(x) - \mathfrak{N}(-x)$ приведены в этих таблицах с 15 знаками после запятой. При $0 < x < 1$ значения x берутся через 0,0001, при $x > 1$ — через 0,001.

[2]) Здесь и в дальнейшем знак $\sim$ означает, что отношение величин, соединенных этим знаком, стремится к 1.

Таблица 1

Нормальная функция распределения $\mathfrak{N}(x)$

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5159	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8380
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8718	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9083	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9758	0,9762	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9980	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9984	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9986	0,9987	0,9987	0,9988	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995

При $x<0$ используется соотношение $\mathfrak{N}(-x)=1-\mathfrak{N}(x)$.

Замечание о терминологии. Термин *функция распределения* используется в математической литературе для неубывающих функций от x, стремящихся к 0 при $x\to-\infty$ и к 1 при $x\to\infty$. В настоящее время статистики предпочитают термин *кумулятивная функция распределения*, но прилагательное «кумулятивная» является излишним. *Плотность распределения* — это неотрицательная функция $f(x)$, интеграл от которой по всей оси x равен единице. Интеграл от $-\infty$ до x от любой плотности распределения является функцией распределения. Применявшийся ранее термин *функция частот* (frequency function) представляет собой синоним для плотности распределения.

Нормальное распределение часто называют *гауссовским распределением*, но оно использовалось в теории вероятностей еще Муавром и Лапласом. Если изменить начало отсчета и единицу измерения, то $\mathfrak{N}(x)$ преобразуется в $\mathfrak{N}((x-a)/b)$. Эта функция называется *функцией распределения нормального закона со средним значением* a *и дисперсией* b^2 (или стандартным отклонением $|b|$). Функцию $2\mathfrak{N}(x\sqrt{2})-1$ часто называют *функцией ошибок*.

§ 2. СИММЕТРИЧНЫЕ РАСПРЕДЕЛЕНИЯ

Рассмотрим, как используется нормальное распределение в качестве приближения для биномиального распределения с $p=1/2$.

Есть две причины, по которым частный случай $p=1/2$ выделяется особо. Во-первых, вычисления здесь намного проще, и, следовательно, легче объяснить, как в этой задаче возникает нормальное распределение. Во-вторых, данный частный случай рассматривался в связи со случайными блужданиями (см. гл. III, 2), и, следовательно, желательно иметь соответствующие выкладки, не отягощенные техническими деталями, необходимыми для несимметричных распределений.

Для определенности возьмем $n=2\nu$ четным и для упрощения положим

$$a_k=b(\nu+k;\,2\nu,\,1/2), \tag{2.1}$$

т. е. a_k — члены симметричного биномиального распределения, перенумерованные в зависимости от их расстояния до максимальной вероятности, a_0 — максимальная вероятность и k изменяется от $-\nu$ до ν. Поскольку $a_{-k}=a_k$, рассмотрим только $k\geqslant 0$.

(*В обозначениях гл.* III *имеем* $a_k=p_{2\nu,\,2k}$. Приведенное ниже доказательство могло быть там же и помещено, так как не использует идеи, развитые после § 2 гл. III.)

Чтобы понять, как ведет себя последовательность a_0, a_1, a_2, ..., сравним ее общий член с a_0, используя соотношение

$$a_k=a_0\frac{\nu(\nu-1)\ldots(\nu-k+1)}{(\nu+1)(\nu+2)\ldots(\nu+k)}, \tag{2.2}$$

которое легко получить из определения (2.1).

Нас интересует поведение последовательности только при больших значениях ν, и поэтому будем рассматривать лишь те значения k, при которых отношение k/ν мало, так как при остальных k члены a_k пренебрежимо малы. Разделив числитель и знаменатель дроби в (2.2) на ν^k, получим дробь с членами вида $1+j/\nu$, где j изменяется от $-(k-1)$ до k. Далее,

$$1+j/\nu=e^{j/\nu+\cdots}, \tag{2.3}$$

где многоточие означает члены, сумма которых меньше, чем $(j/\nu)^2$. Для такого приближения дробь в (2.2) сводится к экспоненте с показателем

$$-(2/\nu)\,[1+\ldots+(k-1)]-k/\nu=-k^2/\nu$$

и погрешность приближения меньше, чем k^3/ν^2. Соответственно при $\nu\to\infty$ и k из интервала $0<k<K_\nu$, где

$$K_\nu^3/\nu^3\to 0, \tag{2.4}$$

имеем приближение

$$a_k\sim a_0 e^{-k^2/\nu}. \tag{2.5}$$

Если выразить биномиальные коэффициенты через факториалы, то, используя формулу Стирлинга (9.1) гл. II[1]), видим, что

$$a_0=\binom{2\nu}{\nu}2^{-2\nu}\sim\frac{1}{\sqrt{\pi\nu}}. \tag{2.6}$$

Подставляя это в (2.5), получаем

$$a_k\sim h\mathfrak{n}(kh),\quad \text{где } h=\sqrt{2/\nu}=2/\sqrt{n}. \tag{2.7}$$

Это основное соотношение справедливо при $\nu\to\infty$ и при k, не превосходящих чисел K_ν, которые удовлетворяют условию (2.4). Мы будем использовать (2.7) только для значений k порядка $\sqrt{\nu}$; в этом случае (2.4) тривиально выполняется.

На практике нам обычно требуются приближения для вероятностей попадания в различные интервалы, т. е. для сумм вида[2])

$$A(x_1,\ x_2)=\sum_{x_1\leqslant k\leqslant x_2}a_k, \tag{2.8}$$

где суммирование проводится по всем целым числам между x_1 и x_2 включительно. Сейчас мы покажем, как можно найти прибли-

---

[1]) *Замечание о постоянной в формуле Стирлинга.* Напомним, что при доказательстве формулы Стирлинга в § 9 гл. II не было показано, что постоянная в ней равна $\sqrt{2\pi}$. Восполним этот пробел следующим образом. Постоянную π в (2.6) мы должны заменить на неизвестную постоянную. В теореме об аппроксимации (см. ниже) это приведет лишь к тому, что правую часть (2.10) следует умножить на неизвестную постоянную c, и нужно доказать, что $c=1$. Возьмем в (2.10) $z_1=0$. Отношение обеих частей (2.10) стремится к 1 при $n\to\infty$. Но из оценки «хвоста» (3.5) гл. VI вытекает, что левая часть лежит между 1/2 и $1/2-4z_3^{-2}$, в то время как для правой части из (1.8) следует двойное неравенство

$$c>c\,[\mathfrak{N}(z_2)-1/2]=(1/2)\,c-c\,[1-\mathfrak{N}(z_2)]>(1/2)\,c-c\mathfrak{n}(z_2)/z_2.$$

Для достаточно больших z_2 обе части (2.10) сколь угодно близки соответственно к 1/2 и $c/2$, и поэтому $c=1$, как и утверждалось.

[2]) Мы не стали обозначать сумму в (2.8) через $\mathbf{S}_n$, так как этот символ в гл. III и VI имеет другой смысл. На языке теории случайных блужданий $A(x_1,x_2)$ есть вероятность того, что в момент $n=2\nu$ частица находится между $2x_1$ и $2x_2$; в терминологии настоящей главы $A(x_1,x_2)$ есть вероятность того, что число успехов в $n=2\nu$ испытаниях лежит между $\nu+x_1$ и $\nu+x_2$. В следующем параграфе это число снова будет обозначаться через $\mathbf{S}_n$.

жение для $A(x)$ через площадь области, лежащей под графиком $\mathfrak{n}(x)$, а эту площадь, в свою очередь, можно выразить через интеграл $\mathfrak{N}$. В силу монотонности $\mathfrak{n}$ ясно, что площадь области, лежащей под графиком $\mathfrak{n}$ между kh и $(k+1)h$, меньше, чем $h\mathfrak{n}(kh)$, но больше, чем $h\mathfrak{n}((k+1)h)$, откуда следует, что

$$\int\limits_{x_1h}^{x_2h+h} \mathfrak{n}(s)\,ds < \sum_{x_1 \leqslant k \leqslant x_2} h\mathfrak{n}(kh) < \int\limits_{x_1h-h}^{x_2h} \mathfrak{n}(s)\,ds. \tag{2.9}$$

Отсюда вытекает, что средний член в (2.9) есть приближение для $A(x_1, x_2)$, которое будет хорошим, когда ν велико и отношение k^2/ν не мало и не велико, т. е. когда h мало, а xh не мало и не велико. Два крайних члена в (2.9) равны соответственно $\mathfrak{N}(x_2h+h)-\mathfrak{N}(x_1h)$ и $\mathfrak{N}(x_2h)-\mathfrak{N}(x_1h-h)$. Их разность стремится к 0 при $h \to 0$, и поэтому их можно заменить на $\mathfrak{N}(x_2h)-\mathfrak{N}(x_1h)$.

Сформулируем полученный результат в виде предельной теоремы, заменив при этом переменную x на $z=xh$.

Теорема об аппроксимации. *Для фиксированных $z_1 < z_2$ справедливо соотношение*

$$\sum_{(1/2)z_1\sqrt{n} \leqslant k \leqslant (1/2)z_2\sqrt{n}} a_k \to \mathfrak{N}(z_2)-\mathfrak{N}(z_1). \tag{2.10}$$

Вскоре мы увидим, что эта теорема существенно обобщается в определенных случаях, когда z_1 и z_2 могут меняться вместе с n, не оставаясь ограниченными. Заметим, что предельная теорема (2.7) гл. III содержится в (2.10), которая, в свою очередь, есть лишь частный случай общей теоремы из следующего параграфа.

Оценки погрешности. Нам не нужно беспокоиться о погрешности, допускаемой при замене суммы интегралом, поскольку (2.9) содержит верхние и нижние оценки.

Чтобы оценить погрешность приближения в (2.7) положим

$$a_k = a_0 e^{-k^2/\nu+\varepsilon_1} = h\mathfrak{n}(kh)\, e^{\varepsilon_1-\varepsilon_2}, \tag{2.11}$$

так что ε_1 представляет собой погрешность, допускаемую при опускании членов более высокого порядка в (2.3), а величина ε_2 определяется из (2.6). Из наших рассуждений видно, что

$$\varepsilon_1 = \sum_{j=1}^{k-1}\left(\log\frac{1+j/\nu}{1-j/\nu}-\frac{2j}{\nu}\right)+\left(\log\left(1+\frac{k}{\nu}\right)-\frac{k}{\nu}\right). \tag{2.12}$$

Оценки погрешности приближения наиболее интересны для относительно малых ν, и для рассмотрения этого случая будем предполагать только, что $k < (1/3)\nu$. Сравнивая разложение (8.11) гл. II с суммой геометрической прогрессии со знаменателем $1/3$, мы видим, что каждый член ряда (2.12) положителен и меньше, чем $(j/\nu)^3$. Следовательно, сумма ряда положительна и меньше, чем $k^4/(4\nu)^3$. Из (8.9) гл. II аналогично выводится, что последнее слагаемое в правой части (2.12) отрицательно и больше, чем $-3k^2/(4\nu^2)$. Таким образом,

$$-3k^2/n^2 < \varepsilon_1 < 2k^4/n^3, \quad \text{если } k < (1/6)\,n. \tag{2.13}$$

Как правило, в приложениях величины k и $\sqrt{n}$ сравнимы, и тогда условие $k < n/6$ тривиально выполняется. При данных обстоятельствах неравенства (2.13) достаточно точны.

Что касается (2.6), то, как следует из уточненной формулы Стирлинга (9.15) гл. II, для a_0 получается лучшее приближение, если умножить правую часть (2.6) на $e^{1/(4n)}$, и в любом случае

$$1/(4n) - 1/(20n^3) < \varepsilon_2 < 1/(4n) + 1/(360n^3). \tag{2.14}$$

Итак, найдены *точные оценки погрешностей приближений* (2.7) *и* (2.10). Эти оценки применимы даже для относительно небольших значений n.

Основной результат наших исследований заключается в том, что относительная погрешность приближения (2.7) имеет порядок k^2/n^2 или k^4/n^3 в зависимости от того, какая из этих величин больше. На практике эти оценки обычно используются, когда k^2/n велико, и в этом случае относительная погрешность имеет порядок k^4/n^3. Наши оценки указывают также путь, как улучшить приближение с помощью соответствующего изменения членов (см. задачу 14).

§ 3. ПРЕДЕЛЬНАЯ ТЕОРЕМА МУАВРА — ЛАПЛАСА

Сейчас будет показано, как полученные выше приближения можно обобщить на произвольное биномиальное распределение с $p \neq 1/2$. Рассуждения те же, но потребуется больше вычислений. Первая сложность возникает в связи с максимальной вероятностью распределения. Как следует из (3.2) гл. VI, индекс m максимальной вероятности—это единственное целое число вида

$$m = np + \delta, \quad \text{где } -q < \delta \leqslant p. \tag{3.1}$$

Величиной δ в конечном счете мы будем пренебрегать, но в вычислениях она появится. (В случае $p = 1/2$ δ не было, так как мы предполагали, что $n = 2\nu$ четно.)

Как и в предыдущем параграфе, перенумеруем члены биномиального распределения и запишем

$$a_k = b(m+k;\ n,\ p) = \binom{n}{m+k} p^{m+k} q^{n-m-k}. \tag{3.2}$$

Для определенности предположим, что $k > 0$; для $k < 0$ рассуждения аналогичны. (В случае $k < 0$ нужно лишь поменять местами p и q.) Аналогично (2.2) имеем

$$a_k = a_0 \frac{(n-m)(n-m-1)\ldots(n-m-k+1)\,p^k}{(m+1)(m+2)\ldots(m+k)\,q^k}. \tag{3.3}$$

Это можно записать иначе:

$$a_k = a_0 \frac{(1-pt_0)(1-pt_1)\ldots(1-pt_{k-1})}{(1+qt_0)(1+qt_1)\ldots(1+qt_{k-1})}, \tag{3.4}$$

где для сокращения записи использовано обозначение

$$t_j = (j + \delta + q)/[(n+1)\,pq]. \tag{3.5}$$

Воспользуемся теперь представлением (3.4) только для тех значений k, для которых t_k мало, скажем $t_k < 1/2$. Из разложения Тейлора для логарифма (8.9) гл. II видно, что

$$(1-pt_j)/(1+qt_j)=e^{-t_j+\cdots}, \tag{3.6}$$

где опущенный член по абсолютной величине не превосходит t_j^2. Таким образом,

$$a_k = a_0 e^{-(t_0+\ldots+t_{k-1})+\cdots}, \tag{3.7}$$

где многоточие соответствует члену, абсолютная величина которого[1]) меньше, чем $kt_{k-1}^2 < k^3/(npq)^2$. Далее,

$$t_0+t_1+\ldots+t_{k-1}=\frac{(1/2)\,k\,(k-1)+k\,(\delta+q)}{(n+1)\,pq}. \tag{3.8}$$

Для простоты заменим правую часть (3.8) на $k^2/(2npq)$. При этом величина допускаемой погрешности меньше, чем $2k/(npq)$. Итак, если мы запишем

$$a_k = a_0 e^{-k^2/(2npq)+\rho_k}, \tag{3.9}$$

то член ρ_k, характеризующий погрешность приближения, удовлетворяет неравенству

$$|\rho_k| < k^3/(npq)^2 + 2k/(npq). \tag{3.10}$$

Покажем теперь, что

$$a_0=\frac{n!}{m!\,(n-m)!}\,p^m q^{n-m} \sim \frac{1}{\sqrt{2\pi npq}}, \tag{3.11}$$

что обобщает аналогичное соотношение (2.6) для симметричного случая. В идеальном случае, когда $p=m/n$, оценка (3.11) является немедленным следствием формулы Стирлинга (9.1) гл. II. Непосредственное дифференцирование показывает, что средний член (3.11) достигает максимума при $p=m/n$. При заданном m нужно рассмотреть только те значения p, для которых выполнено (3.1), и тогда минимум a_0 достигается в одной из крайних точек, а именно для $p=m/(n+1)$ или $p=(m+1)/(n+1)$. Для этих значений p непосредственное использование формулы Стирлинга снова приводит к (3.11), только n заменяется на $n+1$. Следовательно, (3.11) справедливо для всех возможных значений p. Если для сокращения записи положить

$$h=1/\sqrt{npq}, \tag{3.12}$$

то из (3.9) вытекает соотношение

$$a_k \sim h\mathfrak{n}(kh), \tag{3.13}$$

---

[1]) Мы удовлетворимся очень грубыми оценками погрешности приближения

но при условии, что k изменяется вместе с n таким образом, что $\rho_k \to 0$. Итак, доказана следующая теорема.

Теорема 1. *Если $n \to \infty$ и k изменяется в интервале $k < K_n$, причем $K_n^3/n^2 \to 0$, то соотношение* (3.13) *справедливо*[1]) *равномерно по k, т. е. для любого $\varepsilon > 0$ и достаточно больших n*

$$1-\varepsilon < a_k/[h\mathfrak{n}(kh)] < 1+\varepsilon. \tag{3.14}$$

Пример. Рис. 3 иллюстрирует случай $n=10$ и $p=1/5$, когда $npq=1,6$. С учетом того, что n очень мало, приближение кажется поразительно хорошим. Для $k=0, \ldots, 6$ вероятности $b(k; n, p)$ рав-

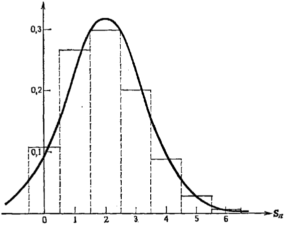

Рис. 3. Нормальное приближение для биномиального распределения. Ступенчатая функция дает вероятности $b(k; 10, 1/5)$ k успехов в 10 испытаниях Бернулли с $p=1/5$. Непрерывная кривая дает для каждого целого k соответствующее нормальное приближение.

ны 0,1074, 0,2684, 0,3020, 0,2013, 0,0880, 0,0264, 0,0055. Соответствующие приближения, согласно (3.13), равны 0,0904, 0,2307, 0,3154, 0,2307, 0,0904, 0,0189, 0,0021. ►

Теорема 1 используется в основном для вывода приближений для вероятностей вида

$$\mathbf{P}\{\alpha \leqslant \mathbf{S}_n \leqslant \beta\} = \sum_{\nu=\alpha}^{\beta} b(\nu; n, p) = \sum_{k=\alpha-m}^{\beta-m} a_k. \tag{3.15}$$

[1]) Если k изменяется вместе с n так, что $k^3/n^2 \to \infty$, то нормальное приближение заменяется приближением другого вида, см. задачи 13 и 15.

В области применимости теоремы 1 хорошее приближение получается при замене a_k на $h\mathfrak{n}(kh)$. Последнее произведение можно интерпретировать как площадь прямоугольника высотой $\mathfrak{n}(kh)$, основанием которого является интервал длиной h с центром в kh (см. рис. 3). Как обычно, заменим площадь прямоугольника площадью соответствующей области, заключенной между осью x и графиком функции $\mathfrak{n}$; известно, что получающаяся при этом погрешность в пределе пренебрежимо мала при $h\to 0$. Таким образом, для целых α и β получаем приближение

$$\mathbf{P}\{\alpha\leqslant \mathbf{S}_n\leqslant\beta\}\approx\mathfrak{N}((\alpha-m+1/2)h)-\mathfrak{N}((\beta-m-1/2)h). \quad (3.16)$$

Нормальное приближение в такой форме целесообразно использовать, когда h лишь относительно мало и требуется наилучшая возможная точность. Однако в окончательной формулировке предпочтительно заменить аргументы в правой части (3.16) более простыми выражениями $z_1=(\alpha-np)h$ и $z_2=(\beta-np)h$; погрешность, получающаяся при этом упрощении, очевидно, стремится к нулю при $h\to 0$. Итак, доказана следующая основная теорема.

Теорема 2. (Предельная теорема Муавра — Лапласа.) *Для фиксированных* [1] z_1 *и* z_2 *при* $n\to\infty$ *справедливо соотношение*

$$\mathbf{P}\{np+z_1\sqrt{npq}\leqslant \mathbf{S}_n\leqslant np+z_2\sqrt{npq}\}\to\mathfrak{N}(z_2)-\mathfrak{N}(z_1). \quad (3.17)$$

Помимо большой теоретической важности эта теорема интересна тем, что оправдывает использование правой части (3.17) в качестве приближения для левой части. Из (3.10) легко получить хорошие оценки погрешности приближения, но мы не будем останавливаться на этом вопросе. Практические примеры можно найти в следующем параграфе.

Предельное соотношение (3.17) принимает более приятный вид, если $\mathbf{S}_n$ заменить нормированным числом успехов $\mathbf{S}_n^*$, определяемым по формуле

$$\mathbf{S}_n^*=(\mathbf{S}_n-np)/\sqrt{npq}. \quad (3.18)$$

Это равносильно выбору $\sqrt{npq}$ в качестве единицы измерения отклонения $\mathbf{S}_n$ от np. В терминах случайных величин (гл. IX) np будет называться *математическим ожиданием*, а npq *дисперсией* случайной величины $\mathbf{S}_n$ (квадратный корень $\sqrt{npq}$ есть стандартное отклонение). Неравенство в левой части (3.17) эквивалентно следующему: $z_1\leqslant \mathbf{S}_n^*\leqslant z_2$, и поэтому (3.17) можно *переписать в виде*

$$\mathbf{P}\{z_1\leqslant \mathbf{S}_n^*\leqslant z_2\}\to\mathfrak{N}(z_2)-\mathfrak{N}(z_1). \quad (3.19)$$

[1] Из теоремы 1 видно, что это условие можно ослабить. См. также § 6 и задачи 14 и 16.

Как правило, мы будем использовать предельную теорему в этом виде. Из (3.19), в частности, следует, что для больших n вероятность, стоящая в левой части, практически не зависит от p. Это позволяет сравнивать случайные флуктуации в различных последовательностях испытаний Бернулли, просто относя их к нашим стандартным единицам.

Замечание об остановке в произвольный момент времени. Необходимо отметить, что наши предельные теоремы и приближенные формулы справедливы тогда, когда число n испытаний фиксировано заранее независимо от исхода испытаний. Если игрок имеет право остановить игру в благоприятный для него момент, то его общий выигрыш нельзя оценить при помощи нормального приближения, так как продолжительность игры теперь случайна. Для любого фиксированного n очень маловероятно, что $\mathbf{S}_n^*$ велико, однако при большом числе испытаний даже самое маловероятное событие обязательно произойдет. Мы увидим, что в длительной игре $\mathbf{S}_n^*$ практически наверняка будет иметь последовательность максимумов порядка величины $\sqrt{2\log\log n}$ (закон повторного логарифма, см. гл. VIII, 5).

§ 4. ПРИМЕРЫ

а) Пусть $p=1/2$ и $n=200$. Рассмотрим $\mathbf{P}\{95\leqslant \mathbf{S}_n\leqslant 105\}$, т. е. вероятность того, что при 200 бросаниях моменты число выпадений герба отличается от 100 не более чем на 5. Здесь $h=1/\sqrt{50}=$ $=0{,}141421\ldots$ относительно велико, и нужно быть внимательным при определении границ интервала. Использование (3.16) приводит к приближению

$$\mathbf{P}\{95\leqslant \mathbf{S}_n\leqslant 105\}\approx \mathfrak{N}(5{,}5h)-\mathfrak{N}(-5{,}5h)= \\ =2\mathfrak{N}(0{,}7778\ldots)-1=0{,}56331.$$

Истинное значение вероятности равно 0,56325... . Малостью погрешности мы в основном обязаны симметричности распределения.

б) Пусть $p=1/10$ и $n=500$. Здесь $h=1/\sqrt{45}=0{,}14907\ldots$. Рассуждая, как и выше, получаем

$$\mathbf{P}\{50\leqslant \mathbf{S}_n\leqslant 55\}\approx \mathfrak{N}(5{,}5h)-\mathfrak{N}(-0{,}5h)= \\ =\mathfrak{N}(5{,}5h)+\mathfrak{N}(0{,}5h)-1=0{,}3235\ldots ,$$

в то время как истинное значение вероятности равно 0,3176... . Погрешность составляет около 2%.

в) Вероятность того, что $\mathbf{S}_n$ лежит в пределах $np\pm 2\sqrt{npq}$, приблизительно равна $\mathfrak{N}(2)-\mathfrak{N}(-2)=0{,}9545$; для $np\pm 3\sqrt{npq}$ соответствующая вероятность приближенно равна 0,9973. Удивительно, в сколь узких границах с большой вероятностью лежат случайные флуктуации. Например, вероятность того, что при 10^6 бросаниях монеты число выпадений герба будет отличаться от среднего значения 500 000 более чем на 1000, меньше 0,0455.

Таблица 2

Сравнение биномиального распределения при $n=100$, $p=0,3$ и нормального приближения

Число успехов	Вероятность	Нормальное приближение	Погрешность в процентах
$9\leqslant S_n\leqslant 11$	0,000006	0,00003	+400
$12\leqslant S_n\leqslant 14$	0,00015	0,00033	+100
$15\leqslant S_n\leqslant 17$	0,00201	0,00283	+40
$18\leqslant S_n\leqslant 20$	0,01430	0,01599	+12
$21\leqslant S_n\leqslant 23$	0,05907	0,05895	0
$24\leqslant S_n\leqslant 26$	0,14887	0,14447	−3
$27\leqslant S_n\leqslant 29$	0,23794	0,23405	−2
$31\leqslant S_n\leqslant 33$	0,23013	0,23405	+2
$34\leqslant S_n\leqslant 36$	0,14086	0,14447	+3
$37\leqslant S_n\leqslant 39$	0,05889	0,05895	0
$40\leqslant S_n\leqslant 42$	0,01702	0,01599	−6
$43\leqslant S_n\leqslant 45$	0,00343	0,00283	−18
$46\leqslant S_n\leqslant 48$	0,00049	0,00033	−33
$49\leqslant S_n\leqslant 51$	0,00005	0,00003	−40

г) Пусть $n=100$ и $p=0,3$. В табл. 2 приведен типичный пример (для относительно небольших n), показывающий, как изменяется точность нормального приближения при удалении интервала (α, β) от центрального члена.

д) Найдем число a, такое, что для больших n неравенство $|S_n^*|>a$ выполняется с вероятностью, близкой к 1/2. Для этого необходимо, чтобы

$$\mathfrak{N}(a)-\mathfrak{N}(-a)=1/2,$$

или $\mathfrak{N}(a)=3/4$. По таблицам для нормального распределения находим $a=0,6745$, и поэтому неравенства

$$|S_n-np|<0,6745\sqrt{npq} \quad \text{и} \quad |S_n-np|>0,6745\sqrt{npq} \qquad (4.1)$$

приближенно равновероятны. В частности, вероятность того, что при n бросаниях монеты число выпадений герба лежит в пределах $(1/2)\,n\pm0,337\sqrt{n}$, приближенно равна 1/2, как и вероятность того, что при n бросаниях кости число выпадений одного очка лежит в интервале $(1/6)\,n\pm0,251\sqrt{n}$.

е) *Задача о конкуренции*. Этот пример иллюстрирует практические применения формулы (3.17). Две конкурирующие железнодорожные компании имеют на участке между Чикаго и Лос-Анджелесом по одному поезду, которые отправляются и прибывают одновременно и оборудованы примерно одинаково. Предположим, что n пассажиров выбирают поезд наугад и независимо друг от друга, так что число пассажиров в каждом поезде есть результат n испытаний Бернулли с $p=1/2$. Если число мест в поезде $s<n$, то с положительной вероятностью в нем окажется больше s пассажиров и всем мест не хватит. Используя приближение (3.17), находим

$$f(s) \approx 1-\mathfrak{N}((2s-n)/\sqrt{n}). \tag{4.2}$$

Если s столь велико, что $f(s)<0{,}01$, то число мест будет достаточным в 99 случаях из 100. Вообще, компания может установить произвольный уровень риска α и определить s из условия $f(s)<\alpha$. Для этого достаточно положить

$$s \geqslant (1/2)(n+t_\alpha \sqrt{n}), \tag{4.3}$$

где t_α — корень уравнения $\alpha=1-\mathfrak{N}(t_\alpha)$, который можно найти по таблицам. Например, если $n=1000$ и $\alpha=0{,}01$, то $t_\alpha \approx 2{,}33$ и достаточно $s=537$ мест. Если обе компании примут уровень риска $\alpha=0{,}01$, то два поезда будут иметь в общей сложности 1074 места, из которых 74 будут пустыми. Потери из-за конкуренции (или случайных флуктуаций) поразительно малы. Аналогично 514 мест было бы достаточно в 80% всех случаев и 549 мест — в 999 из 1000 случаев.

Подобные соображения применимы и в других работах об обслуживании с конкуренцией. Например, если m кинотеатров соперничают из-за одних и тех же n зрителей, то для вероятности успеха для каждого кинотеатра следует положить $p=1/m$ и (4.3) нужно заменить неравенством $s \geqslant m^{-1}[n+t_\alpha \sqrt{n(m-1)}]$. Общее число пустых мест при такой системе равно $ms-n \approx t_\alpha \sqrt{n(m-1)}$. Для $\alpha=0{,}01$, $n=1000$ и $m=2$, 3, 4, 5 это число соответственно приближенно равно 74, 105, 126 и 147. Потери из-за конкуренции снова малы.

ж) *Случайные цифры.* В примере гл. II,3,з) мы рассмотрели $n=1200$ испытаний с $p=0{,}3024$ и частотой успехов 0,3142. Отклонение частоты от вероятности равно $\varepsilon=0{,}0118$. Далее,

$$\mathbf{P}\{|S_n/n-p|>\varepsilon\}=\mathbf{P}\{|S_n-np|>\varepsilon n\}\approx$$
$$\approx \mathbf{P}\{|S_n-np|>0{,}880\sqrt{npq}\}\approx 2(1-\mathfrak{N}(0{,}88))\approx 0{,}379.$$

Это означает, что приблизительно в 38 из 100 подобных случаев частота успехов отклоняется от p больше, чем в нашем примере.

з) *Выборка.* В некотором обществе часть p людей курит. Число p неизвестно, и для его определения производится выборка с возвращением. Желательно найти p с погрешностью, не превосходящей 0,005. Каким должен быть объем выборки n?

Обозначим долю курящих в выборке через p'. Ясно, что, каким ни был объем выборки, нельзя быть уверенным, что $|p'-p|<0{,}005$, ибо может случиться, что выборка будет состоять только из курящих. Лучшее, чего можно добиться, это сделать маловероятным событие «величина погрешности больше заданного значения 0,005». С этой целью установим произвольный доверительный уровень α, например $\alpha=0{,}95$, и выберем n настолько большим, чтобы вероятность события $|p'-p|<0{,}005$ была не меньше α. Поскольку np' можно интерпретировать как число успехов в n испытаниях, имеем

$$\mathbf{P}\{|p'-p|<0{,}005\}=\mathbf{P}\{|\mathbf{S}_n-np|<0{,}005n\}, \qquad (4.4)$$

и мы хотим выбрать n столь большим, чтобы эта вероятность была не меньше α. Сначала по таблицам находим число z_α, для которого $\mathfrak{N}(z_\alpha)-\mathfrak{N}(-z_\alpha)=\alpha$. Затем, используя нормальное приближение, необходимо выбрать n столь большим, что $0{,}005\sqrt{n}/\sqrt{pq}\geqslant z_\alpha$, или $n\geqslant 40\,000pqz_\alpha^2$. Правая часть последнего неравенства содержит неизвестную вероятность p, но при любых обстоятельствах $pq\leqslant 1/4$, и поэтому *объем выборки* $n\geqslant 10\,000z_\alpha^2$ *будет достаточным.*

Для доверительного уровня $\alpha=0{,}95$ находим $z_\alpha=1{,}960$, и поэтому объем выборки $n=40\,000$ вполне достаточен. Осуществить выборку такого объема достаточно сложно, но требование $|p'-p|<0{,}005$ является чрезвычайно сильным. Если бы мы потребовали лишь, что $|p'-p|<0{,}01$, то достаточен был бы объем выборки 10 000 (при том же доверительном уровне). Так называемая четырехпроцентная точность означает, что $|p'-p|<0{,}045$, и в этом случае объем выборки равен лишь 475: в среднем только 5 из 100 случайных выборок такого объема дадут оценку с большей погрешностью. (На практике обычно трудно бывает получить представительную выборку любого объема.) ►

§ 5. СВЯЗЬ С ПУАССОНОВСКИМ ПРИБЛИЖЕНИЕМ

Ошибка нормального приближения мала, если npq велико. С другой стороны, при больших n и малых p вероятности $b(k;\,n,\,p)$ будут близки к пуассоновским вероятностям $p(k;\,\lambda)$ с $\lambda=np$. При малых λ можно пользоваться только пуассоновским приближением, но при больших λ возможно применение как пуассоновского, так и нормального приближений. Отсюда следует, что при больших значениях λ, видимо, можно приближать распределение Пуассона нормальным распределением, и в примере гл. X, 1, в) мы увидим, что это действительно так (см. также задачу 9). Здесь мы ограничимся иллюстрацией этого утверждения на одном численном и на одном практическом примере.

Примеры. а) Распределение Пуассона с $\lambda=100$ приписывает множеству целых чисел $a, a+1, \ldots, b$ вероятность

$$P(a,\,b)=p(a;\,100)+p(a+1;\,100)+\ldots+p(b;\,100).$$

Это распределение Пуассона можно рассматривать как приближение для биномиального распределения с $n=100\,000\,000$ и $p=10^{-6}$. Тогда $npq\approx 100$, поэтому естественно аппроксимировать это биномиальное распределение нормальным хотя бы для значений, близких к центральному члену 100. Но это означает, что вероятность $P(a, b)$ близка к разности

$$\mathfrak{N}((b-99{,}5)/10)-\mathfrak{N}((a-100{,}5)/10).$$

О степени приближения можно судить по следующей таблице.

	Точное значение	Нормальное приближение
$P(85, 90)$	0,11384	0,11049
$P(90, 95)$	0,18485	0,17950
$P(95, 105)$	0,41763	0,41768
$P(90, 110)$	0,70652	0,70628
$P(110, 115)$	0,10738	0,11049
$P(115, 120)$	0,05323	0,05335

б) *Задача о телефонных линиях.* Следующая задача с некоторыми упрощениями взята из практики [1]). Телефонная станция A должна соединить 2000 абонентов с соседней станцией B. Было бы слишком дорого и бессмысленно проводить 2000 линий из A в B. Достаточно сделать число линий N настолько большим, чтобы при обычных условиях только на один из каждых ста вызовов не сразу находилась свободная линия. Предположим, что в течение наиболее напряженного часа дня каждому абоненту требуется связь с B в среднем на 2 минуты. Мы можем сравнить положение в любой фиксированный момент этого часа с группой из 2000 испытаний каждое с вероятностью $p=1/30$ того, что потребуется линия. При обычных условиях можно считать эти испытания независимыми (хотя это перестает быть верным, когда такие события, как неожиданный ливень или землетрясение, заставляют многих людей вызывать такси или звонить в местную газету; тогда эта теория неприменима и линии будут перегружены). Итак, имеем 2000 испытаний Бернулли с $p=1/30$, и нужно найти такое наименьшее число N, при котором вероятность более N «успехов» меньше 0,01; в наших обозначениях $\mathbf{P}\{\mathbf{S}_{2000}\geqslant N\}<0{,}01$.

При *пуассоновском приближении* следует взять $\lambda=2000/30\approx 66{,}67$. По таблицам находим, что вероятность не менее 87 успехов приближенно равна 0,0097, а вероятность 86 или более успехов — примерно 0,013. Это означает, что было бы *достаточно* 87 *линий.* При использовании *нормального приближения* сначала найдем ко-

[1]) Molina E. C., Probability in engineering, Electrical Engineering, 54 (1935), 423—427; Bell Telephone System Technical Publications Monograph B-854. В этих работах задача решается используемым нами методом Пуассона, более удобным для инженеров.

рень x уравнения $1-\mathfrak{N}(x)=0,01$, который равен $x=2{,}327$. Далее, должно выполняться неравенство

$$(N-1/2-np)/\sqrt{npq}\geqslant 2{,}327.$$

Поскольку $n=2000$ и $p=1/30$, отсюда получается $N\geqslant 67{,}17+(2{,}327)(8{,}027)\approx 85{,}8$. Таким образом, при использовании нормального приближения вытекало бы, что *достаточно* 86 *линий*.

Результаты двух решений практически совпадают. Кроме того, из них можно извлечь и другую полезную информацию. Например, может показаться, что число линий будет меньше, если 2000 абонентов разбить на две группы по 1000 в каждой и уже для этих двух групп раздельно провести линии между A и B. Однако, пользуясь описанным выше методом, найдем, что в действительности дополнительно потребуется 10 линий, так что первоначальный подход предпочтительней. ▶

§ 6*). БОЛЬШИЕ ОТКЛОНЕНИЯ

Теорема Муавра — Лапласа описывает асимптотическое поведение вероятности $\mathbf{P}\{z_1<\mathbf{S}_n^*<z_2\}$ при фиксированных z_1 и z_2. Из ее доказательства видно, что теорема применима и в случае, когда z_1 и z_2 могут меняться вместе с n, причем, возможно, $z_1\to\infty$ при условии, что рост происходит достаточно медленно. В этом случае обе части (3.17) стремятся к нулю и теорема содержательна только тогда, когда *отношение* обеих частей стремится к единице. Следующая теорема показывает, при каких условиях это верно. Для упрощения формулировки двойное неравенство $z_1<\mathbf{S}_n^*<z_2$ заменим неравенством $\mathbf{S}_n^*>z_1$. Возможность такой замены обосновывается следующей леммой, в которой показано, что при $z_1\to\infty$ верхний предел z_2 не играет никакой роли.

Лемма. *Если $x_n\to\infty$, то для любого фиксированного*[1] $\eta>0$

$$\mathbf{P}\{\mathbf{S}_n^*>x_n+\eta\}/\mathbf{P}\{\mathbf{S}_n^*>x_n\}\to 0, \tag{6.1}$$

или

$$\{\mathbf{P}\{x_n<\mathbf{S}_n^*\leqslant x_n+\eta\}\sim\mathbf{P}\{\mathbf{S}_n^*>x_n\}. \tag{6.2}$$

Иначе говоря, если $\mathbf{S}_n^*$ превосходит x_n, то скорее всего $\mathbf{S}_n^*$ очень близко к x_n, и в этом предельном соотношении бо́льшие значения $\mathbf{S}_n^*$ не играют никакой роли.

---

*) Теорема этого параграфа полезна, но в этом томе она применяется лишь в гл. VII,4 и в гл. VIII,5.

[1]) Из доказательства будет видно, что достаточно условия $x_n\eta\to\infty$. Более сильный и интересный вариант см. в задаче 18.

Доказательство. Используя обозначение (3.2), для биномиального распределения имеем

$$\mathbf{P}\{S_n^* > x_n\} = \sum_{\nu=0}^{\infty} a_{r_n+\nu}, \quad \mathbf{P}\{S_n^* > x_n + \eta\} = \sum_{\nu=0}^{\infty} a_{s_n+\nu}, \tag{6.3}$$

где r_n и s_n — целые числа, которые отличаются соответственно от $x_n\sqrt{npq}$ и $(x_n+\eta)\sqrt{npq}$ не более чем на единицу. Далее, из (3.4) очевидно, что для больших n

$$a_{k+1}/a_k < 1 - pt_k < 1 - k/n < e^{-k/n}, \tag{6.4}$$

и поэтому

$$a_{s_n+\nu}/a_{r_n+\nu} < e^{-(s_n-r_n)\,r_n/n} < e^{-(1/2)\,\eta x_n\,pq}. \tag{6.5}$$

По предположению $x_n \to \infty$, и, следовательно, члены второго ряда в (6.3) становятся пренебрежимо малыми по сравнению с соответствующими членами первого ряда. ►

Теперь мы готовы обобщить предельную теорему следующим образом.

Теорема. *Если* $x_n \to \infty$ *так, что* $x_n^3/\sqrt{n} \to 0$, *то*

$$\mathbf{P}\{S_n^* > x_n\} \sim 1 - \mathfrak{N}(x_n). \tag{6.6}$$

В силу (1.7) асимптотическое соотношение (6.6) *полностью эквивалентно* следующему:

$$\mathbf{P}\{S_n^* > x_n\} \sim (1/\sqrt{2\pi})\,(1/x_n)\,e^{-(1/2)x_n^2}. \tag{6.7}$$

Доказательство. В силу предыдущей леммы и теоремы 1 § 3 имеем

$$\mathbf{P}\{S_n^* > x_n\} \sim \sum_{k=r_n}^{\infty} h\mathfrak{n}(kh), \tag{6.8}$$

где r_n — целое число, такое, что $|r_nh - x_n| < h$. Следовательно, сумма в правой части (6.8) лежит между $1-\mathfrak{N}(x_n-2h)$ и $1-\mathfrak{N}(x_n+2h)$. Используя (1.7), для разности этих двух величин получаем

$$\mathfrak{N}(x_n+2h) - \mathfrak{N}(x_n-2h) < 4h\mathfrak{n}(x_n-2h) \to 0, \tag{6.9}$$

так что сумма в (6.8) $\sim 1-\mathfrak{N}(x_n)$, что и утверждалось. ►

Дальнейшие обобщения см. в задачах 14 и 16.

§ 7. ЗАДАЧИ

1. Обобщая (1.7), доказать, что

$$1-\mathfrak{N}(x) \sim \frac{1}{\sqrt{2\pi}}\,e^{-(1/2)x^2}\left\{\frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \frac{1\cdot 3\cdot 5}{x^7} + -\ldots + (-1)^k\frac{1\cdot 3\ldots(2k-1)}{x^{2k+1}}\right\}. \tag{7.1}$$

и что при $x > 0$ правая часть *превосходит* $1-\mathfrak{N}(x)$, если k четно, и *оказывается меньше* $1-\mathfrak{N}(x)$, если k нечетно.

2. Доказать, что для произвольной постоянной $a > 0$ при $x \to \infty$ справедливо соотношение

$$(1-\mathfrak{N}(x+a/x))/(1-\mathfrak{N}(x)) \to e^{-a}. \tag{7.2}$$

3. Найти вероятность того, что среди 10 000 случайных цифр цифра 7 появится не более 968 раз.

4. Найти приближенное значение вероятности того, что число выпадений единицы при 12 000 бросаний кости лежит между 1900 и 2150.

5. Найти число k, при котором с вероятностью 0,5 число выпадений герба при 1000 бросаниях монеты будет лежать между 490 и k.

6. Для определения доли женщин f в некотором обществе производится выборка. Определить объем выборки, при котором с вероятностью, не меньшей 0,99, погрешность составит менее 0,005.

7. При 10 000 бросаниях монеты герб появлялся 5400 раз. Есть ли основания считать, что монета несимметрична?

---

8. Найти приближение для максимальной вероятности следующего триномиального распределения:

$$\frac{n!}{k!\,r!\,(n-k-r)!}\,p_1^k p_2^r\,(1-p_1-p_2)^{n-k-r}.$$

9. *Нормальное приближение для распределения Пуассона.* Используя формулу Стирлинга, показать, что для фиксированных $\alpha < \beta$ при $\lambda \to \infty$

$$\sum_{\lambda+\alpha\sqrt{\lambda}<k<\lambda+\beta\sqrt{\lambda}} p(k;\lambda) \to \mathfrak{N}(\beta)-\mathfrak{N}(\alpha). \tag{7.3}$$

10. *Нормальное приближение для гипергеометрического распределения.* Пусть n, m, k — положительные целые числа, которые стремятся к бесконечности так, что

$$\frac{r}{n+m} \to t, \quad \frac{n}{n+m} \to p, \quad \frac{m}{n+m} \to q, \quad h\{k-rp\} \to x, \tag{7.4}$$

где $h = 1/\sqrt{(n+m)\,pqt\,(1-t)}$. Доказать, что

$$\binom{n}{k}\binom{m}{r-k}\binom{n+m}{r}^{-1} \sim h\mathfrak{n}(x). \tag{7.5}$$

Указание. Лучше использовать нормальное приближение для биномиального распределения, а не формулу Стирлинга.

11. *Нормальное распределение и серии в комбинаторных задачах*[1]). Формула (11.19) гл. II дает вероятность иметь ровно k серий альф при размещении n альф и m бет. Эта вероятность равна

$$p_k = \binom{n-1}{k-1}\binom{m+1}{k}\binom{n+m}{n}^{-1}. \tag{7.6}$$

Пусть $n \to \infty$, $m \to \infty$ так, что соотношения (7.4) выполнены. При фиксированных $\alpha < \beta$ вероятность того, что число серий альф лежит между $nq+\alpha q\sqrt{pn}$ и $nq+\beta q\sqrt{pn}$, стремится к $\mathfrak{N}(\beta)-\mathfrak{N}(\alpha)$.

---

[1]) Wald A., Wolfowitz J., On a test whether two samples are from the same population, Ann. Math. Statist., 11(1940), 147—162. Более общие результаты см. в статье Mood A. M. The distribution theory of runs, там же, 367—392.

12. Новый вывод закона больших чисел. Вывести закон больших чисел, рассмотренный в § 4 гл. VI, из предельной теоремы Муавра—Лапласа.

Предельные теоремы для вероятностей больших отклонений

13. Используя обозначения § 3, показать, что если k изменяется вместе с n так, что $k^4/n^3 \longrightarrow 0$, то

$$a_k = b\,(k+m;\ n,\ p) \sim h\mathfrak{n}\,(kh)\,e^{-(p-q)\,k^3h^4/6}, \quad h = 1/\sqrt{npq}. \qquad (7.7)$$

Это обобщение теоремы 1 § 3.

14. Используя результат предыдущей задачи и лемму из § 6, доказать, что верна следующая теорема.

Теорема. *Если x_n изменяется вместе с n так, что $x_n^4/n \longrightarrow 0$, но $x_n \longrightarrow \infty$, то*

$$\mathbf{P}\{\mathbf{S}_n^* > x_n\} \sim [1 - \mathfrak{N}\,(x_n)]\,e^{-(p-q)\,x_n^3/\sqrt{npq}}. \qquad (7.8)$$

15. *Обобщение задачи* 13. Положим

$$f(x) = \sum_{\nu=3}^{\infty} \frac{p^{\nu-1} - q^{\nu-1}}{\nu\,(\nu-1)} h^{\nu-2} x^{\nu} = \frac{p^2 - q^2}{6} x^3 h + \frac{p^3 + q^3}{12} x^4 h^2 + \ldots, \qquad (7.9)$$

где $h = 1/\sqrt{npq}$. Если k изменяется вместе с n так, что $k/n \longrightarrow 0$, то

$$a_k \sim h\mathfrak{n}\,(kh) \cdot e^{-f\,(kh)}. \qquad (7.10)$$

[Если $k^3/n^2 \longrightarrow 0$, то этот результат сводится к теореме 3 § 1; если $k^4/n^3 \longrightarrow 0$, то получаем (7.7); если $k^5/n^4 \longrightarrow 0$, то получается (7.7) с добавлением в показатель члена четвертого порядка, и т. д.]

16. *Обобщение задачи* 14. *Если x_n изменяется вместе с n так, что $x_n \longrightarrow \infty$, но $x_n/\sqrt{n} \longrightarrow 0$, то*

$$\mathbf{P}\{\mathbf{S}_n^* > x_n\} \sim [1 - \mathfrak{N}\,(x_n)]\,e^{-f\,(x_n)}. \qquad (7.11)$$

Если $x_n^4/n \longrightarrow 0$, то этот результат сводится к (7.8). При $x_n^5/n^{3/2} \to 0$ можно заменить $f\left(x_n^{3/2}\right)$ многочленом четвертой степени из правой части (7.9) и т. д.

17. Если $p > q$, то $\mathbf{P}\{\mathbf{S}_n^* > x\} > \mathbf{P}\{\mathbf{S}_n^* < -x\}$ для всех больших x. *Указание.* Использовать результат задачи 15.

18. Показать, что при $x_n \longrightarrow \infty$ и $x_n/\sqrt{n} \longrightarrow 0$ справедливо соотношение

$$\mathbf{P}\{x_n < \mathbf{S}_n^* < x_n + a/x_n\} \sim (1 - e^{-a})\,\mathbf{P}\{\mathbf{S}_n^* > x_n\}. \qquad (7.12)$$

Словесно это выражается так: условная вероятность события $\{\mathbf{S}_n^* \geqslant x_n + a/x_n\}$ при условии, что $\mathbf{S}_n^* > x_n$, стремится к e^{-a}. (Более слабый вариант этой теоремы был доказан А. Я. Хинчиным[1].)

---

[1] Khintchin A. Über einen Grenzwertsatz der Wahrscheinlichkeitsrechnung, Math. Annalen, 10(1929), 745—752.— *Прим. перев.*

НЕОГРАНИЧЕННЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ ИСПЫТАНИЙ БЕРНУЛЛИ

В этой главе обсуждаются некоторые свойства случайности и важный закон повторного логарифма для испытаний Бернулли. Различные аспекты теории случайных флуктуаций, связанных с испытаниями Бернулли (по крайней мере для $p=1/2$), рассмотрены в гл. III.

§ 1. БЕСКОНЕЧНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ ИСПЫТАНИЙ

В предыдущей главе мы имели дело с вероятностями, связанными с n испытаниями Бернулли, и изучали их асимптотическое поведение при $n\to\infty$. Обратимся теперь к более общему типу задач, в которых события не могут быть определены в конечном пространстве элементарных событий.

Пример. *Задача о сериях.* Пусть α и β — положительные целые числа. Рассмотрим какую-нибудь, вообще говоря, неограниченную последовательность испытаний Бернулли, например последовательность бросаний монеты или кости. Предположим, что Павел поспорил с Петром о том, что серия из α последовательных успехов появится раньше, чем серия из β последовательных неудач. Интуитивно ясно, что означает событие «Павел выиграл», но нужно помнить, что в математической теории событие — это «множество элементарных событий» и оно не имеет смысла, пока не определено соответствующее пространство элементарных событий. Модель с конечным числом испытаний для наших целей недостаточна, но эта трудность преодолевается простым переходом к пределу. После n испытаний Петр либо выиграл, либо проиграл, либо игра осталась незаконченной. Пусть соответствующие вероятности будут x_n, y_n, z_n ($x_n+y_n+z_n=1$). При увеличении числа испытаний n вероятность ничейного результата z_n может только убывать, а x_n и y_n, очевидно, возрастают. Следовательно, существуют пределы $x=\lim x_n$, $y=\lim y_n$ и $z=\lim z_n$. Никто не усомнится в том, что эти пределы можно назвать вероятностями того, что в конце концов Петр или выиграет, или проиграет, или игра закончится вничью. Однако

*) Эта глава непосредственно не связана с материалом последующих глав и может быть опущена при первом чтении.

соответствующие три события определены лишь в пространстве элементарных событий, построенном по бесконечным последовательностям испытаний, и это пространство не является дискретным.

Этот пример был приведен только в качестве иллюстрации. Численные значения x_n, y_n, z_n нас пока не интересуют. Мы вернемся к их вычислению в примере гл. XIII,8,б). Пределы x, y, z можно найти более простым методом, который применим и в более общих случаях. Этот метод мы изложим здесь из-за его важности и того интереса, который он представляет сам по себе.

Пусть A—событие, состоящее в том, что *серия из α последовательных успехов появится раньше, чем серия из β последовательных неудач*. Событие A означает выигрыш Павла и $x=\mathbf{P}(A)$. Если u и v—условные вероятности события A при условии, что первое испытание закончилось соответственно успехом или неудачей, то $x=pu+qv$ [см. формулу (1.8) гл. V]. Сначала предположим, что первое испытание закончилось успехом. В этом случае событие A может произойти в результате α взаимно исключающихся исходов: 1) следующие $\alpha-1$ испытаний закончатся успехами; вероятность этого равна $p^{\alpha-1}$. 2) Первая неудача произойдет в ν-м испытании, где $2\leqslant\nu\leqslant\alpha$. Обозначим это событие через H_ν. Тогда $\mathbf{P}\{H_\nu\}=p^{\nu-2}q$ и $\mathbf{P}\{A\mid H_\nu\}=v$. Следовательно (еще раз используя формулу полной вероятности), получаем

$$u=p^{\alpha-1}+qv\,(1+p+\ldots+p^{\alpha-2})=p^{\alpha-1}+v\,(1-p^{\alpha-1}), \tag{1.1}$$

Если первое испытание закончилось неудачей, то аналогичные рассуждения приводят к выражению

$$v=pu\,(1+q+\ldots+q^{\beta-2})=u\,(1-q^{\beta-1}), \tag{1.2}$$

Итак, для неизвестных величин u и v имеем два уравнения и для $x=pu+qv$ находим

$$x=p^{\alpha-1}\,(1-q^{\beta})/(p^{\alpha-1}+q^{\beta-1}-p^{\alpha-1}q^{\beta-1}). \tag{1.3}$$

Чтобы получить y, нужно лишь поменять местами p и q, α и β. Итак,

$$y=q^{\beta-1}\,(1-p^{\alpha})/(p^{\alpha-1}+q^{\beta-1}-p^{\alpha-1}q^{\beta-1}), \tag{1.4}$$

Поскольку $x+y=1$, имеем $z=0$; *вероятность ничьей равна нулю*.

Например, при бросании монеты ($p=1/2$) вероятность того, что серия из двух гербов появится раньше, чем серия из трех решеток, равна 0,7; два последовательных герба появляются раньше четырех последовательных решеток с вероятностью 5/6, три последовательных герба появляются раньше четырех последовательных решеток с вероятностью 15/22. При бросании кости вероятность того, что два последовательных выпадения единицы произойдут раньше, чем пять последовательных ее невыпадений, равна 0,1753, и т. д. ▶

В этом томе мы ограничиваемся теорией дискретных пространств элементарных событий, что значительно уменьшает изящество математических рассуждений. В общей теории n испытаний Бернулли рассматриваются лишь как начало бесконечной последовательности испытаний. В этом случае элементарным событием является бесконечная последовательность букв $У$ и $Н$ и пространство элементарных событий состоит из всех таких последовательностей. Конечная последовательность, например $УУНУ$, означает совокупность всех элементарных событий с таким началом, т. е. составное событие, состоящее в том, что в бесконечной последовательности испытаний первые четыре исхода соответственно дадут $У$, $У$, $Н$, $У$. В бесконечных пространствах элементарных событий рассмотренную выше игру можно интерпретировать без предельных переходов. Возьмем произвольное элементарное событие, т. е. любую последовательность вида УНУНН.... В ней может быть серия из α последовательных успехов $У$, а может и не быть такой серии. Если такая серия есть, то она может произойти до серии из β последовательных неудач или не произойти до нее. Таким образом, мы разбиваем все элементарные события на три класса, представляющие собой события «Петр выиграл», «Петр проиграл», «нет никакого результата». Их вероятностями являются числа x, y, z, полученные выше. Единственным затруднением является то, что это пространство элементарных событий недискретно, а мы еще не определяли вероятности для общих пространств элементарных событий.

Заметим, что мы обсуждаем сейчас скорее вопросы терминологии, чем действительные трудности. В нашем примере надлежащее определение или интерпретация числа x не вызывает вопросов. Трудность состоит лишь в том, что, желая быть последовательным, мы должны решить, говорить ли о x как о «пределе вероятностей x_n того, что Петр победил после n испытаний», или говорить о событии «Петр победил», которое соответствует недискретному пространству элементарных событий. Мы предлагаем поступать обоими способами. Для упрощения терминологии будем говорить о событиях даже тогда, когда они определены в бесконечных пространствах элементарных событий. Однако для точности теоремы будут формулироваться в терминах конечных пространств элементарных событий и предельных переходов.

События, которые изучаются в этой главе, как и предыдущий пример, обладают следующей особенностью. Событие «Петр победил» хотя и определено в бесконечном пространстве, является объединением событий «Петр победил после n испытаний ($n=1, 2, \ldots$)», каждое из которых зависит только от конечного числа испытаний. Искомая вероятность x есть предел монотонной последовательности вероятностей x_n, которые зависят только от конечного числа испытаний. Нам не требуется теория, идущая дальше модели n испытаний Бернулли, мы лишь берем на себя смелость упрощать неуклюжие выражения [1]), называя некоторые числа вероятностями, а не пользуясь термином «пределы вероятностей».

§ 2. СИСТЕМЫ ИГРЫ

Печальный опыт многих игроков учит нас, что никакая система игры не увеличивает шансы игрока на выигрыш. Если теория ве-

[1]) Для читателя, знакомого с общей теорией меры, ситуацию можно описать следующим образом. Мы рассматриваем только те события, которые либо зависят от конечного числа испытаний, либо являются пределами *монотонных* последовательностей таких событий. Мы вычисляем эти простые пределы вероятностей, и для этого нам, очевидно, не нужна теория меры. Но только общая теория меры показывает, что наши пределы не зависят от способа перехода к пределу и являются вполне аддитивными функциями множеств.

роятностей правильно отображает жизнь, то этому опыту должно соответствовать некое доказуемое утверждение.

Для определенности рассмотрим, вообще говоря, неограниченную последовательность испытаний Бернулли и предположим, что перед каждым испытанием игрок может выбрать, участвует ли он в игре или нет. «Система» представляет собой фиксированные правила, согласно которым выбираются те испытания, при которых игрок делает ставку. Например, игрок может включаться в игру при каждом седьмом испытании или после каждого вступления в игру ждать, пока 7 раз не появится герб. Он может вступать в игру только после появления серии из 13 гербов или первый раз включиться в игру после появления первого герба, второй раз — после первой серии из двух последовательных гербов и вообще k-й раз — сразу после появления k последовательных гербов. В последнем случае он будет делать ставки все реже и реже. Нам не нужно рассматривать размеры ставок в отдельных испытаниях, мы хотим показать, что не существует «системы», изменяющей положение игрока, и что результат будет тот же, как если бы он постоянно участвовал в игре. Подразумевается, что это утверждение можно доказать лишь для систем в обычном смысле, когда игрок не знает результата следующего испытания (вопрос о том, существует или нет предвидение, нас не интересует). Необходимо также признать, что правило «уходи домой после трех проигрышей» меняет положение, но мы исключим такие неинтересные системы.

Определим систему как множество фиксированных правил, которые для каждого испытания однозначно определяют, вступать в игру или нет. Перед k-м испытанием это решение может зависеть от исхода первых $k-1$ испытаний, но не зависит от исхода k-го, $(k+1)$-го, $(k+2)$-го, ... испытаний. Наконец, правила должны обеспечивать бесконечное продолжение игры. Поскольку множество правил фиксировано, событие «в n испытаниях игрок делал ставки более чем r раз» корректно определено и можно найти его вероятность. Из последнего предположения в определении системы вытекает, что для любого r эта вероятность стремится к 1 при $n \to \infty$.

Сформулируем теперь нашу основную теорему о том, что *при любой системе последовательные испытания, когда игрок делает ставку, образуют последовательность испытаний Бернулли с неизменной вероятностью успеха.* При соответствующем изменении формулировки эта теорема справедлива для всех видов независимых испытаний. Последовательность испытаний, когда игрок делает ставку, является точной копией первоначальных испытаний, так что никакая система не может повлиять на удачу игрока. Важность этого утверждения впервые осознал Мизес, который ввел в качестве основной аксиомы невозможность успешной системы игры. Приведенная здесь формулировка и доказательство принадлежат Дубу [1]). Для простоты предположим, что $p=1/2$.

[1]) Doob J. L., Note on probability, Annals of Mathematics, 37 (1936), 363—367.

Пусть A_k — событие «первое вступление в игру происходит при k-м испытании». Из определения системы вытекает, что при $n \to \infty$ вероятность того, что первая ставка была сделана до n-го испытания, стремится к 1. Это означает, что

$$\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}+ \ldots + \mathbf{P}\{A_n\} \to 1,$$

или

$$\sum \mathbf{P}\{A_k\} = 1. \tag{2.1}$$

Далее, пусть B_k — событие «при k-м испытании выпал герб» и B — событие «при первом вступлении в игру выпал герб». Тогда событие B есть объединение взаимно исключающих друг друга событий A_1B_1, A_2B_2, A_3B_3, ... Далее, A_k зависит только от исхода первых $k-1$ испытаний, а B_k зависит только от k-го испытания. Следовательно, A_k и B_k независимы и $\mathbf{P}\{A_kB_k\} = \mathbf{P}\{A_k\}\mathbf{P}\{B_k\} = (1/2)\,\mathbf{P}\{A_k\}$. Итак, $\mathbf{P}\{B\} = \sum \mathbf{P}\{A_kB_k\} = (1/2)\sum \mathbf{P}\{A_k\} = (1/2)$. Отсюда вытекает, что при нашей системе вероятность выпадения герба при первом вступлении в игру равна $1/2$ и то же самое верно и для каждого из последующих вступлений в игру.

Остается показать, что моменты вступления в игру стохастически независимы. Это означает, что вероятность выпадения герба как при первом, так и при втором вступлении в игру должна равняться $1/4$ (и аналогично для всех других комбинаций, а также для последующих вступлений в игру). Для проверки этого утверждения обозначим через A_k^* событие, состоящее в том, что второе вступление в игру произошло на k-м испытании. Пусть E — событие «при первых двух вступлениях в игру выпадали гербы»; оно является объединением всех событий $A_jB_jA_k^*B_k$, где $j<k$ (если $j \geqslant k$, то A_j и A_k^* несовместны и $A_jA_k^* = 0$). Следовательно,

$$\mathbf{P}\{E\} = \sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty} \mathbf{P}\{A_jB_jA_k^*B_k\}. \tag{2.2}$$

Как и ранее, видим, что при фиксированных j и $k>j$ события B_k (герб в k-м испытании) и $A_jB_jA_k^*$ (которое зависит только от результатов первых $k-1$ испытаний) независимы. Поэтому

$$\begin{aligned}\mathbf{P}\{E\} &= (1/2)\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty} \mathbf{P}\{A_jB_jA_k^*\} = \\ &= (1/2)\sum_{j=1}^{\infty} \mathbf{P}\{A_jB_j\}\sum_{k=j+1}^{\infty} \mathbf{P}\{A_k^* \mid A_jB_j\}\end{aligned} \tag{2.3}$$

(см. формулу (1.8) гл. V). Далее, когда бы ни произошло первое вступление в игру и каков бы ни был его исход, игра, безусловно, продолжается, т. е. рано или поздно произойдет второе вступление в игру. Это означает, что при заданном событии A_jB_j, таком, что $\mathbf{P}\{A_jB_j\}>0$, условные вероятности вступления в игру на k-м испытании в сумме должны давать единицу. Следовательно, сумма второго ряда в (2.3) равна единице. Раньше мы видели, что $\sum \mathbf{P}\{A_jB_j\}=1/2$. Итак, $\mathbf{P}\{E\}=1/4$, что и утверждалось. Аналогичные рассуждения применимы для любых комбинаций испытаний.

Заметим, что положение будет другим, если игроку разрешено изменять размер ставки. В этом случае существуют выигрышные стратегии и результат игры зависит от выбора стратегии. Мы вернемся к этому вопросу в гл. XIV, 2.

§ 3. ЛЕММЫ БОРЕЛЯ — КАНТЕЛЛИ

Две простые леммы о бесконечных последовательностях испытаний используются столь часто, что заслуживают особого внимания. Мы сформулируем их для испытаний Бернулли, но они применимы и в более общем случае.

Снова рассмотрим бесконечную последовательность испытаний Бернулли. Пусть $A_1, A_2, \ldots$ — бесконечная последовательность событий, каждое из которых зависит лишь от конечного числа испытаний, иначе говоря, мы предполагаем, что существует целое число n_k, такое, что A_k является событием из пространства первых n_k испытаний Бернулли. Положим

$$a_k=\mathbf{P}\{A_k\}. \tag{3.1}$$

(Например, A_k может быть событием, состоящим в том, что $2k$-е испытание завершает серию из по крайней мере k последовательных успехов. Тогда $n_k=2k$ и $a_k=p^k$.)

Для каждой бесконечной последовательности букв $У$ и H можно установить, осуществилось ли 0, 1, 2,... или бесконечно много событий $\{A_k\}$. Это означает, что имеет смысл говорить о событии U_r, состоящем в том, что в бесконечной последовательности испытаний осуществилось более r событий $\{A_k\}$, а также о событии U_∞, состоящем в том, что произошло бесконечно много событий из совокупности $\{A_k\}$. Событие U_r определено только в бесконечном пространстве элементарных событий, и его вероятность есть предел вероятности $\mathbf{P}\{U_{n,r}\}$ того, что в результате n испытаний среди событий $\{A_k\}$ произошло более r событий. Наконец, $\mathbf{P}\{U_\infty\}=\lim \mathbf{P}\{U_r\}$; этот предел существует, так как $\mathbf{P}\{U_r\}$ убывает при возрастании r.

Лемма 1. *Если ряд $\sum a_k$ сходится, то с вероятностью единица произойдет только конечное число событий A_k.* Точнее, утверждается, что $\mathbf{P}\{U_r\}<\varepsilon$ для достаточно больших r, или: *для любого $\varepsilon>0$ существует целое число r, такое, что вероятность осуществления при n испытаниях одного или более событий $A_{r+1}, A_{r+2}, \ldots$ меньше ε для всех n.*

Доказательство. Выберем r так, что $a_{r+1}+a_{r+2}+\ldots\leqslant\varepsilon$; в силу сходимости ряда $\sum a_k$ это возможно. Не ограничивая общности, можем предполагать, что A_k упорядочены таким образом, что $n_1 \leqslant n_2 \leqslant n_3 \leqslant \dots$. Пусть N — наибольший индекс, при котором $n_N \leqslant n$. Тогда $A_1, \dots, A_N$ определены в пространстве n испытаний, и лемма утверждает, что вероятность осуществления одного или нескольких событий $A_{r+1}, A_{r+2}, \dots A_N$ меньше ε. Это верно, так как, согласно основному неравенству (7.6) гл. I, имеем

$$\mathbf{P}\{A_{r+1} \cup A_{r+2} \cup \dots \cup A_N\} \leqslant a_{r+1} + a_{r+2} + \dots + a_N \leqslant \varepsilon, \quad (3.2)$$

что и утверждалось. ▶

Удовлетворительное обращение этой леммы известно только в частном случае взаимно независимых A_k. Такое положение возникает, когда испытания разбиты на неперекрывающиеся группы и A_k зависит только от испытаний k-й группы (например, A_k может быть событием, состоящим в том, что в k-й тысяче испытаний произошло более 600 успехов).

Лемма 2. *Если события A_k взаимно независимы и ряд $\sum a_k$ расходится, то с вероятностью единица осуществится бесконечно много событий A_k.* Иначе говоря, для любого r вероятность того, что в n испытаниях произойдет более чем r событий A_k, стремится к 1 при $n \to \infty$.

Доказательство. Предположим противное. Тогда существует n, такое, что с положительной вероятностью u ни одно событие A_k, где $k > n$, не произойдет. Однако

$$u \leqslant (1-a_n)(1-a_{n+1}) \dots (1-a_{n+r}), \quad (3.3)$$

так как произведение в правой части равно вероятности того, что события A_k, где $n \leqslant k \leqslant n+r$, не произойдут. Поскольку $1-x \leqslant e^{-x}$, произведение в правой части не превосходит $e^{-(a_n+\dots+a_{n+r})}$, но сумма в показателе может быть сделана сколь угодно большой за счет выбора достаточно больших r. Итак, $u=0$, что противоречит предположению. ▶

Примеры. а) Чему равна вероятность того, что в последовательности испытаний Бернулли комбинация *УНУ* появится бесконечное число раз? Пусть A_k — событие, состоящее в том, что в результате k-го, $k+1$-го и $(k+2)$-го испытаний получилась комбинация *УНУ*. События A_k не являются взаимно независимыми, но последовательность $A_1, A_4, A_7, A_{10}, \dots$ содержит только взаимно независимые события (поскольку никакие два из них не зависят от исхода одного и того же испытания). Так как $a_k = p^2q$ не зависит от k, ряд $a_1+a_4+a_7+\dots$ расходится, и поэтому с вероятностью единица комбинация *УНУ* осуществляется бесконечное число раз. Аналогичные рассуждения, очевидно, применимы для произвольной комбинации любой длины.

б) *Книги, написанные путем бросания монеты.* Рассмотрим, например, фразу «вероятность — это забавная штука», записанную в азбуке Морзе в виде конечной последовательности точек и тире. Если мы вместо точки напишем Γ и вместо тире — P, то эта фраза примет вид конечной последовательности из гербов и решеток. Из предыдущего примера следует, что при длительном бросании монеты рано или поздно наверняка получится заданная фраза и это повторится бесконечное число раз. Кстати, в записи результатов длительных бросаний монеты должна содержаться закодированная в азбуке Морзе любая мыслимая книга от «Гамлета» до восьмизначных логарифмических таблиц. Предлагалось научить большую группу обезьян наугад колотить по клавишам пишущих машинок, надеясь в конце концов получить великие литературные произведения. Используя для этой цели монету, можно не тратиться на обучение и корм и предоставить обезьянам заниматься своими обезьяньими делами.

§ 4. УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

Интуитивное понятие вероятности основывается на представлении о правильности следующего утверждения: если $\mathbf{S}_n$ — число успехов в первых n испытаниях последовательности испытаний Бернулли, то

$$\mathbf{S}_n/n \to p. \tag{4.1}$$

В абстрактной теории это не может быть верным для *каждой* последовательности испытаний. Действительно, наше пространство элементарных событий содержит элементарное событие, представляющее собой реализацию логической возможности появления бесконечной последовательности только одних успехов, и для этого события $\mathbf{S}_n/n=1$. Однако можно показать, что (4.1) выполняется с вероятностью 1, поэтому случаи, когда (4.1) не выполняется, являются исключением, которым можно пренебречь.

Заметим, что мы рассматриваем сейчас утверждение намного более сильное, чем закон больших чисел (4.1) гл. VI. Последний утверждает, что для любого достаточно большого *фиксированного* n среднее $\mathbf{S}/n$ скорее всего близко к p, но он ничего не говорит о том, должно ли отношение $\mathbf{S}_n/n$ оставаться близким к p при увеличении числа испытаний. Обычный закон больших чисел оставляет открытой возможность того, что в n дополнительных испытаний осуществится хотя бы одно из событий $k^{-1}\,\mathbf{S}_k<p-\varepsilon$, где $n<k\leqslant 2n$. Вероятность этого равна сумме большого числа вероятностей, о каждой из которых мы лишь знаем, что она мала. Сейчас мы докажем, что с вероятностью единица разность $\mathbf{S}_n/n-p$ становится *и остается* малой.

Усиленный закон больших чисел. *Для любого $\varepsilon>0$ с вероятностью единица осуществится лишь конечное число событий*

$$|S_n/n-p|>\varepsilon. \tag{4.2}$$

Отсюда вытекает, что соотношение (4.1) справедливо с вероятностью единица. В терминологии конечных пространств элементарных событий утверждается, что для любых $\varepsilon>0$, $\delta>0$ найдется такое r, что для всех ν вероятность одновременного выполнения ν неравенств

$$|S_{r+k}/(r+k)-p|<\varepsilon, \quad k=1, 2, \ldots, \nu, \tag{4.3}$$

больше, чем $1-\delta$.

Доказательство. Докажем более сильное утверждение. Пусть A_k — событие, состоящее в том, что

$$|S_k^*|=|(S_k-kp)/\sqrt{kpq}|\geqslant\sqrt{2a\log k}, \tag{4.4}$$

где $a>1$. Тогда из (6.7) гл. VII, очевидно, следует, что по крайней мере для всех достаточно больших k имеем

$$P\{A_k\}<e^{-a\log k}=1/k^a. \tag{4.5}$$

Поэтому ряд $\sum P\{A_k\}$ сходится, и по лемме 1 из предыдущего параграфа *с вероятностью единица выполняется только конечное число неравенств* (4.4). С другой стороны, если справедливо (4.2), то

$$|(S_n-np)/\sqrt{npq}|>(\varepsilon/\sqrt{pq})\sqrt{n} \tag{4.6}$$

и для больших n правая часть больше $\sqrt{2a\log n}$. Следовательно, выполнение бесконечно многих неравенств (4.2) влечет за собой осуществление бесконечного числа событий A_k и имеет поэтому нулевую вероятность. ▶

Усиленный закон больших чисел был впервые сформулирован Кантелли (1917); до этого Борель и Хаусдорф рассмотрели некоторые частные случаи. Усиленный закон, так же как и обычный закон больших чисел, является лишь весьма частным случаем одной общей теоремы о случайных величинах. Рассмотренный совместно с нашей теоремой о неэффективности систем игры закон больших чисел влечет за собой существование предела (4.1) не только для исходной последовательности испытаний, но и для всех подпоследовательностей, полученных в соответствии с правилами из § 2. *Таким образом, эти две теоремы вместе описывают основные свойства случайности, которые присущи интуитивному представлению о вероятности* и важность которых особенно подчеркивал Мизес.

§ 5. ЗАКОН ПОВТОРНОГО ЛОГАРИФМА

Как и в гл. VII, снова введем нормированное число успехов в n испытаниях

$$S_n^* = (S_n - np)/\sqrt{npq}. \qquad (5.1)$$

Предельная теорема Лапласа утверждает, что $P\{S_n^* > x\} \sim 1 - \mathfrak{N}(x)$. Следовательно, для любого фиксированного n величина S_n^* не может быть большой, но интуитивно ясно, что в длительной последовательности испытаний S_n^* рано или поздно примет сколь угодно большое значение. Умеренные значения S_n^* наиболее вероятны, но максимум будет медленно возрастать. Насколько быстро? При доказательстве усиленного закона больших чисел из (4.5) мы получили, что с вероятностью единица неравенство $S_n^* < \sqrt{2a \log n}$ выполняется для любого $a > 1$ и всех достаточно больших n. Это дает нам верхнюю оценку для флуктуаций S_n^*, но эта оценка плохая. Чтобы установить это, применим те же самые рассуждения к подпоследовательности $S_2^*, S_4^*, S_8^*, S_{16}^*, \ldots$, т. е. определим событие A_k соотношением $S_{2^k}^* \geqslant \sqrt{2a \log k}$. Из неравенства (4.5) вытекает, что неравенство $S_{2^k}^* < \sqrt{2a \log k}$ выполняется для $a > 1$ и всех достаточно больших k. Но для $n = 2^k$ имеем $\log k \sim \log\log n$, и поэтому для $a > 1$ и всех n вида $n = 2^k$ неравенство

$$S_n^* < \sqrt{2a \log\log n} \qquad (5.2)$$

выполняется, начиная с некоторого k. Теперь естественно предположить, что в действительности (5.2) справедливо для *всех* достаточно больших n, и, в самом деле, это одно из утверждений закона повторного логарифма. Согласно этой замечательной теореме[1]), $\sqrt{2 \log\log n}$ есть *точная* верхняя оценка в том смысле, что при любом $a < 1$ для бесконечно многих n выполнено неравенство, противоположное (5.2).

Теорема. *С вероятностью единица справедливо соотношение*

$$\limsup_{n \to \infty} \left(S_n^* / \sqrt{2 \log\log n}\right) = 1. \qquad (5.3)$$

Это означает, что при $\lambda > 1$ с вероятностью единица происходит только конечное число событий

$$S_n > np + \lambda \sqrt{2npq \log\log n}; \qquad (5.4)$$

---

[1]) Khintchin A. Über einen Satz der Wahrscheinlichkeitsrechnung, Fundamenta Mathematicae, 6 (1924), 9—20. Этому открытию предшествовали частные результаты других авторов. Настоящее доказательство построено таким образом, что оно допускает непосредственное обобщение на более общие случайные величины,

при $\lambda < 1$ *с вероятностью единица неравенство* (5.4) *справедливо для бесконечного числа значений* n.

По соображениям симметрии из (5.3) вытекает

$$\liminf_{n \to \infty} (S_n^*/\sqrt{2 \log \log n}) = -1. \tag{5.3a}$$

Доказательство. Начнем с двух предварительных замечаний.

1. Существует постоянная $c>0$, зависящая от p, но не зависящая от n и такая, что

$$\mathbf{P}\{S_n > np\} > c \tag{5.5}$$

при всех n. Действительно, рассмотрение биномиального распределения показывает, что левая часть (5.5) никогда не равна нулю, а по предельной теореме Лапласа она стремится к 1/2 при $n \to \infty$. Следовательно, как и утверждалось, левая часть больше некоторой положительной постоянной.

2. Нам потребуется также следующая *лемма.* Пусть x фиксировано и A — событие, заключающееся в том, что хотя бы для одного k, $k \leqslant n$, выполняется неравенство

$$S_k - kp > x. \tag{5.6}$$

Тогда

$$\mathbf{P}\{A\} \leqslant c^{-1}\, \mathbf{P}\{S_n - np > x\}. \tag{5.7}$$

Для доказательства этой леммы обозначим через A_ν событие, состоящее в том, что (5.6) выполняется при $k=\nu$, но не выполняется при $k=1, 2, \ldots, \nu-1$ (здесь $1 \leqslant \nu \leqslant n$). События $A_1, A_2, \ldots, A_n$ попарно несовместны, и их объединение дает A; следовательно,

$$\mathbf{P}\{A\} = \mathbf{P}\{A_1\} + \ldots + \mathbf{P}\{A_n\}. \tag{5.8}$$

Далее, пусть при $\nu < n$ U_ν — событие, состоящее в том, что общее число успехов в испытаниях с номерами $\nu+1, \nu+2, \ldots, n$ больше, чем $(n-\nu)p$. Если оба события A_ν и U_ν произошли, то $S_n > S_\nu + (n-\nu)p > np + x$, и, так как события $A_\nu U_\nu$ попарно несовместны, отсюда вытекает, что

$$\mathbf{P}\{S - np > x\} \geqslant \mathbf{P}\{A_1 U_1\} + \ldots + \mathbf{P}\{A_{n-1} U_{n-1}\} + \mathbf{P}\{A_n\}. \tag{5.9}$$

Далее, A_ν зависит только от первых ν испытаний, а U_ν — только от последующих $n-\nu$ испытаний. Поэтому события A_ν и U_ν независимы и $\mathbf{P}\{A_\nu U_\nu\} = \mathbf{P}\{A_\nu\}\,\mathbf{P}\{U_\nu\}$. Из предварительного замечания (5.5) мы знаем, что $\mathbf{P}\{U_\nu\} > c > 0$, и из (5.9) и (5.8) получаем

$$\mathbf{P}\{S_n - np > x\} \geqslant c \sum \mathbf{P}\{A_\nu\} = c\mathbf{P}\{A\}. \tag{5.10}$$

Это и доказывает (5.7).

3. Докажем теперь утверждение теоремы, относящееся к (5.4) при $\lambda > 1$. Пусть γ — число, такое, что

$$1 < \gamma < \lambda, \tag{5.11}$$

а n_r — целое число, ближайшее к γ^r. Обозначим через B_r событие состоящее в том, что неравенство

$$\mathbf{S}_n - np > \lambda \sqrt{2n_r pq \log\log n_r} \tag{5.12}$$

выполняется хотя бы для одного n из интервала $n_r \leqslant n < n_{r+1}$. Очевидно, (5.4) справедливо для бесконечного числа значений n только тогда, когда осуществится бесконечно много событий B_r. Следовательно, используя первую лемму Бореля — Кантелли, видим, что достаточно доказать утверждение

$$\text{ряд } \sum \mathbf{P}\{B_r\} \text{ сходится.} \tag{5.13}$$

Согласно неравенству (5.7), имеем

$$\mathbf{P}\{B_r\} \leqslant c^{-1}\mathbf{P}\{\mathbf{S}_{n_{r+1}} - n_{r+1}p > \lambda \sqrt{2n_r pq \log\log n_r}\} = \\ = c^{-1}\mathbf{P}\{\mathbf{S}^*_{n_{r+1}} > \lambda \sqrt{2(n_r/n_{r+1})\log\log n_r}\}. \tag{5.14}$$

Поскольку при достаточно больших r $n_{r+1}/n_r \sim \gamma < \lambda$, получаем

$$\mathbf{P}\{B_r\} \leqslant c^{-1}\mathbf{P}\{\mathbf{S}^*_{n_{r+1}} > \sqrt{2\lambda \log\log n_r}\}. \tag{5.15}$$

Следовательно, из (6.7) гл. VII для больших r имеем

$$\mathbf{P}\{B_r\} \leqslant c^{-1}e^{-\lambda\log\log n_r} = 1/[c(\log n_r)^\lambda] \sim 1/[c(r\log\gamma)^\lambda]. \tag{5.16}$$

Так как $\lambda > 1$, утверждение (5.13) доказано.

4. Докажем, наконец, утверждение, относящееся к (5.4) при $\lambda < 1$. В качестве γ выберем на этот раз целое число, столь большое, что

$$(\gamma - 1)/\gamma > \eta > \lambda, \tag{5.17}$$

где η — постоянная, которую определим позднее. Положим $n_r = \gamma^r$.

Вторая лемма Бореля — Кантелли применима только к независимым событиям, и по этой причине определим

$$\mathbf{D}_r = \mathbf{S}_{n_r} - \mathbf{S}_{n_{r-1}}; \tag{5.18}$$

здесь $\mathbf{D}_r$ — общее число успехов, происшедших в испытаниях, следующих за n_{r-1}-м вплоть до n_r-го включительно. Величина $\mathbf{D}_r$ имеет биномиальное распределение $b(k;\ n,\ p)$, где $n = n_r - n_{r-1}$. Пусть A_r — событие, состоящее в том, что

$$\mathbf{D}_r - (n_r - n_{r-1})p > \eta \sqrt{2pqn_r \log\log n_r}. \tag{5.19}$$

Мы утверждаем, что *с вероятностью единица осуществится бесконечно много событий* A_r. Поскольку различные события A_r зависят от неперекрывающихся групп испытаний (а именно A_r зависит от испытаний с номерами $n_{r-1} < n \leqslant n_r$), они взаимно независимы, и, согласно второй лемме Бореля — Кантелли, достаточно доказать, что ряд $\sum \mathbf{P}\{A_r\}$ расходится. Имеем

$$\mathbf{P}\{A_r\} = \mathbf{P}\left\{\frac{\mathbf{D}_r - (n_r - n_{r-1})p}{\sqrt{(n_r - n_{r-1})pq}} > \eta \sqrt{2\frac{n_r}{n_r - n_{r-1}} \log\log n_r}\right\}. \quad (5.20)$$

Здесь $n_r/(n_r - n_{r-1}) = \gamma/(\gamma - 1) < \eta^{-1}$ в силу (5.17). Следовательно,

$$\mathbf{P}\{A_r\} \geqslant \mathbf{P}\left\{\frac{\mathbf{D}_r - (n_r - n_{r-1})p}{\sqrt{(n_r - n_{r-1})pq}} > \sqrt{2\eta \log\log n_r}\right\}. \quad (5.21)$$

Опять используя оценку (6.7) гл. VII, для больших r находим

$$\mathbf{P}\{A_r\} > \frac{1}{\log\log n_r} e^{-\eta \log\log n_r} = \frac{1}{(\log\log n_r)(\log n_r)^{\eta}}. \quad (5.22)$$

Так как $n_r = \gamma^r$ и $\eta < 1$, для больших r имеем $\mathbf{P}\{A_r\} > 1/r$, что и доказывает расходимость ряда $\sum \mathbf{P}\{A_r\}$.

На последнем шаге доказательства нужно показать, что величиной $\mathbf{S}_{n_{r-1}}$ в (5.18) можно пренебречь. Из первой, уже доказанной части теоремы знаем, что для любого $\varepsilon > 0$ найдется такое N, что с вероятностью не меньше $1-\varepsilon$ для всех $r > N$ выполняется неравенство

$$|\mathbf{S}_{n_{r-1}} - n_{r-1}p| < 2\sqrt{2pqn_{r-1} \log\log n_{r-1}}. \quad (5.23)$$

Далее, предположим, что число η выбрано столь близким к 1, что

$$1 - \eta < [(\eta - \lambda)/2]^2. \quad (5.24)$$

Тогда из (5.17) вытекает, что

$$4n_{r-1} = 4n_r\gamma^{-1} < n_r(\eta - \lambda)^2, \quad (5.25)$$

и поэтому из (5.23) получаем

$$\mathbf{S}_{n_{r-1}} - n_{r-1}p > -(\eta - \lambda)\sqrt{2pqn_r \log\log n_r}. \quad (5.26)$$

Сложив (5.26) и (5.19), получим (5.4) с $n = n_r$. Отсюда следует, что с вероятностью не меньше $1 - \varepsilon$ неравенство (5.4) выполнено для бесконечного числа значений r, что и завершает доказательство. ►

Закон повторного логарифма для испытаний Бернулли является частным случаем более общей теоремы, впервые сформулированной А. Н. Колмогоровым [1]). В настоящее время доказаны и более сильные теоремы (см. задачи 7 и 8).

---

[1]) Kolmogoroff A., Das Gesetz des iterierten Logarithmus, Mathematische Annalen, 101 (1929), 126–135.

§ 6. ИНТЕРПРЕТАЦИЯ НА ЯЗЫКЕ ТЕОРИИ ЧИСЕЛ

Пусть x — действительное число из интервала $0 \leqslant x \leqslant 1$ и

$$x = 0, a_1\, a_2\, a_3 \ldots \tag{6.1}$$

есть его десятичное разложение (так что каждое a_j обозначает одну из цифр $0, 1, \ldots, 9$). Это разложение единственно для всех чисел, за исключением чисел вида $a/10^n$ (где a — целое), а последние могут быть записаны либо в виде разложения, содержащего бесконечно много нулей, либо в виде разложения, содержащего бесконечно много девяток. Во избежание неопределенности условимся не пользоваться разложениями второго вида.

Десятичные разложения связаны с испытаниями Бернулли с $p = 1/10$, в которых 0 соответствует успеху, а остальные цифры — неудаче. Если заменить в (6.1) все нули буквой $У$ и все другие цифры буквой $Н$, то (6.1) будет представлять собой возможный исход бесконечной последовательности испытаний Бернулли с $p = 1/10$. Обратно, произвольная последовательность букв $У$ и $Н$ может быть получена описанным выше способом из разложения некоторого числа x.

Таким образом, каждое событие в пространстве элементарных событий, связанном с испытаниями Бернулли, представляется некоторой совокупностью чисел x. Например, событие «успех в n-м испытании» представляется всеми числами x, у которых n-й десятичный знак равен нулю. Эти числа образуют совокупность 10^{n-1} интервалов, длина каждого 10^{-n}, и общая длина этих интервалов равна $1/10$, т. е. вероятности нашего события. Каждому конкретному исходу n испытаний Бернулли соответствует совокупность некоторых интервалов, например последовательность $УНУ$ представляется девятью интервалами $0{,}01 \leqslant x < 0{,}011$, $0{,}02 \leqslant x < 0{,}021, \ldots$ $\ldots, 0{,}09 \leqslant x < 0{,}091$. Вероятность каждого исхода конечной последовательности испытаний равна общей длине соответствующих интервалов на оси x. Вероятности более сложных событий всегда выражаются через вероятности исходов конечной последовательности испытаний, и вычисления проводятся в соответствии с правилом сложения, которое верно для обычной меры Лебега на оси x. Следовательно, наши вероятности всегда будут совпадать с мерой соответствующего множества точек на оси x. Итак, у нас есть способ перевода всех предельных теорем для испытаний Бернулли с $p = 1/10$ в теоремы о десятичных разложениях. Выражение «с вероятностью единица» эквивалентно следующим: «для почти всех x» и «почти всюду».

Мы рассматривали случайную величину $\mathbf{S}_n$, равную числу успехов в n испытаниях. Здесь уместно подчеркнуть тот факт, что $\mathbf{S}_n$ является функцией от элементарных событий. Обозначим через $\mathbf{S}_n(x)$ *число нулей среди первых n десятичных знаков числа x*. Очевидно, график $\mathbf{S}_n(x)$ представляет собой ступенчатую ломаную с разрывами в точках вида $a/10^n$, где a — целое. Отношение $\mathbf{S}_n(x)/n$ называется *частотой нулей* среди первых n десятичных знаков числа x.

На языке обычной теории меры закон больших чисел утверждает, что $S_n(x)/n \to 1/10$ по мере, а из усиленного закона больших чисел следует, что $S_n(x)/n \to 1/10$ почти всюду. Закон повторного логарифма Хинчина показывает, что

$$\limsup \left[(S_n(x) - n/10) / \sqrt{n \log \log n} \right] = 0{,}3 \sqrt{2} \tag{6.2}$$

для почти всех x. Это дает ответ на проблему, изучавшуюся в ряде работ, начиная с книги Хаусдорфа (1913) [1]) и статьи Харди и Литлвуда (1914) [2]). Дальнейшие уточнения этого результата см. в задачах 7 и 8.

Вместо нуля можно было рассмотреть любую другую цифру и сформулировать следствие усиленного закона больших чисел, состоящее в том, что частота каждой из десяти цифр стремится к 1/10 для почти всех x. Аналогичные теоремы справедливы и в случае, когда десятичная система счисления заменяется любой другой. Этот факт был открыт Борелем [3]) и обычно выражается утверждением, что почти все числа «нормальны» [4]).

§ 7. ЗАДАЧИ

1. Найти целое число β, для которого вероятность того, что при бросании кости серия из трех последовательных единиц появится раньше серии из β неединиц, примерно равна половине.

2. Рассмотрим повторные независимые испытания с тремя возможными исходами A, B, C и соответствующими вероятностями p, q, r ($p+q+r=1$). Найти вероятность того, что серия A длиной α появится раньше, чем серия B длиной β.

3. *Продолжение.* Найти вероятность того, что серия A длиной α появится раньше, чем либо серия B длиной β, либо серия C длиной γ.

4. Пусть в последовательности испытаний Бернулли A_n — событие, состоящее в том, что серия из n последовательных успехов произойдет между 2^n-м и 2^{n+1}-м испытаниями. Доказать, что если $p \geqslant 1/2$, то с вероятностью единица осуществится бесконечно много событий A_n, а если $p < 1/2$, то с вероятностью единица осуществится только конечное число A_n.

5 [5]). Обозначим через N_n длину серии успехов, начинающуюся в n-м испытании (т. е. $N_n = 0$, если n-е испытание закончилось неудачей и т. д.). Доказать, что с вероятностью единица

$$\limsup (N_n / \operatorname{Log} n) = 1, \tag{7.1}$$

где Log — логарифм по основанию $1/p$.

---

[1]) Hausdorff F., Grundzüge der Mengenlehre, Leipzig, 1913.

[2]) Hardy G. H. Littlewood J. E. Some problems of Diophantine approximation, Acta Mathematica, 37 (1914), 155—239.

[3]) Borel E., Les probabilités dénombrables et leurs applications arithmetiques, Rend. Circ. Math. Palermo, 27 (1909), 247—271.— *Прим. перев.*

[4]) Более подробное изложение вопросов, затронутых в этом параграфе, см. в работах Хинчин А. Я., Метрические задачи теории иррациональных чисел, Успехи матем. наук, 1 (1936), 7—32 и Kac M. Statistical independence in probability analysis and number theory.— New York, 1959. [Имеется перевод: Кац М. Статистическая независимость в теории вероятностей, анализе и теории чисел.— М.: ИЛ, 1963.] — *Прим. перев.*

[5]) Задача подсказана сообщением, полученным от Д. Дж. Ньюмена.

Указание. Рассмотреть событие A_n, состоящее в том, что за n-м испытанием следует серия из более чем $a \operatorname{Log} n$ успехов. Для $a > 1$ вычисления простые. При $a < 1$ рассмотреть подпоследовательность испытаний с номерами $a_1, a_2, \ldots$, где a_n — целое число, ближайшее к $n \operatorname{Log} n$.

6. Получить из закона повторного логарифма утверждение: с вероятностью единица для бесконечного числа значений n все величины S_k^*, где $n < k < 17n$, положительны. (*Замечание.* При помощи результатов гл. III можно доказать значительно более сильные утверждения.)

7. Пусть $\varphi(t)$ — положительная монотонно возрастающая функция и n_r — целое число, ближайшее к $e^{r/\log r}$. Если ряд

$$\sum [1/\varphi(n_r)]\, e^{-(1/2)\varphi^2(n_r)} \tag{7.2}$$

сходится, то с вероятностью единица неравенство

$$S_n > np + \sqrt{npq}\,\varphi(n) \tag{7.3}$$

справедливо лишь для конечного числа значений n. Заметим, что без потери общности можно предполагать, что $\varphi(n) < 10\sqrt{\log\log n}$; для больших значений $\varphi(n)$ утверждение вытекает из закона повторного логарифма.

8. Доказать [1]), что ряд (7.2) сходится тогда и только тогда, когда сходится ряд

$$\sum [\varphi(n)/n]\, e^{-(1/2)\varphi^2(n)}. \tag{7.4}$$

Указание. Собрать члены с номерами n, где $n_{r-1} < n < n_r$, и заметить, что $n_r - n_{r-1} \sim n_r(1 - 1/\log r)$; кроме того, ряд (7.4) может сходиться только тогда, когда $\varphi^2(n) > 2\log\log n$.

9. Используя результат предыдущей задачи, доказать, что с вероятностью единица

$$\limsup\, [S_n^* - \sqrt{2\log\log n}]\, \frac{\sqrt{2\log\log n}}{\log\log\log n} = \frac{3}{2}. \tag{7.5}$$

[1]) Объединяя результаты задач 7 и 8, видим, что в случае сходимости ряда (7.4) неравенство (7.3) выполняется с вероятностью единица только для конечного числа значений n. Обратно, если ряд (7.4) расходится, то неравенство (7.3) выполняется с вероятностью единица для бесконечного числа значений n. Это обратное утверждение доказать значительно труднее, см. Feller W., The general form of the so-called law of the iterated logarithm, Trans. Amer. Math. Soc., 54 (1943), 373—402, где доказаны более общие теоремы для произвольных случайных величин. Относительно частного случая испытаний Бернулли с $p = 1/2$ см. Erdős P., On the law of the iterated logarithm, Ann. of Math. (2), 43 (1942), 419—436. Закон повторного логарифма получается как частный случай при $\varphi(t) = \lambda\sqrt{2\log\log t}$.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ; МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

§ 1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Согласно определению, которое дается в курсах математического анализа, величина y называется *функцией* действительного числа x, если каждому значению x соответствует некоторое значение y. Это определение можно расширить на случай, когда независимая переменная не является действительным числом. Так, например, расстояние — это функция пары точек, периметр треугольника — функция, определенная на множестве треугольников, последовательность $\{a_n\}$ является функцией, определенной для всех положительных целых чисел, биномиальный коэффициент $\binom{x}{k}$ — функция, определенная для пар чисел (x, k), из которых второе является неотрицательным целым числом. В этом же смысле мы можем говорить о числе $\mathbf{S}_n$ успехов в n испытаниях Бернулли как о функции, определенной на пространстве элементарных событий, каждой из 2^n точек которого соответствует некоторое число $\mathbf{S}_n$.

Функция, определенная на пространстве элементарных событий, называется случайной величиной. В предыдущих главах мы занимались случайными величинами, не употребляя этого термина. Типичными случайными величинами являются: число тузов у одного игрока при игре в бридж, число совпадающих дней рождения в группе из n человек, число серий успехов в n испытаниях Бернулли. В каждом случае задано правило, по которому каждому элементарному событию соответствует единственное число $\mathbf{X}$. Классическая теория вероятностей была посвящена в основном нахождению выигрыша игрока, а этот выигрыш также является случайной величиной. Действительно, каждую случайную величину можно интерпретировать как выигрыш реального или воображаемого игрока в соответствующей игре. Положение частицы при диффузии, энергия, температура и т. п. физических систем являются случайными величинами, но они определены на недискретных пространствах элементарных событий, и поэтому их изучение откладывается. В случае дискретного пространства элементарных событий теоретически можно задать любую случайную величину $\mathbf{X}$, занумеровав в каком-либо порядке все точки пространства и связав с каждой из них соответствующее значение $\mathbf{X}$.

Термин «случайная величина» несколько неточен, более подходящим был бы термин «функция случая» (независимой переменной является точка в пространстве элементарных событий, т. е. исход эксперимента или случай).

Пусть $\mathbf{X}$ — случайная величина, а x_1, x_2,...— ее значения; в дальнейшем x_j, как правило, будут целыми числами. Совокупность всех элементарных событий, на которых $\mathbf{X}$ принимает фиксированное значение x_j, образует событие $\mathbf{X}=x_j$; его вероятность обозначается через $\mathbf{P}\{\mathbf{X}=x_j\}$. *Функция*

$$\mathbf{P}\{\mathbf{X}=x_j\}=f(x_j) \quad (j=1,\ 2,\dots) \tag{1.1}$$

называется распределением (вероятностей) [1]) *случайной величины* $\mathbf{X}$. Ясно, что

$$f(x_j)\geqslant 0, \quad \sum f(x_j)=1. \tag{1.2}$$

В этой терминологии можно говорить, что число успехов $\mathbf{S}_n$ в испытаниях Бернулли есть случайная величина с распределением вероятностей $\{b(k;\ n,\ p)\}$, а число испытаний до первого успеха включительно является случайной величиной с распределением $\{q^{k-1}p\}$.

Рассмотрим теперь две случайные величины $\mathbf{X}$ и $\mathbf{Y}$, определенные на одном пространстве элементарных событий, и обозначим их значения соответственно через x_1, x_2,... и y_1, y_2,... . Пусть соответствующие распределения вероятностей будут $\{f(x_j)\}$ и $\{g(y_k)\}$. Совокупность точек, для которых выполнены оба условия $\mathbf{X}=x_j$ и $\mathbf{Y}=y_k$, образует событие, вероятность которого обозначим через $\mathbf{P}\{\mathbf{X}=x_j,\ \mathbf{Y}=y_k\}$. *Функция*

$$\mathbf{P}\{\mathbf{X}=x_j,\ \mathbf{Y}=y_k\}=p(x_j,\ y_k) \quad (j,\ k=1,\ 2,\dots) \tag{1.3}$$

называется совместным распределением вероятностей случайных величин $\mathbf{X}$ *и* $\mathbf{Y}$. Это распределение лучше всего представить в виде таблицы с двумя входами, как сделано ниже (табл. 1 и 2). Ясно, что

$$p(x_j,\ y_k)\geqslant 0, \quad \sum_{j,\ k} p(x_j,\ y_k)=1. \tag{1.4}$$

Кроме того, для каждого фиксированного j

$$p(x_j,\ y_1)+p(x_j,\ y_2)+p(x_j,\ y_3)+\ldots=\mathbf{P}\{\mathbf{X}=x_j\}=f(x_j) \tag{1.5}$$

[1]) Распределение вероятностей дискретной случайной величины $\mathbf{X}$ представляет собой функцию $f(x_j)$, определенную на множестве значений x_j величины $\mathbf{X}$. Термин «распределение вероятностей» нужно отличать от термина «функция распределения», который применяется к неубывающим функциям, стремящимся к 0 при $x\to-\infty$ и к 1 при $x\to\infty$. Функция распределения $F(x)$ дискретной случайной величины $\mathbf{X}$ определяется формулой

$$F(x)=\mathbf{P}\{\mathbf{X}\leqslant x\}=\sum_{x_j\leqslant x} f(x_j),$$

в которой сумма берется по всем значениям x_j, не превосходящим x. Таким образом, функцию распределения случайной величины можно вычислить по распределению вероятностей этой величины, и наоборот. В этом томе мы вообще не будем касаться функций распределения.

и для каждого фиксированного k

$$p(x_1,\ y_k)+p(x_2,\ y_k)+p(x_3,\ y_k)+\ldots=\mathbf{P}\{\mathbf{Y}=y_k\}=g(y_k). \quad (1.6)$$

Иначе говоря, суммируя вероятности по строкам и столбцам, мы получаем распределения вероятностей величин $\mathbf{X}$ и $\mathbf{Y}$. Их можно изобразить так, как показано в табл. 1 и 2, и тогда они называются *маргинальными (частными) распределениями*. Прилагательное «маргинальный» связано с тем, что распределения величин $\mathbf{X}$ и $\mathbf{Y}$ записываются по краям таблицы, соответствующей их совместному распределению, и используется также для стилистической ясности, когда совместное распределение двух величин и их частные (маргинальные) распределения появляются в одном и том же контексте. Строго говоря, прилагательное «маргинальный» излишне.

Понятие совместного распределения переносится также на *системы из более чем двух случайных величин*.

Примеры. а) *Случайные размещения* 3 *шаров по* 3 *ящикам*. Обратимся к пространству элементарных событий которое состоит из 27 точек и формально задано в табл. 1, приведенной в примере гл. I, 2, а), и припишем каждой точке вероятность 1/27. Пусть $\mathbf{N}$ — число занятых ящиков, а $\mathbf{X}_i$ ($i=1,\ 2,\ 3$) — число шаров в ящике с номером i. Это наглядные описания. Формально же $\mathbf{N}$ есть функция, принимающая значение 1 в точках с номерами 1—3, значение 2 в точках с номерами 4—21 и значение 3 в точках с номерами 22—27. Следовательно, распределение вероятностей величины $\mathbf{N}$ определяется равенствами $\mathbf{P}\{\mathbf{N}=1\}=1/9$, $\mathbf{P}\{\mathbf{N}=2\}=2/3$, $\mathbf{P}\{\mathbf{N}=3\}=2/9$. Совместные распределения величин $(\mathbf{N},\ \mathbf{X}_1)$ и $(\mathbf{X}_1,\ \mathbf{X}_2)$ приведены в табл. 1 и 2.

Таблица 1

Совместное распределение $(\mathbf{N},\ \mathbf{X}_1)$ в примере «а»

$\mathbf{N}$ \ $\mathbf{X}_1$	0	1	2	3	Распределение $\mathbf{N}$
1	2/27	0	0	1/27	1/9
2	6/27	6/27	6/27	0	2/3
3	0	6/27	0	0	2/9
Распределение $\mathbf{X}_1$	8/27	12/27	6/27	1/27	

$\mathbf{E}(\mathbf{N})=19/9$, $\quad \mathbf{E}(\mathbf{N}^2)=129/27$, $\quad \mathrm{Var}(\mathbf{N})=26/81$,

$\mathbf{E}(\mathbf{X}_1)=1$, $\quad \mathbf{E}(\mathbf{X}_1^2)=45/27$, $\quad \mathrm{Var}(\mathbf{X}_1)=2/3$,

$\mathbf{E}(\mathbf{N}\mathbf{X}_1)=19/9$, $\quad \mathrm{Cov}(\mathbf{N},\ \mathbf{X}_1)=0$.

$\mathbf{N}$ — число занятых ящиков, $\mathbf{X}_1$ — число шаров в первом ящике, когда три шара случайно размещаются по трем ящикам.

Таблица 2

Совместное распределение (X_1, X_2) в примере «а»

X_2 \ X_1	0	1	2	3	Распределение X_2
0	1/27	3/27	3/27	1/27	8/27
1	3/27	6/27	3/27	0	12/27
2	3/27	3/27	0	0	6/27
3	1/27	0	0	0	1/27
Распределение X_1	8/27	12/27	6/27	1/27	

$E(X_i)=1$, $E(X_i^2)=45/27$, $\mathrm{Var}(X_i)=2/3$,
$E(X_1X_2)=2/3$, $\mathrm{Cov}(X_1, X_2)=-1/3$.

X_i — число шаров в i-м ящике, когда 3 шара случайно размещаются по 3 ящикам.

б) *Полиномиальное распределение.* Существует множество ситуаций, в которых совместное распределение трех случайных величин задается полиномиальным распределением (см. гл. VI, 9), т. е.

$$\mathbf{P}\{X_1=k_1,\ X_2=k_2,\ X_3=k_3\}=\\ =\frac{n!\,p_1^{k_1}p_2^{k_2}p_3^{k_3}(1-p_1-p_2-p_3)^{n-k_1-k_2-k_3}}{k_1!\ k_2!\ k_3!\ (n-k_1-k_2-k_3)!}; \qquad (1.7)$$

здесь k_1, k_2 и k_3 — неотрицательные целые числа, такие, что $k_1+k_2+k_3\leqslant n$. Например, если X_1, X_2 и X_3 представляют собой числа выпадений единицы, двойки и тройки при n бросаниях правильной кости, то их совместное распределение задается формулой (1.7) с $p_1=p_2=p_3=1/6$. Снова предположим, что производится выборка с возвращением из генеральной совокупности, элементы которой разбиты на несколько групп. Если X_j обозначает число попавших в выборку элементов из j-й группы, то совместное распределение (X_1, X_2, X_3) имеет вид (1.7).

Чтобы получить (маргинальное) распределение величин (X_1, X_2), нужно зафиксировать k_1 и k_2 и просуммировать вероятности (1.7) по всем возможным значениям k_3, т. е. $k_3=0, \ldots, n-k_1-k_2$. Используя формулу бинома Ньютона, получаем триномиальное распределение

$$\mathbf{P}\{X_1=k_1,\ X_2=k_2\}=\frac{n!\,p_1^{k_1}p_2^{k_2}(1-p_1-p_2)^{n-k_1-k_2}}{k_1!\,k_2!\,(n-k_1-k_2)!}. \qquad (1.8)$$

Суммируя по $k_2=0, \ldots, n-k_1$, находим распределение величины X_1: оно сводится к биномиальному распределению с $p=p_1$.

в) *Геометрические распределения.* Рассмотрим последовательность испытаний Бернулли, которые продолжаются по меньшей мере до появления двух успехов. Пусть $\mathbf{X}_1$ — число неудач до первого успеха, а $\mathbf{X}_2$ — число неудач между первыми двумя успехами. Совместное распределение величин $(\mathbf{X}_1, \mathbf{X}_2)$ дается формулой (см. гл. VI, 8)

$$\mathbf{P}\{\mathbf{X}_1 = j,\ \mathbf{X}_2 = k\} = q^{j+k}p^2. \tag{1.9}$$

Суммируя по k, получаем геометрическое распределение для $\mathbf{X}_1$. (Этот пример попутно показывает, как использование случайной величины позволяет избегать трудностей, связанных с несчетными пространствами элементарных событий.)

г) *Выборка случайного объема.* Несколько удивительный результат получается из одного варианта примера б). Предположим, что число испытаний не фиксировано заранее, а зависит от исхода случайного эксперимента таким образом, что вероятность осуществления ровно n испытаний равно $e^{-\lambda}\lambda^n/n!$. Иначе говоря, число испытаний теперь само является случайной величиной с распределением Пуассона $\{e^{-\lambda}\lambda^n/n!\}$. При заданном числе испытаний n событие $\{\mathbf{X}_1 = k_1,\ \mathbf{X}_2 = k_2,\ \mathbf{X}_3 = k_3\}$ имеет (условную) вероятность, равную правой части (1.7). Чтобы найти абсолютную (безусловную) вероятность этого события, нужно умножить правую часть в (1.7) на $e^{-\lambda}\lambda^n/n!$ и просуммировать по всем возможным значениям n. Для данных k_j, конечно, необходимо, чтобы

$$n \geqslant k_1 + k_2 + k_3.$$

Вводя разность $r = n - (k_1 + k_2 + k_3)$ в качестве нового индекса суммирования, получаем

$$\mathbf{P}\{\mathbf{X}_1 = k_1,\ \mathbf{X}_2 = k_2,\ \mathbf{X}_3 = k_3\} = \\ = e^{-\lambda}\frac{(\lambda p_1)^{k_1}(\lambda p_2)^{k_2}(\lambda p_3)^{k_3}}{k_1!\,k_2!\,k_3!}\sum_{r=0}^{\infty}\frac{\lambda^r(1-p_1-p_2-p_3)^r}{r!}. \tag{1.10}$$

В правой части мы видим степенной ряд и можем записать окончательный результат в виде

$$\mathbf{P}\{\mathbf{X}_1 = k_1,\ \mathbf{X}_2 = k_2,\ \mathbf{X}_3 = k_3\} = \\ = e^{-\lambda p_1}\frac{(\lambda p_1)^{k_1}}{k_1!}\cdot e^{-\lambda p_2}\frac{(\lambda p_2)^{k_2}}{k_2!}\cdot e^{-\lambda p_3}\frac{(\lambda p_3)^{k_3}}{k_3!}. \tag{1.11}$$

После суммирования по k_2 и k_3 исчезают второй и третий сомножители, и находим, что величина $\mathbf{X}_1$ сама имеет распределение Пуассона. Любопытный факт заключается в том, что совместное распределение имеет форму таблицы умножения. О такой ситуации будем говорить, что *три величины* $\mathbf{X}_j$ *взаимно независимы.* (Этот пример является по существу иной формулировкой задачи 27 гл. VI,10.) ►

В обозначениях (1.3) условная вероятность события $\mathbf{Y}=y_k$ при условии, что $\mathbf{X}=x_j$ [и $f(x_j)>0$], записывается в виде

$$\mathbf{P}\{\mathbf{Y}=y_k|\mathbf{X}=x_j\}=p(x_j, y_k)/f(x_j). \qquad (1.12)$$

Таким образом, каждому значению $\mathbf{X}$ ставится в соответствие некоторое число, и поэтому (1.12) определяет функцию от $\mathbf{X}$, которая называется *условным распределением* $\mathbf{Y}$ *при заданном* $\mathbf{X}$ и обозначается $\mathbf{P}\{\mathbf{Y}=y_k|\mathbf{X}\}$. Из табл. 1 и 2 видно, что условная вероятность (1.12), вообще говоря, отличается от $g(y_k)$. Это означает, что по известным значениям $\mathbf{X}$ можно сделать некоторые выводы о значениях $\mathbf{Y}$, и наоборот; две случайные величины являются (стохастически) *зависимыми*. Наибольшая степень зависимости будет тогда, когда $\mathbf{Y}$ является функцией от $\mathbf{X}$, т. е. когда значение $\mathbf{X}$ однозначно определяет $\mathbf{Y}$. Например, если монета бросается n раз, а $\mathbf{X}$ и $\mathbf{Y}$ являются числами выпадений гербов и решеток, то $\mathbf{Y}=n-\mathbf{X}$. Аналогично, когда $\mathbf{Y}=\mathbf{X}^2$, по значению $\mathbf{X}$ можно вычислить $\mathbf{Y}$. Для совместного распределения это означает, что все члены каждой строки, за исключением одного, нулевые. С другой стороны, если $p(x_j, y_k)=f(x_j)\,g(y_k)$ для всех комбинаций x_j, y_k, то события $\mathbf{X}=x_j$ и $\mathbf{Y}=y_k$ независимы; совместное распределение принимает форму таблицы умножения. В этом случае говорят о *независимых* случайных величинах. Они получаются, в частности, в связи с независимыми испытаниями, например числа, появляющиеся при двух бросаниях кости, независимы. Иллюстрацию другого рода дает пример г).

Заметим, что совместное распределение $\mathbf{X}$ и $\mathbf{Y}$ определяет распределения величин $\mathbf{X}$ и $\mathbf{Y}$, но мы не можем найти совместное распределение $\mathbf{X}$ и $\mathbf{Y}$ по их маргинальным распределениям. Если две величины $\mathbf{X}$ и $\mathbf{Y}$ имеют одинаковое распределение, то они могут быть независимыми, а могут и не быть таковыми. Например, две величины $\mathbf{X}_1$ и $\mathbf{X}_2$ в табл. 2 имеют одинаковые распределения и независимы.

Все наши замечания применимы и в случае более чем двух величин. Подведем итог формальным определением.

Определение. *Случайная величина* $\mathbf{X}$ *есть функция, определенная на данном пространстве элементарных событий, т. е.* $\mathbf{X}$ *задает правило, по которому каждому элементарному событию соответствует некоторое действительное число. Распределение вероятностей случайной величины* $\mathbf{X}$ *есть функция, определенная в* (1.1). *Если две случайные величины* $\mathbf{X}$ *и* $\mathbf{Y}$ *определены на одном и том же пространстве элементарных событий, то их совместное распределение задается равенствами* (1.3), *в которых всем комбинациям* (x_j, y_k) *значений, принимаемых* $\mathbf{X}$ *и* $\mathbf{Y}$, *приписываются определенные вероятности. Это понятие очевидным образом переносится на произвольное конечное множество величин* $\mathbf{X}, \mathbf{Y}, \ldots, \mathbf{W}$, *определенных на одном и том же пространстве элементарных событий. Эти величины называются взаимно независимыми, если для любой комбинации их значений*,

$(x, y, \ldots, w)$ справедливо равенство

$$\mathbf{P}\{\mathbf{X}=x,\ \mathbf{Y}=y,\ \ldots,\ \mathbf{W}=w\}= \\ =\mathbf{P}\{\mathbf{X}=x\}\,\mathbf{P}\{\mathbf{Y}=y\}\ldots\mathbf{P}\{\mathbf{W}=w\}. \quad (1.13)$$

В гл. V, 4 было определено пространство элементарных событий, соответствующее n взаимно независимым испытаниям. Сравнивая это определение с (1.13), видим, что *если $\mathbf{X}_k$ зависит только от исхода k-го испытания, то величины $\mathbf{X}_1, \ldots, \mathbf{X}_n$ взаимно независимы.* Вообще, если случайная величина $\mathbf{U}$ зависит лишь от исходов первых k испытаний, а другая величина $\mathbf{V}$ зависит только от исходов остальных $n-k$ испытаний, то $\mathbf{U}$ и $\mathbf{V}$ независимы (см. задачу 39).

Мы можем представлять себе случайную величину как набор числовых отметок на точках пространства элементарных событий. Это представление знакомо по ситуации с игральной костью, где грани занумерованы, и мы говорим о числах как о возможных исходах отдельных испытаний. Пользуясь обычной математической терминологией, можно сказать, что случайная величина $\mathbf{X}$ является отображением исходного пространства элементарных событий в новое пространство, состоящее из точек $x_1, x_2, \ldots$. Следовательно,

Если последовательность $\{f(x_j)\}$ удовлетворяет очевидным условиям (1.2), то можно говорить о случайной величине $\mathbf{X}$, принимающей значения $x_1, x_2, \ldots$ с вероятностями $f(x_1), f(x_2)\ldots$, без дальнейших указаний на исходное пространство элементарных событий; новое пространство образуется точками $x_1, x_2, \ldots$. Задание распределения вероятностей эквивалентно заданию пространства элементарных событий, точками которого являются действительные числа. Говорить о двух независимых случайных величинах $\mathbf{X}$ и $\mathbf{Y}$ с распределениями $\{f(x_j)\}$ и $g\ \{(y_k)\}$ — значит говорить о пространстве элементарных событий, точками которого являются пары чисел (x_j, y_k) с вероятностями, заданными по правилу $\mathbf{P}\{(x_j, y_k)\}=f(x_j)g(y_k)$. Аналогично за пространство элементарных событий, соответствующее множеству n случайных величин $(\mathbf{X}, \mathbf{Y}, \ldots, \mathbf{W})$, можно взять совокупность точек $(x, y, \ldots, w)$ в n-мерном пространстве с вероятностями, определяемыми совместным распределением. Случайные величины взаимно независимы, если их совместное распределение задается формулой (1.13).

Пример. д) *Испытания Бернулли с переменными вероятностями.* Рассмотрим n *независимых* испытаний, каждое из которых имеет лишь два возможных исхода, $У$ и H. Вероятность исхода $У$ в k-м испытании равна p_k, для исхода H она равна $q_k=1-p_k$. Если $p_k=p$, то эта схема сводится к испытаниям Бернулли. Проще всего описать нашу схему, приписав исходам $У$ и H значения 1 и 0 и сказав, что у нас есть n взаимно независимых случайных величин $\mathbf{X}_k$ с распределениями $\mathbf{P}\{\mathbf{X}_k=1\}=p_k$, $\mathbf{P}\{\mathbf{X}_k=0\}=q_k$. Модель теперь полностью описана. Эта схема также известна под названием «испыта-

ния Пуассона» (или схемы Пуассона), которое может вызвать путаницу [1]). (См. примеры 5, б) и гл. XI, 6, б).) ►

Ясно, что одно и то же распределение может появляться при различных пространствах элементарных событий. Если мы говорим, что случайная величина **X** принимает значения 0 и 1 с вероятностями 1/2, то мы молчаливо подразумеваем пространство элементарных событий, состоящее из двух точек 0 и 1. Но величина **X** может быть определена тем условием, что она равна 0 или 1 в зависимости от того, выпадает ли при десятом бросании монеты герб или решетка; в этом случае **X** определена на пространстве последовательностей (*ГГР*...), и это пространство элементарных событий состоит из 2^{10} точек.

В принципе можно ограничить теорию вероятностей изучением пространств элементарных событий, определенных в терминах распределений вероятностей случайных величин. При таком подходе не нужно будет обращаться к абстрактному пространству элементарных событий и использовать термины типа «испытания» и «исходы экспериментов». Сведение теории вероятностей к случайным величинам является кратчайшим путем к применению анализа и упрощает многие стороны теории. Однако оно имеет и недостаток, состоящий в затушевывании вероятностных оснований. Представление о случайной величине как о «чем-то принимающем разные значения с различными вероятностями» легко может остаться неясным. Но случайные величины являются обычными функциями, и в этом понятии нет ничего специфического, принадлежащего только теории вероятностей.

Пример. е) Пусть **X** — случайная величина с возможными значениями $x_1, x_2, \ldots$ и соответствующими вероятностями $f(x_1), f(x_2), \ldots$. Читатель всегда сумеет, если это поможет его воображению, построить мысленный эксперимент, приводящий к **X**. Например, разделим колесо рулетки на дуги $l_1, l_2, \ldots$, длины которых относятся как $f(x_1) : f(x_2) : \ldots$. Представим себе игрока, получающего (положительную или отрицательную) сумму x_j, если рулетка остановится в точке дуги l_j. Тогда **X** есть выигрыш игрока. В n испытаниях выигрыши соответствуют n независимым величинам с одинаковым распределением $\{f(x_j)\}$. Чтобы получить две величины с заданным совместным распределением $\{p(x_j, y_k)\}$, возьмем дуги, соответствующие каждой комбинации (x_j, y_k), и вообразим двух игроков, получающих соответственно суммы x_j и y_k. ►

Если **X**, **Y**, **Z**,...,— случайные величины, определенные на одном и том же пространстве элементарных событий, то произвольная функ-

---

[1]) Как правило, эти названия применяют к случаю, когда $p_1, \ldots, p_n$ *зависят* от n («треугольная схема»; см, конец примера гл. XI, 6, б)). — *Прим. перев.*

ция $F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots)$ снова является случайной величиной. Ее распределение может быть получено из совместного распределения величин $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$ просто путем сложения тех вероятностей, которые соответствуют комбинациям $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots)$, приводящим к одинаковым значениям функции $F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots)$.

Примеры. ж) В примере, который иллюстрировался табл. 2, сумма $\mathbf{X}_1+\mathbf{X}_2$ является случайной величиной, принимающей значения 0, 1, 2, 3 с вероятностями 1/27, 6/27, 12/27 и 8/27. Произведение $\mathbf{X}_1 \mathbf{X}_2$ принимает значения 0, 1 и 2 с вероятностями 15/27, 6/27 и 6/27.

з) Вернемся к примеру в) и рассмотрим различные функции от $\mathbf{X}_1$ и $\mathbf{X}_2$. Наиболее интересна сумма $\mathbf{S}=\mathbf{X}_1+\mathbf{X}_2$. Для нахождения $\mathbf{P}\{\mathbf{S}=\nu\}$ нужно просуммировать (1.9) по всем значениям j, k, для которых $j+k=\nu$. Существует $\nu+1$ таких пар, и в рассматриваемом случае все они имеют одинаковые вероятности $p^\nu q^2$. Итак, $\mathbf{P}\{\mathbf{X}=\nu\}=(\nu+1)\,p^\nu q^2$, что является частным случаем формулы (8.1) гл. VI.

Пусть теперь $\mathbf{U}$ определяется как наименьшая из двух величин $\mathbf{X}_1$, $\mathbf{X}_2$; иначе говоря, $\mathbf{U}=\mathbf{X}_1$, если $\mathbf{X}_2 \geqslant \mathbf{X}_1$, и $\mathbf{U}=\mathbf{X}_2$, если $\mathbf{X}_2 \leqslant \mathbf{X}_1$. Для нахождения $\mathbf{P}\{\mathbf{U}=\nu\}$ нужно просуммировать (1.9) по всем парам (j, k), для которых $j=\nu$ и $k \geqslant \nu$ или $j>\nu$ и $k=\nu$. Это приводит к суммам двух геометрических прогрессий

$$\mathbf{P}\{\mathbf{U}=\nu\}=\frac{q^{2\nu}p^2}{1-q}+\frac{q^{2\nu+1}p^2}{1-q}=q^{2\nu}(1+q)\,p, \qquad (1.14)$$

где $\nu=0, 1, \ldots$.

Аналогичные вычисления показывают, что

$$\mathbf{P}\{\mathbf{X}_1-\mathbf{X}_2=\nu\}=q^{|\nu|}p/(1+q), \quad \nu=0, \pm 1, \pm 2, \ldots . \qquad (1.15)$$

►

Замечание о попарной независимости. В примере гл. V, 3, д) как о любопытном факте говорится о том, что три попарно независимых события могут не быть взаимно независимыми. Чтобы сформулировать аналогичный результат для случайных величин, рассмотрим простейший случай, а именно пространство элементарных событий, состоящее из девяти точек, каждая из которых имеет вероятность 1/9. Шесть из этих точек отождествим с различными перестановками чисел 1, 2 и 3, а остальные три точки поставим в соответствие тройкам (1, 1, 1), (2, 2, 2) и (3, 3, 3). Введем теперь три случайные величины $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, такие, что $\mathbf{X}_k$ равна числу, стоящему на k-м месте. Возможными значениями этих величин являются 1, 2, 3, и легко проверить, что их частные и совместные распределения даются соотношениями

$$\mathbf{P}\{\mathbf{X}_j=r\}=1/3, \quad \mathbf{P}\{\mathbf{X}_j=r,\ \mathbf{X}_k=s\}=1/9. \qquad (1.16)$$

(Здесь отличие от выводов в примере гл. V, 3, д) состоит только в обозначениях.) Из (1.16) следует, что три введенные случайные величины попарно независимы. С другой стороны, зная $\mathbf{X}_1$ и $\mathbf{X}_2$, мы однозначно находим $\mathbf{X}_3$, так что эти величины не являются взаимно независимыми.

Определим теперь тройку $(\mathbf{X}_4, \mathbf{X}_5, \mathbf{X}_6)$ так же, как тройку $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, но независимо от нее. Итак, получаем шесть попарно независимых величин, удовлетворяющих соотношениям (1.16). Продолжая подобным образом, приходим к *последовательности величин* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$, *удовлетворяющих* (1.16) *и таких, что* $\mathbf{X}_k$ *попарно независимы, но не взаимно независимы*[1]). Мы вернемся к этому в примере гл. XV, 13, е).

§ 2. МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

Чтобы добиться разумной простоты, часто приходится описывать распределения вероятностей при помощи немногих «типичных значений». В качестве примера можно привести максимальную вероятность биномиального распределения и медиану, которая использовалась в задачах о времени ожидания в гл. II, 7. Самым важным из этих «типичных значений» является, безусловно, математическое ожидание, или среднее. Оно наиболее удобно для аналитических преобразований и обладает полезным для статистиков свойством, называемым выборочной устойчивостью. Определение математического ожидания вытекает из обычного понятия среднего значения. Если в некоторой группе семей n_k семей имеют ровно по k детей, то общее число семей равно $n=n_0+n_1+n_2+\ldots$, а общее число детей составляет $m=n_1+2n_2+3n_3+\ldots$. Среднее число детей в семье равно m/n. Аналогия между вероятностями и частотами подсказывает следующее определение.

Определение. *Пусть* $\mathbf{X}$ — *случайная величина, принимающая значения* $x_1, x_2, \ldots$ *с вероятностями* $f(x_1), f(x_2), \ldots$. *Математическое ожидание, или среднее, случайной величины* $\mathbf{X}$ *определяется формулой*

$$\mathrm{E}(\mathbf{X})=\sum x_k f(x_k) \tag{2.1}$$

при условии, что ряд в правой части абсолютно сходится. В этом случае говорят, что $\mathbf{X}$ *имеет конечное математическое ожидание. Если ряд* $\sum |x_k| f(x_k)$ *расходится, то говорят, что* $\mathbf{X}$ *не имеет конечного математического ожидания.*

Иногда удобно представить себе вероятности как пределы частот, наблюдаемых в повторных опытах. Такое представление приводит к следующей интуитивной трактовке понятия математического ожидания. Пусть некоторый опыт повторяется «при неизменных условиях» n раз, и пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — наблюдающиеся в действительности значения величины $\mathbf{X}$. При больших значениях n среднее значение $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$ должно быть близко к $\mathrm{E}(\mathbf{X})$. Законы больших чисел придают этому неясному интуитивному описанию точный смысл.

Само собой разумеется, что большинство обычно встречающихся случайных величин имеют конечные математические ожидания, в про-

[1]) Это построение можно изменить так, чтобы любые три последовательные случайные величины не были независимыми. Дальнейшие модификации приводят к различным контрпримерам в теории случайных процессов. См. Feller W., Non-Markovian processes with the semi-group property, Ann. Math. Statist., 30 (1959), 1252—1253.

тивном случае это понятие было бы бесплодным (хотя величины, не имеющие конечного математического ожидания, возникают в связи с важными физическими задачами о возвращении). Термины *математическое ожидание* и *среднее* являются синонимами. Говорят также о *среднем распределения*, не указывая соответствующей случайной величины. Обозначение $\mathbf{E}(\mathbf{X})$ общепринято в математике и статистике. В физике обозначение $\mathbf{E}(\mathbf{X})$ обычно заменяется обозначениями $\overline{\mathbf{X}}$, $\langle\mathbf{X}\rangle$ и $\langle\mathbf{X}\rangle_{\mathrm{Av}}$.

Займемся теперь вычислением математических ожиданий для таких функций, как, например, $\mathbf{X}^2$. Эта функция является новой случайной величиной, принимающей значения x_k^2; вообще говоря, вероятность того, что $\mathbf{X}^2=x_k^2$, равна не $f(x_k)$, а сумме $f(x_k)+f(-x_k)$, и $\mathbf{E}(\mathbf{X}^2)$ определяется как сумма $x_k^2\{f(x_k)+f(-x_k)\}$. Очевидно, в любом случае

$$\mathbf{E}(\mathbf{X}^2)=\sum x_k^2 f(x_k) \tag{2.2}$$

при условии, что ряд сходится. Тем же способом можно получить общую теорему.

Теорема 1. *Каждая функция* $\varphi(x)$ *определяет новую случайную величину* $\varphi(\mathbf{X})$. *Если* $\varphi(\mathbf{X})$ *имеет конечное математическое ожидание, то*

$$\mathbf{E}(\varphi(\mathbf{X}))=\sum \varphi(x_k) f(x_k). \tag{2.3}$$

Ряд в правой части (2.3) *сходится тогда и только тогда, когда* $\mathbf{E}(\varphi(\mathbf{X}))$ *существует. Для любой постоянной* a *справедливо соотношение* $\mathbf{E}(a\mathbf{X})=a\mathbf{E}(\mathbf{X})$.

Если несколько случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$ определены на одном и том же пространстве элементарных событий, то их сумма $\mathbf{X}_1+\ldots+\mathbf{X}_n$ является новой случайной величиной. Ее возможные значения и соответствующие вероятности нетрудно найти по совместному распределению величин $\mathbf{X}_\nu$, и, таким образом, можно вычислить $\mathbf{E}(\mathbf{X}_1+\ldots+\mathbf{X}_n)$. Более простой способ вычисления этого математического ожидания дается следующей важной теоремой.

Теорема 2. *Если случайные величины* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *имеют математические ожидания, то математическое ожидание их суммы существует и равно сумме их математических ожиданий:*

$$\mathbf{E}(\mathbf{X}_1+\ldots+\mathbf{X}_n)=\mathbf{E}(\mathbf{X}_1)+\ldots+\mathbf{E}(\mathbf{X}_n). \tag{2.4}$$

Доказательство. Достаточно доказать (2.4) для двух случайных величин $\mathbf{X}$ и $\mathbf{Y}$. Используя обозначения (1.3), можем записать

$$\mathbf{E}(\mathbf{X})+\mathbf{E}(\mathbf{Y})=\sum_{j,k} x_j p(x_j, y_k)+\sum_{j,k} y_k p(x_j, y_k), \tag{2.5}$$

где суммирование производится по всем возможным значениям x_j, y_k (которые не обязательно различны). Оба ряда сходятся абсолютно.

следовательно, их сумму можно переписать в виде $\sum_{j,k}(x_j+y_k)\,p(x_j, y_k)$, что по определению является математическим ожиданием $\mathbf{X}+\mathbf{Y}$. Тем самым доказательство закончено. ►

Ясно, что аналогичная общая теорема для произведений неверна. Например, $\mathbf{E}(\mathbf{X}^2)$, вообще говоря, отличается от $(\mathbf{E}(\mathbf{X}))^2$. Так, если $\mathbf{X}$ — число очков, выпавшее при бросании правильной игральной кости, то

$$\mathbf{E}(\mathbf{X}) = 7/2, \quad \text{но } \mathbf{E}(\mathbf{X}^2) = (1+4+9+16+25+36)/6 = 91/6.$$

Однако для независимых случайных величин справедливо простое правило умножения.

Теорема 3. *Если $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с конечными математическими ожиданиями, то их произведение является случайной величиной с конечным математическим ожиданием и*

$$\mathbf{E}(\mathbf{X}\mathbf{Y}) = \mathbf{E}(\mathbf{X})\ \mathbf{E}(\mathbf{Y}). \tag{2.6}$$

Доказательство. Для нахождения $\mathbf{E}(\mathbf{X}\mathbf{Y})$ нужно умножить каждое возможное значение $x_j y_k$ на соответствующую вероятность. Следовательно,

$$\mathbf{E}(\mathbf{X}\mathbf{Y}) = \sum_{j,k} x_j y_k f(x_j)\, g(y_k) = \left\{\sum_j x_j f(x_j)\right\}\left\{\sum_k y_k g(y_k)\right\}, \tag{2.7}$$

причем перестановка членов возможна в силу абсолютной сходимости рядов. ►

По индукции это правило умножения доказывается для произвольного числа взаимно независимых случайных величин.

Для математического ожидания условного распределения вероятностей удобно иметь свое обозначение. Если $\mathbf{X}$ и $\mathbf{Y}$ — две случайные величины с совместным распределением (1.3), то *условное математическое ожидание* $\mathbf{E}(\mathbf{Y}|\mathbf{X})$ *величины* $\mathbf{Y}$ *при заданном* $\mathbf{X}$ *есть функция, принимающая в точке x_j значение*

$$\sum_k y_k \mathbf{P}\{\mathbf{Y}=y_k \mid \mathbf{X}=x_j\} = \left(\sum_k y_k p(x_j, y_k)\right)\Big/ f(x_j). \tag{2.8}$$

Это определение имеет смысл только в том случае, когда ряд в правой части сходится абсолютно и $f(x_j)>0$ при всех j.

Условное математическое ожидание $\mathbf{E}(\mathbf{Y}|\mathbf{X})$ является новой случайной величиной. Для вычисления ее математического ожидания нужно умножить (2.8) на $f(x_j)$ и просуммировать по x_j. В результате получим

$$\mathbf{E}(\mathbf{E}(\mathbf{Y}\mid\mathbf{X})) = \mathbf{E}(\mathbf{Y}). \tag{2.9}$$

§ 3. ПРИМЕРЫ И ПРИЛОЖЕНИЯ

а) *Биномиальное распределение.* Пусть $\mathbf{S}_n$ — число успехов в n испытаниях Бернулли с вероятностью успеха p. Мы знаем, что $\mathbf{S}_n$ имеет биномиальное распределение $\{b(k;\ n,\ p)\}$; следовательно, $\mathbf{E}(\mathbf{S}_n)=\sum kb(k;\ n,\ p)=np\sum b(k-1;\ n-1,\ p)$. Последняя сумма состоит из всех членов биномиального распределения для $n-1$ испытаний и поэтому равна 1. Итак, *математическое ожидание биномиального распределения* равно

$$\mathbf{E}(\mathbf{S}_n)=np. \tag{3.1}$$

Этот результат можно получить без вычислений с помощью метода, который часто оказывается полезным. Пусть $\mathbf{X}_k$ — число успехов в k-м испытании. Эта случайная величина принимает только значения 0 и 1 с соответствующими вероятностями q и p. Следовательно,

$$\mathbf{E}(\mathbf{X}_k)=0\cdot q+1\cdot p=p,$$

а так как

$$\mathbf{S}_n=\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n, \tag{3.2}$$

то мы получаем (3.1) непосредственно из (2.4).

б) *Распределение Пуассона.* Если $\mathbf{X}$ имеет распределение Пуассона $p(k;\ \lambda)=e^{-\lambda}\lambda^k/k!$, где $k=0,\ 1,\ \ldots,$ то

$$\mathbf{E}(\mathbf{X})=\sum kp(k;\ \lambda)=\lambda\sum p(k-1;\ \lambda).$$

Последний ряд содержит все члены распределения, и поэтому его сумма равна единице. Следовательно, λ *есть математическое ожидание распределения Пуассона* $\{e^{-\lambda}\lambda^k/k!\}$.

в) *Отрицательное биномиальное распределение.* Пусть $\mathbf{X}$ — случайная величина, имеющая *геометрическое распределение* $\mathbf{P}\{\mathbf{X}=k\}=q^kp$, где $k=0,\ 1,\ 2,\ \ldots$. Тогда $\mathbf{E}(\mathbf{X})=qp(1+2q+3q^2+\ldots)$. Справа в скобках стоит производная геометрического ряда, поэтому $\mathbf{E}(\mathbf{X})=qp(1-q)^{-2}=q/p$. В гл. VI, 8 мы видели, что $\mathbf{X}$ можно трактовать как число неудач до первого успеха в последовательности испытаний Бернулли. В более общем случае рассматривается пространство элементарных событий, соответствующее испытаниям Бернулли, которые продолжаются до n-го успеха. Для $r<n$ положим $\mathbf{X}_1=\mathbf{X}$ и обозначим через $\mathbf{X}_r$ число неудач между $(r-1)$-м и r-м успехами. Тогда каждая величина $\mathbf{X}_\nu$ имеет геометрическое распределение $\{q^kp\}$ и $\mathbf{E}(\mathbf{X}_\nu)=q/p$. Сумма

$$\mathbf{Y}_r=\mathbf{X}_1+\ldots+\mathbf{X}_r$$

дает число неудач до r-го успеха. Иначе говоря, $\mathbf{Y}_r$ есть случайная величина с отрицательным биномиальным распределением, определяемым одной из двух эквивалентных формул (8.1) или (8.2) гл. VI. Отсюда следует, что *математическое ожидание отрицательного* *биномиального распределения равно rq/p.* Это можно проверить непосредственными вычислениями. Из формулы (8.2) гл. VI ясно, что

$$kf(k;\ r,\ p)=rp^{-1}\ qf(k-1;\ r+1,\ p),$$

и сумма членов распределения $\{f(k-1;\ r+1,\ p)\}$ дает единицу. Эти непосредственные вычисления хороши тем, что они применимы и в случае *нецелых* значений r. С другой стороны, первоначальные рассуждения позволяют получить тот же результат, не зная точного вида распределения $\mathbf{X}_1+\ldots+\mathbf{X}_r$.

г) *Время ожидания при выборе.* Из генеральной совокупности, содержащей N различных элементов, производится выбор с возвращением. Вследствие возможных повторений случайная выборка объема r в общем случае будет состоять из менее чем r различных элементов. При увеличении объема выборки новые элементы будут появляться в ней все реже и реже. Нас интересует, каков должен быть объем выборки $\mathbf{S}_r$, чтобы в ней содержалось r различных элементов. (Как частный случай рассмотрим генеральную совокупность из $N=$ $=365$ возможных дней рождения; здесь $\mathbf{S}_r$ представляет собой число людей, попавших в выборку до момента, когда в выборке окажется r различных дней рождения. Аналогичная трактовка возможна и при случайном размещении шаров по ящикам. Рассматриваемая задача представляет особый интерес для тех, кто собирает купоны и другие предметы, так как новое приобретение можно сравнить со случайным выбором [1]).)

Для упрощения изложения назовем очередное извлечение элемента успешным, если выбранный элемент появляется в выборке впервые. Тогда $\mathbf{S}_r$ представляет собой число извлечений до r-го успеха включительно. Положим $\mathbf{X}_k=\mathbf{S}_{k+1}-\mathbf{S}_k$. Тогда $\mathbf{X}_k-1$ есть число неудачных извлечений между k-м и $(k+1)$-м успехами. Во время этих извлечений в генеральной совокупности было $N-k$ элементов, которые еще не появлялись в выборке, и поэтому $\mathbf{X}_k-1$ равно числу неудач до первого успеха в испытаниях Бернулли с $p=$ $=(N-k)/N$. Следовательно, согласно примеру в), имеем $\mathrm{E}(\mathbf{X}_k)=$ $=1+q/p=N/(N-k)$. Наконец, поскольку $\mathbf{S}_r=1+\mathbf{X}_1+\ldots+\mathbf{X}_r$, получаем

$$\mathrm{E}(\mathbf{S}_r)=N\ \{1/N+1/(N-1)+\ldots+1/(N-r+1)\}. \tag{3.3}$$

В частности, $\mathrm{E}(\mathbf{S}_N)$ есть ожидаемое число извлечений, необходимых для того, чтобы в выборку вошли все элементы генеральной совокупности. При $N=10$ имеем $\mathrm{E}(\mathbf{S}_5)\approx 6{,}5$ и $\mathrm{E}(\mathbf{S}_{10})\approx 29{,}3$. Это означает, что для выборки первой половины совокупности из 10 элементов в

[1]) Polya G., Eine Warscheinlichkeitsaufgabe zur Kundenwerbung. Zeitschrift für Angewandte Mathematik und Mechanik. 10 (1930), 96—97. Пойа другими методами разбирает несколько более общую задачу. Существует обширная литература, посвященная различным вариантам задачи о собирании купонов. (См. задачи 24, 25, а также задачи 12—14 в гл. XI, 7 и задачу 12 в гл. II, 11.)

среднем достаточно семи извлечений, а на вторую половину потребуется еще 23 извлечения.

Чтобы получить приближенную формулу для (3.3), будем трактовать $(N-k)^{-1}$ как площадь прямоугольника, основанием которого является интервал единичной длины с центром в точке $N-k$, а высотой — значение функции x^{-1} в этой точке. Заменяя площадь этого прямоугольника на площадь области, лежащей ниже графика функции x^{-1}, получаем приближение

$$\mathbf{E}(\mathbf{S}_r) \approx N \int\limits_{N-r+1/2}^{N+1/2} x^{-1}\,dx = N \log[(N+1/2)/(N-r+1/2)]. \quad (3.4)$$

В качестве примера использования этого приближения возьмем произвольное $\alpha<1$ и рассмотрим *ожидаемое число извлечений, необходимых для получения в выборке доли α всей генеральной совокупности*. Оно равно $\mathbf{E}(\mathbf{S}_r)$, где r — наименьшее целое $\geqslant \alpha N$. При $N\to\infty$ погрешность, допускаемая при использовании (3.4), стремится к нулю и искомое математическое ожидание в пределе равно $N\log(1-\alpha)^{-1}$. Заметим, что все эти результаты получены без использования самого распределения вероятностей величины $\mathbf{S}_r$. [Его легко найти по вероятностям (2.3) гл. IV, полученным в задаче о размещении.]

д) *Задача об оценке.* Некоторая урна содержит шары с номерами от 1 до N. Пусть $\mathbf{X}$ — *наибольший номер, полученный при n извлечениях*, если производится случайный выбор с возвращением. Событие $\mathbf{X}\leqslant k$ означает, что каждый из n вынутых номеров не превышает k, и, следовательно, $\mathbf{P}\{\mathbf{X}\leqslant k\}=(k/N)^n$. Таким образом, распределение вероятностей случайной величины $\mathbf{X}$ определяется формулой

$$\begin{aligned} p_k = \mathbf{P}\{\mathbf{X}=k\} &= \mathbf{P}\{\mathbf{X}\leqslant k\} - \mathbf{P}\{\mathbf{X}\leqslant k-1\} = \\ &= \{k^n-(k-1)^n\}N^{-n}. \end{aligned} \quad (3.5)$$

Отсюда следует, что

$$\begin{aligned} \mathbf{E}(\mathbf{X}) &= \sum_{k=1}^{N} kp_k = N^{-n}\sum_{k=1}^{N}\{k^{n+1}-(k-1)^{n+1}-(k-1)^n\} = \\ &= N^{-n}\left\{N^{n+1}-\sum_{k=1}^{N}(k-1)^n\right\}. \end{aligned} \quad (3.6)$$

При больших N последняя сумма приближенно равна площади области, лежащей ниже кривой $y=x^n$ и ограниченной прямыми $x=0$ и $x=N$, т. е. $N^{n+1}/(n+1)$. Отсюда следует, что при больших N

$$\mathbf{E}(\mathbf{X}) \approx [n/(n+1)]\,N. \quad (3.7)$$

Если в городе имеется $N=1000$ машин и производится случайная выборка объема $n=10$, то математическое ожидание максимального наблюденного регистрационного номера приближенно равно 910. В прикладной статистике наблюденный максимум выборки используется для оценки неизвестного истинного значения N. Этот метод применялся во время последней войны для оценки объема производства противника (см. задачи 8 и 9).

е) *Использование в статистическом критерии.* Этот пример [1]) иллюстрирует практическое применение понятия математического ожидания, позволяющее избежать сложных вычислений, связанных с нахождением вероятностных распределений.

Споры гриба *Sordaria* образуются цепочками по восемь элементов. Каждая цепочка может разделиться на несколько частей, и в конечном счете споры отбрасываются группами, содержащими от 1 до 8 спор. Есть основания предполагать, что разрывы семи соединений между спорами в цепочке стохастически независимы и вероятность разрыва одинакова для всех соединений и равна p. При этих предположениях теоретически возможно найти совместное распределение групп из одной споры, из двух и т. д., но потребуются утомительные вычисления. С другой стороны, для эмпирической проверки гипотез достаточно знать ожидаемые числа групп из одной споры, из двух и т. д., а их легко найти. Например, споры, находящиеся на концах цепочки, с вероятностью p становятся одиночными, в то время как для всех других спор эта вероятность равна p^2. Следовательно, по правилу сложения математическое ожидание числа одиночных спор, образованных из одной цепочки, определяется формулой $\varepsilon_1=2p+6p^2$. Аналогичные рассуждения показывают, что математическое ожидание числа групп из двух спор равно $\varepsilon_2=2qp+5qp^2$, где $q=1-p$. Подобным образом получаем $\varepsilon_3=2q^2p+4q^2p^2, \ldots, \varepsilon_8=q^7$. Математическое ожидание числа всех возможных групп равно $\varepsilon_1+\ldots+\varepsilon_8=1+7p$. (Это очевидно без вычислений, так как математическое ожидание числа разрывов соединений равно $7p$ и каждый разрыв увеличивает количество групп на 1.)

При реальных наблюдениях в полевых условиях было найдено 7251 спор, образовавшихся, очевидно, из $N=907$ цепочек (5 спор не было обнаружено). Если применима наша вероятностная модель, то должно иметь место равенство $(1+7p)\ N\approx 7251$, или $p=0{,}168$. (Этот довод опирается на интуитивное представление о математическом ожидании и обосновывается законом больших чисел.) Наблюденное число f_k групп должно быть близко к математическому ожиданию $N\varepsilon_k$. Как показывает табл. 3, различия невелики и нет оснований отвергать нашу модель. ►

[1]) Этот пример взят из речи Д. Р. Кокса при его вступлении в должность в Бербек-Колледже (Лондон) в 1961 г. Кокс ссылается на работу Ingold C. T., Hadland S. A., New Phytologist, 58 (1959), 46—57.

Таблица 3

Наблюденные числа f_k и математические ожидания Ne_k чисел групп по k спор в примере «е»

k	f_k	Ne_k	k	f_k	Ne_k
1	490	458,3	5	200	170,6
2	343	360,8	6	134	131,7
3	265	281,8	7	72	101,1
4	199	219,7	8	272	250,3

§ 4. ДИСПЕРСИЯ

Пусть $\mathbf{X}$ — случайная величина с распределением $\{f(x_j)\}$, и пусть $r \geqslant 0$ — некоторое целое число. *Если математическое ожидание случайной величины* $\mathbf{X}^r$, *т. е.*

$$\mathbf{E}(\mathbf{X}^r) = \sum x_j^r f(x_j) \tag{4.1}$$

существует, то оно называется моментом порядка r случайной величины $\mathbf{X}$. Если ряд (4.1) не является абсолютно сходящимся, то говорят, что момент порядка r не существует. Поскольку $|\mathbf{X}|^{r-1} \leqslant |\mathbf{X}|^r + 1$, *из существования момента порядка r вытекает существование момента порядка* $r-1$ *и, следовательно, всех моментов меньшего порядка.*

Моменты играют важную роль в общей теории, но в этом томе будет использоваться только момент второго порядка. Если он существует, то существует и математическое ожидание

$$\mu = \mathbf{E}(\mathbf{X}). \tag{4.2}$$

Тогда естественно заменить случайную величину $\mathbf{X}$ ее *отклонением от математического ожидания* $\mathbf{X}-\mu$. Поскольку $(x-\mu)^2 \leqslant 2(x^2 + \mu^2)$, момент второго порядка величины $\mathbf{X}-\mu$ существует, если существует $\mathbf{E}(\mathbf{X}^2)$, и определяется по формуле

$$\mathbf{E}((\mathbf{X}-\mu)^2) = \sum_j (x_j^2 - 2\mu x_j + \mu^2) f(x_j). \tag{4.3}$$

Разбивая правую часть на три отдельные суммы, получаем эквивалентную запись в виде $\mathbf{E}(\mathbf{X}^2) - 2\mu\mathbf{E}(\mathbf{X}) + \mu^2 = \mathbf{E}(\mathbf{X}^2) - \mu^2$.

Определение. *Пусть* $\mathbf{X}$ *— случайная величина, имеющая момент второго порядка* $\mathbf{E}(\mathbf{X}^2)$, *и пусть* $\mu = \mathbf{E}(\mathbf{X})$ *— ее математическое ожидание. Определим величину, называемую дисперсией случайной величины* $\mathbf{X}$, *формулой*

$$\operatorname{Var}(\mathbf{X}) = \mathbf{E}((\mathbf{X}-\mu)^2) = \mathbf{E}(\mathbf{X}^2) - \mu^2. \tag{4.4}$$

Положительное значение корня квадратного из этого числа (или нуль) называется стандартным отклонением случайной величины **X**.

Для простоты будем часто говорить о дисперсии распределения, не указывая соответствующей случайной величины.

Примеры. а) Если **X** принимает значение $\pm c$, каждое с вероятностью 1/2, то $\operatorname{Var}(\mathbf{X})=c^2$.

б) Если **X** — число очков, выпавших при бросании симметричной игральной кости, то $\operatorname{Var}(\mathbf{X})=(1/6)(1^2+2^2+\ldots+6^2)-(7/2)^2=35/12$.

в) Для *распределения Пуассона* $p(k;\lambda)$ среднее равно λ (см. пример 3, б)), и поэтому для дисперсии получаем

$$\sum k^2 p(k;\lambda)-\lambda^2=\lambda\sum kp(k-1;\lambda)-\lambda^2=$$
$$=\lambda\sum(k-1)\,p(k-1;\lambda)+\lambda\sum p(k-1;\lambda)-\lambda^2=\lambda^2+\lambda-\lambda^2=\lambda.$$

В этом случае математическое ожидание и дисперсия совпадают.

г) Для *биномиального распределения* (см. пример 3, а)) аналогичные вычисления показывают, что дисперсия равна

$$\sum k^2 b(k;n,p)-(np)^2=np\sum kb(k-1;n-1,p)-(np)^2=$$
$$=np\{(n-1)p+1\}-(np)^2=npq.$$ ▶

Полезность понятия дисперсии выявится постепенно, в частности, в связи с предельными теоремами в гл. X. Здесь мы отметим, что дисперсия является, хотя и довольно грубой, *мерой разброса* значений случайной величины. Действительно, если дисперсия $\operatorname{Var}(\mathbf{X})=\sum(x_j-\mu)^2 f(x_j)$ мала, то каждый член этой суммы тоже мал. Следовательно, значение x_j, при котором $|x_j-\mu|$ велико, должно иметь малую вероятность. Иначе говоря, при малой дисперсии большие отклонения **X** от среднего μ маловероятны. Обратно, большая дисперсия указывает на то, что не все значения случайной величины **X** лежат вблизи от математического ожидания.

Возможно, некоторым читателям поможет следующая механическая интерпретация. Предположим, что механическая система состоит из единичной массы, распределенной вдоль оси x так, что в точке x_j сконцентрирована масса $f(x_j)$. Тогда среднее μ есть абсцисса *центра тяжести*, а дисперсия — *момент инерции*. Ясно, что различные распределения массы могут иметь одинаковые абсциссу центра тяжести и момент инерции, но хорошо известно, что этими двумя величинами могут быть описаны некоторые важные механические свойства такой системы.

Если **X** представляет собой некоторую величину, которую можно измерить, типа длины или температуры, то ее численные значения зависят от начала отсчета и единицы измерения. Изменение последних означает переход от **X** к новой величине $a\mathbf{X}+b$, где a и b — постоянные. Ясно, что $\operatorname{Var}(\mathbf{X}+b)=\operatorname{Var}(\mathbf{X})$, и поэтому

$$\operatorname{Var}(a\mathbf{X}+b)=a^2\operatorname{Var}(\mathbf{X}). \tag{4.5}$$

Выбор начала отсчета и единицы измерения является в значительной степени произвольным, и часто за начало отсчета удобнее всего брать среднее, а за единицу измерения — стандартное отклонение. Так мы поступили в гл. VII, 3, когда ввели нормированное число успехов $S_n^* = (S_n - np)/\sqrt{npq}$. В общем случае, если $\mathbf{X}$ имеет среднее μ и дисперсию σ^2, то $\mathbf{X}-\mu$ имеет нулевое среднее и дисперсию σ^2, и поэтому у *случайной величины*

$$\mathbf{X}^* = (\mathbf{X}-\mu)/\sigma \quad (\sigma>0) \tag{4.6}$$

среднее равно 0 *и дисперсия* 1. *Она называется нормированной случайной величиной, соответствующей* $\mathbf{X}$. На языке физики переход от $\mathbf{X}$ к $\mathbf{X}^*$ трактовался бы как введение безразмерной величины.

§ 5. КОВАРИАЦИЯ; ДИСПЕРСИЯ СУММЫ

Пусть $\mathbf{X}$ и $\mathbf{Y}$ — две случайные величины, определенные на одном пространстве элементарных событий. Тогда $\mathbf{X}+\mathbf{Y}$ и $\mathbf{XY}$ тоже случайные величины, и их распределения могут быть найдены путем простой группировки членов совместного распределения $\mathbf{X}$ и $\mathbf{Y}$. Наша цель состоит теперь в вычислении $\operatorname{Var}(\mathbf{X}+\mathbf{Y})$. Для этого введем понятие ковариации, которое будет подробнее проанализировано в § 8. Если $\{p(x_j, y_k)\}$ — совместное распределение $\mathbf{X}$ и $\mathbf{Y}$, то математическое ожидание $\mathbf{XY}$ определяется формулой

$$\mathrm{E}(\mathbf{XY}) = \sum x_j y_k p(x_j, y_k) \tag{5.1}$$

при условии, конечно, что ряд сходится абсолютно. Далее, $|x_j y_k| \leqslant (x_j^2+y_k^2)/2$, и, следовательно, $\mathrm{E}(\mathbf{XY})$ обязательно существует, если существуют $\mathrm{E}(\mathbf{X}^2)$ и $\mathrm{E}(\mathbf{Y}^2)$. В этом случае существуют также математические ожидания

$$\mu_x = \mathrm{E}(\mathbf{X}), \quad \mu_y = \mathrm{E}(\mathbf{Y}) \tag{5.2}$$

и величины $\mathbf{X}-\mu_x$ и $\mathbf{Y}-\mu_y$ имеют нулевые средние. Для их произведения по теореме сложения из § 2 имеем

$$\begin{aligned}\mathrm{E}((\mathbf{X}-\mu_x)(\mathbf{Y}-\mu_y)) &= \mathrm{E}(\mathbf{XY}) - \mu_x \mathrm{E}(\mathbf{Y}) - \mu_y \mathrm{E}(\mathbf{X}) + \mu_x\mu_y = \\ &= \mathrm{E}(\mathbf{XY}) - \mu_x\mu_y.\end{aligned} \tag{5.3}$$

Определение. *Ковариация случайных величин* $\mathbf{X}$ *и* $\mathbf{Y}$ *определяется формулой*

$$\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = \mathrm{E}((\mathbf{X}-\mu_x)(\mathbf{Y}-\mu_y)) = \mathrm{E}(\mathbf{XY}) - \mu_x\mu_y. \tag{5.4}$$

Это определение имеет смысл, если $\mathbf{X}$ *и* $\mathbf{Y}$ *имеют конечные дисперсии.*

Из § 2 мы знаем, что для независимых случайных величин $\mathrm{E}(\mathbf{XY}) = \mathrm{E}(\mathbf{X})\ \mathrm{E}(\mathbf{Y})$. Следовательно, из (5.4) вытекает следующая теорема.

Теорема 1. *Если* $\mathbf{X}$ *и* $\mathbf{Y}$ *независимы, то* $\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = 0$.

Заметим, что *обратное утверждение неверно*. Табл. 1 дает пример двух зависимых случайных величин, ковариация которых тем не менее равна нулю. Мы вернемся к этому вопросу в § 8. Следующая теорема важна, и правило сложения (5.6) для независимых случайных величин постоянно применяется.

Теорема 2. *Если* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — *случайные величины с конечными дисперсиями* $\sigma_1^2, \ldots, \sigma_n^2$ *и* $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$, *то*

$$\mathrm{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} \sigma_k^2 + 2\sum_{j,k} \mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_k), \tag{5.5}$$

причем последняя сумма состоит из $\binom{n}{2}$ *пар* $(\mathbf{X}_j, \mathbf{X}_k)$ *с* $j < k$.

В частности, если $\mathbf{X}_j$ *взаимно независимы, то*

$$\mathrm{Var}(\mathbf{S}_n) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2. \tag{5.6}$$

Доказательство. Положим $\mu_k = \mathbf{E}(\mathbf{X}_k)$ и $m_n = \mu_1 + \ldots + \mu_n = \mathbf{E}(\mathbf{S}_n)$. Тогда $\mathbf{S}_n - m_n = \sum (\mathbf{X}_k - \mu_k)$ и

$$(\mathbf{S}_n - m_n)^2 = \sum (\mathbf{X}_k - \mu_k)^2 + 2\sum (\mathbf{X}_j - \mu_j)(\mathbf{X}_k - \mu_k). \tag{5.7}$$

Вычисляя математическое ожидание от обеих частей (5.7), получаем (5.5). ►

Примеры. а) *Биномиальное распределение* $\{b(k;\, n, p)\}$. В примере 3, а) случайные величины $\mathbf{X}_k$ взаимно независимы. Имеем

$$\mathbf{E}(\mathbf{X}_k^2) = 0^2 \cdot q + 1^2 \cdot p = p$$

и $\mathbf{E}(\mathbf{X}_k) = p$. Следовательно, $\sigma_k^2 = p - p^2 = pq$, и из (5.6) видно, что *дисперсия биномиального распределения* равна npq. Этот же результат был получен в примере 4, г) непосредственным вычислением.

б) *Испытания Бернулли с переменными вероятностями успеха.* Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины, такие, что $\mathbf{X}_k$ принимает значения 1 и 0 с вероятностями соответственно p_k и $q_k = 1 - p_k$. Тогда $\mathbf{E}(\mathbf{X}_k) = p_k$ и $\mathrm{Var}(\mathbf{X}_k) = p_k - p_k^2 = p_k q_k$; снова полагая

$$\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n,$$

из (5.6) имеем

$$\mathrm{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} p_k q_k. \tag{5.8}$$

Как в примере 1, д), величину $\mathbf{S}_n$ можно интерпретировать как общее число успехов в n независимых испытаниях, каждое из которых заканчивается либо успехом, либо неудачей. Тогда $p = (p_1 + \ldots + p_n)/n$ есть средняя вероятность успеха, и вполне естественно сравнить описанную ситуацию с испытаниями Бернулли с постоянной вероятностью успеха p. Такое сравнение приводит к поразительному результату. Равенство (5.8) можно переписать в виде

$$\operatorname{Var}(\mathbf{S}_n) = np - \sum p_k^2.$$

Далее, легко показать (элементарными вычислениями или по индукции), что из всех комбинаций $\{p_k\}$, таких, что $\sum p_k = np$, сумма $\sum p_k^2$ имеет наименьшее значение, когда все p_k равны. Отсюда следует, что если средняя вероятность успеха p остается постоянной, то дисперсия $\operatorname{Var}(\mathbf{S}_n)$ *достигает максимума при* $p_1 = \ldots = p_n = p$. Итак, получен удивительный результат: *изменение* p_k, *или их неодинаковость, уменьшает величину случайных флуктуаций*, которая характеризуется дисперсией[1]). Например, число пожаров в городе за год можно рассматривать как случайную величину; при фиксированном среднем значении дисперсия максимальна, если вероятности пожара для всех строений одинаковы. Если для n машин задана средняя производительность p, *то производительность всего комплекса будет наименее равномерной тогда, когда все машины одинаковы*. (Приложение к современному образованию очевидно, но оно не обнадеживает.)

в) *Совпадение карт*. Колода из n занумерованных карт раскладывается в случайном порядке так, что все $n!$ возможных расположений равновероятны. Число совпадений (число карт, попавших на место, соответствующее их номеру) представляет собой случайную величину $\mathbf{S}_n$ со значениями $0, 1, \ldots, n$. Ее распределение вероятностей найдено в гл. IV, 4. Среднее и дисперсию можно вычислить, исходя из этого распределения, однако следующий способ проще и поучительней.

Определим случайную величину $\mathbf{X}_k$, принимающую значения 1 или 0 в зависимости от того, попала ли карта с номером k на k-е место или нет. Тогда $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Далее, каждая карта с вероятностью $1/n$ оказывается на k-м месте. Следовательно, $\mathbf{P}\{\mathbf{X}_k = 1\} = 1/n$ и $\mathbf{P}\{\mathbf{X}_k = 0\} = (n-1)/n$. Поэтому $\mathbf{E}(\mathbf{X}_k) = 1/n$, и отсюда следует, что $\mathbf{E}(\mathbf{S}_n) = 1$: в среднем на колоду приходится одно совпадение.

Чтобы найти $\operatorname{Var}(\mathbf{S}_n)$, вычислим сначала дисперсию σ_k^2 случайной величины $\mathbf{X}_k$:

$$\sigma_k^2 = 1/n - (1/n)^2 = (n-1)/n^2. \tag{5.9}$$

Затем найдем $\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)$. Произведение $\mathbf{X}_j\mathbf{X}_k$ равно 0 или 1, причем последнее имеет место в том случае, когда обе карты с номерами j и k оказываются на своих местах, а вероятность этого события равна

[1]) Более сильные результаты в том же направлении см. в работе Hoeffding W., On the distribution of the number of successes in independent trials, Ann. Math. Statist., 27 (1956), 713—721. О приближении распределением Пуассона см. пример гл. XI, 6, б).

$1/n(n-1)$. Следовательно,

$$\begin{gathered}\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)=1/[n(n-1)],\\ \operatorname{Cov}(\mathbf{X}_j,\ \mathbf{X}_k)=1/[n(n-1)]-1/n^2=1/[n^2(n-1)].\end{gathered}\tag{5.10}$$

Итак, наконец,

$$\operatorname{Var}(\mathbf{S}_n)=n\frac{n-1}{n^2}+2\binom{n}{2}\frac{1}{n^2(n-1)}=1.\tag{5.11}$$

Мы видим, что и среднее, и дисперсия числа совпадений равны единице. Этот результат можно использовать в задаче об *угадывании карт*, которая обсуждалась в гл. IV, 4. Там рассматривались три метода угадывания, первый из которых соответствовал совпадению карт. Второй можно описать как последовательность n испытаний Бернулли с вероятностью $p=1/n$; в этом случае математическое ожидание числа верных догадок равно $np=1$ и дисперсия равна $npq=(n-1)/n$. Математические ожидания в обоих случаях одинаковы, но бо́льшая дисперсия при первом методе указывает на бо́льшие случайные флуктуации около среднего и тем самым обещает несколько более острую игру. (При колодах более сложного состава различие между двумя дисперсиями станет несколько больше, хотя никогда не будет действительно значительным.) При последнем способе угадывания испытуемый все время называет одну и ту же карту; число верных догадок, очевидно, равно единице, а случайных колебаний вообще нет (нулевая дисперсия). Мы видим, что способ угадывания не может повлиять на математическое ожидание числа верных догадок, но от него зависит величина случайных флуктуаций.

г) *Выбор без возвращения*. Предположим, что генеральная совокупность состоит из b черных и g зеленых элементов и производится случайная выборка объема r (без возвращения). Число $\mathbf{S}_k$ черных элементов в выборке есть случайная величина с *гипергеометрическим распределением* (см. гл. II, 6), поэтому математическое ожидание и дисперсия могут быть найдены прямым вычислением. Однако предпочтительнее другой метод. Введем случайную величину $\mathbf{X}_k$, принимающую значения 1 или 0 в зависимости от того, является ли k-й элемент в выборке черным или нет ($k\leqslant r$). Из соображений симметрии вероятность того, что $\mathbf{X}_k=1$, равна $b/(b+g)$, и поэтому

$$\mathbf{E}(\mathbf{X}_k)=b/(b+g),\quad \operatorname{Var}(\mathbf{X}_k)=bg/(b+g)^2.\tag{5.12}$$

Далее, при $j\neq k$ произведение $\mathbf{X}_j\mathbf{X}_k=1$, если j-й и k-й элементы выборки черные, в противном случае $\mathbf{X}_j\mathbf{X}_k=0$. Вероятность того, что $\mathbf{X}_j\mathbf{X}_k=1$, равна $b(b-1)/[(b+g)(b+g-1)]$, и поэтому

$$\begin{gathered}\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)=b(b-1)/[(b+g)(b+g-1)],\\ \operatorname{Cov}(\mathbf{X}_j\mathbf{X}_k)=-bg/[(b+g)^2(b+g-1)].\end{gathered}\tag{5.13}$$

Таким образом,

$$\mathbf{E}(\mathbf{S}_r) = rb/(b+g), \quad \operatorname{Var}(\mathbf{S}_r) = \frac{rbg}{(b+g)^2}\left\{1 - \frac{r-1}{b+g-1}\right\}. \qquad (5.14)$$

При выборке с возвращением мы имели бы такое же среднее и немного бо́льшую дисперсию, а именно $rbg/(b+g)^2$. ►

§ 6. НЕРАВЕНСТВО ЧЕБЫШЕВА [1])

Мы видим, что малость дисперсии указывает на малую вероятность больших отклонений от математического ожидания. Это утверждение уточняется неравенством Чебышева, которое является чрезвычайно полезным. Неравенство Чебышева предполагает существование момента второго порядка.

Теорема. *Для произвольного* $t>0$

$$\mathbf{P}\{|\mathbf{X}|\geqslant t\} \leqslant t^{-2}\mathbf{E}(\mathbf{X}^2). \qquad (6.1)$$

В частности, если $\mathbf{E}(\mathbf{X})=\mu$, *то*

$$\mathbf{P}\{|\mathbf{X}-\mu|\geqslant t\} \leqslant t^{-2}\operatorname{Var}(\mathbf{X}). \qquad (6.2)$$

Доказательство. Если первое неравенство применить к случайной величине $\mathbf{X}-\mu$, то получаем второе неравенство. Используя обозначения § 4, имеем

$$\mathbf{P}\{|\mathbf{X}|>t\} = \sum_{|x_j|\geqslant t} f(x_j) \leqslant t^{-2} \sum_{|x_j|\geqslant t} x_j^2 f(x_j), \qquad (6.3)$$

суммирование производится по всем x_j, абсолютная величина которых не меньше t. Последняя сумма не превышает $\mathbf{E}(\mathbf{X}^2)$, и поэтому (6.1) справедливо.

Неравенство Чебышева нужно рассматривать скорее как теоретический инструмент, чем как практический метод оценивания. Важность этого неравенства вытекает из его универсальности, но от утверждений большой общности нельзя ожидать точных результатов в отдельных случаях.

Примеры. а) Если $\mathbf{X}$ — число очков, выпавших при бросании правильной игральной кости, то (см. пример 4, б)) $\mu=7/2$, $\sigma^2=35/12$. Максимум отклонения $\mathbf{X}$ от μ равен $2{,}5\approx 3\sigma/2$. Вероятность больших отклонений равна нулю, в то время как неравенство Чебышева утверждает лишь, что эта вероятность меньше чем 0,47.

б) Для биномиального распределения $\{b(k;\ n,\ p)\}$ имеем (см. пример 5, а)) $\mu=np$, $\sigma^2=npq$. При больших n мы знаем, что

$$\mathbf{P}\{|\mathbf{S}_n-np|>x\sqrt{npq}\} \approx 1-\mathfrak{N}(x)+\mathfrak{N}(-x). \qquad (6.4)$$

Неравенство Чебышева утверждает лишь, что левая часть меньше чем x^{-2}; это, очевидно, намного более слабая оценка, чем (6.4).

[1]) П. Л. Чебышев (1821—1894).

§ 7*). НЕРАВЕНСТВО КОЛМОГОРОВА

В качестве примера более точной оценки докажем следующее утверждение.

Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины с математическими ожиданиями $\mu_k = \mathbf{E}(\mathbf{X}_k)$ и дисперсиями σ_k^2. Положим

$$\mathbf{S}_k = \mathbf{X}_1 + \ldots + \mathbf{X}_k, \tag{7.1}$$

$$\begin{gathered} m_k = \mathbf{E}(\mathbf{S}_k) = \mu_1 + \ldots + \mu_k, \\ s_k^2 = \operatorname{Var}(\mathbf{S}_k) = \sigma_1^2 + \ldots + \sigma_k^2. \end{gathered} \tag{7.2}$$

Для каждого $t > 0$ вероятность того, что одновременно выполняются n неравенств

$$|\mathbf{S}_k - m_k| < ts_n, \quad k = 1, 2, \ldots, n, \tag{7.3}$$

не меньше, чем $1 - t^{-2}$.

При $n = 1$ эта теорема сводится к неравенству Чебышева. При $n > 1$ неравенство Чебышева дает ту же самую оценку для вероятности единственного соотношения $|\mathbf{S}_n - m_n| < ts_n$, так что неравенство Колмогорова значительно сильнее.

Доказательство. Оценим вероятность x того, что хотя бы одно из неравенств (7.3) не выполняется. Теорема утверждает, что $x \leqslant t^{-2}$.

Определим n случайных величин $\mathbf{Y}_\nu$ следующим образом: $\mathbf{Y}_\nu = 1$, если

$$|\mathbf{S}_\nu - m_\nu| \geqslant ts_n, \tag{7.4}$$

но

$$|\mathbf{S}_k - m_k| < ts_n \quad \text{для } k = 1, 2, \ldots, \nu - 1; \tag{7.5}$$

$\mathbf{Y}_\nu = 0$ во всех остальных случаях. Словесно это выражается так: $\mathbf{Y}_\nu$ равно 1 для всех тех элементарных событий, для которых первым из невыполняющихся неравенств (7.3) является ν-е. Тогда для каждого элементарного события не более чем одна из величин $\mathbf{Y}_k$ равна 1 и сумма $\mathbf{Y}_1 + \mathbf{Y}_2 + \ldots + \mathbf{Y}_n$ может принимать только значения 0 и 1. Эта сумма равна 1 тогда и только тогда, когда хотя бы одно из неравенств (7.3) не выполняется, и поэтому

$$x = \mathbf{P}\{\mathbf{Y}_1 + \ldots + \mathbf{Y}_n = 1\}. \tag{7.6}$$

Поскольку $\mathbf{Y}_1 + \ldots + \mathbf{Y}_n$ равно 0 или 1, имеем $\sum \mathbf{Y}_k \leqslant 1$. Умножая обе части последнего неравенства на $(\mathbf{S}_n - m_n)^2$ и вычисляя математическое ожидание, получаем

$$\sum_{k=1}^{n} \mathbf{E}(\mathbf{Y}_k (\mathbf{S}_n - m_n)^2) \leqslant s_n^2. \tag{7.7}$$

*) В этом параграфе рассматривается специальный вопрос, и при первом чтении его можно пропустить.

Для оценки слагаемых в левой части положим

$$\mathbf{U}_k = (\mathbf{S}_n - m_n) - (\mathbf{S}_k - m_k) = \sum_{\nu=k+1}^{n} (\mathbf{X}_\nu - \mu_\nu); \tag{7.8}$$

тогда

$$\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_n - m_n)^2) = \mathbf{E}(\mathbf{Y}_k(\mathbf{S}_k - m_k)^2) + 2\mathbf{E}(\mathbf{Y}_k\mathbf{U}_k(\mathbf{S}_k - m_k)) + \mathbf{E}(\mathbf{Y}_k\mathbf{U}_k^2). \tag{7.9}$$

Далее, $\mathbf{U}_k$ зависит только от $\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n$, в то время как $\mathbf{Y}_k$ и $\mathbf{S}_k$ зависят только от $\mathbf{X}_1, \ldots, \mathbf{X}_k$. Следовательно, $\mathbf{U}_k$ не зависит от $\mathbf{Y}_k(\mathbf{S}_k - m_k)$, и поэтому $\mathbf{E}(\mathbf{Y}_k\mathbf{U}_k(\mathbf{S}_k - m_k)) = \mathbf{E}(\mathbf{Y}_k(\mathbf{S}_k - m_k)) \times \mathbf{E}(\mathbf{U}_k) = 0$, так как $\mathbf{E}(\mathbf{U}_k) = 0$. Таким образом, из (7.9) находим

$$\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_n - m_n)^2) \geqslant \mathbf{E}(\mathbf{Y}_k(\mathbf{S}_k - m_k)^2). \tag{7.10}$$

Но $\mathbf{Y}_k \neq 0$ только тогда, когда $|\mathbf{S}_k - m_k| \geqslant ts_n$, так что $\mathbf{Y}_k(\mathbf{S}_k - m_k)^2 \geqslant t^2 s_n^2 \mathbf{Y}_k$. Следовательно, объединяя (7.7) и (7.10), получаем

$$s_n^2 \geqslant t^2 s_n^2 \mathbf{E}(\mathbf{Y}_1 + \ldots + \mathbf{Y}_n). \tag{7.11}$$

Поскольку $\mathbf{Y}_1 + \ldots + \mathbf{Y}_n$ равняется либо 0, либо 1, математическое ожидание правой части равно вероятности x, определяемой формулой (7.6). Итак, $xt^2 \leqslant 1$, что и утверждалось. ►

§ 8*). КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Пусть $\mathbf{X}$ и $\mathbf{Y}$ — две произвольные случайные величины со средними μ_x и μ_y и положительными дисперсиями σ_x^2 и σ_y^2. Введем соответствующие нормированные случайные величины $\mathbf{X}^*$ и $\mathbf{Y}^*$, определяемые формулой (4.6). Их ковариация называется *коэффициентом корреляции случайных величин* $\mathbf{X}, \mathbf{Y}$ *и обозначается* $\rho(\mathbf{X}, \mathbf{Y})$. Итак, используя (5.4), имеем

$$\rho(\mathbf{X}, \mathbf{Y}) = \mathrm{Cov}(\mathbf{X}^*, \mathbf{Y}^*) = \mathrm{Cov}(\mathbf{X}, \mathbf{Y}) / (\sigma_x \sigma_y). \tag{8.1}$$

Ясно, что коэффициент корреляции не зависит от выбора начала отсчета и единицы измерения, т. е. для произвольных постоянных a_1, a_2, b_1, b_2, где $a_1 > 0$, $a_2 > 0$, имеем $\rho(a_1\mathbf{X} + b_1, a_2\mathbf{Y} + b_2) = \rho(\mathbf{X}, \mathbf{Y})$.

Использование коэффициента корреляции равносильно особой форме записи ковариации[1]. К сожалению, термин «корреляция» создает неправильное представление о свойствах этого коэффициента. Из § 5 известно, что $\rho(\mathbf{X}, \mathbf{Y}) = 0$, если $\mathbf{X}$ и $\mathbf{Y}$ независимы. Важно понять, что обратное утверждение неверно. На самом деле *коэффициент корреляции* $\rho(\mathbf{X}, \mathbf{Y})$ *может быть равен нулю тогда, когда* $\mathbf{Y}$ *является функцией от* $\mathbf{X}$.

*) В этом параграфе рассматривается специальный вопрос, и при первом чтении его можно пропустить.

[1]) Физик определил бы коэффициент корреляции как «безразмерную ковариацию».

Примеры. а) Пусть X принимает значения ± 1, ± 2 с вероятностью 1/4 каждое. Предположим, что $\mathbf{Y}=\mathbf{X}^2$. Совместное распределение определяется соотношениями $p(-1,1)=p(1,1)=p(2,4)=$ $=p(-2,4)=1/4$. Из соображений симметрии $\rho(\mathbf{X}, \mathbf{Y})=0$, хотя между $\mathbf{Y}$ и $\mathbf{X}$ существует функциональная зависимость.

б) Пусть $\mathbf{U}$ и $\mathbf{V}$ имеют одинаковые распределения и $\mathbf{X}=\mathbf{U}+\mathbf{V}$, $\mathbf{Y}=\mathbf{U}-\mathbf{V}$. Тогда $\mathbf{E}(\mathbf{X}\mathbf{Y})=\mathbf{E}(\mathbf{U}^2)-\mathbf{E}(\mathbf{V}^2)=0$ и $\mathbf{E}(\mathbf{Y})=0$. Следовательно, $\operatorname{Cov}(\mathbf{X}, \mathbf{Y})=0$, и поэтому также $\rho(\mathbf{X}, \mathbf{Y})=0$. Например, $\mathbf{X}$ и $\mathbf{Y}$ могут быть соответственно суммой и разностью очков, выпавших на двух костях. Тогда величины $\mathbf{X}$ и $\mathbf{Y}$ либо обе четны, либо обе нечетны и, следовательно, зависимы. ►

Отсюда вытекает, что коэффициент корреляции никоим образом не является исчерпывающей мерой зависимости между $\mathbf{X}$ и $\mathbf{Y}$. Однако $\rho(\mathbf{X}, \mathbf{Y})$ связано с *линейной* зависимостью между $\mathbf{X}$ и $\mathbf{Y}$.

Теорема. *Всегда* $|\rho(\mathbf{X}, \mathbf{Y})|\leqslant 1$, *причем* $\rho(\mathbf{X}, \mathbf{Y})=\pm 1$ *только тогда, когда существуют постоянные a и b, такие, что справедливо равенство* $\mathbf{Y}=a\mathbf{X}+b$, *за исключением, быть может, тех значений* $\mathbf{X}$, *которые принимаются с нулевой вероятностью.*

Доказательство. Пусть $\mathbf{X}^*$ и $\mathbf{Y}^*$ — нормированные случайные величины. Тогда

$$\operatorname{Var}(\mathbf{X}^*\pm\mathbf{Y}^*)=\operatorname{Var}(\mathbf{X}^*)\pm 2\operatorname{Cov}(\mathbf{X}^*, \mathbf{Y}^*)+\operatorname{Var}(\mathbf{Y}^*)= \\ =2(1\pm\rho(\mathbf{X}, \mathbf{Y})). \tag{8.2}$$

Левая часть не может быть отрицательной, следовательно, $|\rho(\mathbf{X}, \mathbf{Y})|\leqslant 1$. Для $\rho(\mathbf{X}, \mathbf{Y})=1$ необходимо, чтобы $\operatorname{Var}(\mathbf{X}^*-\mathbf{Y}^*)=0$, а это значит, что случайная величина $\mathbf{X}^*-\mathbf{Y}^*$ принимает с вероятностью единица одно значение. В этом случае $\mathbf{X}^*-\mathbf{Y}^*=\text{const}$, и поэтому $\mathbf{Y}=a\mathbf{X}+\text{const}$, где $a=\sigma_y/\sigma_x$. Аналогичное доказательство применимо и к случаю $\rho(\mathbf{X}, \mathbf{Y})=-1$. ►

§ 9. ЗАДАЧИ

1. Семь шаров случайно распределяются по семи ящикам. Пусть $\mathbf{X}_i$ — число ящиков, содержащих ровно i шаров. Используя вероятности из таблицы в § 5 гл. II, выписать совместное распределение $\mathbf{X}_2$ и $\mathbf{X}_3$.

2. Бросаются две правильные кости. Пусть $\mathbf{X}$ — число очков на первой кости и $\mathbf{Y}$ — бо́льшее из двух выпавших чисел. а) Выписать совместное распределение $\mathbf{X}$ и $\mathbf{Y}$. б) Найти математические ожидания, дисперсии и ковариацию.

3. Пусть $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ — соответственно число выпадений герба, число серий гербов и длина максимальной из этих серий при пяти бросаниях монеты. Составить таблицу из 32 элементарных событий с соответствующими значениями $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{Z}$.

Простым подсчетом найти совместное распределение пар $(\mathbf{X}, \mathbf{Y})$, $(\mathbf{X}, \mathbf{Z})$, $(\mathbf{Y}, \mathbf{Z})$ и распределения величин $\mathbf{X}+\mathbf{Y}$ и $\mathbf{X}\mathbf{Y}$. Вычислить математические ожидания, дисперсии и ковариации этих случайных величин.

4. Пусть $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{Z}$ — независимые случайные величины, имеющие одно и то же геометрическое распределение $\{q^k p\}$. Найти а) $\mathbf{P}\{\mathbf{X}=\mathbf{Y}\}$; б) $\mathbf{P}\{\mathbf{X}\geqslant 2\mathbf{Y}\}$; в) $\mathbf{P}\{\mathbf{X}+\mathbf{Y}\leqslant\mathbf{Z}\}$.

5. *Продолжение.* Пусть $\mathbf{U}$ наименьшая из величин $\mathbf{X}$ и $\mathbf{Y}$. Положим $\mathbf{V}=\mathbf{X}-\mathbf{Y}$. Показать, что $\mathbf{U}$ и $\mathbf{V}$ независимы[1]).

6. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ — независимые случайные величины, имеющие распределения Пуассона $\{p(k;\lambda_1)\}$ и $\{p(k;\lambda_2)\}$.

а) Доказать, что $\mathbf{X}_1+\mathbf{X}_2$ имеет распределение Пуассона $\{p(k;\lambda_1+\lambda_2)\}$.

б) Доказать, что *условное распределение* $\mathbf{X}_1$ *при заданной сумме* $\mathbf{X}_1+\mathbf{X}_2$ *есть биномиальное распределение*, а именно что

$$\mathbf{P}\{\mathbf{X}_1=k\mid \mathbf{X}_1+\mathbf{X}_2=n\}=b(k;\ n,\ \lambda_1/(\lambda_1+\lambda_2)). \tag{9.1}$$

7. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ независимы и имеют одно и то же геометрическое распределение $\{q^kp\}$ (как в задаче 4). Не вычисляя, показать, что *условное распределение* $\mathbf{X}_1$ *при заданной сумме* $\mathbf{X}_1+\mathbf{X}_2$ *равномерно*, т. е. что

$$\mathbf{P}\{\mathbf{X}_1=k\mid \mathbf{X}_1+\mathbf{X}_2=n\}=1/(n+1),\quad k=0,\ \ldots,\ n. \tag{9.2}$$

8. Пусть $\mathbf{X}_1,\ \mathbf{X}_2,\ \ldots,\ \mathbf{X}_r$ — взаимно независимые случайные величины, каждая из которых имеет *равномерное распределение* $\mathbf{P}\{\mathbf{X}_j=k\}=1/N$ для $k=1,\ 2,\ \ldots,\ N$. Пусть $\mathbf{U}_n$ — наименьшая из величин $\mathbf{X}_1,\ \ldots,\ \mathbf{X}_n$, а $\mathbf{V}_n$ — наибольшая из них. Найти распределения $\mathbf{U}_n$ и $\mathbf{V}_n$. Как связана эта задача с задачей об оценке (пример 3, д))?

9. *Дополнение к задаче об оценке* (пример 3, д)). а) Найти совместное распределение максимального и минимального из наблюденных значений. Особо разобрать случай $n=2$. (*Указание.* Сначала вычислить $\mathbf{P}\{\mathbf{X}\leqslant r,\ \mathbf{Y}\geqslant s\}$.)

б) Найти условную вероятность того, что первые два наблюдения дали j и k, если известно, что $\mathbf{X}=r$.

в) Найти $\mathrm{E}(\mathbf{X}^2)$ и затем асимптотическое выражение для $\mathrm{Var}(\mathbf{X})$ при $N\to\infty$ (при фиксированном n).

10. *Имитация симметричной монеты.* Для данной несимметричной монеты, такой, что вероятность выпадения герба равна α, имитируем правильную монету следующим образом. Бросаем несимметричную монету дважды. Считаем комбинацию $ГР$ за успех, а $РГ$ — за неудачу; если ни одно из этих событий не произошло, то повторяем бросания до осуществления любого из них. а) Показать, что эта модель приводит к испытаниям Бернулли с $p=1/2$. б) Найти распределение и математическое ожидание числа бросаний до появления одного из событий $ГР$ или $РГ$.

11. *Задача Банаха о спичечных коробках* (пример гл. VI, 8, а)). Показать, что математическое ожидание распределения $\{u_r\}$ определяется по формуле $\mu=(2N+1)u_0-1$. Используя формулу Стирлинга, доказать, что оно приближенно равно $2\sqrt{N/\pi}-1$. (При $N=50$ среднее примерно равно 7,04.)

Указание. Начать с соотношения

$$(N-r)\,u_r=(1/2)\,(2N+1)\,u_{r+1}-(1/2)\,(r+1)\,u_{r+1}.$$

Использовать тот факт[2]), что $\sum u_r=1$.

12. *Выборочный контроль.* Предположим, что изделия, каждое из которых с вероятностью p стандартно, подвергаются контролю с вероятностью p' каждое. Имеем четыре класса изделий, а именно «стандартные и проверенные», «стандартные, но не проверенные» и т. д. с соответствующими вероятностями $pp',\ pq',\ p'q,\ qq'$, где $q=1-p,\ q'=1-p'$. Итак, мы имеем дело со сложными испытаниями Бернулли [см. пример гл. VI, 9, в)]. Пусть $\mathbf{N}$ — число изделий, прошедших стол контроля (как проверенных, так и непроверенных), прежде чем обнаружено первое бракованное изделие, и пусть $\mathbf{K}$ — число

---

[1]) Геометрическое распределение является единственным определенным на целых числах распределением вероятностей, для которого это утверждение верно. См. Ferguson T. S., A characterization of the geometric distribution, Amer. Math. Monthly, 72 (1965), 256—260.

[2]) Этот факт аналитически не очевиден; его можно доказать индукцией по N.

необнаруженных бракованных изделий среди них. Найти совместное распределение $\mathbf{N}$ и $\mathbf{K}$ и их маргинальные распределения.

13. *Продолжение.* Найти $\mathrm{E}(\mathbf{K}/(\mathbf{N}+1))$ и $\mathrm{Cov}(\mathbf{K}, \mathbf{N})$. [При промышленном производстве обнаруженное бракованное изделие заменяется стандартным, так что $\mathbf{K}/(\mathbf{N}+1)$ представляет собой долю брака в партии и является мерой ее качества. Заметьте, что $\mathrm{E}(\mathbf{K}/(\mathbf{N}+1))$ не равно $\mathrm{E}(\mathbf{K})/\mathrm{E}(\mathbf{N}+1)$.]

14. Пусть $\mathbf{X}$ — длина серии (успехов или неудач), начавшейся в первом испытании, в последовательности испытаний Бернулли. а) Найти распределение $\mathbf{X}$, $\mathrm{E}(\mathbf{X})$, $\mathrm{Var}(\mathbf{X})$. б) Пусть $\mathbf{Y}$ — длина *второй* серии. Найти распределение $\mathbf{Y}$, $\mathrm{E}(\mathbf{Y})$, $\mathrm{Var}(\mathbf{Y})$ и совместное распределение $\mathbf{X}$, $\mathbf{Y}$.

15. Пусть $\mathbf{X}$ и $\mathbf{Y}$ имеют одно и то же отрицательное биномиальное распределение. Найти условную вероятность $\mathrm{P}\{\mathbf{X}=j \mid \mathbf{X}+\mathbf{Y}=k\}$ и показать, что равенство (12.16) гл. II теперь очевидно без всяких вычислений[1]).

16. Доказать, что если две случайные величины $\mathbf{X}$ и $\mathbf{Y}$ принимают только два значения каждая и $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})=0$, то $\mathbf{X}$ и $\mathbf{Y}$ независимы.

---

17. *Дни рождения.* Для группы из n человек найти математическое ожидание числа дней в году, на каждый из которых приходятся дни рождения ровно k человек. (Предполагается, что в году 365 дней и все размещения дней рождения равновероятны.)

18. *Продолжение.* Найти математическое ожидание числа дней в году, являющихся днем рождения по меньшей мере двух человек. Как велико должно быть n, чтобы это среднее было больше 1?

19. Человек, имеющий n ключей, хочет отпереть свою дверь, испытывая ключи независимо один от другого и в случайном порядке. Найти математическое ожидание и дисперсию числа испытаний, а) если неподошедшие ключи не исключаются из дальнейших испытаний, б) если они исключаются. (Предполагается, что к двери подходит только один ключ. Точное распределение приведено в гл. II, 7, но для решения этой задачи оно не требуется.)

20. Пусть $(\mathbf{X}, \mathbf{Y})$ — случайные величины, совместное распределение которых является триномиальным, т. е. определяется формулой (1.8). Найти $\mathrm{E}(\mathbf{X})$, $\mathrm{Var}(\mathbf{X})$ и $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})$ а) непосредственным вычислением, б) представляя $\mathbf{X}$ и $\mathbf{Y}$ как суммы n случайных величин и используя методы § 5.

21. Найти ковариацию числа выпадений единицы и шестерки при n бросаниях кости.

22. В задаче об отлове животных (задача 24 гл. VI, 10) доказать, что математическое ожидание числа животных, пойманных при ν-м отлове, равно $nqp^{\nu-1}$.

23. Пусть $\mathbf{X}$ имеет *геометрическое* распределение $\mathrm{P}\{\mathbf{X}=k\}=q^k p$, где $k=0, 1, \ldots$. Показать, что $\mathrm{Var}(\mathbf{X})=qp^{-2}$. Вывести отсюда, что дисперсия *отрицательного* биномиального распределения $\{f(k; r, p)\}$ равна rqp^{-2} при условии, что r — положительное целое число. Непосредственным вычислением доказать, что это утверждение остается справедливым при всех $r>0$.

24. В *задаче о времени ожидания* (пример 3, г)) доказать, что

$$\mathrm{Var}(\mathbf{S}_r)=N\left\{\frac{1}{(N-1)^2}+\frac{2}{(N-2)^2}+\ldots+\frac{r-1}{(N-r+1)^2}\right\}.$$

Проверить, что $N^{-2}\mathrm{E}(\mathbf{S}_N)\sim\sum k^{-2}$. (Сумма этого ряда равна $\pi^2/6$.) *Указание.* Использовать выражение для дисперсии геометрического распределения, найденное в предыдущей задаче.

---

[1]) Этот вывод допускает обобщения на более чем две случайные величины. Он принадлежит Wisniewski T. K. M., Amer. Statistician, 20 (1966), 25.

25. *Продолжение.* Пусть $\mathbf{Y}_r$ — число извлечений, необходимых для получения заранее намеченных r элементов (а не любых r различных элементов, как в § 3). Найти $\mathbf{E}(\mathbf{Y}_r)$ и $\operatorname{Var}(\mathbf{Y}_r)$. (*Замечание.* Точное распределение $\mathbf{Y}_r$ найдено в задаче 12 гл. II, 11, но для решения данной задачи оно не требуется.)

26. *Задача об анализе крови* [1]). Большому числу N людей нужно сделать анализ крови. Это можно организовать двумя способами. (i) Кровь каждого человека исследуется отдельно. В этом случае потребуется N анализов. (ii). Кровь k человек смешивается, и анализируется смесь. Если результат анализа *отрицателен*, то одного этого анализа достаточно для k человек. Если же он *положителен*, то кровь каждого из k человек нужно исследовать отдельно, и для k человек всего потребуется $k+1$ анализов.

Предположим, что вероятность p положительного результата анализа одна и та же для всех людей и что результаты анализов для различных людей стохастически независимы.

а) Найти вероятность того, что анализ смешанной крови k человек даст положительный результат.

б) Чему равно математическое ожидание числа $\mathbf{X}$ анализов при втором способе исследования?

в) Найти уравнение для такого значения k, при котором минимально математическое ожидание числа анализов при втором способе исследования. (Не пытайтесь решить это уравнение.)

г) Показать, что это значение k близко к $1/\sqrt{p}$ и поэтому минимум математического ожидания числа анализов примерно равен $2N\sqrt{p}$. (Это замечание принадлежит М. С. Раффу.)

27. *Структура выборки.* Генеральная совокупность состоит из r классов, количества элементов в которых относятся как $p_1:p_2:\ldots:p_r$. Производится случайная выборка объема n с возвращением. Найти математическое ожидание числа классов, *не* представленных в этой выборке.

28. Пусть $\mathbf{X}$ — число серий альф при случайном размещении r_1 альф и r_2 бет. Распределение $\mathbf{X}$ дается в задаче 23 гл. II, 11. Найти $\mathbf{E}(\mathbf{X})$ и $\operatorname{Var}(\mathbf{X})$.

29. Пусть в *урновой схеме Пойа* (пример гл. V, 2, в)) $\mathbf{X}_n$ равно единице или нулю в зависимости от того, какой шар извлечен при n-м испытании — черный или красный. Доказать, что $\rho(\mathbf{X}_n, \mathbf{X}_m)=c/(b+r+c)$ при $n\neq m$.

30. *Продолжение.* Пусть $\mathbf{S}_n$ — общее число черных шаров, извлеченных при первых n испытаниях (т. е. $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$). Найти $\mathbf{E}(\mathbf{S}_n)$ и $\operatorname{Var}(\mathbf{S}_n)$. Проверить результат с помощью рекуррентной формулы из задачи 22 гл. V, 8. *Указание.* Использовать результаты задач 19 и 20 гл. V, 8.

31. *Выбор по группам.* Город разбит на n кварталов, в n_j из которых насчитывается по x_j жителей $(n_1+n_2+\ldots=n)$. Пусть $m=\sum n_jx_j/n$ — среднее число жителей в квартале. Положим $a^2=n^{-1}\sum n_jx_j^2-m^2$. Выбором без возвращения случайно отбираются r кварталов, и в каждом из них подсчитывается число жителей. Пусть эти числа равны $\mathbf{X}_1,\ldots,\mathbf{X}_r$ соответственно. Показать, что

$$\mathbf{E}(\mathbf{X}_1+\ldots+\mathbf{X}_r)=mr,\qquad \operatorname{Var}(\mathbf{X}_1+\ldots+\mathbf{X}_r)=\frac{a^2r\,(n-r)}{n-1}.$$

(При выборе с возвращением дисперсия была бы больше, а именно a^2r.)

[1]) Эта задача основана на методике, разработанной Р. Дорфманом во время второй мировой войны. При использовании своей методики в армейской практике Дорфман добился экономии на 80%. Когда эта задача появилась в первом издании, она привлекла широкое внимание и привела к различным обобщениям, а также к новым приложениям в биологии и промышленности. Основное улучшение заключается во введении более чем двух этапов. См., например, Sobel M., Groll P. A., Group testing to eliminate efficiently all defectives in a binomial sample, The Bell System Journal, 38 (1959), 1179—1252; Watson G. S., A study of the group screening method, Technometrics, 3 (1961). 371—388; Finucan H. M., The blood-testing problem, Applied Statistics, 13 (1964), 43—50.

32. *Длина случайной цепи* [1]). Цепь на плоскости x, y состоит из n звеньев единичной длины. Угол между двумя последовательными звеньями равен $\pm\alpha$, где α — положительная постоянная. Каждая из возможностей имеет вероятность 1/2, и величины последовательных углов взаимно независимы. Расстояние $\mathbf{L}_n$ от начала до конца цепи есть случайная величина, и мы хотим доказать, что

$$\mathrm{E}(\mathbf{L}_n^2)=n\frac{1+\cos\alpha}{1-\cos\alpha}-2\cos\alpha\frac{1-\cos^n\alpha}{(1-\cos\alpha)^2}. \tag{9.3}$$

Не ограничивая общности, можно предполагать, что первое звено расположено в положительном направлении оси x. Угол между k-м звеном и положительным направлением оси x есть случайная величина $\mathbf{S}_{k-1}$, где $\mathbf{S}_0=0$, $\mathbf{S}_k=\mathbf{S}_{k-1}+\mathbf{X}_k\alpha$, а $\mathbf{X}_k$ — взаимно независимые случайные величины, принимающие значения ± 1 с вероятностями 1/2. Проекциями k-го звена на оси x и y будут соответственно $\cos\mathbf{S}_{k-1}$ и $\sin\mathbf{S}_{k-1}$. Следовательно, для $n\geqslant 1$

$$\mathbf{L}_n^2=\left(\sum_{k=0}^{n-1}\cos\mathbf{S}_k\right)^2+\left(\sum_{k=0}^{n-1}\sin\mathbf{S}_k\right)^2. \tag{9.4}$$

Доказать последовательно по индукции, что при $m<n$

$$\mathrm{E}(\cos\mathbf{S}_n)=\cos^n\alpha,\quad \mathrm{E}(\sin\mathbf{S}_n)=0, \tag{9.5}$$

$$\mathrm{E}((\cos\mathbf{S}_m)\cdot(\cos\mathbf{S}_n))=\cos^{n-m}\alpha\cdot\mathrm{E}(\cos^2\mathbf{S}_m), \tag{9.6}$$

$$\mathrm{E}((\sin\mathbf{S}_m)\cdot(\sin\mathbf{S}_n))=\cos^{n-m}\alpha\cdot\mathrm{E}(\sin^2\mathbf{S}_m), \tag{9.7}$$

$$\mathrm{E}(\mathbf{L}_n^2)-\mathrm{E}(\mathbf{L}_{n-1}^2)=1+2\cos\alpha\cdot(1-\cos^{n-1}\alpha)/(1-\cos\alpha) \tag{9.8}$$

(где $\mathbf{L}_0=0$) и отсюда получить наконец (9.3).

---

33. Испытания Бернулли проводятся до тех пор, пока число успехов не достигнет r, где r — фиксированное целое число. Пусть $\mathbf{X}$ — число испытаний, которые для этого потребовались. Найти [2]) $\mathrm{E}(r/\mathbf{X})$. (Формула для этого математического ожидания содержит бесконечный ряд, для которого может быть получено конечное выражение.)

34. При случайном размещении r шаров по n ящикам вероятность того, что m ящиков окажутся пустыми, удовлетворяет реккурентной формуле (11.8) гл. II. Пусть m_r — математическое ожидание числа пустых ящиков. *Используя реккурентную формулу*, доказать, что

$$m_{r+1}=(1-n^{-1})\,m_r$$

и вывести отсюда, что $m_r=n(1-1/n)^r$.

35. Пусть $\mathbf{S}_n$ — число успехов в n испытаниях Бернулли. Доказать, что

$$\mathrm{E}(|\mathbf{S}_n-np|)=2\nu q b(\nu;\ n,\ p),$$

где ν — целое число, такое, что $np<\nu\leqslant np+1$.

---

[1]) Это двухмерный аналог химической задачи о длине *длинной полимерной молекулы*. Задача иллюстрирует применение к случайным величинам, которые не представляются в виде суммы простых случайных величин.

[2]) Этот пример иллюстрирует влияние *остановки в произвольный момент времени*. Если число испытаний n фиксировано, то отношение числа успехов N к числу испытаний n есть случайная величина со средним p. Часто ошибочно предполагают, что то же самое верно и в нашем примере, где число успехов r фиксировано, а число испытаний зависит от случая. Если $p=1/2$ и $r=2$, то $\mathrm{E}(2/\mathbf{X})=0{,}614$ вместо 0,5; при $r=3$ находим $\mathrm{E}(3/\mathbf{X})=0{,}579$.

Указание. Величина в левой части равна $2\sum_{k=0}^{\nu-1}(np-k)\binom{n}{k}p^kq^{n-k}$.

Другая возможность: использовать формулу (10.7) гл. VI.

38. Пусть $\{\mathbf{X}_k\}$ — последовательность взаимно независимых одинаково распределенных случайных величин. Предположим, что $\mathbf{X}_k$ принимает только положительные значения и что существуют математические ожидания $\mathrm{E}(\mathbf{X}_k)=a$ и $\mathrm{E}(\mathbf{X}_k^{-1})=b$. Положив $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$, доказать, что $\mathrm{E}(\mathbf{S}_n^{-1})$ конечно и $\mathrm{E}(\mathbf{X}_k\mathbf{S}_n^{-1})=n^{-1}$ при $k=1, 2, \ldots, n$.

37. *Продолжение* [1]. Доказать, что

$$\mathrm{E}(\mathbf{S}_m/\mathbf{S}_n)=m/n, \qquad \text{если } m\leqslant n,$$

$$\mathrm{E}(\mathbf{S}_m/\mathbf{S}_n)=1+(m-n)\,a\mathrm{E}(\mathbf{S}_n^{-1}), \qquad \text{если } m\geqslant n.$$

38. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые одинаково распределенные случайные величины со средним m и дисперсией σ^2. Положив $\overline{\mathbf{X}}=(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$, доказать, что [2]

$$[1/(n-1)]\,\mathrm{E}\left(\sum_{k=1}^{n}(\mathbf{X}_k-\overline{\mathbf{X}})^2\right)=\sigma^2.$$

39. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины, $\mathbf{U}$ — функция от $\mathbf{X}_1, \ldots, \mathbf{X}_k$ и $\mathbf{V}$ — функция от $\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n$. Доказать, что $\mathbf{U}$ и $\mathbf{V}$ — взаимно независимые случайные величины.

---

40. *Обобщенное неравенство Чебышева.* Пусть $\varphi(x)>0$ монотонно возрастает при $x>0$; предположим, что существует $\mathrm{E}(\varphi(|\mathbf{X}|))=M$. Доказать, что

$$\mathrm{P}\{|\mathbf{X}|\geqslant t\}\leqslant M/\varphi(t).$$

41. *Неравенство Шварца.* Для любых двух случайных величин $\mathbf{X}$ и $\mathbf{Y}$ с конечными дисперсиями справедливо неравенство $\mathrm{E}^2(\mathbf{XY})\leqslant\mathrm{E}(\mathbf{X}^2)\,\mathrm{E}(\mathbf{Y}^2)$. Доказать это, используя тот факт, что квадратичная форма $\mathrm{E}((t\mathbf{X}+\mathbf{Y})^2)$ неотрицательна.

---

[1] Чжун Кайлай заметил, что результаты этой задачи можно получить из результата задачи 36.

[2] Иначе это можно выразить так: $\sum(\mathbf{X}_k-\overline{\mathbf{X}})^2/(n-1)$ является *несмещенной оценкой* для σ^2.

ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ

§ 1. ОДИНАКОВО РАСПРЕДЕЛЕННЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Предельные теоремы для испытаний Бернулли, установленные в гл. VII и VIII, являются частными случаями общих предельных теорем, которые в данном томе не рассматриваются. Однако для того, чтобы установить новую точку зрения на математическое ожидание случайной величины, мы рассмотрим здесь хотя бы некоторые формулировки закона больших чисел.

Связь между испытаниями Бернулли и теорией случайных величин станет яснее, если мы рассмотрим зависимость числа успехов $\mathbf{S}_n$ от числа испытаний n. При каждом испытании $\mathbf{S}_n$ возрастает на 1 или 0, и мы можем записать это в виде

$$\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n, \tag{1.1}$$

где случайная величина $\mathbf{X}_k$ принимает значение 1 в случае успеха k-го испытания и значение 0 в случае неудачи. Следовательно, $\mathbf{S}_n$ есть сумма n взаимно независимых случайных величин, каждая из которых принимает значение 1 или 0 с вероятностями p и q соответственно. Отсюда остается только один шаг до того, чтобы рассмотреть суммы вида (1.1), где $\mathbf{X}_k$ — взаимно независимые случайные величины с произвольными распределениями. Согласно (слабому) закону больших чисел из гл. VI, 4, для больших n весьма правдоподобно, что среднее число успехов $\mathbf{S}_n/n$ близко к p. Это утверждение является частным случаем следующей теоремы.

Закон больших чисел. *Пусть* $\{\mathbf{X}_k\}$ — *последовательность взаимно независимых одинаково распределенных случайных величин. Если математическое ожидание* $\mu=\mathbf{E}(\mathbf{X}_k)$ *существует, то для любого* $\varepsilon>0$ *при* $n\to\infty$

$$\mathbf{P}\{|(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n-\mu|>\varepsilon\}\to 0. \tag{1.2}$$

Иначе говоря, вероятность того, что среднее $\mathbf{S}_n/n$ будет отличаться от математического ожидания меньше чем на произвольно заданное ε, стремится к 1.

В таком общем виде теорема впервые была доказана А. Я. Хинчиным[1]). Предыдущие доказательства проводились при излишнем

[1]) Khintchin A. Comptes rendus de l'Académie des Sciences, Paris, 189 (1929), 477—479. Между прочим, читатель должен иметь в виду предостережение, сделанное в связи с законом больших чисел для испытаний Бернулли в конце § 4 гл. VI.

ограничении, состоявшем в требовании конечности дисперсии [1]) $\mathrm{Var}(\mathbf{X}_k)$. Однако для этого случая существует гораздо более точный результат, обобщающий предельную теорему Муавра — Лапласа для испытаний Бернулли, а именно следующая теорема.

Центральная предельная теорема. *Пусть* $\{\mathbf{X}_k\}$ — *последовательность взаимно независимых одинаково распределенных случайных величин. Предположим, что математическое ожидание* $\mu = \mathbf{E}(\mathbf{X}_k)$ *и дисперсия* $\sigma^2 = \mathrm{Var}(\mathbf{X}_k)$ *существуют, и положим* $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. *Тогда для любого фиксированного* β

$$\mathbf{P}\{(\mathbf{S}_n - n\mu)/(\sigma\sqrt{n}) < \beta\} \to \mathfrak{N}(\beta). \tag{1.3}$$

Здесь $\mathfrak{N}(x)$ — нормальное распределение, определенное в гл. VII, 1. Эта теорема принадлежит Линдебергу [2]); А. М. Ляпунов и другие авторы доказывали ее раньше при более ограничительных условиях. Следует отметить, что эта теорема является только частным случаем гораздо более общей теоремы, формулировка и доказательство которой приводятся в томе 2. Здесь мы отметим, что утверждение (1.3) сильнее, чем (1.2), так как (1.3) дает оценку для вероятности того, что разность $|n^{-1}\mathbf{S}_n - \mu|$ больше чем $\sigma/\sqrt{n}$. С другой стороны, закон больших чисел (1.2) справедлив даже в том случае, когда случайные величины $\mathbf{X}_k$ не имеют конечной дисперсии, и, следовательно, он применим к более общему случаю, чем центральная предельная теорема. По этой причине мы дадим независимое доказательство закона больших чисел, но сначала проиллюстрируем эти две теоремы примерами.

Примеры. а) Пусть в последовательности независимых бросаний симметричной кости $\mathbf{X}_k$ — число очков, выпавших при k-м бросании. Тогда

$$\mathbf{E}(\mathbf{X}_k) = (1+2+3+4+5+6)/6 = 3{,}5$$

и $\mathrm{Var}(\mathbf{X}_k) = (1^2+2^2+3^2+4^2+5^2+6^2)/6 - (3{,}5)^2 = 35/12$. Закон больших чисел утверждает: правдоподобно, что для больших n среднее число очков $\mathbf{S}_n/n$ будет близко к 3,5. Согласно центральной предельной теореме,

$$\mathbf{P}\{|\mathbf{S}_n - 3{,}5n| < \alpha\sqrt{35n/12}\} \approx \mathfrak{N}(\alpha) - \mathfrak{N}(-\alpha). \tag{1.4}$$

В случае $n=1000$ и $\alpha=1$ получаем $\mathbf{P}\{3450 < \mathbf{S}_n < 3550\} \approx 0{,}68$. Для $\alpha = 0{,}6744\ldots$ правая часть в (1.4) равна 1/2, поэтому для $\mathbf{S}_n$ шансы находиться внутри интервала 3500 ± 36 или вне его примерно одинаковы.

---

[1]) А. А. Марков показал, что достаточно существования $\mathbf{E}(|\mathbf{X}_k|^{1+\alpha})$ для некоторого $\alpha > 0$.

[2]) Lindeberg J. W., Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift, 15 (1922), 211—225.

б) *Выборка.* Предположим, что в генеральной совокупности, состоящей из N семей, N_k семей имеют ровно по k детей ($k=0, 1, \ldots$; $\sum N_k = N$). Если семья выбирается случайным образом, то число детей в ней является случайной величиной, принимающей значение ν с вероятностью $p_\nu = N_\nu/N$. При выборе с возвращением выборку объема n можно рассматривать как n независимых случайных величин или «наблюдений» $\mathbf{X}_1, \ldots, \mathbf{X}_n$, имеющих одинаковое распределение; $\mathbf{S}_n/n$ является *средним значением выборки.* Согласно закону больших чисел, среднее значение достаточно большой случайной выборки будет, вероятно, близким к $\mu = \sum \nu p_\nu = \sum \nu N_\nu/N$, т. е. к среднему значению генеральной совокупности. Центральная предельная теорема позволяет оценить вероятную величину расхождения между этими величинами и определить размер выборки, необходимый для надежных оценок. На практике μ и σ^2 неизвестны, однако во многих случаях нетрудно получить предварительную оценку для σ^2, и всегда можно заключить σ^2 в надежные границы. Если необходимо, чтобы с вероятностью не меньшей 0,99 среднее значение выборки $\mathbf{S}_n/n$ отличалось от неизвестного среднего значения генеральной совокупности μ меньше, чем на 1/10, то объем выборки должен быть взят таким, чтобы

$$\mathbf{P}\{|(\mathbf{S}_n - n\mu)/n| < 1/10\} \geqslant 0{,}99. \tag{1.5}$$

Корень уравнения $\mathfrak{N}(x) - \mathfrak{N}(-x) = 0{,}99$ равен $x = 2{,}57\ldots$ и, следовательно, число n должно удовлетворять неравенству $\sqrt{n}/(10\sigma) \geqslant 2{,}57$, или $n \geqslant 660\sigma^2$. Осторожная предварительная оценка для σ^2 дает возможность найти требуемый объем выборки. Аналогичные ситуации встречаются довольно часто. Так, когда экспериментатор определяет среднее из n измерений, он также полагается на закон больших чисел и использует среднее значение выборки в качестве оценки для неизвестного математического ожидания. Надежность оценки может быть выражена только через σ^2, и обычно мы вынуждены использовать довольно грубые оценки для σ^2.

в) *Распределение Пуассона.* В гл. VII, 5 мы показали, что при больших λ распределение Пуассона $\{p(k; \lambda)\}$ можно аппроксимировать нормальным распределением. На самом деле этот факт является непосредственным следствием центральной предельной теоремы. Предположим, что случайные величины $\mathbf{X}_k$ имеют распределение Пуассона $\{p(k; \gamma)\}$. Тогда $\mathbf{S}_n$ имеет распределение Пуассона $\{p(k; n\gamma)\}$ с математическим ожиданием и дисперсией, равными $n\gamma$. Написав λ вместо $n\gamma$, мы приходим к выводу, что при $n \to \infty$

$$\sum_{k < \lambda + \beta\sqrt{\lambda}} e^{-\lambda}\lambda^k/k! \to \mathfrak{N}(\beta), \tag{1.6}$$

где суммирование производится по всем k, меньшим чем $\lambda + \beta\sqrt{\lambda}$. Теперь очевидно, что утверждение (1.6) справедливо и в том случае, когда $\lambda \to \infty$ произвольным образом. Эта теорема используется в теории суммирования расходящихся рядов и представляет общий интерес. Оценки разности выражений, стоящих справа и слева в равенстве (1.6), можно получить из общей теории. ▶

Замечание о случайных величинах, не имеющих математического ожидания

Если математическое ожидание μ не существует, то и закон больших чисел, и центральная предельная теорема становятся бессмысленными, но их можно заменить более общими теоремами, дающими аналогичную информацию. В современной теории случайные величины, не имеющие математических ожиданий, играют важную роль, и многие времена ожидания и возвращения в физике оказываются величинами такого типа. Это справедливо даже для простой игры с бросанием монеты.

Предположим, что n монет подбрасываются поодиночке. Обозначим через $\mathbf{X}_k$ время ожидания до первого момента, когда у k-й монеты число выпавших решеток будет равно числу выпавших гербов. Случайные величины $\mathbf{X}_k$ взаимно независимы и одинаково распределены; каждая случайная величина $\mathbf{X}_k$ принимает только положительные четные значения и $\mathbf{P}\{\mathbf{X}_k=2r\}=f_{2r}$, причем распределение вероятностей $\{f_{2r}\}$ определяется формулой (3.7) гл. III. Сумма $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ имеет такое же распределение, как и время ожидания до n-го момента, когда число выпавших гербов равно числу выпавших решеток. Такое же распределение имеет момент n-го возвращения в начало в симметричном случайном блуждании. Как показано в теореме 4 гл. III, 7, распределение $\mathbf{S}_n$ имеет вид

$$\mathbf{P}\{\mathbf{S}_n < n^2x\} \to 2[1-\mathfrak{N}(1/\sqrt{x})]. \tag{1.7}$$

Итак, мы получили предельную теорему того же типа, что и центральная предельная теорема, с той заметной разницей, что теперь *предельное распределение имеет случайная величина* $\mathbf{S}_n/n^2$, *а не* $\mathbf{S}_n/n$, *как ранее*. С точки зрения физики случайные величины $\mathbf{X}_k$ можно интерпретировать как независимые измерения некоторого параметра, и наша предельная теорема утверждает, что *среднее*

$$(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$$

возрастает по вероятности линейно вместе с n. Этот парадоксальный результат нельзя отбросить как патологический случай, поскольку оказывается, что $\mathbf{X}_k$ — типичные времена ожидания, получающиеся при рассмотрении многих физических и экономических процессов. Предельная теорема (1.7) является также типичным представителем многих современных предельных теорем для случайных величин, не имеющих математического ожидания [1]).

[1]) Аналоги закона больших чисел для случайных величин, не имеющих математического ожидания, приводятся в § 4 и в задаче 13. Неожиданные следствия утверждения (1.7) подробно обсуждались в гл. III.

§ 2*). ДОКАЗАТЕЛЬСТВО ЗАКОНА БОЛЬШИХ ЧИСЕЛ

Без ограничения общности можно предположить, что $\mu=\mathbf{E}(\mathbf{X}_k)=0$. В противном случае вместо $\mathbf{X}_k$ следует рассматривать $\mathbf{X}_k-\mu$, что привело бы просто к замене обозначений. В частном случае, когда дисперсия $\sigma^2=\operatorname{Var}(\mathbf{X}_k)$ существует, закон больших чисел является тривиальным следствием неравенства Чебышева (6.1), гл. IX, согласно которому

$$\mathbf{P}\{|\mathbf{S}_n|>t\}\leqslant n\sigma^2/t^2. \tag{2.1}$$

При $t=\varepsilon n$ правая часть стремится к 0 при $n\to\infty$, следовательно, утверждение (1.2) справедливо.

Более сложным является случай, когда дисперсия не существует. При этом предположении доказательство основывается на универсальном *методе усечения*, который обычно используется при выводе различных предельных теорем. Пусть δ — фиксированное положительное число, которое мы определим позже. Для каждого n определим следующие n пар случайных величин:

$$\begin{aligned} &\mathbf{U}_k=\mathbf{X}_k, \quad \mathbf{V}_k=0, \qquad \text{если } |\mathbf{X}_k|\leqslant\delta n,\\ &\mathbf{U}_k=0, \quad \mathbf{V}_k=\mathbf{X}_k, \qquad \text{если } |\mathbf{X}_k|>\delta n. \end{aligned} \tag{2.2}$$

Здесь $k=1, \ldots, n$ и следует помнить о зависимости случайных величин $\mathbf{U}_k$ и $\mathbf{V}_k$ от n. Тогда

$$\mathbf{X}_k=\mathbf{U}_k+\mathbf{V}_k. \tag{2.3}$$

Для доказательства закона больших чисел достаточно показать, что для заданного $\varepsilon>0$ постоянная δ может быть выбрана таким образом, что при $n\to\infty$

$$\mathbf{P}\{|\mathbf{U}_1+\ldots+\mathbf{U}_n|>(1/2)\,\varepsilon n\}\to 0 \tag{2.4}$$

и

$$\mathbf{P}\{|\mathbf{V}_1+\ldots+\mathbf{V}_n|>(1/2)\,\varepsilon n\}\to 0. \tag{2.5}$$

Обозначим возможные значения случайных величин $\mathbf{X}_j$ через $x_1, x_2, \ldots$, а вероятности, с которыми $\mathbf{X}_j$ принимают эти значения, через $f(x_j)$. Пусть $a=\mathbf{E}(|\mathbf{X}_j|)$, т. е.

$$a=\sum_j |x_j|\, f(x_j). \tag{2.6}$$

Случайная величина $\mathbf{U}_1$ ограничена величиной δn, и, следовательно,

$$\mathbf{E}(\mathbf{U}_1^2)<a\delta n. \tag{2.7}$$

Так как случайные величины $\mathbf{U}_1, \ldots, \mathbf{U}_n$ взаимно независимы и имеют одинаковые распределения, то

$$\operatorname{Var}(\mathbf{U}_1+\ldots+\mathbf{U}_n)=n\operatorname{Var}(\mathbf{U}_1)\leqslant n\mathbf{E}(\mathbf{U}_1^2)\leqslant a\delta n^2. \tag{2.8}$$

*) Этот параграф может быть опущен при первом чтении.

С другой стороны, из определения случайных величин U_k следует, что при $n \to \infty$

$$E(U_1) \to E(X_1)=0. \qquad (2.9)$$

Отсюда получаем, что при достаточно большом n имеет место неравенство

$$E((U_1+\ldots+U_n)^2) \leqslant 2a\delta n^2. \qquad (2.10)$$

Теперь соотношение (2.4) сразу следует из неравенства Чебышева (6.1) гл. IX, согласно которому

$$P\{|U_1+\ldots+U_n| > (1/2)\varepsilon n\} \leqslant 8a\delta/\varepsilon^2. \qquad (2.11)$$

Выбирая δ достаточно малым, мы можем сделать правую часть сколь угодно малой, и поэтому соотношение (2.4) справедливо.

Что касается соотношения (2.5), то заметим, что, согласно основному неравенству (7.6) гл. I,

$$P\{V_1+\ldots+V_n \neq 0\} \leqslant nP\{V_1 \neq 0\}. \qquad (2.12)$$

Для произвольного $\delta > 0$ имеем

$$P\{V_1 \neq 0\} = P\{|X_1| > \delta n\} = \sum_{|x_j| > \delta n} f(x_j) \leqslant \leqslant [1/(\delta n)] \sum_{|x_j| > \delta n} |x_j| f(x_j). \qquad (2.13)$$

Последняя сумма стремится к 0 при $n \to \infty$. Следовательно, выражение, стоящее в левой части неравенства (2.12), также стремится к 0. Этим утверждением, более сильным, чем (2.5), и завершается доказательство теоремы. ►

§ 3. ТЕОРИЯ «БЕЗОБИДНЫХ» ИГР

При дальнейшем анализе смысла закона больших чисел мы будем пользоваться традиционной терминологией игроков, хотя наши рассуждения в равной степени допускают и менее легкомысленные приложения, а два наших основных предположения более реальны в статистике и физике, чем в игорном доме. Во-первых, предположим, что игрок обладает *неограниченным капиталом*, поэтому никакой проигрыш не приведет к окончанию игры. (Отказ от этого предположения приводит к задаче о *разорении* игрока, которая всегда интригует изучающих теорию вероятностей. Эта задача играет важную роль в последовательном анализе Вальда и в теории стохастических процессов, и мы вернемся к ней в гл. XIV.) Во-вторых, предположим, что игрок *не имеет права прервать игру в произвольный момент; число испытаний n должно быть фиксировано заранее* и не зависеть от хода игры. (В действительности игрок, обладающий неограниченным капиталом, может дождаться серии удач и в подходящий момент прекратить игру. Такого игрока интересует не вероятное состояние в заданный момент, а только максимальные флуктуации, которые можно считать правдоподобными при большом числе партий и которые описываются законом повторного логарифма, а не законом больших чисел (см. гл. VIII, 5).)

В дальнейшем случайная величина $\mathbf{X}_k$ будет обозначать выигрыш (положительный или отрицательный) игрока в k-й партии. Тогда сумма $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ является общим выигрышем в n независимых партиях. Если за право играть игрок уплачивает (не обязательно положительный) взнос μ' перед каждой игрой, то общий уплаченный взнос в n партиях и *общий чистый выигрыш* соответственно равны $n\mu'$ и $\mathbf{S}_n-n\mu'$. Когда существует математическое ожидание $\mu=\mathbf{E}(\mathbf{X}_k)$, применим закон больших чисел, и поэтому, грубо говоря, при больших n весьма правдоподобно, что разность $\mathbf{S}_n-n\mu$ окажется малой по сравнению с n. Поэтому, если μ' меньше μ, то при больших n игрок, вероятно, будет иметь выигрыш порядка $n(\mu-\mu')$. По тем же соображениям взнос $\mu'>\mu$ практически наверняка приводит к убытку. Короче говоря, случай $\mu'<\mu$ *благоприятен* для игрока, тогда как случай $\mu'>\mu$ *неблагоприятен*.

Заметим, что мы еще ничего не говорили о случае $\mu'=\mu$. В этом случае *единственно* возможный вывод состоит в том, что при достаточно большом n общий выигрыш или проигрыш $\mathbf{S}_n-n\mu$ будет с очень большой вероятностью малым по сравнению с n. Но неизвестно, что вероятнее для $\mathbf{S}_n-n\mu$: оказаться положительным или отрицательным, т. е. неизвестно, будет ли игра выгодной или разорительной. Это не было учтено в классической теории, которая называла взнос $\mu'=\mu$ «безобидной» ценой и игру с таким взносом «*безобидной*». Столь многообещающее и привлекательное название приводило к многочисленным заблуждениям. Следует понимать, что «*безобидная» игра может быть явно разорительной для игрока*. ►

В игровых и других простых ситуациях, в которых случайная величина $\mathbf{X}_k$ имеет конечную дисперсию, можно оправдать понятие «безобидная игра», но, когда дисперсия бесконечна, употребление этого термина совершенно не оправдано. Нет оснований считать, что общий чистый выигрыш $\mathbf{S}_n-n\mu'$ будет колебаться около нуля. Действительно, *можно привести примеры «безобидных» игр* [1], *в которых вероятность того, что игрок потерпит чистый убыток, стремится к единице*. Закон больших чисел утверждает, что, вероятно, этот убыток будет величиной меньшего порядка, чем n. Однако, кроме этого, ничего более утверждать нельзя. Если a_n — произвольная последовательность, такая, что $a_n/n\to 0$, то можно придумать «безобидную» игру, в которой вероятность того, что в n партиях общий чистый убыток превысит a_n, стремится к единице. В задаче 15 приводится пример игры, в которой игрок практически

[1]) Feller W., Note on the law of large numbers and «fair» games, Ann. Math. Statist., 16 (1945), 301—304.

может быть уверен, что его убыток превысит $n/\log n$. Эта игра является «безобидной», и взнос за участие в каждой партии равен единице. Трудно себе представить, что игрок будет считать игру «безобидной», если он практически уверен в том, что будет нести постоянно увеличивающийся убыток.

Было бы ошибкой считать эти явления неестественными или не имеющими практического значения. Пренебрежение случайными величинами, не имеющими математического ожидания, нанесло большой ущерб приложениям, так как эти случайные величины играют существенную роль даже в простейших случайных процессах. Например, случайное блуждание (или игра с бросанием монеты), обсуждавшееся в гл. III, служит прототипом многих стохастических процессов в физике и экономике. Как было показано в гл. III, время ожидания и время первого возвращения в этом случайном блуждании не имеют математических ожиданий и поэтому подвержены случайным флуктуациям, так что возникает парадокс, не согласующийся с нашей интуицией. Несовершенная интуиция, так же как и многие современные приложения теории вероятностей, находятся под сильным влиянием традиционного недопонимания смысла закона больших чисел и распространенного представления о так называемом законе о среднем. Это наследство классической теории, в которой математический анализ неизбежно переплетался с эмпирическими и метафизическими соображениями и в которой с различными предельными теоремами связывалось нечто таинственное [1]).

Вернемся к «нормальной» ситуации, при которой существует не только математическое ожидание $E(\mathbf{X}_k)$, но и дисперсия $\mathrm{Var}(\mathbf{X}_k)$. В этом случае закон больших чисел дополняется центральной предельной теоремой, из которой следует, что при «безобидной» игре весьма правдоподобно, что чистый выигрыш $\mathbf{S}_n-n\mu$ в результате продолжительной игры будет иметь величину порядка $\sqrt{n}$ и при больших n он с равными шансами будет положительным или отрицательным. Таким образом, если центральная предельная теорема применима, то термин «безобидная игра» оказывается оправданным, но даже в этом случае мы имеем дело с предельной теоремой, что подчеркивается словами «в результате продолжительной игры».

Чтобы проиллюстрировать это, рассмотрим игровой аппарат, опустив в который доллар игрок может с вероятностью 10^{-6} выиграть 10^6-1 долларов либо потерять опущенный доллар. Здесь мы имеем испытания Бернулли, и игра «безобидна». Сыграв миллион раз, игрок уплатит за это миллион долларов. Он может выиграть 0, 1, 2,... раз. Согласно приближению Пуассона для би-

[1]) Изучающие современную теорию вероятностей, возможно, будут удивлены, узнав, что еще в 1934 г. ведущие специалисты могли сомневаться в возможности сформулировать основные предельные теоремы теории вероятностей чисто аналитически.

номиального распределения, вероятность выиграть ровно k раз с точностью до нескольких десятичных знаков равна $e^{-1}/k!$. Таким образом, с вероятностью 0,368... игрок потеряет миллион и с той же вероятностью он окупит свои расходы; с вероятностью 0,184... он может выиграть один миллион и т. д. Здесь 10^6 испытаний эквивалентны одному-единственному испытанию при игре с выигрышем, имеющим распределение Пуассона. Такая игра может быть осуществлена сравнением двух больших колод карт, как это описано в гл. IV, 4.

На практике бессмысленно ожидать выполнимость закона больших чисел после трех или четырех сравнений. По той же причине не имеет смысла применять закон больших чисел в примере с игровым аппаратом, пока не произведено много миллионов испытаний. К этой же схеме относится страхование от пожара, автомобильных катастроф и т. д.; риску подвергается огромная сумма, но соответствующая вероятность очень мала. Кроме того, страхование происходит только один раз в год, так что число испытаний n никогда не становится большим. Для застрахованного игра будет «небезобидной», но тем не менее экономически выгодной; закон больших чисел здесь ни при чем. Что касается страховой компании, то она имеет дело с большим числом игр; но так как дисперсия велика, то возникают случайные флуктуации. Страховые премии должны быть такими, чтобы предотвратить огромный убыток в отдельные годы, и поэтому компания имеет дело скорее с задачей о разорении, чем с законом больших чисел.

§ 4*). ПЕТЕРБУРГСКАЯ ИГРА

В классической теории понятие математического ожидания не было четко отделено от определения вероятности и в обращении с ним не было достаточной математической строгости. Поэтому изучение случайных величин, не имеющих математических ожиданий, сталкивалось с непреодолимыми трудностями, и даже сравнительно недавние дискуссии кажутся странными тому, кто изучает современную теорию вероятностей. Важность случайных величин, не имеющих математических ожиданий, была подчеркнута в предыдущем параграфе, и вполне естественно привести здесь пример аналога закона больших чисел для этих величин. С этой целью мы рассмотрим известный парадокс петербургской игры [1]).

В петербургской игре каждое испытание состоит в бросании правильной монеты до тех пор, пока не выпадет герб; если это случится при r-м бросании, то игрок получает 2^r долларов. Иначе говоря, мы имеем дело с независимыми случайными величинами, ко-

*) Этот параграф может быть опущен при первом чтении.

[1]) Этот парадокс был рассмотрен Даниилом Бернулли (1700—1782). Отметим, что испытания Бернулли были названы в честь Якова Бернулли.

торые принимают значения $2^1, 2^2, 2^3, \ldots$ с вероятностями $2^{-1}, 2^{-2}, 2^{-3}, \ldots$ соответственно. Математическое ожидание формально определяется суммой $\sum x_r f(x_r)$, в которой $x_r = 2^r$ и $f(x_r) = 2^{-r}$, так что каждое слагаемое равно единице. Таким образом, в этой игре не существует конечного математического ожидания и закон больших чисел неприменим. Ясно, что игра станет менее благоприятной для игрока, если изменить правила и установить, что игрок не получает ничего, если желаемый результат не будет достигнут после N бросаний (т. е. решетка выпадет N раз подряд). В этой измененной игре выигрыш имеет конечное математическое ожидание, равное N, и закон больших чисел применим. Из этого следует, что исходная игра будет «благоприятной» для игрока, даже если он платит взнос, равный N, за право играть в одной партии. Это справедливо для каждого N; однако чем больше N, тем больше должно быть число испытаний, чтобы был вероятен положительный выигрыш, так что бессмысленно говорить о «благоприятной» игре. Классическая теория утверждала, что $\mu' = \infty$ является «безобидным» взносом, но современный студент с трудом поймет туманные рассуждения, приводившие к этому «парадоксу».

Оказывается вполне возможным определить взнос за право участия в петербургской игре таким образом, чтобы она имела все свойства «безобидной» игры в классическом смысле, за исключением того, что этот взнос будет зависеть от номера испытания, вместо того чтобы оставаться постоянным. Переменный взнос неудобен в игорном доме, однако петербургская игра и без того неосуществима вследствие ограниченности имеющихся денежных средств. В случае конечного математического ожидания $\mu = \mathbf{E}(\mathbf{X}_k) > 0$ игра называется «безобидной», если при больших n отношение общего выигрыша $\mathbf{S}_n$ к общему уплаченному взносу e_n будет, вероятно, близким к единице (т. е. если разность $\mathbf{S}_n - e_n$ будет, вероятно, иметь порядок, меньший чем e_n). Если математическое ожидание $\mathbf{E}(\mathbf{X}_k)$ не существует, то нельзя сохранять взнос за право участия в игре постоянным и надо определить e_n иначе. Мы будем говорить, что *игра с общим взносом e_n является безобидной в классическом смысле, если для каждого $\varepsilon > 0$*

$$\mathbf{P}\{|\mathbf{S}_n/e_n - 1| > \varepsilon\} \to 0. \tag{4.1}$$

Это полный аналог закона больших чисел, где $e_n = n\mu'$. Закон больших чисел понимается физиком в том смысле, что среднее из n независимых измерений оказывается близким к μ. В данном примере среднее из n измерений оказывается близким к e_n/n. В тех случаях, когда предельная теорема (4.1) применима, она имеет такое же теоретическое и практическое значение, что и закон больших чисел.

Покажем теперь [1]), что *петербургская игра становится «без-*

[1]) Это частный случай закона больших чисел, из которого легко вывести необходимые и достаточные условия выполнения (4.1); см. Feller W., Acta Scientiarum Litterarum Univ. Szeged, 8 (1937), 191—201.

обидной» в классическом смысле, если положить $e_n = n \operatorname{Log} n$, где логарифм берется по основанию 2, т. е. $2^{\operatorname{Log} n} = n$.

Доказательство. Используем метод усечения из § 2, однако на этот раз определим случайные величины $\mathbf{U}_k$ и $\mathbf{V}_k$ ($k=1, 2, \ldots, n$) следующим образом:

$$\begin{aligned} &\mathbf{U}_k = \mathbf{X}_k, \quad \mathbf{V}_k = 0, \quad \text{если } \mathbf{X}_k \leqslant n \operatorname{Log} n, \\ &\mathbf{U}_k = 0, \quad \mathbf{V}_k = \mathbf{X}_k, \quad \text{если } \mathbf{X}_k > n \operatorname{Log} n. \end{aligned} \tag{4.2}$$

Тогда

$$\mathbf{P}\{|e_n^{-1}\mathbf{S}_n - 1| > \varepsilon\} \leqslant \mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - e_n| > \varepsilon e_n\} + \\ + \mathbf{P}\{\mathbf{V}_1 + \ldots + \mathbf{V}_n \neq 0\}, \tag{4.3}$$

так как событие, стоящее в левой части неравенства, не может произойти до тех пор, пока не произойдет хотя бы одно из событий, стоящих в правой части. Итак,

$$\mathbf{P}\{\mathbf{V}_1 + \ldots + \mathbf{V}_n \neq 0\} \leqslant n\mathbf{P}\{\mathbf{X}_1 > n \operatorname{Log} n\} \leqslant 2/\operatorname{Log} n \to 0. \tag{4.4}$$

Поэтому, чтобы проверить (4.1), достаточно доказать, что

$$\mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - n \operatorname{Log} n| > \varepsilon n \operatorname{Log} n\} \to 0. \tag{4.5}$$

Положим $\mu_n = \mathbf{E}(\mathbf{U}_k)$ и $\sigma_n^2 = \operatorname{Var}(\mathbf{U}_k)$. Эти величины зависят от n, но одинаковы для $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$. Если r — наибольшее из целых чисел, удовлетворяющих неравенству $2^r \leqslant n \operatorname{Log} n$, то $\mu_n = r$, и, следовательно, для достаточно больших n

$$\operatorname{Log} n < \mu_n \leqslant \operatorname{Log} n + \operatorname{Log}\operatorname{Log} n. \tag{4.6}$$

Аналогично

$$\sigma_n^2 < \mathbf{E}(\mathbf{U}_k^2) = 2 + 2^2 + \ldots + 2^r < 2^{r+1} \leqslant 2n \operatorname{Log} n. \tag{4.7}$$

Так как сумма $\mathbf{U}_1 + \ldots + \mathbf{U}_n$ имеет математическое ожидание $n\mu_n$ и дисперсию $n\sigma_n^2$, из неравенства Чебышева получаем

$$\mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - n\mu_n| > \varepsilon n\mu_n\} \leqslant \frac{n\sigma_n^2}{\varepsilon^2 n^2 \mu_n^2} < \frac{2}{\varepsilon^2 \operatorname{Log} n} \to 0. \tag{4.8}$$

Теперь, согласно (4.6), имеем $\mu_n \sim \operatorname{Log} n$, и, следовательно, (4.8) эквивалентно (4.5). ▶

§ 5. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ С РАЗЛИЧНЫМИ РАСПРЕДЕЛЕНИЯМИ

До сих пор мы рассматривали только случайные величины $\mathbf{X}_k$, имеющие одинаковые распределения. Эта ситуация соответствует повторению одной и той же игры; однако еще интереснее выяснить, что случится, если вид игры будет изменяться на каждом шаге. Для этого не обязательно представлять себе игорный дом; статистик, применяющий статистические критерии, занят вполне достойным видом «игры», а в его случае распределение случайных величин меняется от одного шага к другому.

Для определенности представим себе, что задана бесконечная последовательность распределений вероятностей, так что при каждом n мы имеем n взаимно независимых случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$ с заданными распределениями. Предположим, что математические ожидания и дисперсии существуют, и положим

$$\mu_k = \mathbf{E}(\mathbf{X}_k), \quad \sigma_k^2 = \operatorname{Var}(\mathbf{X}_k). \tag{5.1}$$

Сумма $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ имеет математическое ожидание m_n и дисперсию s_n^2, определяемые формулами

$$m_n = \mu_1 + \ldots + \mu_n, \quad s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2 \tag{5.2}$$

(см. формулы (2.4) и (5.6) гл. IX). В частном случае одинаковых распределений мы имели $m_n = n\mu$, $s_n^2 = n\sigma^2$.

Говорят, что для последовательности $\{\mathbf{X}_k\}$ выполняется (слабый) закон больших чисел, если при каждом $\varepsilon > 0$

$$\mathbf{P}\{|\mathbf{S}_n - m_n|/n > \varepsilon\} \to 0. \tag{5.3}$$

О последовательности $\{\mathbf{X}_k\}$ говорят, что она удовлетворяет центральной предельной теореме, если при любых фиксированных $\alpha < \beta$

$$\mathbf{P}\{\alpha < (\mathbf{S}_n - m_n)/s_n < \beta\} \to \mathfrak{N}(\beta) - \mathfrak{N}(\alpha). \tag{5.4}$$

Одна из характерных особенностей теории вероятностей заключается в том, что и закон больших чисел, и центральная предельная теорема выполняются для поразительно широкого класса последовательностей $\{\mathbf{X}_k\}$. В частности, *закон больших чисел справедлив каждый раз, когда случайные величины $\mathbf{X}_k$ равномерно ограничены*, т. е. существует такая константа A, что $|\mathbf{X}_k| < A$ при всех k. *Более общее условие, достаточное для того, чтобы выполнялся закон больших чисел, состоит в том, что*

$$s_n/n \to 0. \tag{5.5}$$

Это утверждение является прямым следствием неравенства Чебышева, и доказательство, приведенное в самом начале § 2, здесь полностью применимо. Заметим, однако, что условие (5.5) не является необходимым (см. задачу 14).

Известны различные достаточные условия для центральной предельной теоремы, но все они следуют из *теоремы Линдеберга* [1]), *согласно которой центральная предельная теорема выполняется, если при каждом $\varepsilon > 0$ усеченные случайные величины $\mathbf{U}_k$, определенные формулами*

$$\begin{aligned} \mathbf{U}_k &= \mathbf{X}_k - \mu_k, && \text{если } |\mathbf{X}_k - \mu_k| \leqslant \varepsilon s_n, \\ \mathbf{U}_k &= 0, && \text{если } |\mathbf{X}_k - \mu_k| > \varepsilon s_n, \end{aligned} \tag{5.6}$$

[1]) См. примечание 2 на стр. 258.

удовлетворяют условиям $s_n \to \infty$ *и*

$$(1/s_n^2) \sum_{k=1}^{n} \mathbf{E}(\mathbf{U}_k^2) \to 1. \tag{5.7}$$

Если случайные величины $\mathbf{X}_k$ равномерно ограничены, т. е. если $|\mathbf{X}_k| < A$, то случайные величины $\mathbf{U}_k = \mathbf{X}_k - \mu_k$ для всех n настолько больших, что $s_n > 2A\varepsilon^{-1}$. Тогда левая часть (5.7) равна 1. Поэтому из теоремы Линдеберга следует, что *любая равномерно ограниченная последовательность* $\{\mathbf{X}_k\}$ *взаимно независимых случайных величин удовлетворяет центральной предельной теореме* в предположении, конечно, что $s_n \to \infty$. Позднее было показано [1]), что условие Линдеберга также является и необходимым для справедливости (5.4). Доказательство этого утверждения будет дано в томе 2; там же приводятся оценки разности между величинами, стоящими в правой и левой частях соотношения (5.4).

Мы показали, что в случае, когда случайные величины $\mathbf{X}_k$ имеют одинаковые распределения, центральная предельная теорема сильнее, чем закон больших чисел. Однако в общем случае это не так, и мы увидим, что центральная предельная теорема применима и к последовательностям, не удовлетворяющим закону больших чисел.

Примеры. а) Пусть $\lambda > 0$ фиксировано, и пусть $\mathbf{X}_k = \pm k^\lambda$, причем каждое значение принимается с вероятностью 1/2 (например, ставка при k-м бросании монеты равна $\pm k^\lambda$). Здесь $\mu_k = 0$, $\sigma_k^2 = k^{2\lambda}$ и

$$s_n^2 = 1^{2\lambda} + 2^{2\lambda} + 3^{2\lambda} + \ldots + n^{2\lambda} \sim n^{2\lambda+1}/(2\lambda + 1). \tag{5.8}$$

Условие (5.5) справедливо при $\lambda < 1/2$. Следовательно, закон больших чисел выполняется при $\lambda < 1/2$; покажем, что при $\lambda \geqslant 1/2$ закон больших чисел не выполняется.

При $k = 1, 2, \ldots, n$ имеем $|\mathbf{X}_k| = k^\lambda \leqslant n^\lambda$, поэтому при $n > (2\lambda + 1)\varepsilon^{-2}$ усеченные переменные $\mathbf{U}_k$ совпадают с $\mathbf{X}_k$. Таким образом, условие Линдеберга выполняется и

$$\mathbf{P}\{\alpha < \sqrt{(2\lambda + 1)/n^{2\lambda+1}}\, \mathbf{S}_n < \beta\} \to \mathfrak{N}(\beta) - \mathfrak{N}(\alpha). \tag{5.9}$$

Следовательно, вероятнее всего, что сумма $\mathbf{S}_n$ будет иметь порядок роста $n^{\lambda+1/2}$, так что при $\lambda \geqslant 1/2$ закон больших чисел не выполняется. В этом примере *центральная предельная теорема*

[1]) Feller W., Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift, 40 (1935), 521—559. Получены также обобщения центральной предельной теоремы, применимые к случайным величинам, не имеющим математических ожиданий. Отметим, что здесь мы рассматриваем только независимые случайные величины; для зависимых случайных величин условие Линдеберга не является ни необходимым, ни достаточным.

применима для всех $\lambda > 0$, *а закон больших чисел — только при* $\lambda < 1/2$.

б) Рассмотрим две независимые последовательности по 1000 бросаний монеты (или извлечение по 1000 монет из двух урн) и исследуем *разность* **D** чисел выпадения герба. Пусть бросания двух последовательностей занумерованы числами от 1 до 1000 и от 1001 до 2000 соответственно, а 2000 случайных величин определены следующим образом: если при k-м бросании выпала решетка, то $\mathbf{X}_k = 0$, если же выпал герб, то $\mathbf{X}_k = 1$ для $k \leqslant 1000$ и $\mathbf{X}_k = -1$ для $k > 1000$. Тогда $\mathbf{D} = \mathbf{X}_1 + \ldots + \mathbf{X}_{2000}$. Случайные величины $\mathbf{X}_k$ имеют среднее значение $\mu_k = \pm 1/2$ и дисперсию $\sigma_k^2 = 1/4$, поэтому $\mathbf{E}(\mathbf{D}) = 0$ и $\mathrm{Var}(\mathbf{D}) = 500$. Следовательно, вероятность того, что разность **D** будет иметь значение в пределах $\pm\sqrt{500}\alpha$, приблизительно равна $\mathfrak{N}(\alpha) - \mathfrak{N}(-\alpha)$, и случайная величина **D** сравнима с отклонением $\mathbf{S}_{2000} - 1000$ числа выпадений герба в 2000 бросаний от среднего числа выпадений 1000.

в) Приложения к *теории наследственности* иллюстрируют большое разнообразие выводов, основанных на центральной предельной теореме. В гл. V, 5 изучались свойства организма, которые по существу зависят только от одной пары генов (аллелей). Мы предполагали, что другие свойства (такие, например, как рост), зависят от совместного действия многих пар генов. Для простоты изложения предположим, что для каждой отдельной пары генов существуют три генотипа *AA*, *Aa* или *aa*. Пусть x_1, x_2 и x_3 — соответствующие эффекты этих генотипов. Генотип индивидуума является случайным событием, и вклад отдельной пары генов в конечный результат, рост, есть случайная величина **X**, принимающая три значения: x_1, x_2, x_3 с некоторыми вероятностями. Рост зависит от совместных вкладов многих таких случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, и так как каждый из этих вкладов мал, то в первом приближении можно считать, что рост равен *сумме* $\mathbf{X}_1 + \ldots + \mathbf{X}_n$. Конечно, не *все* $\mathbf{X}_k$ взаимно независимы. Но центральная предельная теорема справедлива также и для широкого класса зависимых случайных величин, и, кроме того, большинство случайных величин $\mathbf{X}_k$, вероятнее всего, можно считать независимыми. Эти рассуждения можно уточнить; они приведены здесь только для того, чтобы показать, как центральная предельная теорема объясняет тот факт, что эмпирические распределения многих биометрических показателей, таких, как рост, близки к нормальному распределению. Эта теория позволяет также предсказывать свойства, которые передаются по наследству, например оценивать средний рост детей, зная рост их родителей. Такие биометрические исследования были начаты Ф. Гальтоном и К. Пирсоном [1].

[1] Френсис Гальтон (1822—1911); Карл Пирсон (1857—1936).

§ 6*). ПРИЛОЖЕНИЯ К КОМБИНАТОРНОМУ АНАЛИЗУ

Приведем два примера приложений центральной предельной теоремы к задачам, которые не связаны непосредственно с теорией вероятностей. Оба примера относятся к $n!$ перестановкам элементов $a_1, \ldots, a_n$, причем каждой перестановке приписывается вероятность $1/n!$.

а) **Инверсии.** Говорят, что элемент a_k образует в данной перестановке r инверсий, если он стоит впереди r элементов с меньшими номерами (т. е. элементов, которые предшествуют a_k в натуральном порядке). Например, в перестановке $(a_3a_6a_1a_5a_2a_4)$ элементы a_1 и a_2 не образуют ни одной инверсии, элемент a_3 образует две, a_4 — ни одной, a_5 — две и a_6 — четыре инверсии. В перестановке $(a_6a_5a_4a_3a_2a_1)$ элемент a_k образует $k—1$ инверсий, а общее число инверсий равно 15. Число инверсий $\mathbf{X}_k$, образованных элементом a_k, является случайной величиной, а сумма $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ равна общему числу инверсий. Здесь случайные величины $\mathbf{X}_k$ принимают значения $0, 1, \ldots, k—1$, каждое с вероятностью $1/k$; поэтому

$$\mu_k = (k-1)/2,$$
$$\sigma_k^2 = [1 + 2^2 + \ldots + (k-1)^2]/k - [(k-1)/2]^2 = (k^2-1)/12. \tag{6.1}$$

Число инверсий, образованных элементом a_k, не зависит от относительного расположения элементов $a_1, a_2, \ldots, a_{k-1}$, и, следовательно, случайные величины $\mathbf{X}_k$ взаимно независимы. Из (6.1) получаем

$$m_n = [1 + 2 + \ldots + (n-1)]/2 = n(n-1)/4 \sim n^2/4 \tag{6.2}$$

и

$$s_n^2 = (1/12) \sum_{k=1}^{n} (k^2-1) = (2n^3 + 3n^2 - 5n)/72 \sim n^3/36. \tag{6.3}$$

При больших n имеем $\varepsilon s_n > n \geqslant \mathbf{U}_k$, и поэтому случайные величины $\mathbf{U}_k$, входящие в условие Линдеберга, совпадают с $\mathbf{X}_k$. Следовательно, применима центральная предельная теорема, и мы приходим к выводу, что *число* $\mathbf{N}_n$ *перестановок, для которых число инверсий лежит внутри интервала* $n^2/4 \pm (\alpha/6)\sqrt{n^3}$, *асимптотически равно* $n!\{\mathfrak{N}(\alpha) - \mathfrak{N}(-\alpha)\}$. В частности, примерно для половины всех перестановок число инверсий лежит внутри интервала $n^2/4 \pm 0{,}11\sqrt{n^3}$.

б) **Циклы.** Каждую перестановку можно разбить на циклы, т. е. группы элементов, переставляемых между собой. Так, в перестановке $(a_3a_6a_1a_5a_2a_4)$ элементы a_1 и a_3 меняются местами и оставшиеся четыре элемента переставляются между собой; эта перестановка содержит два цикла. Если элемент стоит на своем естественном месте, то он образует цикл, так что тождественная перестановка $(a_1,$

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

$a_2, \ldots, a_n)$ содержит столько же циклов, сколько и элементов. С другой стороны, циклические перестановки $(a_2, a_3, \ldots, a_n, a_1)$, $(a_3, a_4, \ldots, a_n, a_1, a_2)$ и т. д. содержат по одному циклу каждая. При изучении циклов удобно описывать перестановки, указывая стрелками, на какие места переходят элементы. Например, запись $1 \to 3 \to 4 \to 1$ означает, что элемент a_1 переходит на третье место, элемент a_3 — на четвертое место, a_4 — на первое, и третий шаг, таким образом, завершает цикл. Чтобы продолжить это описание, надо начинать со следующего в естественном порядке элемента, т. е. с a_2. В новых обозначениях перестановку $(a_4, a_8, a_1, a_3, a_2, a_5, a_7, a_6)$ можно записать в виде $1 \to 3 \to 4 \to 1$; $2 \to 5 \to 6 \to 8 \to 2$; $7 \to 7$. Иначе говоря, мы строим перестановку $(a_1, \ldots, a_n)$ при помощи n последовательных выборов. Сначала мы находим место i, занятое a_1, затем — место, занятое a_i, и т. д. На первом, втором, ..., n-м шаге имеется $n, n-1, \ldots, 1$ возможностей, и только одна из них завершает цикл.

Пусть случайная величина $\mathbf{X}_k$ равна 1, если при таком построении цикл завершается на k-м шаге, и равна 0 в противном случае. (В предыдущем примере $\mathbf{X}_3=\mathbf{X}_7=\mathbf{X}_8=1$ и $\mathbf{X}_1=\mathbf{X}_2=\mathbf{X}_4=\mathbf{X}_5=\mathbf{X}_6=0$.) Очевидно, что $\mathbf{X}_1=1$ тогда и только тогда, когда элемент a_1 стоит на первом месте. Из нашего построения следует, что $\mathbf{P}\{\mathbf{X}_k=1\}=1/(n-k+1)$, $\mathbf{P}\{\mathbf{X}_k=0\}=(n-k)/(n-k+1)$ и случайные величины $\mathbf{X}_k$ взаимно независимы [1]). Математические ожидания и дисперсии величин $\mathbf{X}_k$ соответственно равны

$$\mu_k=1/(n-k+1), \quad \sigma_k^2=(n-k)/(n-k+1)^2, \tag{6.4}$$

откуда следует, что

$$m_n=1+1/2+1/3+\ldots+1/n \sim \log n \tag{6.5}$$

и

$$s_n^2=\sum_{k=1}^{n}(n-k)/(n-k+1)^2 \sim \log n. \tag{6.6}$$

Сумма $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ равна общему числу циклов. *Среднее число циклов равно* m_n; *число перестановок, для которых число циклов заключено между* $\log n+\alpha\sqrt{\log n}$ *и* $\log n+\beta\sqrt{\log n}$, *приблизительно равно* $n!\{\mathfrak{N}(\beta)-\mathfrak{N}(\alpha)\}$. Более точные оценки даются усовершенствованными вариантами центральной предельной теоремы [2]).

[1]) Формально распределение случайной величины $\mathbf{X}_k$ зависит не только от k, но также и от n. Однако достаточно переупорядочить $\mathbf{X}_k$ в обратном порядке, чтобы получить случайные величины с распределением, зависящим только от индекса. (См. также пример гл. XI, 2, д).)

[2]) Много различных асимптотических оценок, относящихся к комбинаторному анализу, получено в работе Гончаров В. Л., Из области комбинаторного анализа, Изв. АН СССР, серия матем., 8 (1944), 3—48. Использованный здесь метод проще, но имеет меньшую область применимости; см. Feller W., The fundamental limit theorems in probability, Bull. Amer. Math. Soc., 51 (1945), 800—832.

§ 7*). УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

Из (слабого) закона больших чисел (5.3) следует, что для каждого достаточно большого n весьма правдоподобно, что отклонение $|\mathbf{S}_n - m_n|$ окажется малым по сравнению с n. В связи с испытаниями Бернулли уже было указано (гл. VIII), что из этого не следует, что $|\mathbf{S}_n - m_n|/n$ остается малым для всех больших n; может случиться, что закон больших чисел применим, но $|\mathbf{S}_n - m_n|/n$ продолжает колебаться между конечными или бесконечными границами. Закон больших чисел позволяет утверждать только то, что большие значения $|\mathbf{S}_n - m_n|/n$ появляются очень редко.

Будем говорить, что последовательность $\mathbf{X}_k$ *удовлетворяет усиленному закону больших чисел, если каждой паре* $\varepsilon > 0$, $\delta > 0$ *соответствует* N, *для которого с вероятностью* $1-\delta$ *или большей для любого* $r > 0$ *выполняются все* $r+1$ *неравенств*

$$|\mathbf{S}_n - m_n|/n < \varepsilon, \quad n = N,\ N+1,\ \ldots,\ N+r. \tag{7.1}$$

Грубо говоря, (7.1) означает, что с вероятностью, близкой к единице, $|\mathbf{S}_n - m_n|/n$ *остается* малым[1]) при всех $n > N$.

Критерий Колмогорова. *Сходимость ряда*

$$\sum \sigma_k^2/k^2 \tag{7.2}$$

является достаточным условием для того, чтобы усиленный закон больших чисел был применим к последовательности взаимно независимых случайных величин $\mathbf{X}_k$ *с дисперсией* σ_k^2.

Доказательство. Пусть событие A_ν состоит в том, что неравенство (7.1) *не* выполняется хотя бы при одном n из интервала $2^{\nu-1} < n \leqslant 2^\nu$. Очевидно достаточно доказать, что при достаточно большом ν и при всех r

$$\mathbf{P}\{A_\nu\} + \mathbf{P}\{A_{\nu+1}\} + \ldots + \mathbf{P}\{A_{\nu+r}\} < \delta,$$

т. е. доказать сходимость ряда $\sum \mathbf{P}\{A_\nu\}$. Но если событие A_ν произошло, то это означает, что при некотором n из интервала $2^{\nu-1} < n \leqslant 2^\nu$ выполняется неравенство

$$|\mathbf{S}_n - m_n| \geqslant \varepsilon \cdot 2^{\nu-1}, \tag{7.3}$$

и по неравенству Колмогорова (гл. IX, 7) имеем

$$\mathbf{P}\{A_\nu\} \leqslant 4\varepsilon^{-2} \cdot s_{2^\nu}^2 \cdot 2^{-2\nu}. \tag{7.4}$$

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

[1]) В общей теории вводится пространство элементарных событий, соответствующее бесконечной последовательности $\{\mathbf{X}_k\}$. Усиленный закон больших чисел утверждает тогда, что $|\mathbf{S}_n - m_n|/n$ стремится к нулю с вероятностью единица. В терминологии теории функций действительного переменного усиленный закон больших чисел означает сходимость почти всюду, а слабый закон больших чисел эквивалентен сходимости по мере.

Поэтому

$$\sum_{\nu=1}^{\infty} \mathbf{P}\{A_\nu\} \leqslant 4\varepsilon^{-2} \sum_{\nu=1}^{\infty} 2^{-2\nu} \sum_{k=1}^{2^\nu} \sigma_k^2 = 4\varepsilon^{-2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{2^\nu \geqslant k} 2^{-2\nu} \leqslant$$
$$\leqslant 8\varepsilon^{-2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2}, \qquad (7.5)$$

чем и завершается доказательство. ▶

Как пример типичного приложения критерия Колмогорова докажем следующую теорему.

Теорема. *Если взаимно независимые случайные величины* $\mathbf{X}_k$ *имеют одинаковые распределения вероятностей* $\{f(x_j)\}$ *с конечным математическим ожиданием* $\mu = \mathbf{E}(\mathbf{X}_k)$, *то последовательность* $\{\mathbf{X}_k\}$ *удовлетворяет усиленному закону больших чисел.*

Данная теорема, конечно, сильнее, чем слабый закон больших чисел из § 1. Эти две теоремы приводятся отдельно, так как их доказательства интересны с методологической точки зрения. Относительно обратных утверждений см. задачи 17 и 18.

Доказательство. Снова используем метод усечения. Определим две новые последовательности случайных величин, полагая

$$\begin{array}{lll} \mathbf{U}_k = \mathbf{X}_k, & \mathbf{V}_k = 0, & \text{если } |\mathbf{X}_k| < k, \\ \mathbf{U}_k = 0, & \mathbf{V}_k = \mathbf{X}_k, & \text{если } |\mathbf{X}_k| \geqslant k. \end{array} \qquad (7.6)$$

Случайные величины $\mathbf{U}_k$ взаимно независимы, и мы покажем, что они удовлетворяют критерию Колмогорова. Если $\sigma_k^2 = \mathrm{Var}(\mathbf{U}_k)$, то

$$\sigma_k^2 \leqslant \mathbf{E}(\mathbf{U}_k^2) = \sum_{|x_j| < k} x_j^2 f(x_j). \qquad (7.7)$$

Для краткости положим

$$a_\nu = \sum_{\nu-1 \leqslant |x_j| < \nu} |x_j| f(x_j). \qquad (7.8)$$

Так как математические ожидания $\mathbf{E}(\mathbf{X}_k)$ существуют, ряд $\sum a_\nu$ сходится. Кроме того, из (7.7) следует, что

$$\sigma_k^2 \leqslant a_1 + 2a_2 + 3a_3 + \ldots + ka_k \qquad (7.9)$$

и

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{\nu=1}^{k} \nu a_\nu = \sum_{\nu=1}^{\infty} \nu a_\nu \sum_{k=\nu}^{\infty} \frac{1}{k^2} < 2 \sum_{\nu=1}^{\infty} a_\nu < \infty. \qquad (7.10)$$

Таким образом, для последовательности $\{\mathbf{U}_k\}$ ряд (7.2) сходится. Поскольку

$$\mathbf{E}(\mathbf{U}_k) = \mu_k = \sum_{|x_j| < k} x_j f(x_j), \qquad (7.11)$$

то $\mu_k \to \mu$ и, следовательно, $(\mu_1 + \ldots + \mu_n)/n \to \mu$. Поэтому, применяя к последовательности $\{U_k\}$ усиленный закон больших чисел, убеждаемся, что неравенство

$$\left| n^{-1} \sum_{k=1}^{n} U_k - \mu \right| < \varepsilon \tag{7.12}$$

выполняется с вероятностью, не меньшей чем $1-\delta$ при *всех* $n > N$, если N выбрано достаточно большим. Остается доказать, что аналогичное утверждение будет справедливо, если U_k заменить на X_k. Для этого, очевидно, достаточно показать, что N можно выбрать столь большим, что событие $U_k = X_k$ для всех $k > N$ имеет вероятность сколь угодно близкую к единице. Это в свою очередь сводится к утверждению, состоящему в том, что с вероятностью единица только конечное число случайных величин V_k отлично от нуля. Согласно первой лемме Бореля — Кантелли из гл. VIII, 3, для этого нужно доказать сходимость ряда $\sum P\{V_k \neq 0\}$. Легко видеть, что

$$P\{V_n \neq 0\} = \sum_{|x_j| \geq n} f(x_j) \leq \frac{a_{n+1}}{n} + \frac{a_{n+2}}{n+1} + \frac{a_{n+3}}{n+2} + \ldots, \tag{7.13}$$

и, следовательно,

$$\sum P\{V_n \neq 0\} \leq \sum_{n=1}^{\infty} \sum_{\nu=n}^{\infty} \frac{a_{\nu+1}}{\nu} = \sum_{\nu=1}^{\infty} \frac{a_{\nu+1}}{\nu} \sum_{n=1}^{\nu} 1 = \sum_{\nu} a_{\nu+1} < \infty, \tag{7.14}$$

что и утверждалось.

§ 8. ЗАДАЧИ

1. Доказать, что закон больших чисел применим в примере 5, а) также и при $\lambda \leq 0$. Центральная предельная теорема выполняется, если $\lambda \geq -1/2$.
2. Установить, будут ли выполнены закон больших чисел и центральная предельная теорема для последовательности взаимно независимых случайных величин X_k с распределениями, заданными следующим образом ($k \geq 1$):
   а) $P\{X_k = \pm 2^k\} = 1/2$;
   б) $P\{X_k = \pm 2^k\} = 2^{-(2k+1)}$, $P\{X_k = 0\} = 1 - 2^{-2k}$;
   в) $P\{X_k = \pm k\} = 1/(2\sqrt{k})$, $P\{X_k = 0\} = 1 - 1/\sqrt{k}$.
3. *Условие Ляпунова* (1901). Показать, что выполняется условие Линдеберга, если при некотором $\delta > 0$

$$(1/s_n^{2+\delta}) \sum_{k=1}^{n} E(|X_k|^{2+\delta}) \to 0.$$

4. Пусть X_k — взаимно независимые случайные величины, причем X_k принимает $2k+1$ значений $0, \pm L_k, \pm 2L_k, \ldots, \pm kL_k$ с вероятностью $1/(2k+1)$ каждое. Найти условия для констант L_k, обеспечивающие выполнение закона больших чисел и/или центральной предельной теоремы для последовательности $\{X_k\}$.
5. Решить ту же задачу, если X_k принимает значения $a_k, -a_k$ и 0 с вероятностями $p_k, p_k, 1-2p_k$ соответственно.

Замечание. В следующих семи задачах рассматривается закон больших чисел для зависимых случайных величин.

6. В задаче 13 гл. V, 8, положим $X_k = 1$, если при k-м бросании выпала красная грань, и $X_k = 0$ — в противном случае. Показать, что закон больших чисел неприменим.

7. Пусть $\{X_k\}$ взаимно независимы и имеют одинаковые распределения с математическим ожиданием μ и конечной дисперсией. Доказать, что если $S_n = X_1 + \ldots + X_n$, то закон больших чисел не выполняется для последовательности $\{S_n\}$, но выполняется для последовательности $a_n S_n$, если $na_n \to 0$. *Указание.* Вычислить $\operatorname{Var}(S_1, \ldots, S_n)/n$.

8. Пусть $\{X_k\}$ — последовательность случайных величин, такая, что X_k может зависеть только от X_{k-1} и X_{k+1}, но не зависит от всех других X_j. Показать, что закон больших чисел выполняется, если X_k имеют конечные дисперсии.

9. Доказать, что если совместное распределение величин $(X_1, \ldots, X_n)$ определено при каждом n, причем дисперсии ограничены, а ковариации отрицательны, то применим закон больших чисел.

10. *Продолжение.* Заменим условие $\operatorname{Cov}(X_j, X_k) \leqslant 0$ предположением, что $\operatorname{Cov}(X_j, X_k) \to 0$ равномерно при $|j-k| \to \infty$. Доказать, что выполняется закон больших чисел.

11. Доказать, что если $|S_n| < cn$, а $\operatorname{Var}(S_n) > \alpha n^2$, то закон больших чисел не применим к $\{X_k\}$.

12. В урновой схеме Пойа (пример гл. V, 2, в)) положим, что X_k равно 1 или 0 в зависимости от того, был ли k-й извлеченный шар черным или красным. Тогда S_n — число черных шаров при n извлечениях. Доказать, что закон больших чисел к $\{X_k\}$ не применим. *Указание.* Использовать результат предыдущей задачи и задачи 30 гл. IX, 9.

---

13. Взаимно независимые случайные величины X_k принимают значения $r = 2, 3, 4, \ldots$ с вероятностями $p_r = c/(r^2 \log r)$, где постоянная c выбрана так, что $\sum p_r = 1$. Показать, что выполняется обобщенный закон больших чисел (4.1), если положить $e_n = c \cdot n \log \log n$.

14. Пусть $\{X_n\}$ — последовательность взаимно независимых случайных величин, таких, что $X_n = \pm 1$ с вероятностями $(1-2^{-n})/2$ и $X_n = \pm 2^n$ с вероятностями 2^{-n-1}. Доказать, что к $\{X_k\}$ применимы как слабый закон больших чисел, так и усиленный закон больших чисел. (*Замечание.* Это означает, что условие (5.5) не является необходимым.)

15. *Пример разорительной «безобидной» игры.* Пусть возможные значения выигрыша при каждом испытании будут 0, 2, 2^2, 2^3, ...; вероятность того, что выигрыш равен 2^k, равна

$$p_k = 1/[2^k k(k+1)], \tag{8.1}$$

а вероятность нулевого выигрыша равна $p_0 = 1 - (p_1 + p_2 + \ldots)$. Тогда математическое ожидание величины выигрыша равно

$$\mu = \sum 2^k p_k = (1 - 1/2) + (1/2 - 1/3) + (1/3 - 1/4) + \ldots = 1. \tag{8.2}$$

Предположим, что игрок при каждом испытании уплачивает за право участия в игре доллар, так что после n испытаний его чистый выигрыш или проигрыш равен $S_n - n$. Показать, что при каждом $\varepsilon > 0$ *вероятность того, что в n испытаниях игрок проиграет более чем $(1-\varepsilon)\, n/\mathrm{Log}\, n$ долларов, стремится к единице*, причем Log означает логарифм по основанию 2. Иначе говоря, надо доказать, что

$$P\{S_n - n < -(1-\varepsilon)\, n/\mathrm{Log}\, n\} \to 1. \tag{8.3}$$

Указание. Использовать метод усечения из § 4, но в (4.2) заменить границу $n\,\mathrm{Log}\, n$ границей $n/\mathrm{Log}\, n$. Показать, что вероятность того, что $U_k = X_k$

при всех $k \leqslant n$, стремится к единице, и доказать, что

$$P\{|U_1 + \ldots + U_n - nE(U_1)| < \varepsilon n/\mathrm{Log}\, n\} \to 1; \tag{8.4}$$

$$1 - 1/\mathrm{Log}\, n \geqslant E(U_1) \geqslant 1 - (1+\varepsilon)/\mathrm{Log}\, n. \tag{8.5}$$

Подробности см. в статье, указанной в примечании к § 3.

16. Пусть $\{X_n\}$ — последовательность взаимно независимых одинаково распределенных случайных величин. Предположим, что X_n не имеет конечного математического ожидания и что A — некоторая положительная постоянная. Доказать, что с вероятностью единица осуществится бесконечно много событий $|X_n| > An$.

17. *Обращение усиленного закона больших чисел.* Доказать, что в предположениях задачи 16 с вероятностью единица $|S_n| > An$ для бесконечно многих n.

18. *Обращение критерия Колмогорова.* Доказать, что если ряд $\sum \sigma_k^2/k^2$ расходится, то существует последовательность $\{X_k\}$ взаимно независимых случайных величин с $\mathrm{Var}\{X_k\} = \sigma_k^2$, к которой неприменим закон больших чисел. *Указание.* Сначала показать, что сходимость ряда $\sum P\{|X_n| > \varepsilon n\}$ является необходимым условием для выполнения усиленного закона больших чисел.

ГЛАВА XI ЦЕЛОЧИСЛЕННЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ. ПРОИЗВОДЯЩИЕ ФУНКЦИИ

§ 1. ОБЩИЕ ПОЛОЖЕНИЯ

Среди дискретных случайных величин особенно важны величины, принимающие только целые значения $k=0, 1, 2, \ldots$. Наиболее общим методом изучения таких случайных величин является метод производящих функций. Как мы убедимся в дальнейшем, этот метод — частный случай метода характеристических функций, который служит основой для решения многих задач теории вероятностей. С более общей точки зрения метод производящих функций относится к области операционных методов, широко применяемых в теории дифференциальных и интегральных уравнений. В теории вероятностей производящие функции применялись со времен Муавра и Лапласа, но в полной мере возможности этого метода использовались редко.

Определение. *Пусть a_0, a_1, a_2, ...— последовательность действительных чисел. Если ряд*

$$A(s)=a_0+a_1s+a_2s^2+\ldots \tag{1.1}$$

сходится в некотором интервале $-s_0<s<s_0$, то функция $A(s)$ называется производящей функцией последовательности $\{a_j\}$.

Переменная s сама по себе ничего не обозначает. Если последовательность $\{a_j\}$ ограничена, то сравнение с геометрической прогрессией показывает, что ряд (1.1) сходится хотя бы при $|s|<1$.

Примеры. Если $a_j=1$ для всех j, то $A(s)=1/(1-s)$. Производящей функцией последовательности $(0, 0, 1, 1, 1, \ldots)$ будет функция $s^2/(1-s)$. Последовательность $a_j=1/j!$ имеет производящую функцию e^s. При фиксированном n последовательность $a_j=\binom{n}{j}$ имеет производящую функцию $(1+s)^n$. Если $\mathbf{X}$ — число очков, выпавших при бросании правильной кости, то распределение вероятностей случайной величины $\mathbf{X}$ имеет производящую функцию $(s+s^2+s^3+s^4+s^5+s^6)/6$. ►

Пусть случайная величина $\mathbf{X}$ принимает значения $0, 1, 2, \ldots$. Удобно ввести специальные обозначения как для распределения величины $\mathbf{X}$, так и для «хвостов» распределения. Мы будем писать

$$\mathbf{P}\{\mathbf{X}=j\}=p_j,\quad \mathbf{P}\{\mathbf{X}>j\}=q_j; \tag{1.2}$$

тогда

$$q_k=p_{k+1}+p_{k+2}+\ldots,\quad k\geqslant 0. \tag{1.3}$$

Производящими функциями последовательностей $\{p_j\}$ и $\{q_k\}$ будут функции

$$P(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \ldots, \tag{1.4}$$

$$Q(s) = q_0 + q_1 s + q_2 s^2 + q_3 s^3 + \ldots . \tag{1.5}$$

Так как $P(1)=1$, *то ряд* (1.4) *сходится абсолютно хотя бы при* $-1 \leqslant s \leqslant 1$. Коэффициенты ряда (1.5) меньше единицы, поэтому *он сходится хотя бы при* $-1 < s < 1$.

Теорема 1. *При* $-1 < s < 1$ *имеем*

$$Q(s) = (1 - P(s))/(1 - s). \tag{1.6}$$

Доказательство. Коэффициент при s^n в разложении $(1-s)\cdot Q(s)$ равен $q_n - q_{n-1} = -p_n$, если $n \geqslant 1$, и равен $q_0 = p_1 + p_2 + \ldots = 1 - p_0$, если $n = 0$. Поэтому $(1-s)\cdot Q(s) = 1 - P(s)$, что и утверждалось. ►

Рассмотрим теперь производную

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}. \tag{1.7}$$

Здесь ряд сходится хотя бы при $-1 < s < 1$. При $s=1$ правая часть формально сводится к $\sum k p_k = \mathbf{E}(\mathbf{X})$. Если математическое ожидание существует, то производная $P'(s)$ будет непрерывной в замкнутом интервале $-1 \leqslant s \leqslant 1$. Если же ряд $\sum k p_k$ расходится, то $P'(s) \to \infty$ при $s \to 1$. В этом случае мы будем говорить, что $\mathbf{X}$ *имеет бесконечное математическое ожидание*, и писать $P'(1) = \mathbf{E}(\mathbf{X}) = \infty$. (Если все рассматриваемые величины положительны, то употребление символа ∞ не приводит к противоречиям). Применяя теорему о среднем к числителю правой части соотношения (1.6), получаем, что $Q(s) = P'(\sigma)$, где число σ заключено между s и 1. Так как обе функции монотонно возрастают, из этого следует, что $P'(s)$ и $Q(s)$ имеют одинаковые конечные или бесконечные пределы, которые мы будем обозначать через $P'(1)$ или $Q(1)$. Итак, доказана следующая теорема.

Теорема 2. *Математическое ожидание* $\mathbf{E}(\mathbf{X})$ *удовлетворяет соотношениям*

$$\mathbf{E}(\mathbf{X}) = \sum_{j=1}^{\infty} j p_j = \sum_{k=0}^{\infty} q_k, \tag{1.8}$$

или в терминах производящих функций

$$\mathbf{E}(\mathbf{X}) = P'(1) = Q(1). \tag{1.9}$$

Дифференцируя формулу (1.7) и используя соотношение $P'(s) = Q(s) - (1-s)Q'(s)$, тем же способом находим

$$\mathbf{E}(\mathbf{X}(\mathbf{X}-1)) = \sum k(k-1)p_k = P''(1) = 2Q'(1). \tag{1.10}$$

Для того чтобы получить дисперсию $\mathbf{X}$, к этому выражению надо прибавить $\mathbf{E}(\mathbf{X})-\mathbf{E}^2(\mathbf{X})$, что приводит нас к следующей теореме.

Теорема 3. *Справедливо равенство*

$$\operatorname{Var}(\mathbf{X}) = P''(1)+P'(1)-P'^2(1) = \\ = 2Q'(1)+Q(1)-Q^2(1). \tag{1.11}$$

В случае бесконечной дисперсии $P''(s)\to\infty$ при $s\to 1$.

Формулы (1.9) и (1.11) часто дают самый простой способ вычисления $\mathbf{E}(\mathbf{X})$ и $\operatorname{Var}(\mathbf{X})$.

§ 2. СВЕРТКИ

Если случайная величина $\mathbf{X}$ принимает только неотрицательные целые значения, то $s^{\mathbf{X}}$ является новой случайной величиной и производящую функцию распределения величины $\mathbf{X}$ можно компактно записать в виде $\mathbf{E}(s^{\mathbf{X}})$. Если случайные величины $\mathbf{X}$ и $\mathbf{Y}$ независимы, то независимыми будут и случайные величины $s^{\mathbf{X}}$ и $s^{\mathbf{Y}}$; поэтому

$$\mathbf{E}(s^{\mathbf{X}+\mathbf{Y}}) = \mathbf{E}(s^{\mathbf{X}})\,\mathbf{E}(s^{\mathbf{Y}}).$$

Мы приведем другое доказательство этого важного результата, так как оно позволит нам получить полезное обобщение.

Пусть $\mathbf{X}$ и $\mathbf{Y}$ — неотрицательные независимые целочисленные случайные величины с распределениями вероятностей $\mathbf{P}\{\mathbf{X}=j\}=a_j$ и $\mathbf{P}\{\mathbf{Y}=j\}=b_j$. Событие $(\mathbf{X}=j,\ \mathbf{Y}=k)$ имеет вероятность a_jb_k. Сумма $\mathbf{S}=\mathbf{X}+\mathbf{Y}$ есть новая случайная величина, и событие $\mathbf{S}=r$ есть объединение несовместных событий

$$(\mathbf{X}=0,\ \mathbf{Y}=r),\ (\mathbf{X}=1,\ \mathbf{Y}=r-1),\ \ldots,\ (\mathbf{X}=r,\ \mathbf{Y}=0).$$

Поэтому распределение $c_r=\mathbf{P}\{\mathbf{S}=r\}$ задается формулой

$$c_r = a_0b_r+a_1b_{r-1}+a_2b_{r-2}+\ldots+a_{r-1}b_1+a_rb_0. \tag{2.1}$$

Операция (2.1), сопоставляющая двум последовательностям $\{a_k\}$ и $\{b_k\}$ новую последовательность $\{c_k\}$, встречается так часто, что удобно ввести для нее специальное название и обозначение.

Определение. *Пусть $\{a_k\}$ и $\{b_k\}$ — две произвольные числовые последовательности (не обязательно распределения вероятностей). Новая последовательность $\{c_r\}$, определяемая формулой (2.1), называется сверткой* [1]) *последовательностей $\{a_k\}$ и $\{b_k\}$ и обозначается*

$$\{c_k\}=\{a_k\}*\{b_k\}. \tag{2.2}$$

Пример. а) Если $a_k=b_k=1$ для всех $k\geqslant 0$, то $c_k=k+1$. Если $a_k=k$, $b_k=1$, то $c_k=1+2+\ldots+k=k(k+1)/2$. Наконец, если $a_0=a_1=1/2$, $a_k=0$ для $k\geqslant 2$, то $c_k=(b_k+b_{k-1})/2$ и т. д. ►

[1]) Некоторые авторы предпочитают немецкий термин Faltung; французский термин — composition. [В оригинале «convolution» — основной термин, принятый в англоязычной литературе.— *Перев.*]

Пусть последовательности $\{a_k\}$ и $\{b_k\}$ имеют производящие функции $A(s)=\sum a_k s^k$ и $B(s)=\sum b_k s^k$ соответственно. Произведение $A(s)B(s)$ можно получить почленным перемножением степенных рядов для $A(s)$ и $B(s)$. Собирая члены с одинаковыми степенями s, убеждаемся, что коэффициент c_r при s^r в разложении $A(s)B(s)$ задается формулой (2.1). Таким образом, имеет место следующая теорема.

Теорема. *Если $\{a_k\}$ и $\{b_k\}$ — последовательности с производящими функциями $A(s)$ и $B(s)$, а $\{c_k\}$ — их свертка, то производящая функция $C(s)=\sum c_k s^k$ равна произведению*

$$C(s)=A(s)B(s). \tag{2.3}$$

Если $\mathbf{X}$ и $\mathbf{Y}$ — неотрицательные целочисленные независимые случайные величины с производящими функциями $A(s)$ и $B(s)$, то их сумма $\mathbf{X}+\mathbf{Y}$ имеет производящую функцию $A(s)B(s)$.

Пусть теперь $\{a_k\}$, $\{b_k\}$, $\{c_k\}$, $\{d_k\}$, ... — какие-либо последовательности. Мы можем образовать свертку $\{a_k\}*\{b_k\}$, затем свертку этой новой последовательности и последовательности $\{c_k\}$ и т. д. Производящая функция $\{a_k\}*\{b_k\}*\{c_k\}*\{d_k\}$ равна $A(s)B(s)\times C(s)D(s)$, и из этого следует, что порядок, в котором образуются свертки, безразличен. Например, $\{a_k\}*\{b_k\}*\{c_k\}=\{c_k\}*\{b_k\}*\{a_k\}$ и т. д. *Таким образом, свертка является ассоциативной и коммутативной операцией* (точно так же, как сложение случайных величин).

При изучении сумм независимых случайных величин $\mathbf{X}_n$ особый интерес представляет тот частный случай, когда эти величины имеют одинаковые распределения. *Если $\{a_j\}$ — общее распределение вероятностей величин $\mathbf{X}_n$, то распределение суммы $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ будем обозначать $\{a_j\}^{n*}$.* Таким образом,

$$\{a_j\}^{2*}=\{a_j\}*\{a_j\},\quad \{a_j\}^{3*}=\{a_j\}^{2*}*\{a_j\},\ \ldots; \tag{2.4}$$

и вообще

$$\{a_j\}^{n*}=\{a_j\}^{(n-1)*}*\{a_j\}. \tag{2.5}$$

Иначе говоря, *$\{a_j\}^{n*}$ есть последовательность чисел с производящей функцией $A^n(s)$*. В частности, $\{a_j\}^{1*}$ совпадает с $\{a_j\}$, а *$\{a_j\}^{0*}$ определяется как последовательность с производящей функцией $A^0(s)=1$, т. е. как последовательность* $(1, 0, 0, 0, \ldots)$.

Примеры. б) *Биномиальное распределение.* Производящая функция биномиального распределения $b(k;\,n,\,p)=\binom{n}{k}p^kq^{n-k}$ равна

$$\sum_{k=0}^{n}\binom{n}{k}(ps)^kq^{n-k}=(q+ps)^n. \tag{2.6}$$

Тот факт, что эта производящая функция есть n-я степень бинома $q+ps$, означает, что $\{b(k;\,n,\,p)\}$ есть распределение суммы $\mathbf{S}_n=$

$=\mathbf{X}_1+\ldots+\mathbf{X}_n$ n независимых случайных величин с одинаковыми производящими функциями $q+ps$; каждая величина принимает значение нуль с вероятностью q и значение единица с вероятностью p. Таким образом,

$$\{b(k;\ n,\ p)\}=\{b(k;\ 1,\ p\}^{n*}. \tag{2.7}$$

Представление $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ уже несколько раз использовалось нами (см. примеры гл. IX, 3, а) и 5, а)). Предыдущее рассуждение можно обратить и использовать для нового вывода биномиального распределения. Из мультипликативного свойства $(q+ps)^m(q+ps)^n=(q+ps)^{m+n}$ следует равенство

$$\{b(k;\ m,\ p)\}*\{b(k;\ n,\ p)\}=\{b(k;\ m+n,\ p)\}, \tag{2.8}$$

которое эквивалентно формуле (10.4) гл. VI. Дифференцируя производящую функцию $(q+ps)^n$, можно совсем просто получить доказательство того, что $\mathbf{E}(\mathbf{S}_n)=np$ и $\operatorname{Var}(\mathbf{S}_n)=npq$.

в) *Распределение Пуассона.* Производящая функция распределения $p(k;\ \lambda)=e^{-\lambda}\lambda^k/k!$ равна

$$\sum_{k=0}^{\infty} e^{-\lambda}(\lambda s)^k/k!=e^{-\lambda+\lambda s}. \tag{2.9}$$

Отсюда следует равенство

$$\{p(k;\ \lambda)\}*\{p(k;\ \mu)\}=\{p(k;\ \lambda+\mu)\}, \tag{2.10}$$

которое эквивалентно формуле (10.5) гл. VI. Дифференцированием убеждаемся, что и среднее значение, и дисперсия распределения Пуассона равны λ (см. пример гл. IX, 4, в)).

г) *Геометрическое и отрицательное биномиальное распределения.* Пусть случайная величина $\mathbf{X}$ имеет геометрическое распределение

$$\mathbf{P}\{\mathbf{X}=k\}=q^kp, \quad k=0,\ 1,\ 2,\ \ldots, \tag{2.11}$$

где p и q — положительные постоянные, причем $p+q=1$. Соответствующая производящая функция равна

$$p\sum_{k=0}^{\infty}(qs)^k=p/(1-qs). \tag{2.12}$$

Пользуясь результатами § 1, легко убедиться в том, что $\mathbf{E}(\mathbf{X})=q/p$ и $\operatorname{Var}(\mathbf{X})=q/p^2$ в соответствии с результатами, полученными в примере гл. IX, 3, в).

Вероятность того, что в последовательности испытаний Бернулли *первый успех* имеет место после k неудач (т. е. при $(k+1)$-м испытании), равна q^kp, так что $\mathbf{X}$ можно интерпретировать как *время ожидания первого успеха.* Строго говоря, такая интерпретация относится к бесконечному пространству элементарных событий, и преимущество формального определения (2.11) и терминологии случайных величин состоит в том, что мы можем не беспокоиться о структуре исходного пространства элементарных событий. То же самое верно и для *времени ожидания r-го успеха*. Если $\mathbf{X}_k$ — число неудач между $(k-1)$-м и k-м успехами, то $S_r = \mathbf{X}_1 + \ldots + \mathbf{X}_r$ — полное число неудач, предшествующих r-му успеху (а $S_r + r$ — число испытаний до r-го успеха включительно). Поскольку мы имеем дело с испытаниями Бернулли, то величины $\mathbf{X}_k$ взаимно независимы и имеют одинаковое распределение (2.11), так что можно принять это свойство за *определение* величин $\mathbf{X}_k$. Тогда S_r имеет производящую функцию

$$[p/(1-qs)]^r, \tag{2.13}$$

и из формулы разложения бинома (8.7) гл. II сразу получаем, что коэффициент при s^k равен

$$f(k;\ r,\ p) = \binom{-r}{k} p^r (-q)^k, \quad k = 0,\ 1,\ 2,\ \ldots. \tag{2.14}$$

Отсюда следует, что $\mathbf{P}\{S_r = k\} = f(k;\ r,\ p)$ в соответствии с формулой для числа неудач, предшествующих r-му успеху, которая была выведена в гл. VI, 8. Этот результат можно сформулировать иначе, сказав, что *распределение* $\{f(k;\ r,\ p)\}$ *есть r-кратная свертка геометрического распределения*, т. е.

$$\{f(k;\ r,\ p)\} = \{q^k p\}^{r*}. \tag{2.15}$$

До сих пор мы считали r целым, но, как это было доказано в гл. VI, 8, *отрицательное биномиальное распределение* $\{f(k;\ r,\ p)\}$ сохраняет смысл при любом неотрицательном r, не обязательно целом. Производящая функция по-прежнему задается формулой (2.13), и мы видим, что при $r>0$ *математическое ожидание и дисперсия отрицательного биномиального распределения равны* rq/p *и* rq/p^2 *и что*

$$\{f(k;\ r_1,\ p)\} * \{f(k;\ r_2,\ p)\} = \{f(k;\ r_1 + r_2,\ p)\}. \tag{2.16}$$

д) *Циклы*. В примере гл. X, 6, б) мы рассматривали число циклов S_n в случайной перестановке n элементов. Было показано, что эту случайную величину можно представить в виде суммы $S_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ n независимых величин, таких, что каждая случайная величина $\mathbf{X}_k$ может принимать два значения, 1 и 0, с вероятностями $(n-k+1)^{-1}$ и $(n-k)(n-k+1)^{-1}$ соответственно. Из этого сразу следует, что производящая функция случайной величины S_n равна произведению

$$\frac{n-1+s}{n} \cdot \frac{n-2+s}{n-1} \cdots \frac{1+s}{2} \cdot \frac{s}{1} = (-1)^n \binom{-s}{n}. \tag{2.17}$$

Коэффициенты этого многочлена определяют распределение вероятностей случайной величины S_n, однако для их явного представления нужно знать числа Стирлинга. Здесь перед нами пример весьма обычной ситуации, когда производящая функция оказывается проще самого распределения вероятностей. Поэтому весьма благоприятно то обстоятельство, что сама производящая функция может дать много полезной информации. ▶

§ 3. ВОЗВРАЩЕНИЕ В НАЧАЛО И ВРЕМЕНА ОЖИДАНИЙ В ИСПЫТАНИЯХ БЕРНУЛЛИ

В этом параграфе обсуждается несколько важных задач, представляющих методологический интерес, решение которых иллюстрирует силу и гибкость метода производящих функций. Результаты играют важную роль в теории случайных блужданий и могут рассматриваться как прототип аналогичных результатов в теории диффузий. Они будут получены и другими методами в гл. XIV (см., в частности, § 4 и 9). Для частного случая $p=1/2$ эти результаты были получены комбинаторными методами в гл. III. Сравнение этих методов проясняет суть дела [1]).

Рассмотрим схему испытаний Бернулли с вероятностью успеха p. Положим $\mathbf{X}_k=+1$, если k-е испытание привело к успеху, и $\mathbf{X}_k=-1$ в противном случае. Иначе говоря, объектом нашего изучения является последовательность взаимно независимых случайных величин, принимающих значения $+1$ и -1 с соответствующими вероятностями p и q. Это описание является самым простым и наиболее естественным, но приводит к несчетному пространству элементарных событий, поскольку предполагает бесконечные испытания. Фактически мы будем вычислять только некоторые вероятности событий, которые определяются конечным числом испытаний, и поэтому никаких принципиальных трудностей не возникает. Можно говорить о фиксированном числе испытаний N и полагать $N\to\infty$, но это было бы излишней педантичностью, вредной для вероятностной интуиции.

Как обычно, положим

$$\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n,\quad \mathbf{S}_0=0. \tag{3.1}$$

В переводе на образный «игровой» язык это означает, что Петр и Павел играют по единичной ставке, а $\mathbf{S}_n$ — чистый выигрыш Петра после n партий. В терминологии теории случайных блужданий $\mathbf{S}_n$ обозначает положение «частицы», которая передвигается на единицу вправо или влево через одинаковые интервалы времени. Случайное блуждание будет несимметричным, если $p\neq 1/2$.

а. Время ожидания для игры. Событие

$$\mathbf{S}_1\leqslant 0,\ \ldots,\ \mathbf{S}_{n-1}\leqslant 0,\quad \mathbf{S}_n=1 \tag{3.2}$$

в игровых терминах означает, что чистый выигрыш Петра впервые оказался положительным в n-й партии. В терминологии теории случайных блужданий *первое попадание в* $+1$ произошло при n-м испытании; более естественное описание использует терминологию теории физической диффузии и называет (3.2) *первым прохождением* через единицу [2]). Мы ищем вероятность φ_n этого события. Точнее,

---

[1]) Из этого объяснения должно быть ясно, что настоящий параграф приводится в целях иллюстрации, хотя представляет и самостоятельный интерес; в дальнейшем он не используется.

[2]) Или первым достижением единицы; см. примечание 1 на с. 108.—*Прим. перев.*

мы ищем производящую функцию

$$\Phi(s) = \sum_{n=0}^{\infty} \varphi_n s^n, \tag{3.3}$$

где для удобства принято $\varphi_0 = 0$ [1]). По определению $\varphi_1 = p$. Если (3.2) выполняется для некоторого $n > 1$, то $S_1 = -1$ и найдется *наименьший* индекс $\nu < n$, такой, что $S_\nu = 0$. Исходы первых n партий в игровой терминологии описываются следующим образом. (1) В первой партии Петр проигрывает условную единицу. (2) Петру потребуется сыграть еще $\nu - 1$ партий, чтобы восстановить исходную ситуацию. (3) Петру потребуется сыграть еще ровно $n - \nu$ партий, чтобы добиться положительного чистого выигрыша. Эти три события зависят от непересекающейся группы партий и поэтому являются взаимно независимыми. Из определения ясно, что события (2) и (3) имеют вероятности $\varphi_{\nu-1}$ и $\varphi_{n-\nu}$ соответственно и что вероятность одновременного осуществления всех трех событий равна произведению $q\varphi_{\nu-1}\varphi_{n-\nu}$. Итак, событие (3.2) осуществляется тогда и только тогда, когда происходят события (1)—(3) при некотором $\nu < n$. Суммируя по всем возможным ν, получаем

$$\varphi_n = q(\varphi_1\varphi_{n-2} + \varphi_2\varphi_{n-3} + \ldots + \varphi_{n-2}\varphi_1). \tag{3.4}$$

Следует помнить, что это соотношение справедливо только при $n > 1$ и что $\varphi_1 = p$, а $\varphi_0 = 0$. Умножая (3.4) на s^n и суммируя по $n = 2, 3, \ldots$, мы получаем в левой части $\Phi(s) - ps$. Величина, стоящая в скобках, представляет собой $(n-1)$-й член свертки $\{\varphi_n\} * \{\varphi_n\}$ и, следовательно, по теореме из § 2 правая часть равна $qs \cdot \Phi^2(s)$. Таким образом, мы видим, что *производящая функция Φ удовлетворяет квадратному уравнению*

$$\Phi(s) - ps = qs\Phi^2(s). \tag{3.5}$$

Один из двух корней этого уравнения неограничен в окрестности $s=0$, и единственное ограниченное решение имеет вид

$$\Phi(s) = \left(1 - \sqrt{1 - 4pqs^2}\right)/(2qs), \tag{3.6}$$

где $\sqrt{\ }$ означает положительный корень. Формула разложения бинома (8.7) гл. II позволяет записать коэффициенты в виде

$$\varphi_{2k-1} = \frac{(-1)^{k-1}}{2q}\binom{1/2}{k}(4pq)^k, \quad \varphi_{2k} = 0. \tag{3.7}$$

Такнм образом, мы получили явные выражения для вероятностей φ_k, но они имеют лишь второстепенное значение; более поучи-

---

[1]) Как будет показано в дальнейшем, производящую функцию Φ можно получить непосредственно из простых вероятностных рассуждений. Следующий ниже менее изящный вывод приводится только потому, что дает хороший пример того, как нужно обращаться с уравнениями в свертках, которые встречаются в различных контекстах вне теории вероятностей (см., например, задачу 6).

тельно извлечь нужную информацию непосредственно из производящей функции.

Заметим сначала, что сумма $\sum \varphi_n$ равна

$$\Phi(1) = (1 - |p - q|)/(2q), \tag{3.8}$$

и поэтому

$$\sum \varphi_n = \begin{cases} p/q, & \text{если } p < q, \\ 1, & \text{если } p \geqslant q. \end{cases} \tag{3.9}$$

Иначе говоря, *если $p<q$, то вероятность того, что сумма $\mathbf{S}_n$ будет всегда отрицательной, равна* $(q-p)/q$. Если $p \geqslant q$, то эта вероятность равна нулю, поэтому с вероятностью единица сумма $\mathbf{S}_n$ рано или поздно станет положительной. Сколько времени пройдет до этого момента? Простые вычисления показывают, что $\Phi'(1) = (p-q)^{-1}$, если $p>q$, и $\Phi'(1) = \infty$, если $p=q=1/2$. Можно сделать вывод, что *при $p=1/2$ число партий, предшествующих моменту, когда сумма $\mathbf{S}_n$ впервые будет положительной, имеет бесконечное математическое ожидание.*

Имеет смысл прокомментировать на игровом языке этот заслуживающий внимания результат. Из него следует, что в игре с бросанием симметричной монеты Петр теоретически уверен, что рано или поздно добьется положительного чистого выигрыша, но математическое ожидание числа испытаний, необходимых для достижения этой цели, равно бесконечности. Следовательно, игрок с ограниченным капиталом никогда не может быть уверен в том, что он когда-либо добьется чистого положительного выигрыша. Мы вернемся к этому вопросу в гл. XIV в связи с задачей о разорении игрока.

В вероятностных терминах вывод квадратного уравнения (3.5) для Φ можно описать компактнее. Обозначим через $\mathbf{N}$ первый индекс, при котором $\mathbf{S}_\mathbf{N} > 0$. Тогда $\mathbf{N}$ будет случайной величиной в несколько обобщенном смысле, а именно эта случайная величина не определена для события $S_n \leqslant 0$ при *всех n*. (В терминологии гл. XIII нам следовало бы назвать $\mathbf{N}$ *дефектной величиной*.) Производящую функцию Φ можно теперь записать в виде $\Phi(s) = \mathbf{E}(s^\mathbf{N})$. Если $\mathbf{X}_1 = -1$, то $\mathbf{N} = 1 + \mathbf{N}_1 + \mathbf{N}_2$, где $\mathbf{N}_1$ — число испытаний, необходимых для увеличения частных сумм $\mathbf{S}_k$ от -1 до 0, а $\mathbf{N}_2$ — число последующих испытаний, необходимых для увеличения $\mathbf{S}_k$ от 0 до 1. Эти величины независимы и имеют то же распределение, что и $\mathbf{N}$. Поэтому для условных математических ожиданий случайной величины $s^\mathbf{N}$ получаем

$$\mathbf{E}(s^\mathbf{N} \mid \mathbf{X}_1 = -1) = \mathbf{E}(s^{1+\mathbf{N}_1+\mathbf{N}_2} \mid \mathbf{X}_1 = -1) = s\Phi^2(s),$$
$$\mathbf{E}(s^\mathbf{N} \mid \mathbf{X}_1 = 1) = s.$$

Но

$$\mathbf{E}(s^\mathbf{N}) = p\mathbf{E}(s^\mathbf{N} \mid \mathbf{X}_1 = 1) + q\mathbf{E}(s^\mathbf{N} \mid \mathbf{X}_1 = -1), \tag{3.10}$$

что при $\Phi(s) = \mathbf{E}(s^\mathbf{N})$ совпадает с квадратным уравнением (3.5)

6. Возвращение в начало. Равенство сумм числа успехов и неудач имеет место при k-м испытании, если $\mathbf{S}_k = 0$. Заимствуя термин из теории диффузии, мы назовем это событие возвращением в состояние равновесия (т. е. в начало). Число испытаний будет случайным, и вероятность возвращения при $2n$-м испытании равна

$$u_{2n}=\binom{2n}{n}p^nq^n=(-1)^n\binom{-1/2}{n}(4pq)^n. \qquad (3.11)$$

Из формулы разложения бинома (8.7) гл. II получаем производящую функцию

$$U(s)=\sum_{n=0}^{\infty}u_{2n}s^{2n}=1/\sqrt{1-4pqs^2}. \qquad (3.12)$$

Отметим, что $\{u_n\}$ *не* является распределением вероятностей, так как возвращение в начало может происходить неоднократно.

в. Первое возвращение в начало происходит при $2n$-м испытании, если $S_{2n}=0$, но $S_k\neq 0$ при $k=1, \ldots, 2n-1$. Обозначим вероятность этого события через f_{2n}. (Очевидно, $f_{2n-1}=0$.) Рассмотрим отдельно два подсобытия с $X_1=1$ и $X_1=-1$ и обозначим их вероятности через f_{2n}^+ и f_{2n}^- соответственно. Из сказанного в п. «а» ясно, что $f_{2n}^-=q\varphi_{2n-1}$, так как первые $2n-2$ частных сумм $X_2+X_3+\ldots+X_k$ не превышают нуля, но следующая сумма положительна. Поэтому, используя (3.6), получаем

$$F^-(s)=\sum_{n=1}^{\infty}f_{2n}^-s^{2n}=qs\Phi(s)=(1-\sqrt{1-4pqs^2})/2. \qquad (3.13)$$

В силу симметричности производящую функцию последовательности $\{f_n^+\}$ можно получить взаимной перестановкой p и q. Из этого следует, что $F^+=F^-$, и поэтому окончательно [1])

$$F(s)=\sum_{n=1}^{\infty}f_ns^n=1-\sqrt{1-4pqs^2}. \qquad (3.14)$$

Отсюда можно получить интересные выводы, не используя явного вида коэффициентов f_n. Очевидно, что $F(1)$ означает вероятность того, что рано или поздно происходит возвращение в начало. Итак, $F(1)=1-|p-q|$, и поэтому $|p-q|$ *есть вероятность того, что возвращение в начало не происходит никогда*, т. е. $S_k\neq 0$ для всех $k>0$. Только в симметричном случае $p=1/2$ несомненно происходит возвращение в начало. В этом случае последовательность $\{f_n\}$ представляет распределение вероятностей для времени ожидания первого возвращения. Это время имеет *бесконечное математическое ожидание*.

В симметричном случае $p=1/2$ имеем

$$U(s)=(1-F(s))/(1-s^2). \qquad (3.15)$$

Так как U и F являются степенными рядами по s^2, то это соотношение отличается от (1.6) только обозначениями, и по теореме

[1]) Другой вывод можно найти в примере гл. XIII, 4, 6).

1 § 1

$$u_{2n} = f_{2n+2} + f_{2n+4} + \dots, \tag{3.16}$$

т. е. *при $p = 1/2$ вероятность того, что $S_{2n} = 0$, совпадает с вероятностью того, что $2n$ сумм $S_1, \dots, S_{2n}$ отличается от нуля.* Этот результат был получен различными методами в гл. III, 3 и играл основную роль в анализе парадоксальной природы флуктуаций при игре с бросанием монеты.

г. Первые достижения и последующие возвращения. Мы говорим, что *первое достижение точки $r > 0$ происходит при n-м испытании, если $S_n = r$, но $S_k < r$ для всех $k < n$.* Вероятность этого события обозначим через $\varphi_n^{(r)}$. Испытания, следующие за первым достижением точки $\nu > 0$, с вероятностной точки зрения являются копией всей последовательности, и поэтому число испытаний, следующих за первым достижением точки ν до первого достижения точки $\nu + 1$ включительно, имеет такое же распределение $\{\varphi_n\}$, как число испытаний до первого достижения точки 1. Если $p < q$, то сумма φ_n не равна единице, но все же имеет смысл говорить о том, что время ожидания первого достижения является случайной величиной с (возможно дефектным) распределением $\{\varphi_n\}$. Времена ожиданий между последовательными первыми достижениями взаимно независимы, и поэтому полное время ожидания для первого достижения точки r является суммой r независимых случайных величин с общим распределением $\{\varphi_n\}$. *Следовательно, производящая функция последовательности вероятностей первого достижения $\varphi_n^{(r)}$ равна r-й степени Φ.* (Начинающим следует проверить это утверждение непосредственно, получив уравнение свертки для $\varphi_n^{(2)}$, аналогичное (2.4), а затем воспользовавшись индукцией.)

Аналогичные рассуждения справедливы для вероятности $f_n^{(r)}$ того, что *r-е возвращение в начало происходит при n-м испытании. Производящая функция последовательности $\{f_n^{(r)}\}$ равна r-й степени F.* Сравнивая (3.6) и (3.14), сразу же замечаем, что

$$f_n^{(r)} = (2q)^r \varphi_{n-r}^{(r)}. \tag{3.17}$$

В частном случае $p = q = 1/2$ этот результат содержится в теореме 4 гл. III,7.

При помощи производящих функций легко получить различные приближения и предельные теоремы, но при этом используются преобразования Лапласа, которые будут рассматриваться только в гл. XIII тома 2. Систематического способа получения явных выражений для $f_n^{(r)}$ из производящей функции F^r не существует, но правильное предположение легко проверить, исходя из вида производящей функции. Исходя из теоремы 4 гл. III, 7, можно предположить, что

$$f_{n2}^{(r)} = \frac{r}{2n - r} \binom{2n-r}{n} 2^r (pq)^n. \tag{3.18}$$

Для проверки этого предположения достаточно заметить, что из тождества

$$F^r(s) = 2F^{r-1}(s) - 4pqs^2F^{r-2}(s)$$

следует рекуррентное соотношение

$$f_{2n}^{(r)} = 2f_{2n}^{(r-1)} - 4pqf_{2n-2}^{(r-2)}, \tag{3.19}$$

которому также удовлетворяет правая часть (3.18). Таким образом, по индукции убеждаемся в справедливости (3.18). По поводу эквивалентных выражений, имеющих, однако, другой вид, см. задачу 13 гл. XIV, 9.

§ 4. РАЗЛОЖЕНИЕ НА ПРОСТЫЕ ДРОБИ

Если известна производящая функция $P(s) = \sum p_k s^k$, то коэффициенты p_k можно найти дифференцированием по очевидной формуле $p_k = P^{(k)}(0)/k!$. На практике не всегда возможно получить явные выражения для p_k; кроме того, эти выражения порой так сложны, что разумное приближение оказывается предпочтительнее. Наиболее распространенный метод получения таких приближений основан на разложении на простые дроби. Из теории функции комплексного переменного известно, что такие разложения возможны для широкого класса функций, но мы ограничимся рассмотрением простого случая *рациональных функций*.

Предположим, что производящая функция имеет вид

$$P(s) = U(s)/V(s), \tag{4.1}$$

где U и V — многочлены, не имеющие общих корней. Допустим для простоты, что степень U меньше степени V, причем степень V равна m. Кроме того, предположим, что уравнение $V(s)=0$ имеет m различных (действительных или мнимых) корней $s_1, \ldots, s_m$. Тогда

$$V(s) = (s-s_1)(s-s_2)\ldots(s-s_m), \tag{4.2}$$

и из алгебры известно, что $P(s)$ можно разложить на *простые дроби*

$$P(s) = \rho_1/(s_1-s) + \rho_2/(s_2-s) + \ldots + \rho_m/(s_m-s), \tag{4.3}$$

где $\rho_1, \rho_2, \ldots, \rho_m$ — постоянные. Чтобы найти ρ_1, умножим (4.3) на s_1-s; при $s \to s_1$ произведение $(s_1-s)P(s)$ стремится к ρ_1. С другой стороны, из (4.1) и (4.2) получаем

$$(s_1-s)P(s) = -U(s)/[(s-s_2)(s-s_3)\ldots(s-s_m)]. \tag{4.4}$$

При $s \to s_1$ числитель в (4.4) стремится к $-U(s_1)$, а знаменатель — к произведению $(s_1-s_2)(s_1-s_3)\ldots(s_1-s_m)$, равному $V'(s_1)$. Таким образом, $\rho_1 = -U(s_1)/V'(s_1)$. Это рассуждение применимо ко всем корням, поэтому при $k \leqslant m$

$$\rho_k = -U(s_k)/V'(s_k). \tag{4.5}$$

Для заданного ρ_k мы можем легко получить точное выражение для коэффициента при s^n в $P(s)$. Напишем

$$\frac{1}{s_k - s} = \frac{1}{s_k} \cdot \frac{1}{1 - s/s_k}. \tag{4.6}$$

При $|s|<|s_k|$ последнюю дробь можно разложить в геометрический ряд

$$\frac{1}{1-s/s_k}=1+\frac{s}{s_k}+\left(\frac{s}{s_k}\right)^2+\left(\frac{s}{s_k}\right)^3+\ldots. \tag{4.7}$$

Подставляя эти выражения в (4.3), получим *коэффициент p_n при s^n* в виде

$$p_n=\rho_1/s_1^{n+1}+\rho_2/s_2^{n+1}+\ldots+\rho_m/s_m^{n+1}. \tag{4.8}$$

Таким образом, для получения p_n нужно сначала найти корни знаменателя $s_1, s_2, \ldots, s_m$, а затем по (4.5) определить коэффициенты $\rho_1, \ldots, \rho_m$.

Формула (4.8) дает *точное* выражение для вероятности p_n. Вычислить все m корней обычно бывает затруднительно, и поэтому формула (4.8) представляет преимущественно теоретический интерес. К счастью, даже один член в (4.8) почти всегда обеспечивает удовлетворительное приближение. Действительно, пусть s_1 — *наименьший* по абсолютной величине корень. Тогда первый знаменатель в (4.8) является наименьшим. Ясно, что при возрастании n относительная доля остальных членов убывает и преобладает первое слагаемое. Иначе говоря, *если s_1 — наименьший по абсолютной величине корень уравнения $V(s)=0$, то при $n\to\infty$*

$$p_n\sim\rho_1/s_1^{n+1} \tag{4.9}$$

(знак $\sim$ означает, что отношение правой и левой частей стремится к единице). Обычно эта формула дает хорошее приближение даже при сравнительно малых значениях n. Главное преимущество формулы (4.9) состоит в том, что в ней требуется вычислить только один корень алгебраического уравнения.

Легко освободиться от тех ограничений, при которых мы вывели асимптотическую формулу (4.9). Во-первых степень числителя в (4.1) может превосходить степень m знаменателя. Пусть $U(s)$ имеет степень $m+r$ ($r\geqslant 0$); с помощью деления $P(s)$ приводится к сумме многочлена степени r и дроби $U_1(s)/V(s)$, в которой степень многочлена $U_1(s)$ меньше m. Этот многочлен влияет только на первые $r+1$ членов распределения $\{p_n\}$, а $U_1\{s\}/V(s)$ можно, как и выше, разложить на простые дроби. Итак, формула (4.9) остается верной. Во-вторых, ограничение, что $V(s)$ имеет только простые корни, также не является необходимым. Из алгебры известно, что любая рациональная функция допускает разложение на простые дроби. Если s_k — двойной корень $V(s)$, то разложение на простые дроби (4.3) будет содержать дополнительный член вида $a/(s-s_k)^2$, и это внесет член вида $a(n+1)s_k^{-(n+2)}$ в точное выражение (4.8) для p_n. Но если s_1 — простой корень, то это не повлияет на асимптотическое разложение (4.9). Для удобства дальнейших ссылок сформулируем этот результат в виде теоремы.

Теорема. *Если наименьший по абсолютной величине корень s_1 знаменателя рациональной функции $P(s)$ является простым корнем, то коэффициент p_n при s^n асимптотически равен $\rho_1 s_1^{-(n+1)}$ где ρ_1 определяется по формуле* (4.5).

Аналогичное асимптотическое представление существует также и в случае, когда s_1 является кратным корнем (см. задачу 25).

Примеры[1]. а) Пусть a_n — вероятность того, что *число успехов в n испытаниях Бернулли четно*. Это событие может произойти в двух случаях: если за неудачей в первом испытании следует четное число успехов или же за успехом в первом испытании следует нечетное число успехов в остальных испытаниях. Поэтому для $n \geqslant 1$

$$a_n = qa_{n-1} + p(1 - a_{n-1}), \quad a_0 = 1. \tag{4.10}$$

Умножая на s^n и суммируя полученные соотношения по $n = 1$, $2, \ldots$, получаем следующее уравнение для производящей функции:

$$A(s) - 1 = qsA(s) + ps(1-s)^{-1} - psA(s),$$

или

$$2A(s) = [1-s]^{-1} + [1-(q-p)s]^{-1}.$$

Разлагая правую часть последнего соотношения в геометрическую прогрессию, мы, наконец, получаем a_n в явном виде

$$2a_n = 1 + (q-p)^n, \tag{4.11}$$

который с любой точки зрения предпочтительней очевидного ответа

$$a_n = b(0;\ n,\ p) + b(2;\ n,\ p) + \ldots.$$

б) Пусть q_n — вероятность того, что при n бросаниях правильной монеты не появится ни одна серия из трех последовательных выпадений герба. (Заметим, что $\{q_n\}$ не является распределением вероятностей; если p_n — вероятность того, что первая серия из трех последовательных гербов окончится на n-м испытании, то $\{p_n\}$ есть распределение вероятностей, а q_n представляет его «хвосты»: $q_n = p_{n+1} + p_{n+2} + \ldots$.)

Легко показать, что q_n удовлетворяет рекуррентной формуле

$$q_n = (1/2)\, q_{n-1} + (1/4)\, q_{n-2} + (1/8)\, q_{n-3}, \quad n \geqslant 3. \tag{4.12}$$

Действительно, событие, состоящее в том, что при n испытаниях ни разу не появится последовательность $ГГГ$, может произойти только в том случае, если испытания начинаются с $Р$, $ГР$, или $ГГР$. Вероятности того, что последующие испытания не приведут

---

[1]) Хорошую иллюстрацию применения простых дробей для численных приближений дает теория серий успехов (гл. XIII, 7). Явные выражения для вероятности разорения в гл. XIV, 5 и вероятностей перехода в гл. XVI, 1 также получены методом разложения на простые дроби.

к серии $ГГ$, равны соответственно q_{n-1}, q_{n-2} и q_{n-3}, и поэтому правая часть (4.12) содержит вероятности трех взаимно исключающих друг друга возможностей, при которых может произойти событие «ни одной серии $ГГ$».

Очевидно, $q_0=q_1=q_2=1$, и, следовательно, остальные q_n можно последовательно вычислять по формуле (4.12). Чтобы получить производящую функцию $Q(s)=\sum q_n s^n$, умножим обе части равенства (4.12) на s^n и просуммируем по $n\geqslant 3$. Получим

$$Q(s)-1-s-s^2=(1/2)\,s\{Q(s)-1-s\}+(1/4)\,s^2\{Q(s)-1\}+\\+(1/8)\,s^3\,Q(s),$$

или

$$Q(s)=\frac{2s^2+4s+8}{8-4s-2s^2-s^3}. \tag{4.13}$$

Знаменатель имеет действительный корень $s_1=1{,}0873778\ldots$, и два комплексных корня. При $|s|<s_1$ имеем $|4s+2s^2+s^3|<4s_1+ +2s_1^2+s_1^3=8$, и это же неравенство выполняется при $|s|=s_1$, за исключением случая $s=s_1$. Следовательно, другие два корня превосходят s_1 по абсолютной величине. Таким образом, из (4.9) получим

$$q_n\sim 1{,}236840/(1{,}0873778)^{n+1}, \tag{4.14}$$

где числитель равен $(2s_1^2+4s_1+8)/(4+4s_1+3s_1^2)$. Эта формула дает удивительно хорошее приближение даже при небольших значениях n. Вместо точного значения $q_3=0{,}875$ она дает 0,8847, а вместо $q_4=0{,}8125$ дает 0,81360. Относительная ошибка монотонно убывает, и для $q_{12}=0{,}41626\ldots$ точное и приближенное значения совпадают до пятого знака. ►

§ 5. ДВОЙНЫЕ ПРОИЗВОДЯЩИЕ ФУНКЦИИ

Для пары целочисленных случайных величин $\mathbf{X}$, $\mathbf{Y}$ с совместным распределением вида

$$\mathbf{P}\{\mathbf{X}=j,\ \mathbf{Y}=k\}=p_{jk},\quad j,\ k=0,\ 1,\ \ldots \tag{5.1}$$

мы определим производящую функцию от двух переменных формулой

$$P(s_1,\ s_2)=\sum_{j,\,k} p_{jk}s_1^js_2^k. \tag{5.2}$$

Такую производящую функцию для краткости будем называть двойной.

Методы доказательств, которыми мы пользовались в первых двух параграфах, можно применить здесь без существенных изменений, поэтому достаточно отметить три свойства, вытекающие из (5.2).

а) *Производящие функции маргинальных распределений* $\mathbf{P}\{\mathbf{X}=j\}$ *и* $\mathbf{P}\{\mathbf{Y}=k\}$ *равны* $A(s)=p(s, 1)$ *и* $B(s)=P(1, s)$.

б) *Производящая функция суммы* $\mathbf{X}+\mathbf{Y}$ *равна* $P(s, s)$.

в) *Случайные величины* $\mathbf{X}$ *и* $\mathbf{Y}$ *независимы тогда и только тогда, когда* $P(s_1, s_2)=A(s_1)B(s_2)$ *для всех* s_1, s_2.

Примеры. а) *Двумерное распределение Пуассона.* Очевидно, что функция

$$P(s_1, s_2)=e^{-a_1-a_2-b+a_1s_1+a_2s_2+bs_1s_2}, \quad a_i>0,\ b>0 \tag{5.3}$$

разлагается в степенной ряд с положительными коэффициентами, дающими в сумме единицу. Поэтому $P(s_1, s_2)$ представляет собой производящую функцию распределения некоторой пары случайных величин, каждая из которых имеет распределение Пуассона со средним значением, равным соответственно a_1+b и a_2+b. Но сумма $\mathbf{X}+\mathbf{Y}$ имеет производящую функцию $e^{-a_1-a_2-b+(a_1+a_2)s+bs^2}$ и поэтому является случайной величиной, имеющей распределение, отличающееся от распределения Пуассона. (Это так называемое обобщенное распределение Пуассона; см. гл. XII, 2.)

в) *Полиномиальное распределение.* Рассмотрим последовательность n независимых испытаний, каждое из которых приводит к одному из исходов E_0, E_1 или E_2, вероятности которых соответственно равны p_0, p_1 и p_2. Если $\mathbf{X}_i$ — случайная величина, равная числу испытаний, в которых произошло событие E_i, то $(\mathbf{X}_1, \mathbf{X}_2)$ имеет триномиальное распределение с производящей функцией $(p_0+p_1s_1+p_2s_2)^n$. ►

§ 6*). ТЕОРЕМА НЕПРЕРЫВНОСТИ

В гл. VI было показано, что распределение Пуассона $\{e^{-\lambda}\lambda^k/k!\}$ есть предельная форма биномиального распределения для случая, когда вероятность успеха p зависит от n, причем $np\to\lambda$ при $n\to\infty$. В этом случае

$$b(k;\ n,\ p)\to e^{-\lambda}\lambda^k/k!.$$

Производящая функция распределения $\{b(k;\ n,\ p)\}$ равна $(q+ps)^n=\{1-\lambda(1-s)/n\}^n$. Логарифмируя, убеждаемся, что эта производящая функция стремится к функции $e^{-\lambda(1-s)}$, являющейся производящей функцией распределения Пуассона. Покажем, что это свойство производящей функции сохраняется и в общем случае: последовательность распределений вероятностей сходится к предельному распределению тогда и только тогда, когда соответствующие производящие функции сходятся. К сожалению, эта теорема имеет ограниченное применение, так как наиболее интересными предельными

*) Теорема непрерывности будет использоваться только при обсуждении общего вида безгранично делимых распределений в гл. XII, 2 и получении общего числа потомков в ветвящихся процессах в гл. XII, 5.

формами дискретных распределений являются непрерывные распределения (например, нормальное распределение как предельная форма биномиального распределения).

Теорема непрерывности. *Предположим, что при любом фиксированном n последовательность $a_{0,n}$, $a_{1,n}$, $a_{2,n}$, ... является распределением вероятностей, т. е.*

$$a_{k,n} \geqslant 0, \quad \sum_{k=0}^{\infty} a_{k,n} = 1. \tag{6.1}$$

Для того чтобы при любом фиксированном $k \geqslant 0$ существовал предел

$$a_k = \lim_{n \to \infty} a_{k,n}, \tag{6.2}$$

необходимо и достаточно, чтобы при любом s из открытого интервала $0 < s < 1$ существовал предел

$$A(s) = \lim_{n \to \infty} \sum_{k=0}^{\infty} a_{k,n} s^k. \tag{6.3}$$

В этом случае автоматически получается

$$A(s) = \sum_{k=0}^{\infty} a_k s^k. \tag{6.4}$$

Очевидно, что $a_k \geqslant 0$ и $\sum a_k \leqslant 1$. Отметим тем не менее, что сумма может быть строго меньше 1. Например, если $a_{k,n} = f_{k+n}$, то $a_k = 0$ для всех k.

Доказательство[1]). Пусть $A_n(s)$ — ряд, стоящий в правой части (6.3).

(i) Предположим, что (6.2) выполнено, и определим $A(s)$ посредством (6.4). Так как $|a_{k,n} - a_k| \leqslant 1$, то для $0 < s < 1$

$$|A_n(s) - A(s)| \leqslant \sum_{k=0}^{r} |a_{k,n} - a_k| + s^r/(1-s). \tag{6.5}$$

Если выбрать r настолько большим, что $s^r < \varepsilon(1-s)$, то правая часть будет меньше 2ε для всех достаточно больших n. Таким образом, левую часть можно сделать сколь угодно малой, и, следовательно, (6.3) справедливо.

(ii) Предположим, что выполнено (6.3). Очевидно что $A(s)$ является монотонной функцией s, и поэтому $A(0)$ определяется

[1]) Эта теорема представляет собой частный случай теоремы непрерывности для преобразований Лапласа — Стилтьеса и доказывается так же, как и эта более общая теорема непрерывности. В литературе теорема непрерывности для производящих функций часто формулируется и доказывается при излишних ограничениях.

как предел $A(s)$ при $s\to 0$. Теперь

$$a_{0,\,n}\leqslant A_n(s)\leqslant a_{0,\,n}+s/(1-s). \tag{6.6}$$

Из этого следует, что при $n\to\infty$ все предельные значения $a_{0,\,n}$ лежат между $A(0)$ и $A(s)-s/(1-s)$. Устремляя s к 0, видим, что $a_{0,\,n}\to A(0)$, и поэтому (6.2) выполняется при $k=0$.

Это рассуждение можно применить ко всем k. Действительно, для $0<s<1$

$$(A_n(s)-a_{0,\,n})/s\to(A(s)-A(0))/s. \tag{6.7}$$

В левой части мы имеем степенной ряд с неотрицательными коэффициентами, и (6.7) во всех отношениях аналогично (6.3). При помощи рассуждений, подобных проведенным выше, убеждаемся в том, что производная $A'(0)$ существует и что $a_{1,n}\to A'(0)$. По индукции получаем, что (6.2) справедливо для всех k. ►

Примеры. а) *Отрицательное биномиальное распределение.* В примере 2, г), мы видели, что производящая функция распределения $\{f(k;\ r,\ p)\}$ имеет вид $p^r(1-qs)^{-r}$. Пусть теперь λ фиксировано, и пусть $p\to 1$, $q\to 0$ и $r\to\infty$, причем $q\sim\lambda/r$. Тогда

$$[p/(1-qs)]^r=[(1-\lambda/r)/(1-\lambda s/r)]^r. \tag{6.8}$$

Логарифмируя, убеждаемся, что правая часть этого равенства стремится к функции $e^{-\lambda+\lambda s}$, являющейся производящей функцией распределения Пуассона $\{e^{-\lambda}\lambda^k/k!\}$. Поэтому, *если $r\to\infty$ и $rq\to\lambda$, то*

$$f(k;\ r,\ p)\to e^{-\lambda}\lambda^k/k!. \tag{6.9}$$

б) *Испытания Бернулли с переменными вероятностями* [1]). Рассмотрим n независимых испытаний, таких, что k-е испытание оканчивается успехом с вероятностью p_k и неудачей с вероятностью $q_k=1-p_k$. Число успехов $\mathbf{S}_n$ можно записать в виде суммы $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ n взаимно независимых случайных величин $\mathbf{X}_k$, имеющих распределение

$$\mathbf{P}\{\mathbf{X}_k=0\}=q_k,\quad \mathbf{P}\{\mathbf{X}_k=1\}=p_k.$$

Производящая функция случайной величины $\mathbf{X}_k$ равна q_k+p_ks, и, следовательно, производящая функция величины $\mathbf{S}_n$ равна

$$P(s)=(q_1+p_1s)(q_2+p_2s)\ldots(q_n+p_ns). \tag{6.10}$$

В качестве приложения этой схемы допустим, что в каждом доме некоторого города в течение дня может возникнуть пожар с небольшой вероятностью p_k. Сумма $p_1+\ldots+p_n$, где n — число домов в городе, будет ожидаемым числом пожаров. В гл. VI мы видели, что если все p_k равны между собой и случаи возникновения пожара в различных домах стохастически независимы, то число

[1]) См. также примеры гл. IX, 1, д) и гл. IX, 5, б).

пожаров является случайной величиной, имеющей распределение, близкое к распределению Пуассона. Сейчас мы покажем, что это заключение остается в силе также и при более реальном предположении, что вероятности p_k не равны между собой. Этот результат еще больше убеждает нас в том, что распределение Пуассона дает адекватное описание явлений, представляющих собой сочетание нескольких маловероятных событий («успехов»). Примерами служат несчастные случаи и телефонные вызовы.

Воспользуемся уже знакомой нам схемой, в которой число величин n возрастает, причем вероятности p_k меняются вместе с n таким образом, что наибольшая из вероятностей p_k стремится к нулю, а сумма $p_1+p_2+\ldots+p_n=\lambda$ остается постоянной. Тогда, согласно (6.10),

$$\log P(s) = \sum_{k=1}^{n} \log\{1-p_k(1-s)\}. \tag{6.11}$$

Так как $p_k \to 0$, можно воспользоваться тем, что $\log(1-x) = -x-\theta x$, где $\theta \to 0$ при $x \to 0$. Отсюда следует, что

$$\log P(s) = -(1-s)\left\{\sum_{k=1}^{n}(p_k+\theta_k p_k)\right\} \to -\lambda(1-s), \tag{6.12}$$

так что $P(s)$ стремится к производящей функции распределения Пуассона. Следовательно, $\mathbf{S}_n$ *имеет в пределе распределение Пуассона*. Мы заключаем, что при большом n и не очень большом $\lambda = p_1+\ldots+p_n$ распределение величины $\mathbf{S}_n$ может быть приближенно представлено распределением Пуассона. ▶

§ 7. ЗАДАЧИ

1. Пусть $\mathbf{X}$ — случайная величина с производящей функцией $P(s)$. Найти производящие функции случайных величин $\mathbf{X}+1$ и $2\mathbf{X}$.

2. Найти производящие функции следующих последовательностей: а) $\mathbf{P}\{\mathbf{X} \leqslant n\}$; б) $\mathbf{P}\{\mathbf{X} < n\}$; в) $\mathbf{P}\{\mathbf{X} \geqslant n\}$; г) $\mathbf{P}\{\mathbf{X} > n+1\}$; д) $\mathbf{P}\{\mathbf{X} = 2n\}$.

3. Рассмотрим последовательности испытаний Бернулли. Обозначим через u_n вероятность того, что комбинация $УН$ впервые появится при $(n-1)$-м и n-м испытаниях. Найти производящую функцию, среднее значение и дисперсию.

4. Выяснить, какие из формул гл. II, 12 включают свертки и где можно было использовать производящие функции.

5. Пусть a_n — число способов, которыми можно получить сумму в n очков при произвольном числе бросаний правильной кости. Показать, что производящая функция последовательности $\{a_n\}$ равна $\{1-s-s^2-s^3-s^4-s^5-s^6\}^{-1}-1$.

6. Пусть a_n — число способов, которыми $n-2$ (непересекающихся) диагоналей разбивают выпуклый $(n+1)$-угольник $P_0P_1\ldots P_n$ на треугольники[1]. Положим $a_1=1$. Показать, что при $n \geqslant 2$

$$a_n = a_1a_{n-1}+a_2a_{n-2}+\ldots+a_{n-1}a_1.$$

Найти производящую функцию и явные выражения для a_n.

---

[1]) Эта задача обсуждается в книге Polya G. Mathematics and plausible reasoning, Princeton University Press, 1954, p. 102. [Имеется перевод: Пойа Д., Математика и правдоподобные рассуждения.— М.: ИЛ, 1957; М.: Наука, 1975, с. 123.]

Указание. Предположить, что одна из диагоналей проходит через P_0, и принять за k наименьший индекс, такой, что P_0P_k является диагональю.

Замечание. Задачи 7—11 относятся к § 3. Производящие функции Φ, U и F соответствуют первым достижениям точки 1, возвращениям в начало и первым возвращениям; см. формулы (3.6), (3.12) и (3.14). Вычисления здесь не требуются.

7. а) Вероятность того, что возвращение в начало происходит не позже момента n-го испытания, равна $(1-s)^{-1} F(s)$.

б) Следствие. Производящая функция последовательности вероятностей того, что $\mathbf{S}_j \neq 0$, $j=1, \ldots, n$, равна $\sqrt{(1+s)/(1-s)}=(1+s)\,U(s)$.

в) Показать, что это эквивалентно утверждению, приведенному после формулы (3.16).

8. Производящая функция последовательности вероятностей того, что после n-го испытания (исключая само n-е испытание) не происходит возвращения в начало, равна $(1-s)^{-1}\,U(s)\,|\,p-q\,|$.

9. а) Производящая функция последовательности вероятностей $\mathbf{P}\{\mathbf{S}_n=r\}$ (при фиксированном $r>0$) равна $\Phi^r(s)\,U(s)$.

б) При $p=1/2$ это выражение является производящей функцией последовательности вероятностей того, что $\mathbf{S}_k=r$ в точности для одного индекса $k \leqslant n$.

10. а) Найти производящую функцию последовательности вероятностей того, что событие $\mathbf{S}_n=r$ произойдет ровно k раз ($r>0$ и $k>0$ фиксированы).

б) Решить предыдущую задачу, заменив «ровно» на «самое большее».

11. а) Найти производящую функцию последовательности вероятностей того, что первое возвращение в начало, следующее за первым достижением точки $r>0$, происходит при испытании с номером r.

б) Решить предыдущую задачу, опустив слово «первое».

12. Найти производящую функцию величины $\mathbf{S}_r$ (r фиксировано) в *задаче о времени ожидания* (пример гл. IX, 3, г)). Проверить формулу (3.3) гл. IX для среднего значения и вычислить дисперсию.

13. *Продолжение*. Другой метод решения той же задачи. Пусть $p_n(r)=\mathbf{P}\{\mathbf{S}_r=n\}$. Доказать рекуррентную формулу

$$p_{n+1}(r)=[(r-1)/N]\,p_n(r)+[(N-r+1)/N)]\,p_n(r-1). \tag{7.1}$$

Непосредственно из (7.1) вывести производящую функцию.

14. Решить две предыдущие задачи для r заранее выбранных элементов (вместо r произвольных).

15[1]. Назовем *циклом* последовательность испытаний Бернулли до первой неудачи включительно. Найти производящую функцию и распределение вероятностей общего числа $\mathbf{S}_r$ успехов в r циклах.

16. *Продолжение*. а) Пусть $\mathbf{R}$ — число последовательных циклов до ν-го успеха (т. е. ν-й успех принадлежит $\mathbf{R}$-му циклу). Найти $\mathbf{E}(\mathbf{R})$ и $\mathbf{Var}(\mathbf{R})$. Доказать, что

$$\mathbf{P}\{\mathbf{R}=r\}=p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1}.$$

[1]) Задачи 15 и 16 имеют самое непосредственное отношение к игре в бильярд. Вероятность попадания p служит мерой искусства игрока. Игрок продолжает игру до первого промаха. Поэтому число забитых им шаров есть длина «цикла попаданий». Игра заканчивается, когда один из игроков забьет N шаров. Таким образом, в задаче 15 ищется распределение вероятностей для числа циклов, в течение которых игрок забивает k шаров, в задаче 16 — средняя продолжительность игры и вероятность ничьей. Дальнейшие подробности см. в статье Bottema O., Van Veen S. C., Kansberekningen bij het biljartspel, Nieuw Archief voor Wiskunde (на датском языке), 22 (1943), 16—33, 123—158.

б) Рассмотрим *две* последовательности испытаний Бернулли с вероятностями p_1, q_1 и p_2, q_2 соответственно. Найти вероятность совпадения номеров циклов, при которых произойдет N-й успех в каждой последовательности.

---

17. Пусть случайная величина $\mathbf{X}$ принимает значения $0, 1, \ldots, r-1$ каждое с вероятностью $1/r$. Показать, что если r является составным числом, например $r=ab$, то $\mathbf{X}$ можно представить в виде суммы независимых целочисленных случайных величин.

18. Пусть $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ — сумма взаимно независимых случайных величин, каждая из которых принимает значения $1, 2, \ldots, a$ с вероятностями $1/a$. Показать, что производящая функция величины $\mathbf{S}_n$ равна

$$P(s)=\{s(1-s^a)/[a(1-s)]\}^n,$$

откуда для $j\geqslant n$

$$\mathbf{P}\{\mathbf{S}_n=j\}=a^{-n}\sum_{\nu=0}^{\infty}(-1)^{\nu+j-n-a\nu}\binom{n}{\nu}\binom{-n}{j-n-a\nu}=$$
$$=a^{-n}\sum_{\nu=0}^{\infty}(-1)^{\nu}\binom{n}{\nu}\binom{j-a\nu-1}{n-1}.$$

(Лишь конечное число членов этой суммы отлично от нуля.)

Замечание. При $a=6$ мы получим вероятность выпадения $j+n$ очков при бросании n костей. Решение этой задачи восходит к Муавру.

19. *Продолжение.* Вероятность $\mathbf{P}\{\mathbf{S}_n\leqslant j\}$ имеет производящую функцию $P(s)/(1-s)$, и поэтому

$$\mathbf{P}\{\mathbf{S}_n\leqslant j\}=(1/a^n)\sum_{\nu}(-1)^{\nu}\binom{n}{\nu}\binom{j-a\nu}{-n}.$$

20. *Продолжение: предельная форма.* Если $a\to\infty$ и $j\to\infty$, причем $j/a\to x$, то

$$\mathbf{P}\{\mathbf{S}_n\leqslant j\}\to(1/n!)\sum_{\nu}(-1)^{\nu}\binom{n}{\nu}(x-\nu)^n,$$

где сумма берется по всем ν из интервала $0\leqslant\nu<x$.

Замечание. Этот результат принадлежит Лагранжу. В теории геометрических вероятностей правая часть представляет функцию распределения суммы n независимых случайных величин с «равномерным» распределением на интервале $(0, 1)$.

21. Пусть u_n — вероятность того, что число успехов в n испытаниях Бернулли делится на 3. Найти рекуррентное соотношение для u_n, а из него — производящую функцию.

22. *Продолжение: другой метод.* Пусть v_n и w_n — вероятности того, что $\mathbf{S}_n$ имеет соответственно вид $3\upsilon+1$ и $3\upsilon+2$ (так что $u_n+v_n+w_n=1$). Найти три одновременно выполняющиеся рекуррентные соотношения, а из них — три уравнения для производящих функций.

23. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с производящими функциями $U(s)$ и $V(s)$. Показать, что $\mathbf{P}\{\mathbf{X}-\mathbf{Y}=j\}$ равно коэффициенту при s^j в разложении функции $U(s)V(1/s)$, где $j=0, \pm1, \pm2, \ldots$.

24. *Производящие функции моментов.* Пусть $\mathbf{X}$ — случайная величина с производящей функцией $P(s)$, и пусть ряд $\sum p_n s^n$ сходится при некотором $s_0>1$. Тогда все моменты $m_r=\mathbf{E}(\mathbf{X}^r)$ существуют и производящая функция $F(s)$ последовательности $m_r/r!$ сходится по меньшей мере при $|s|<\log s_0$. Кроме

того,

$$F(s) = \sum_{r=0}^{\infty} (m_r/r!)\, s^r = P(e^s).$$

Замечание. $F(s)$ обычно называют *производящей функцией моментов*, хотя на самом деле она производит последовательность $m_r/r!$.

25. Предположим, что $A(s) = \sum a_n s^n$ — рациональная функция $U(s)/V(s)$ и что s_1 — наименьший по абсолютной величине корень многочлена $V(s)$. Показать, что если s_1 имеет кратность r, то

$$a_n \sim \frac{\rho_1}{s_1^{n+r}} \binom{n+r-1}{r-1},$$

где $\rho_1 = (-1)^r\, r!\, U(s_1)/V^{(r)}(s_1)$.

26. *Двойное отрицательное биномиальное распределение.* Показать, что при положительных значениях параметров выражение $p_0^a \{1 - p_1 s_1 - p_2 s_2\}^{-a}$ представляет собой производящую функцию распределения пары случайных величин $(\mathbf{X}, \mathbf{Y})$, таких, что маргинальные распределения $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{X}+\mathbf{Y}$ являются отрицательными биномиальными распределениями [1]).

---

[1]) Распределение этого типа использовали Дж. Бейтс и Дж. Нейман при исследованиях частоты катастроф; см. University of California Publications in Statistics, 1 (1952).

СЛОЖНЫЕ РАСПРЕДЕЛЕНИЯ[1]. ВЕТВЯЩИЕСЯ ПРОЦЕССЫ

Значительная часть теории вероятностей связана с суммами независимых случайных величин, и порою число слагаемых в таких суммах само является случайной величиной. Мы ограничимся здесь рассмотрением ситуации, когда слагаемые являются целочисленными случайными величинами, чтобы показать, как используются производящие функции, и подготовиться к изучению (в томе 2) безгранично делимых распределений и процессов с независимыми приращениями.

В качестве примера особенно привлекательных применений мы опишем некоторые результаты изящной теории ветвящихся процессов.

§ 1. СУММЫ СЛУЧАЙНОГО ЧИСЛА ВЕЛИЧИН

Пусть $\{\mathbf{X}_k\}$ — последовательность взаимно независимых случайных величин с одним и тем же распределением $\mathbf{P}\{\mathbf{X}_k=j\}=f_j$ и производящей функцией $f(s)=\sum f_j s^j$. Мы будем рассматривать суммы

$$\mathbf{S}_\mathbf{N}=\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_\mathbf{N},$$

в которых число $\mathbf{N}$ слагаемых является случайной величиной, не зависящей от $\mathbf{X}_j$. Пусть $\mathbf{N}$ имеет распределение $\mathbf{P}\{\mathbf{N}=n\}=g_n$ и производящую функцию $g(s)=\sum g_n s^n$. Распределение $\{h_j\}$ суммы $\mathbf{S}_\mathbf{N}$ можно найти по основной формуле для условных вероятностей

$$h_j=\mathbf{P}\{\mathbf{S}_\mathbf{N}=j\}=\sum_{n=0}^{\infty}\mathbf{P}\{\mathbf{N}=n\}\,\mathbf{P}\{\mathbf{X}_1+\ldots+\mathbf{X}_n=j\}. \qquad (1.1)$$

Если число возможных значений $\mathbf{N}$ конечно, то случайная величина $\mathbf{S}_\mathbf{N}$ определена на пространстве элементарных событий, соответствующем конечному числу $\mathbf{X}_k$. В противном случае для вероятностного определения случайной величины $\mathbf{S}_\mathbf{N}$ как суммы необходимо

*) Материал этой главы не будет использоваться в дальнейшем.

[1]) В оригинале compound distributions. Здесь имеются в виду распределения, производящие функции которых являются сложными функциями (compound functions) вида $h(s)=g(f(s))$, где g и f — производящие функции (иначе говоря, h представляет собой суперпозицию g и f). — *Прим. перев.*

ввести пространство элементарных событий, соответствующее бесконечной последовательности $\{\mathbf{X}_k\}$. Однако нас будет интересовать только функция распределения $\mathbf{S}_\mathbf{N}$, и мы рассматриваем распределение (1.1) как определение случайной величины $\mathbf{S}_\mathbf{N}$ на пространстве элементарных событий с точками 0, 1, 2, ...

Для фиксированного n распределение $\mathbf{X}_1+\ldots+\mathbf{X}_n$ является n-кратной сверткой $\{f_j\}$ с самим собой, поэтому (1.1) можно записать в компактном виде

$$\{h_j\}=\sum_{n=0}^{\infty} g_n \{f_j\}^{n*}. \tag{1.2}$$

Эту формулу можно упростить, используя производящие функции. Производящей функцией $\{f_j\}^{n*}$ является $f^n(s)$, и из (1.2) следует, что производящая функция суммы $\mathbf{S}_\mathbf{N}$ вычисляется по формуле

$$h(s)=\sum_{j=0}^{\infty} h_j s^j=\sum_{n=0}^{\infty} g_n f^n(s). \tag{1.3}$$

Правая часть представляет собой ряд Тейлора для $g(s)$, в котором s заменено на $f(s)$; значит, она равна $g(f(s))$. Тем самым доказано следующее утверждение.

Теорема. *Производящей функцией суммы* $\mathbf{S}_\mathbf{N}=\mathbf{X}_1+\ldots+\mathbf{X}_\mathbf{N}$ *является сложная функция* $g(f(s))$.

Доказательство можно повторить в терминах условных математических ожиданий. По определению

$$\mathrm{E}(s^{\mathbf{S}_\mathbf{N}}\mid \mathbf{N}=n)=f^n(s), \tag{1.4}$$

и, чтобы получить $h(s)=\mathrm{E}(s^{\mathbf{S}_\mathbf{N}})$, нам нужно умножить (1.4) на $\mathbf{P}\{\mathbf{N}=n\}$ и просуммировать по n (см. формулу (2.9) гл. IX).

Представляют интерес два частных случая.

а) Если $\mathbf{X}_j$—бернуллиевские случайные величины с $\mathbf{P}\{\mathbf{X}_j=1\}=p$ и $\mathbf{P}\{\mathbf{X}_j=0\}=q$, то $f(s)=q+ps$, и поэтому $h(s)=g(q+ps)$.

б) Если $\mathbf{N}$ имеет распределение Пуассона со средним t, то $h(s)=e^{-t+tf(s)}$. Распределение с такой производящей функцией будем называть *обобщенным распределением Пуассона*[1]). В частности, если $\mathbf{X}_j$—бернуллиевские величины и $\mathbf{N}$ имеет распределение Пуассона, то $h(s)=e^{-tp+tps}$; иначе говоря, *сумма* $\mathbf{S}_\mathbf{N}$ *имеет распределение Пуассона со средним* tp.

Примеры. а) В примере гл. VI, 7, в) мы отмечали, что облучение рентгеновскими лучами вызывает перестройку хромосом в клетках; при фиксированных дозе и времени облучения число $\mathbf{N}$ изменений имеет распределение Пуассона. Каждое изменение с

[1]) Мы используем принятый в нашей литературе термин; в оригинале compound Poisson distribution.— *Прим. перев.*

вероятностью q залечивается, а с вероятностью $p=1-q$ приводит к гибели клетки. Здесь S_N — число *наблюдаемых* изменений [1] — имеет распределение Пуассона со средним tp.

б) Пусть в экспериментах с отловом животных [2] g_n означает вероятность того, что численность животных данного вида равна n. Если для каждого животного вероятность быть пойманным равна p, то (в предположении стохастической независимости) число попавших в выборку представителей данного вида есть случайная величина S_N с производящей функцией $g(q+ps)$. Эту схему можно изменять многими способами. Например, пусть g_n — вероятность того, что насекомое откладывает n яиц, а p — вероятность появления из яйца личинки. Тогда S_N есть число появившихся личинок. Или пусть g_n — вероятность наличия в семье n детей, а отношение чисел мальчиков и девочек в популяции равно $p:q$. Тогда S_N — число мальчиков в семье.

в) Каждое растение дает много семян, но каждое семя имеет малую вероятность прорастания, и поэтому разумно предполагать, что число потомков отдельного растения имеет распределение Пуассона. Если $\{g_n\}$ — распределение числа растений, то $g(e^{-\lambda+\lambda s})$ — производящая функция числа проросших семян.

г) *Суммарное время обслуживания.* Рассмотрим телефонную линию, прилавок или любое другое обслуживающее устройство, для которого длительности обслуживания последовательно появляющихся клиентов могут рассматриваться как независимые одинаково распределенные случайные величины $X_1, X_2, \ldots$. Число клиентов (или заявок), поступающих за день, само является случайной величиной N, и суммарное время их обслуживания равно поэтому случайной сумме $X_1+\ldots+X_N$.

§ 2. ОБОБЩЕННОЕ РАСПРЕДЕЛЕНИЕ ПУАССОНА

Среди случайных сумм $S_N=X_1+\ldots+X_N$ наиболее важны для приложений те, в которых N имеет распределение Пуассона. По причинам, которые станут ясны позднее, мы обозначим математическое ожидание N через λt. Если X_j имеют одно и то же распределение $\{f_i\}$, то S_N имеет *обобщенное распределение Пуассона*

$$\{h_i\}_t = e^{-\lambda t}\sum_{n=0}^{\infty}[(\lambda t)^n/n!]\,\{f_i\}^{n*} \tag{2.1}$$

с производящей функцией

$$h_t(s)=e^{-\lambda t+\lambda t f(s)}. \tag{2.2}$$

[1] См. Catcheside D. G., Genetic effects of radiations, Advances in Genetics, edited by M. Demerec, v. 2, Acad. Press, New York, 1948, 271—358, особенно с. 339.

[2] Kendall D. G., On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, Biometrika, 35 (1948), 6—15.

Примеры. а) *Суммарный ущерб.* Предположим, что число ударов молнии в любом интервале времени длительности t является пуассоновской случайной величиной с математическим ожиданием λt. Если $\{f_n\}$ — распределение вероятностей ущерба, причиненного одним ударом молнии, то (при условии стохастической независимости) суммарный ущерб за время t имеет обобщенное распределение Пуассона (2.1).

б) *Ливни в космических лучах.* Принято считать, что число $\mathbf{N}$ ливней в космических лучах в интервале времени длительности t имеет распределение Пуассона с математическим ожиданием λt. Для любого счетчика число зарегистрированных частиц одного ливня является случайной величиной с распределением $\{f_i\}$. Общее число частиц, зарегистрированных за время t, снова есть случайная сумма $\mathbf{S_N}$ с обобщенным распределением Пуассона (2.1).

в) *В экологии* предполагается, что число выводков животных на данном участке имеет распределение Пуассона с математическим ожиданием, пропорциональным площади t участка. Пусть $\{f_k\}$ — распределение числа животных в выводке; предположим, что размеры выводков независимы. При этих условиях число животных на данном участке подчиняется обобщенному распределению Пуассона (2.1). Эта модель широко используется на практике. ►

Заметим, что все три примера тесно связаны с явлением, обсуждавшимся в гл. VI, 6 в связи с распределением Пуассона. В первых двух примерах каждому интервалу времени соответствует случайная величина $\mathbf{S_N}$. (Это верно и для примера в), если мы согласимся рассматривать площадь как своеобразное время.) В такой модели предполагается, что если интервал разбит на два непересекающихся интервала, то их вклады стохастически независимы и в сумме дают $\mathbf{S_N}$. Применительно к производящей функции (2.2) это означает, что

$$h_{t+\tau}(s) = h_t(s)\, h_\tau(s). \tag{2.3}$$

Производящая функция (2.2) любого обобщенного распределения Пуассона удовлетворяет (2.3). Покажем теперь, что верно и обратное: семейство производящих функций h_t, удовлетворяющее условию (2.3), обязательно имеет вид (2.2). (Следует иметь в виду, что это утверждение верно только для целочисленных случайных величин. Понятие обобщенного распределения Пуассона остается осмысленным даже тогда, когда $\mathbf{X}_j$ имеют произвольное распределение, а аналог соотношения (2.3) играет важную роль в общей теории случайных процессов с независимыми приращениями. Однако распределения таких процессов могут и не быть обобщенными распределениями Пуассона.)

Следующие определение и теорема в действительности относятся к распределениям вероятностей на множестве целых чисел

0, 1, . . ., однако для простоты они формулируются в терминах соответствующих производящих функций.

Определение. *Вероятностная производящая функция h называется безгранично делимой, если для любого натурального числа n корень n-й степени $\sqrt[n]{h}$ тоже является вероятностной производящей функцией.*

Следующая теорема показывает, что это утверждение остается справедливым даже когда $n>0$ не является целым числом. Если семейство вероятностных производящих функций удовлетворяет (2.3), то $\sqrt[n]{h_t}=h_{t/n}$, и поэтому h_t безгранично делима. Обратной этому утверждению является следующая теорема.

Теорема [1]. *Все безгранично делимые производящие функции имеют вид* (2.2), *причем* $\{f_i\}$ — *распределение вероятностей на* 0, 1, ...

Доказательство. Положим $h(s)=\sum h_k s^k$ и допустим, что $\sqrt[n]{h}$ является вероятностной производящей функцией при каждом $n\geqslant 1$. Тогда $h_0>0$, поскольку в противном случае свободный член в степенном ряду для $\sqrt[n]{h}$ был бы равен нулю, а это, в свою очередь, означало бы, что $h_0=h_1=\ldots=h_{n-1}=0$. Следовательно, $\sqrt[n]{h(s)}\to 1$ для любого $0\leqslant s\leqslant 1$, и поэтому

$$\log\sqrt[n]{h(s)/h_0}=\log\left[1+\left(\sqrt[n]{h(s)/h_0}-1\right)\right]\sim\sqrt[n]{h(s)/h_0}-1, \quad (2.4)$$

где знак $\sim$ означает, что отношение правой части к левой стремится к 1. Из (2.4) и его частного случая при $s=1$ мы получаем (учитывая, что $h(1)=1$):

$$\frac{\log h(s)-\log h_0}{-\log h_0}=\frac{\log\sqrt[n]{h(s)/h_0}}{\log\sqrt[n]{1/h_0}}\sim\frac{\sqrt[n]{h(s)}-\sqrt[n]{h_0}}{1-\sqrt[n]{h_0}}. \qquad (2.5)$$

Правая часть представляет собой степенной ряд с положительными коэффициентами, и, полагая $s=1$, легко проверить, что сумма этих коэффициентов равна 1. Таким образом, при любом n правая часть — вероятностная производящая функция, и поэтому левая часть является пределом последовательности вероятностных производящих функций. Согласно теореме непрерывности гл. XI,6, это означает, что левая часть — производящая функция неотрицательной последовательности $\{f_j\}$. Полагая $s=1$, убеждаемся в том, что $\sum f_j=1$. Значит, h имеет вид (2.2) с $\lambda t=-\log h_0$. ►

Переформулировкой этой теоремы является следующий критерий.

---

[1]) Это простой частный случай важной теоремы П. Леви.

Критерий. *Функция h является безгранично делимой вероятностной производящей функцией тогда и только тогда, когда $h(1)=1$ и*

$$\log[h(s)/h(0)]=\sum_{k=1}^{\infty} a_k s^k, \quad \text{где } a_k \geqslant 0, \quad \sum a_k=\lambda<\infty. \quad (2.6)$$

Действительно, если выполнено (2.6), то достаточно положить $f_k=a_k/\lambda$, чтобы привести h к канонической форме (2.2) (с $t=1$), а она, в свою очередь, является производящей функцией обобщенного распределения Пуассона, определенного в (2.1).

Примеры. г) Сравнивая (2.2) с теоремой предыдущего параграфа, мы видим, что если распределение $\mathbf{N}$ безгранично делимо, то таким же является и распределение случайной суммы $\mathbf{S_N}$.

д) *Отрицательное биномиальное* распределение с производящей функцией

$$h_t(s)=[p/(1-qs)]^t, \quad p+q=1, \quad (2.7)$$

обладает свойством (2.3) и поэтому безгранично делимо. Переходя к логарифмам, сразу убеждаемся в том, что оно действительно имеет вид (2.2), причем

$$f_n=q^n/(\lambda n), \quad \lambda=\log p^{-1}; \quad (2.8)$$

$\{f_n\}$ называется *логарифмическим распределением* и широко используется статистиками.

е) Из разложений (8.9) и (8.10) гл. II легко получить, что при $q=1-p>p$ функции

$$f(s)=\sqrt{(q-p)\frac{q+ps}{q-ps}}, \quad g(s)=\frac{\sqrt{q-p}}{\sqrt{q^2-p^2s^2}} \quad (2.9)$$

удовлетворяют условию (2.6), и поэтому *как f, так и g являются безгранично делимыми вероятностными производящими функциями.* Интересно заметить, что

$$f(s)=g(s)(q+ps). \quad (2.10)$$

Мы получили здесь *разложение безгранично делимой функции в произведение двух производящих функций, из которых только одна безгранично делима.* Существование таких разложений поначалу воспринималось как большая неожиданность, и эта тема некоторое время оживленно обсуждалась. ►

Замечательное свойство обобщенного распределения Пуассона является поводом для некоторых любопытных умозаключений. Если мы введем для простоты обозначение $\lambda_i=\lambda f_i$, то производящую функцию h_t из (2.2) можно представить в виде произведения

$$h_t(s)=e^{\lambda_1 t(s-1)}e^{\lambda_2 t(s^2-1)}e^{\lambda_3 t(s^3-1)}\ldots, \quad (2.11)$$

Здесь произведение может быть бесконечным, но это не имеет отношения к нашим рассуждениям, и мы можем считать, что лишь конечное число λ_i положительно. Первый сомножитель — это производящая функция обычного распределения Пуассона с математическим ожиданием $\lambda_1 t$. Второй сомножитель — это производящая функция удвоенной пуассоновской случайной величины, т. е. знакомую нам вероятность $e^{-\lambda_2 t}(\lambda_2 t)^n/n!$ имеет точка $2n$, а не n. Аналогично, k-й сомножитель соответствует распределению Пуассона, которое приписывает вероятности числам, кратным k. Таким образом, (2.11) дает новое представление $\mathbf{S}_N$ в виде суммы таких независимых случайных величин $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$, что $\mathbf{Y}_k$ принимает только значения $0, k, 2k, \ldots$ и имеет на них распределение Пуассона. Смысл соотношения (2.11) можно описать следующим образом. Пусть $\mathbf{N}_j$ — число величин $\mathbf{X}_1, \ldots, \mathbf{X}_N$, равных j. Тогда $\mathbf{N}=\mathbf{N}_1+\mathbf{N}_2+\ldots$, и (2.11) означает, что случайные величины $\mathbf{N}_k$ *взаимно независимы и имеют распределения Пуассона.*

Пример. ж) *Автомобильные катастрофы.* Будем рассматривать $\mathbf{X}_n$ как число автомобилей, попавших в n-ю катастрофу. При обычных предположениях о независимости $\mathbf{X}_n$ и о том, что число $\mathbf{N}$ катастроф имеет распределение Пуассона, находим, что общее число автомобилей, попавших в катастрофы, составляет $\mathbf{X}_1+\ldots+\mathbf{X}_N$ и имеет обобщенное распределение Пуассона (2.1). Мы можем теперь рассмотреть отдельно число $\mathbf{N}_k$ катастроф, в каждую из которых попало ровно k автомобилей. Согласно (2.11), случайные величины $\mathbf{N}_k$ взаимно независимы и имеют распределения Пуассона. Практические выводы из этого результата не нуждаются в комментариях. ►

(Еще один вариант обобщенного распределения Пуассона будет приведен в примере гл. XVII, 9, а).)

2а. Процессы с независимыми приращениями

Интерес к предыдущим результатам увеличивает их тесная связь с одним важным классом случайных процессов. Эти процессы будут сейчас неформально описаны, хотя соответствующая теория лежит вне рамок этой книги.

Чтобы начать с простейшего примера, рассмотрим число вызовов, поступающих на телефонную станцию, как функцию времени. Процесс работы станции описывается заданием для каждого t числа $\mathbf{Z}(t)$ вызовов, поступивших между 0 и t. Если последовательные вызовы поступали в моменты $t_1, t_2, \ldots$, то $\mathbf{Z}(t)=0$ при $0\leqslant t<t_1$, и вообще $\mathbf{Z}(t)=k$ при $t_k\leqslant t<t_{k+1}$. Обратно, любую неубывающую функцию, принимающую только значения $0, 1, 2, \ldots$, можно рассматривать как возможный вариант процесса работы телефонной станции. Поэтому вероятностная модель должна строиться на пространстве элементарных событий, точками которого являются функции $\mathbf{Z}(t)$ (а не последовательности, как в случае дискретных испытаний). Вероятности должны быть заданы способом, который позволял бы нам работать с такими сложными событиями, как событие, состоящее в том, что $\mathbf{Z}(t+1)-\mathbf{Z}(t)$ когда-нибудь превзойдет 17 или что $\mathbf{Z}(t)$ в некоторый момент времени превзойдет $at+b$ (последнее событие есть основной объект задачи о разорении

в теории коллективного риска). В дальнейшем мы считаем само собой разумеющимся, что такое задание действительно возможно; наша цель — показать, что из простых и естественных предположений о природе процесса следует, что для каждого фиксированного t случайная величина $\mathbf{Z}(t)$ должна иметь обобщенное распределение Пуассона.

Аналогичные рассуждения справедливы для широкого круга реальных явлений. Случайная величина $\mathbf{Z}(t)$ может соответствовать не числу телефонных вызовов, а суммарной длительности (или стоимости) уже проведенных разговоров, числу автомобилей, попавших в катастрофы, суммарному ущербу, причиненному молниями, общему потреблению электроэнергии, суммарному количеству осадков и т. п. Оставаясь в рамках настоящей главы, мы должны предполагать, что случайные величины $\mathbf{Z}(t)$ принимают только целые неотрицательные значения, но теорию можно обобщить на произвольные случайные величины. Наше внимание будет сосредоточено на процессах, удовлетворяющих двум следующим основным условиям, которые во многих приложениях представляются естественными.

а) Процесс является *однородным по времени*, т. е. распределения приращений $\mathbf{Z}(t+h)-\mathbf{Z}(t)$ зависят лишь от длины интервала времени, но не от его положения [1].

б) Приращения $\mathbf{Z}(t_2)-\mathbf{Z}(t_1)$ и $\mathbf{Z}(t_1)-\mathbf{Z}(t_0)$ на смежных интервалах времени взаимно независимы.

Результаты предыдущего раздела можно теперь переформулировать следующим образом: *если* существует процесс, удовлетворяющий условиям а) и б), то его приращения $\mathbf{Z}(t+h)-\mathbf{Z}(t)$ имеют обобщенные распределения Пуассона. В частности, если $\mathbf{Z}(t)$ имеет скачки только единичного размера, то эти случайные величины имеют обычные распределения Пуассона (см. формулу (2.11)).

Таким образом, мы охарактеризовали обычные и обобщенные распределения Пуассона присущими им свойствами; в отличие от рассуждений, проведенных в гл. VI, распределение Пуассона оказывается не приближением, а вполне самостоятельным распределением (можно сказать, проявлением некоторого естественного закона). Разумеется, теперь перед нами встает обратная задача: каждому ли семейству обобщенных распределений Пуассона соответствует некоторый случайный процесс? Ответ на этот вопрос утвердительный, однако (несколько неожиданно) оказывается, что двух наших условий не достаточно для *однозначного* определения процесса. Для однозначного определения представляющих интерес процессов необходимо усилить условие б), потребовав, чтобы n приращений, соответствующих конечному разбиению $t_0<t_1<\ldots<t_n$, были взаимно независимы при любом n. Это — определяющее свойство *процессов с независимыми приращениями*. Любое семейство обобщенных распределений Пуассона однозначно определяет процесс с независимыми приращениями, и поэтому никаких теоретических трудностей не возникает. Однако мы предполагали независимость только для двух интервалов. Такое ограниченное условие достаточно для определения вида распределений приращений, но можно построить довольно патологические процессы, удовлетворяющие указанному условию [2]. На этом примере видны трудности, присущие построению точной модели случайного процесса.

---

[1] Это условие не столь ограничительно, как может показаться на первый взгляд. Например, вызовы на телефонную станцию в самый напряженный дневной час поступают чаще, чем от полуночи до часу ночи, поэтому процесс является неоднородным по времени. Однако по очевидным причинам связисты рассматривают в основном «час пик», а на этом отрезке времени процесс может считаться однородным. Практика показывает, что в течение часа пик поступление вызовов с удивительной точностью соответствует распределению Пуассона.

[2] В таком процессе приращение $\mathbf{Z}(t_2)-\mathbf{Z}(t_1)$ не зависит ни от $\mathbf{Z}(t_3)-\mathbf{Z}(t_2)$, ни от $\mathbf{Z}(t_1)-\mathbf{Z}(t_0)$, но однозначно определяется последними двумя приращениями. См. Feller W., Non-Markovian processes with the semi-group property, Ann. Math. Statist., 30 (1959), 1252—1253.

§ 3. ПРИМЕРЫ ВЕТВЯЩИХСЯ ПРОЦЕССОВ

Мы опишем случайный процесс, который является упрощенной моделью ряда реальных процессов, а также показывает полезность производящих функций. Словесно этот процесс описывается следующим образом.

Мы рассматриваем частицы, которые могут порождать новые частицы того же вида. Одна начальная частица образует исходное, или нулевое, поколение. Каждая частица с вероятностью p_k ($k=0, 1, 2, \ldots$) порождает ровно k новых частиц; непосредственные потомки частиц n-го поколения образуют $(n+1)$-е поколение. Частицы каждого поколения размножаются независимо одна от другой. Нас интересуют размеры последовательных поколений.

Прежде чем дать строгое описание в терминах случайных величин, имеет смысл привести несколько примеров.

а) *Ядерные цепные реакции.* Это применение стало общеизвестным в связи с атомной бомбой[1]). Частицами являются нейтроны, которые случайным образом сталкиваются с другими частицами. Пусть p — вероятность того, что частица рано или поздно испытает столкновение, породив при этом m частиц; тогда $q=1-p$ — вероятность того, что у частицы не будет потомков, т. е. она останется инертной (удалится или поглотится каким-либо другим способом). В этой схеме единственно возможными значениями числа потомков являются 0 и m, а соответствующие вероятности равны p и q (т. е. $p_0=q$, $p_m=p$, $p_j=0$ для всех остальных j). В худшем случае первая частица останется инертной и процесс никогда не начнется. В лучшем случае в первом поколении будет m частиц, во втором m^2 и т. д. Если p близко к единице, то весьма вероятно, что число частиц будет быстро возрастать. В математической модели это число может расти неограниченно. На самом деле при очень большом числе частиц вероятности расщепления не могут оставаться постоянными и стохастической независимости тоже не может быть. Тем не менее для реальных цепных реакций математическое описание «неограниченно увеличивающегося числа частиц» можно интерпретировать как «взрыв».

б) *Выживание фамилий.* В этом случае (как часто в жизни) учитываются только потомки мужского пола; они играют роль частиц, и p_k — это вероятность того, что новорожденный мальчик будет отцом ровно k мальчиков. Наша схема содержит два искусственных упрощения. Рождаемость меняется от века к веку, и поэтому в действительности распределение $\{p_k\}$ изменяется от поколения к поколению. Более того, общность наследственности и ок-

---

[1]) Приведенное ниже описание следует работе Э. Шредингера (Schrödinger E., Probability problems in nuclear chemistry, Proceedings of the Royal Irish Academy, 51, sect. A, No. 1 (December 1945)). В этой работе пространственная однородность не предполагается.

ружающей среды обусловливают сходство братьев, что противоречит нашему предположению о стохастической независимости. Нашу модель можно уточнить, чтобы учесть эти замечания, но наиболее существенные ее черты изменить не удается. Мы найдем вероятность того, что в n-м поколении будет ровно k обладателей данной фамилии, в частности вероятность исчезновения фамилии. По-видимому, выживание фамилий — первая цепная реакция, изучавшаяся вероятностными методами. Впервые эту задачу рассматривал Ф. Гальтон в 1889 г.; подробности можно найти в книге А. Лотка [1]). Лотка указывает, что американские статистические данные хорошо соответствуют распределению $p_0=0{,}4825$, $p_k=(0{,}2126)\times\times(0{,}5893)^{k-1}$ $(k\geqslant 1)$, которое (если отвлечься от первого члена), является геометрическим.

в) *Гены и мутации.* Каждый ген данного организма (см. гл. V, 5) может быть передан 0, 1, 2, ... непосредственным потомкам, и наша схема описывает этот процесс, пренебрегая, конечно, неоднородностями по популяции и по времени. Особенно полезна эта схема при изучении мутаций, т. е. изменений структуры гена. Случайная мутация приводит к образованию одного гена нового типа, который играет роль частицы нулевого поколения. Теория дает оценки для вероятностей сохранения и распространения мутантного гена. Чтобы пояснить основные идеи, рассмотрим (следуя Р. Э. Фишеру) растение кукурузы, которое является отцовским для 100 семян и материнским для стольких же семян. Если размер популяции не изменяется, то в среднем два из этих 200 семян разовьются в новые растения. Каждое зерно получает данный ген с вероятностью 1/2. Вероятность того, что мутантный ген будет иметься ровно у k новых растений, соответствует поэтому вероятности ровно k успехов в 200 испытаниях Бернулли с вероятностью успеха $p=1/200$, и представляется разумным предположить, что $\{p_k\}$ близко к распределению Пуассона со средним 1. Если мутантный ген обеспечивает какие-либо биологические преимущества, то получается распределение Пуассона со средним $\lambda>1$.

г) *Очереди* [2]). Теория ветвящихся процессов находит интересные применения в теории очередей. Например, покупатель, подошедший к свободному продавцу и сразу приступивший к выбору покупки, называется предком; его непосредственными потомками являются покупатели, пришедшие за время его обслуживания и образовавшие очередь. Этот процесс продолжается до тех пор, пока очередь не будет исчерпана. Мы подробнее рассмотрим этот процесс в примере 5, б), а его более интересный вариант — в примере 5, в). ▶

[1]) Lotka A., Théorie analytique des associations biologiques, v. 2, Actualités scientifiques et industrielles, No. 780 (1939), Paris, Hermann et Cie, 123—136.

[2]) Kendall D. G., Some problems in the theory of queues, J. Roy. Statist. Soc. (Series B), 13 (1951), 151—173; обсуждение с. 173—185.

§ 4. ВЕРОЯТНОСТИ ВЫРОЖДЕНИЯ ВЕТВЯЩИХСЯ ПРОЦЕССОВ

Обозначим через $\mathbf{Z}_n$ размер n-го поколения, а через P_n—производящую функцию его распределения. По предположению $\mathbf{Z}_0=1$ и

$$P_1(s)=P(s)=\sum_{k=0}^{\infty} p_k s^k. \tag{4.1}$$

Можно разделить n-е поколение на $\mathbf{Z}_1$ семейств по их предкам в первом поколении. Это означает, что $\mathbf{Z}_n$ есть сумма $\mathbf{Z}_1$ случайных величин $\mathbf{Z}_n^{(k)}$, каждая из которых представляет собой размер потомства одной частицы первого поколения. По предположению каждая из величин $\mathbf{Z}_n^{(k)}$ имеет то же распределение вероятностей, что $\mathbf{Z}_{n-1}$, и (при фиксированном n) случайные величины $\mathbf{Z}_n^{(k)}$ взаимно независимы. Поэтому производящая функция P_n является сложной функцией:

$$P_n(s)=P(P_{n-1}(s)). \tag{4.2}$$

Это равенство позволяет последовательно находить все производящие функции. Согласно (4.2), мы имеем $P_2(s)=P(P(s))$, $P_3(s)=$ $=P(P_2(s))$ и т. д. Эти вычисления не представляют трудностей, хотя явные выражения для P_n, как правило, найти нелегко. Тем не менее мы вскоре увидим, что из (4.2) можно получить важные следствия.

Пример. Предположим, что число непосредственных потомков имеет геометрическое распределение $\{qp^k\}$, где $p\neq q$. Тогда $P(s)=$ $=q/(1-ps)$, и, вычислив в явном виде P_2, P_3 и т. д. (и приложив известные усилия), мы приходим к общей формуле

$$P_n(s)=q\cdot\frac{p^n-q^n-(p^{n-1}-q^{n-1})\,ps}{p^{n+1}-q^{n+1}-(p^n-q^n)\,ps}. \tag{4.3}$$

Легко проверить, что (4.3) действительно удовлетворяет (4.2).

Если $p=q$, то, полагая $p\to 1/2$, получаем

$$P_n(s)=[n-(n-1)s]/(n+1-ns). \tag{4.4}$$

Заметим, что $P_n(0)\to q/p$, если $p>q$, но $P_n(0)\to 1$, если $p\leqslant q$. Теперь мы обсудим этот результат и найдем его аналог для произвольных распределений $\{p_k\}$. ►

Самый первый вопрос, относящийся к нашему ветвящемуся процессу, состоит в том, будет процесс продолжаться бесконечно или же все потомство вымрет после конечного числа поколений. Положим

$$x_n=\mathbf{P}\{\mathbf{Z}_n=0\}=P_n(0). \tag{4.5}$$

Это вероятность того, что процесс окончится на n-м поколении или еще раньше. По определению $x_1=p_0$, и из (4.2) следует, что

$$x_n=P(x_{n-1}). \tag{4.6}$$

Ввиду тривиальности крайних случаев $p_0=0$ и $p_0=1$ мы далее будем предполагать, что $0<p_0<1$. Из монотонности P выводим, что $x_2=P(p_0)>P(0)=x_1$, и аналогично по индукции, что $x_1<x_2<x_3\ldots$. Значит, эта последовательность имеет предел $x\leqslant 1$, и из (4.6) вытекает, что

$$x=P(x). \tag{4.7}$$

При $0\leqslant s\leqslant 1$ график $P(s)$ — *выпуклая вниз* кривая, начинающаяся в точке $(0, p_0)$ над биссектрисой первого квадранта и заканчивающаяся в точке $(1, 1)$ на этой биссектрисе. Таким образом, возможны только две ситуации.

Случай (i). Весь график находится выше биссектрисы. В этом случае $x=1$ — единственный корень уравнения (4.7), и поэтому $x_n\to 1$. Далее, в этом случае $1-P(s)\leqslant 1-s$ для всех s, и, устремляя s к 1, мы находим, что производная $P'(1)$ удовлетворяет неравенству $P'(1)\leqslant 1$.

Случай (ii). График P пересекает биссектрису в некоторой точке $\sigma<1$. Поскольку выпуклая кривая пересекает прямую линию не более чем в двух точках, в этом случае $P(s)>s$ при $s<\sigma$ и $P(s)<s$ при $\sigma<s<1$. Тогда $x_1=P(0)<P(\sigma)=\sigma$, и по индукции $x_n=P(x_{n-1})<<P(\sigma)=\sigma$. Значит, $x_n\to\sigma$, и, таким образом, $x=\sigma$. С другой стороны, по теореме о среднем значении между σ и 1 существует точка, в которой производная P' равна 1. Из монотонности этой производной следует, что $P'(1)>1$.

Итак, эти два случая характеризуются условиями $P'(1)\leqslant 1$ и $P'(1)>1$ соответственно. Но

$$\mu=P'(1)=\sum_{k=0}^{\infty} kp_k\leqslant\infty \tag{4.8}$$

есть *математическое ожидание числа непосредственных потомков*, т. е. нами доказана интересная теорема.

Теорема. *Если $\mu\leqslant 1$, то процесс вырождается с вероятностью 1. Если же $\mu>1$, то вероятность x_n того, что процесс вырождается в n-м поколении или раньше, стремится к единственному корню $x<1$ уравнения* (4.7).

Как правило, сходимость $x_n\to x$ является быстрой, и поэтому с большой вероятностью процесс либо довольно быстро вырождается, либо продолжается неограниченно. Математическое ожидание размера n-го поколения дается формулой $\mathbf{E}(\mathbf{Z}_n)=P_n'(1)$. Из (4.2) по правилу дифференцирования сложной функции мы получаем, что $P_n'(1)=P'(1)P_{n-1}'(1)=\mu\mathbf{E}(\mathbf{Z}_{n-1})$, и, следовательно[1],

$$\mathbf{E}(\mathbf{Z}_n)=\mu^n. \tag{4.9}$$

[1]) Дальнейшие подробности можно найти в содержательной монографии Harris T. E., The theory of branching processes, Berlin, Springer, 1963. [Имеется перевод: Харрис Т. Е. Теория ветвящихся случайных процессов.— М.: Мир, 1966.] (См. также книгу Севастьянов Б. А. Ветвящиеся процессы.— М.: Наука, 1971.— *Перев.*)

Нет ничего удивительного в том, что процесс неизбежно вырождается при $\mu<1$, однако заранее не было ясно, что устойчивой ситуации не может быть даже при $\mu=1$[1]). Если $\mu>1$, то в соответствии с (4.9) следовало бы ожидать, что $\mathbf{Z}_n$ растет со скоростью геометрической прогрессии. До некоторой степени это верно; однако независимо от того, сколь велико μ, вероятность вырождения может быть отделена от 0. Легко проверить, что $P_n(s)\to x$ при любом $s<1$, а это значит, что коэффициенты при s, s^2, s^3 и т. д. стремятся к 0. Поэтому *с большой вероятностью после достаточно большого числа поколений либо потомков нет совсем, либо их очень много* (вероятности этих событий стремятся соответственно к x и $1-x$).

§ 5. ОБЩЕЕ ЧИСЛО ЧАСТИЦ В ВЕТВЯЩИХСЯ ПРОЦЕССАХ[2])

Рассмотрим теперь случайную величину

$$\mathbf{Y}_n=1+\mathbf{Z}_1+\ldots+\mathbf{Z}_n, \tag{5.1}$$

равную общему числу частиц в поколениях с номерами от 0 до n включительно. Полагая $n\to\infty$, мы получаем общее число частиц, которое может быть конечным или бесконечным. Очевидно, случайная величина $\mathbf{Y}_n$ корректно определена для каждого n; мы обозначим через R_n производящую функцию ее распределения вероятностей. Поскольку $\mathbf{Y}_1=1+\mathbf{Z}_1$, мы имеем $R_1(s)=sP(s)$. Рекуррентную формулу для R_n можно получить при помощи тех же рассуждений, что в предыдущем параграфе; единственное отличие состоит в том, что для получения $\mathbf{Y}_n$ мы должны прибавить к начальной частице сумму потомков $\mathbf{Z}_1$ частиц первого поколения. Следовательно,

$$R_n(s)=sP(R_{n-1}(s)). \tag{5.2}$$

По формуле (5.2) в принципе можно последовательно вычислять R_1, R_2, ..., но этим заниматься не стоит. Асимптотическое поведение R_n можно изучать при помощи тех же геометрических рассуждений, которые использовались в предыдущем параграфе при нахождении вероятности вырождения x.

Заметим прежде всего, что при любом $s<1$

$$R_2(s)=sP(R_1(s))<sP(s)=R_1(s), \tag{5.3}$$

и по индукции находим, что $R_n(s)<R_{n-1}(s)$. Таким образом, $R_n(s)$, монотонно убывая, стремится к пределу $\rho(s)$, который удовлетворяет уравнению

$$\rho(s)=sP(\rho(s)),\quad 0<s<1. \tag{5.4}$$

---

[1]) Если не обращать внимания на случай, когда число непосредственных потомков равно 1 с вероятностью 1.— *Прим. перев.*

[2]) Этот параграф написан под впечатлением работы И. Дж. Гуда (Good I. J., The number of individuals in a cascade process, Proc. Cambridge Philos. Soc., 45 (1949), 360—363).

Согласно теореме непрерывности гл. XI, 6, функция ρ как предел вероятностных производящих функций является производящей функцией такой последовательности ρ_k неотрицательных чисел, что $\sum \rho_k \leqslant 1$.

Из (5.4) следует, что при фиксированном $s<1$ значение $\rho(s)$ является корнем уравнения

$$t = sP(t). \tag{5.5}$$

Покажем, что этот корень является единственным. Для этого снова обозначим буквой x наименьший положительный[1]) корень уравнения $x=P(x)$ (так что $x \leqslant 1$). Заметим, что $y=sP(t)$ (при фиксированном s) является выпуклой вниз функцией t, и поэтому ее график пересекает прямую $y=t$ не более чем в двух точках. Но при $t=0$ правая часть (5.5) больше левой, тогда как при $t=x$ и при $t=1$ верно обратное неравенство; значит, (5.5) имеет ровно один корень между 0 и x и не имеет корней между x и 1. Таким образом, $\rho(s)$ как корень (5.5) определяется однозначно, и мы видим, далее, что $\rho(s)<x$. Но очевидно, $\rho(1)$ является корнем уравнения $t=P(t)$, и, поскольку x — наименьший корень этого уравнения, ясно, что $\rho(1)=x$. Иначе говоря, производящая функция ρ будет вероятностной тогда и только тогда, когда $x=1$. Мы можем подвести итоги этих рассуждений следующим образом.

Пусть ρ_k — вероятность того, что общее число частиц равно k.

а) $\sum \rho_k$ равна вероятности вырождения x (а $1-x$ есть вероятность того, что общее число частиц равно ∞).

б) Производящая функция $\rho(s)=\sum \rho_k s^k$ является единственным решением уравнения (5.5), и $\rho(s) \leqslant x$.

Мы уже знаем, что с вероятностью 1 *общее число частиц конечно, если $\mu \leqslant 1$.* Дифференцирование (5.4) теперь показывает, что *его математическое ожидание равно $1/(1-\mu)$, когда $\mu<1$, и бесконечно, когда $\mu=1$.*

Примеры. а) В примере 4,а) мы нашли, что $P(s)=q/(1-ps)$, и (5.5) сводится к квадратному уравнению $pt^2-t+qs=0$, из которого мы заключаем, что

$$\rho(s) = \left(1-\sqrt{1-4pqs}\right)/(2p). \tag{5.6}$$

(Эта производящая функция появлялась в связи с моментами первого достижения в гл. XI, 3.)

б) *Периоды занятости.* Займемся более подробным анализом задачи об очередях, упоминавшейся в примере 3, г). Предположим для простоты, что покупатели могут прибывать только по одному и только в целочисленные моменты времени. Мы предполагаем, что прибытия регулируются испытаниями Бернулли: в момент n поку-

[1]) При этом предполагается, что $P(0)>0$, так как при $P(0)=0$ случайная величина Y_n стремится к ∞ с вероятностью 1.— *Прим. перев.*

патель появляется с вероятностью p, а с вероятностью $q=1-p$ не появляется. Покупатель, который прибывает, когда продавец свободен, немедленно начинает обслуживаться; в противном случае он становится в очередь. Продавец работает без перерывов, пока в очереди есть покупатели, ожидающие обслуживания. Мы предполагаем, наконец, что последовательные времена обслуживания являются независимыми (целочисленными) случайными величинами с одним и тем же распределением $\{\beta_k\}$ и производящей функцией $\beta(s)=\sum \beta_k s^k$.

Допустим теперь, что покупатель появляется в нулевой момент времени и застает продавца свободным. Немедленно начинается обслуживание. Если оно имеет длительность n, то продавец освобождается в момент n при условии, что в моменты $1, 2, \ldots, n$ не появлялись новые покупатели. В противном случае обслуживание продолжается без перерыва. *Периодом занятости* мы называем длительность непрерывного обслуживания, начавшегося в нулевой момент времени. Мы покажем, как можно использовать теорию ветвящихся процессов при изучении длительности периода занятости.

Покупатель, появившийся в момент 0, начинает период занятости и будет называться предком. Первое поколение «частиц» состоит из покупателей, появившихся до момента окончания обслуживания предка или в этот момент. Если непосредственные потомки отсутствуют, то процесс вырождается. В противном случае непосредственные потомки последовательно обслуживаются, а появляющиеся за их времена обслуживания их непосредственные потомки присоединяются к очереди. Мы имеем здесь такой ветвящийся процесс, в котором *вероятность вырождения x равна вероятности того, что период занятости конечен, а множество всех «частиц» состоит из всех покупателей (включая предка), прибывших в течение периода занятости*. Необходимо отметить, что практически возможны только очереди с $x=1$.

Для применения наших результатов необходимо знать производящую функцию $P(s)$ числа непосредственных потомков. По определению это число определяется случайной суммой $\mathbf{X}_1+\ldots+\mathbf{X}_\mathbf{N}$, где $\mathbf{X}_j$ взаимно независимы и принимают значения 1 и 0 с вероятностями p и q, а $\mathbf{N}$ — время обслуживания предка. Поэтому в рассматриваемой ситуации $P(s)=\beta(ps+q)$, и, следовательно, $\mu=p\sigma$, где $\sigma=\beta'(1)$ — математическое ожидание длительности обслуживания. Значит, *период занятости с вероятностью 1 конечен только при $p\sigma\leqslant 1$. Математическое ожидание числа покупателей, прибывающих за время периода занятости, конечно только при $p\sigma<1$.* Иначе говоря, при $p\sigma=1$ неминуема давка, а если $p\sigma$ несущественно меньше 1, то длинные очереди должны быть обычным явлением.

в) *Длительность периода занятости.* В предыдущем примере изучалось число покупателей, прибывающих за время периода занятости, однако для практики больший интерес представляет длительность этого периода. Ее можно найти при помощи изящного приема [1]), состоящего в рассмотрении целочисленных моментов времени как частиц ветвящегося процесса. Будем говорить, что момент времени n не имеет потомков, если в этот момент не появляется покупатель. Если же покупатель появляется и его обслуживание начинается в момент $m \geqslant n$ и длится r единиц времени, то моменты времени $m+1, \ldots, m+r$ рассматриваются как непосредственные потомки момента n. Предположим, что в момент времени 0 продавец свободен. Несложное рассуждение показывает, что тогда либо покупатель не появляется и ветвящийся процесс не начинается вообще, либо покупатель появляется и участвующие в ветвящемся процессе частицы (моменты времени) образуют период занятости. Производящая функция числа непосредственных потомков имеет вид

$$P(s) = q + p\beta(s). \tag{5.7}$$

Корень x равен вероятности того, что период занятости конечен. Общее число частиц с вероятностью q равно 1 и с вероятностью p равно длительности периода занятости, начавшегося в момент 0. Очевидно, длительность периода занятости имеет производящую функцию $\beta(\rho(s))$. ▶

§ 6. ЗАДАЧИ

1. Распределение (1.1) случайной суммы $\mathbf{S_N}$ имеет среднее $\mathbf{E}(\mathbf{N})\,\mathbf{E}(\mathbf{X})$ и дисперсию $\mathbf{E}(\mathbf{N})\operatorname{Var}(\mathbf{X}) + \operatorname{Var}(\mathbf{N})\,\mathbf{E}^2(\mathbf{X})$. Убедиться в этом: *а*) при помощи производящих функций, б) непосредственно из определения при помощи условных математических ожиданий.

2. *Отлов животных* (пример 1, б)). Доказать следующие утверждения. Если $\{g_n\}$ — геометрическое распределение, то искомое распределение тоже будет геометрическим. Если $\{g_n\}$ — логарифмическое распределение (см. (2.8)), то искомое распределение будет логарифмическим с добавочным членом.

3. Если случайная величина $\mathbf{N}$ имеет распределение Пуассона, то в $\mathbf{N}$ испытаниях Бернулли числа успехов и неудач стохастически независимы. Обобщить этот результат на полиномиальное распределение: а) непосредственно, б) используя производящие функции от нескольких переменных. (Ср. с примером гл. IX, 1, г).)

4. *Рандомизация.* Пусть $\mathbf{N}$ имеет распределение Пуассона с параметром λ, и пусть $\mathbf{N}$ шаров случайно размещаются по n ящикам. Показать без вычислений, что вероятность получить при этом ровно m пустых ящиков равна $\binom{n}{m} e^{-\lambda m/n}(1-e^{-\lambda/n})^{n-m}$.

5. *Продолжение* [2]). Показать, что если по n ящикам случайно размещается фиксированное число r шаров, то вероятность получить ровно m пустых ящиков равна коэффициенту при $e^{-\lambda}\lambda^r/r!$ в разложении приведенного выше выражения в ряд. а) Обсудить связь с производящими функциями моментов (задача 24 гл. XI, 7). б) Использовать этот результат для того, чтобы без всяких усилий получить соотношение (11.7) гл. II.

---

[1]) Этот прием предложил И. Дж. Гуд; см. материалы дискуссии, приведенные после статьи Кендалла, цитированной в примере 3, г).

[2]) Этот изящный способ получения разнообразных комбинаторных формул с помощью рандомизации параметра предложил Домб (Domb C., On the use of a random parameter in combinatorial problems, Proceedings Physical Society, Sec. A, 65 (1952), 305—309).

6. *Смеси вероятностных распределений.* Пусть $\{f_i\}$ и $\{g_i\}$ — два вероятностных распределения, $\alpha > 0$, $\beta > 0$, $\alpha+\beta=1$. Тогда $\{\alpha f_i+\beta g_i\}$ тоже является вероятностным распределением. Обсудить смысл этого преобразования и связь с урновыми схемами гл. V, 2. Обобщить на случай многих распределений. Показать, что такая смесь может быть обобщенным распределением Пуассона.

7. При помощи производящих функций показать, что в ветвящемся процессе $\mathrm{Var}(\mathbf{Z}_{n+1})=\mu\,\mathrm{Var}(\mathbf{Z}_n)+\mu^{2n}\sigma^2$. Используя условные математические ожидания, доказать эквивалентное соотношение $\mathrm{Var}(\mathbf{Z}_{n+1})=\mu^2\,\mathrm{Var}(\mathbf{Z}_n)+\mu^n\sigma^2$. Из каждого из этих соотношений вывести равенство $\mathrm{Var}(\mathbf{Z}_n)=\sigma^2(\mu^{2n-2}+\mu^{2n-3}+\ldots+\mu^{n-1})$.

8. *Продолжение.* Показать, что если $n > m$, то $\mathrm{E}(\mathbf{Z}_n\mathbf{Z}_m)=\mu^{n-m}\mathrm{E}(\mathbf{Z}_m^2)$.

9. *Продолжение.* Показать, что производящей функцией совместного распределения $(\mathbf{Z}_m, \mathbf{Z}_n)$ является $P_m(s_1P_{n-m}(s_2))$. Использовать этот факт для проверки утверждения из задачи 8.

10. Как изменится ветвящийся процесс, если каждая частица перед размножением может исчезнуть с фиксированной вероятностью p?

11. *Ветвящиеся процессы с двумя типами частиц.* Допустим, что каждая частица может иметь потомков двух типов; распределения чисел потомков частиц двух типов задаются производящими функциями $P_1(s_1, s_2)$ и $P_2(s_1, s_2)$. В этом случае существует две вероятности вырождения x, y в зависимости от типа предка. Показать, что пара (x, y) удовлетворяет уравнениям

$$x=P_1(x, y), \quad y=P_2(x, y). \tag{6.1}$$

Доказать, что эти уравнения имеют не более одного решения $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$, отличного от $(1, 1)$. Решение $(1, 1)$ единственно тогда и только тогда, когда $\mu_{11}\leqslant 1$, $\mu_{22}\leqslant 1$ и $(1-\mu_{11})(1-\mu_{22})\geqslant\mu_{12}\mu_{21}$, где $\mu_{ij}=\partial P_i(1, 1)/\partial s_j$.

РЕКУРРЕНТНЫЕ СОБЫТИЯ. ТЕОРИЯ ВОССТАНОВЛЕНИЯ

§ 1. НЕФОРМАЛЬНОЕ ВВЕДЕНИЕ И ПРИМЕРЫ

Мы будем изучать некоторые повторяющиеся, или рекуррентные, наборы исходов, связанные с последовательностями испытаний. Говоря нестрого, набор $\mathscr{E}$ годится для последующей теории, если после каждого появления набора $\mathscr{E}$ испытания начинаются заново в том смысле, что последовательность испытаний, следующих за появлением $\mathscr{E}$, образует копию всей последовательности. Промежутки времени между последовательными появлениями $\mathscr{E}$ — взаимно независимые одинаково распределенные случайные величины.

В простейшем частном случае $\mathscr{E}$ используется как обозначение «успеха» в последовательности испытаний Бернулли. Время ожидания первого успеха имеет геометрическое распределение; когда достигается первый успех, испытания начинаются заново и число испытаний между r-м и $(r+1)$-м успехами имеет то же самое геометрическое распределение. Время ожидания r-го успеха является суммой r независимых случайных величин (пример гл. IX, 3, в)). Эта ситуация повторяется, когда $\mathscr{E}$ обозначает «успех, за которым следует неудача»: появление набора *УН* восстанавливает исходное положение и время до появления следующего *УН* не зависит от исходов предыдущих испытаний. Приведем противоположный пример: пусть к группе людей последовательно присоединяются новые члены, и пусть $\mathscr{E}$ — событие «у двух человек в группе дни рождения совпадают». Такое событие $\mathscr{E}$ не является повторяющимся, потому что после первого его появления оно происходит постоянно. Если мы заменим определение на «$\mathscr{E}$ происходит, когда день рождения нового члена группы уже встречался в ней», то $\mathscr{E}$ может происходить любое число раз, но после появления $\mathscr{E}$ процесс *не* начинается заново. Это происходит потому, что с увеличением размера группы совпадения дней рождения становятся более вероятными, и поэтому большое время ожидания первого повторения дней рождения сочетается с более коротким временем ожидания второго повторения; таким образом, последовательные времена ожидания не являются независимыми и имеют неодинаковые распределения.

Большое значение теории рекуррентных событий объясняется тем, что часто такие события возникают в связи с различными последовательностями случайных величин (случайными процессами). Сложность законов, которым подчиняется последовательность случайных величин, может сделать невозможным ее исчерпывающий анализ, однако существование повторяющихся событий всегда позволяет указать характерные особенности последовательности, доказать существование некоторых пределов и т. п. Такой подход способствует упрощению и стандартизации многих исследований.

Мы рассмотрим теперь несколько типичных примеров; некоторые из них представляют и самостоятельный интерес. Первые примеры связаны с обычными испытаниями Бернулли, последние три касаются более сложных схем. В их описаниях мы используем такие термины, как «продавец» и «покупатель», но каждый раз приводится математическое определение последовательности случайных величин, которое является полным в том смысле, что оно однозначно определяет вероятности всех возможных событий. Часто оказывается, что теория рекуррентных событий позволяет получать содержательные результаты даже в тех случаях, когда основные вероятности не могут быть вычислены в явном виде.

Примеры. а) *Возвращение в начало.* Пусть для последовательности испытаний Бернулли $\mathscr{E}$ обозначает событие «суммарные числа успехов и неудач равны». Как и ранее, мы описываем исходы испытаний последовательностью $\mathbf{X}_1$, $\mathbf{X}_2$, ... взаимно независимых случайных величин, принимающих значения 1 и —1 с вероятностями p и q соответственно. Как обычно полагаем

$$\mathbf{S}_0=0, \quad \mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n. \tag{1.1}$$

Тогда $\mathbf{S}_n$ — это разность между числом успехов и числом неудач, и событие $\mathscr{E}$ *происходит тогда и только тогда, когда* $\mathbf{S}_n=0$. Совершенно ясно, что наступление этого события восстанавливает исходную ситуацию в том смысле, что последующие частные суммы $\mathbf{S}_{n+1}$, $\mathbf{S}_{n+2}$, ... образуют вероятностную копию всей последовательности $\mathbf{S}_1$, $\mathbf{S}_2$, ... (Продолжение см. в примере 4, б).)

б) *Возвращение в начало по отрицательным значениям.* Усложним последний пример: будем считать, что $\mathscr{E}$ происходит при n-м испытании, если

$$\mathbf{S}_n=0, \text{ но } \mathbf{S}_1<0, \ldots, \mathbf{S}_{n-1}<0. \tag{1.2}$$

Очевидно, что и в этом случае наступление события $\mathscr{E}$ означает, что мы начинаем все сначала. (Продолжение см. в примере 4, в).)

в) Другим вариантом примера а) является событие $\mathscr{E}$, состоящее в том, что суммарное число успехов в λ раз больше суммарного числа неудач (где $\lambda>0$ — произвольное, но фиксированное число). Если $\mathscr{E}$ происходит при n-м испытании, то оно происходит еще раз при $(n+m)$-м испытании только тогда, когда число успехов в испытаниях с номерами от $n+1$ до $n+m$ ровно в λ раз больше числа неудач. Промежутки между последовательными появлениями $\mathscr{E}$ являются поэтому независимыми и одинаково распределенными. В качестве частного случая можно рассмотреть событие, состоящее в том, что при $6n$ бросаниях идеальной кости единица появится ровно n раз. (Продолжение см. в задачах 4 и 5.)

г) *Лестничные величины.* Сохраняя обозначения примера а), мы введем новое повторяющееся событие $\mathscr{E}$, *которое происходит при n-м испытании, если* S_n *больше всех предыдущих сумм.* т. е. если

$$S_n > 0, \quad S_n > S_1, \; \ldots, \; S_n > S_{n-1}. \tag{1.3}$$

Если $\mathscr{E}$ происходит при n-м испытании, то процесс начинается заново в следующем смысле. При условии, что (1.3) выполнено, событие $\mathscr{E}$ происходит при $(n+m)$-м испытании тогда и только тогда, когда

$$S_{n+m} > S_n, \; \ldots, \; S_{n+m} > S_{n+m-1}. \tag{1.4}$$

Но разности $S_{n+k}-S_n$ — это частные суммы оставшейся последовательности $X_{n+1}, X_{n+2}, \ldots$, и поэтому новое появление $\mathscr{E}$ для этой оставшейся последовательности определяется точно так же, как событие $\mathscr{E}$ определялось для всей последовательности. Иначе говоря, при изучении события $\mathscr{E}$ все прошлое становится несущественным каждый раз, когда происходит $\mathscr{E}$. (Продолжение см. в примере 4, г).)

д) *Серии успехов в испытаниях Бернулли.* В предыдущих примерах определение события $\mathscr{E}$ не вызывало трудностей; теперь мы рассмотрим случай, когда применение теории рекуррентных событий становится возможным только при весьма осмотрительном выборе определения. Обычно слова «серия успехов длины r» используются для обозначения отрезка последовательности, состоящего ровно из r или не менее чем из r успехов. Ни одно из этих толкований не соответствует рекуррентному событию. В самом деле, если требуется наличие ровно r успехов, то успех в $(n+1)$-м испытании может разрушить серию, завершившуюся к n-му испытанию. С другой стороны, если требуется наличие не менее чем r успехов, то любая серия может продолжаться неограниченно, и ясно, что появление серии не восстанавливает исходного положения.

Классическая теория серий была довольно беспорядочной, и более систематический подход становится возможным, если определять серию длины r так, чтобы она стала рекуррентным событием. *Первая* серия длиной r определена однозначно, и мы условимся начинать наблюдения заново каждый раз, когда появляется серия. При таком соглашении последовательность *УУУ|УНУУУ|УУУ|Н* содержит три серии успехов длины 3 (появляющиеся при испытаниях с номерами 3, 8 и 11). Она содержит пять серий длины 2 (испытания с номерами 2, 4, 7, 9, 11). Формальное определение таково: *последовательность из n букв У и Н содержит столько У-серий длины r, сколько существует непересекающихся отрезков, состоящих в точности из r букв У каждый.* При таком соглашении мы говорим, что $\mathscr{E}$ происходит при n-м испытании, если к последовательности прибавляется новая серия длиной r. Это определяет рекуррентное событие и значительно упрощает теорию, не изменяя ее основных результатов. (Продолжение см. в § 7.)

е) *Продолжение: близкие события.* Очевидно, что рассуждения, проведенные в предыдущем примере, применимы к более общим событиям, например к появлению последовательности *УНУН*. Более интересным является то, что ограничение одним фиксированным набором вовсе не обязательно. Например, появление «двух успехов и трех неудач» определяет повторяющееся событие, как и появление «либо серии успехов длины r, либо серии неудач длины ρ». (Продолжение см. в § 8.)

ж) *Счетчики Гейгера.* Работу счетчиков, используемых при регистрации космических лучей и α-частиц, можно описать следующей упрощенной моделью[1]). Испытания Бернулли проводятся с постоянной скоростью. Счетчик предназначен для регистрации успехов, однако его механизм после каждой регистрации запирается ровно на $r-1$ испытаний. Иначе говоря, успех в n-м испытании регистрируется тогда и только тогда, когда в предыдущих $r-1$ испытаниях успехи не регистрировались. После этого счетчик запирается до тех пор, пока не закончатся испытания с номерами $n, \ldots, n+r-1$, и «освобождается» после $(n+r)$-го испытания, если оно приводит к неудаче. Последовательность на выходе счетчика представляет собой пример зависимых испытаний. За каждой регистрацией следует «мертвое время», однако в моменты, когда счетчик «свободен» (не заперт), состояния счетчика идентичны, и испытания начинаются заново. Обозначая буквой $\mathscr{E}$ событие «после испытания счетчик свободен», мы получаем типичное рекуррентное событие. (Продолжение см. в примере 4,д).)

з) *Простейшая модель очереди* строится по последовательности испытаний Бернулли и последовательности случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots$, принимающих только целые положительные значения. Величины $\mathbf{X}_k$ имеют одно и то же распределение $\{\beta_m\}$, взаимно независимы и не зависят от испытаний Бернулли. Успех в n-м испытании мы интерпретируем как появление в момент времени n покупателя (или вызова на телефонной станции). Случайная величина $\mathbf{X}_k$ равна времени обслуживания k-го покупателя. В любой момент времени продавец либо «свободен», либо «занят», и процесс развивается по следующим правилам. Вначале (в момент времени 0) продавец свободен. Если покупатель появляется, когда продавец свободен, то обслуживание начинается немедленно, и после появления этого покупателя продавец занят в течение времени его обслуживания. Покупатели, которые приходят, когда продавец занят, образуют очередь. Продавец обслуживает покупателей непрерывно до тех пор, пока очередь не рассосется.

Эти правила определяют процесс однозначно, и по заданным

---

[1]) Это дискретный аналог так называемых счетчиков первого типа. Счетчики второго типа описаны в задаче 8. [Более подробную информацию о теории счетчиков можно найти в книге Кокс Д., Смит В., Теория восстановления,— М.: Советское радио, 1967.— *Перев.*]

варианту $\{У, Н, У, У, У, Н, Н, \ldots\}$ процесса прибытия покупателей и варианту $\{3, 1, 17, 2, \ldots\}$ последовательных времен обслуживания нетрудно найти размер очереди в любой момент времени и время ожидания k-го покупателя. Таким образом, в принципе мы могли бы вычислить все интересующие нас вероятности, однако найти практичные методы вычислений нелегко. Ясно, однако, что каждый раз, когда продавец свободен, ситуация является в точности такой же, как в нулевой момент времени. Поэтому в нашей терминологии условие «продавец свободен» задает рекуррентное событие. Мы увидим, что само существование таких рекуррентных событий имеет важные последствия; например, оно означает, что распределеиия вероятностей размера очереди в момент n, времени ожидания n-го покупателя и аналогичных случайных величин стремятся к некоторым пределам, когда $n \to \infty$ (теорема 5.2). Иначе говоря, существование рекуррентных событий позволяет доказывать существование стационарного режима и изучать его основные свойства.

и) *Обслуживание станков.* Представление об области применимости метода рекуррентных событий может дать вариант предыдущего примера, в котором появления покупателей определяются уже не испытаниями Бернулли. Интерпретируем «покупателей» как одинаковые станки, подверженные случайным поломкам, а «продавца» — как ремонтного рабочего. Мы сохраним предположения о порядке обслуживания и образования очереди, но введем новый случайный механизм «появления покупателей», т. е. поломок. Предположим, что общее число станков равно N, и рассмотрим два крайних случая.

i) Предположим сначала, что если станок исправен, то с фиксированной вероятностью p он может сломаться в следующий целочисленный момент времени; когда станок ломается, он заменяется точно таким же новым станком, и время обслуживания — это время, затрачиваемое на установку нового станка. Мы считаем, что станки независимы и что их поломки определяются N независимыми последовательностями испытаний Бернулли. Заметим, что чем больше станков находится в очереди, тем меньше станков исправно, и следовательно, длина очереди в любой момент времени влияет на вероятность новых поломок (или заявок на ремонт). В этом состоит существенное отличие от предыдущего примера, однако условие «продавец свободен» тем не менее определяет рекуррентное событие, поскольку независимо от момента его наступления оно означает наличие одной и той же ситуации.

ii) Допустим теперь, что каждый ремонт обладает последействием, а именно увеличивает вероятность последующих поломок. Это означает, что станки постепенно изнашиваются, и поэтому после поломки хотя бы одного станка повторение благоприятной исходной ситуации оказывается невозможным. В этом случае рекуррентных событий, облегчающих анализ, ие существует.

§ 2. ОПРЕДЕЛЕНИЯ

Рассмотрим последовательность повторяющихся испытаний с возможными исходами E_j ($j=1, 2, \ldots$). Испытания не обязаны быть независимыми (наиболее интересными оказываются применения к цепям Маркова). Как обычно, мы предполагаем, что испытания могут продолжаться неограниченно и что вероятности $\mathbf{P}\{E_{j_1}, E_{j_2}, \ldots, E_{j_n}\}$ определяются однозначно для всех конечных наборов. Пусть $\mathscr{E}$ — некоторое свойство конечных последовательностей, т. е. мы предполагаем, что для любого набора $(E_{j_1}, \ldots, E_{j_n})$ можно сказать, обладает он свойством $\mathscr{E}$ или нет. Будем понимать выражение «$\mathscr{E}$ происходит на n-м месте в (конечной или бесконечной) последовательности $E_{j_1}, E_{j_2}, \ldots$» как синоним слов «подпоследовательность $E_{j_1}, E_{j_2}, \ldots, E_{j_n}$ обладает свойством $\mathscr{E}$». Эта договоренность означает, что появление $\mathscr{E}$ при n-м испытании зависит только от исходов первых n испытаний. Подразумевается также, что, *говоря о «рекуррентном событии $\mathscr{E}$», мы на самом деле имеем в виду класс событий*, определенных условием «$\mathscr{E}$ происходит». Ясно, что $\mathscr{E}$ является скорее набором символов, чем событием. Здесь мы несколько вольно обращаемся с языком, так же как, например, при использовании термина «двумерная задача»: задача сама по себе не имеет размерности.

Определение 1. *Свойство $\mathscr{E}$ определяет рекуррентное событие, если:*

а) *для того чтобы $\mathscr{E}$ происходило на n-м и $(n+m)$-м местах последовательности $(E_{j_1}, E_{j_2}, \ldots, E_{j_{n+m}})$, необходимо и достаточно, чтобы $\mathscr{E}$ происходило на последнем месте каждой из двух подпоследовательностей $(E_{j_1}, E_{j_2}, \ldots, E_{j_n})$ и $(E_{j_{n+1}}, E_{j_{n+2}}, \ldots, E_{j_{n+m}})$;*

б) *если $\mathscr{E}$ происходит на n-м месте, то*

$$\mathbf{P}\{E_{j_1}, \ldots, E_{j_{n+m}}\}=\mathbf{P}\{E_{j_1}, \ldots, E_{j_n}\}\,\mathbf{P}\{E_{j_{n+1}}, \ldots, E_{j_{n+m}}\}.$$

Теперь приобретают очевидный смысл утверждения о том, что $\mathscr{E}$ *впервые* происходит в последовательности $(E_{j_1}, E_{j_2}, \ldots)$ на n-м месте, и т. п. Ясно также, что с каждым рекуррентным событием $\mathscr{E}$ связаны две последовательности чисел, определенные для $n=1, 2, \ldots$ следующим образом:

$$\begin{aligned} u_n &= \mathbf{P}\{\mathscr{E} \text{ происходит при } n\text{-м испытании}\}, \\ f_n &= \mathbf{P}\{\mathscr{E} \text{ впервые происходит при } n\text{-м испытании}\}. \end{aligned} \tag{2.1}$$

Для дальнейшего удобно положить

$$f_0=0, \quad u_0=1 \tag{2.2}$$

и ввести производящие функции

$$F(s)=\sum_{k=1}^{\infty} f_k s^k, \quad U(s)=\sum_{k=0}^{\infty} u_k s^k. \tag{2.3}$$

Заметим, что $\{u_k\}$ не является распределением вероятностей; более того, в типичных случаях $\sum u_k = \infty$. С другой стороны, события «$\mathscr{E}$ впервые происходит при n-м испытании» несовместны, и поэтому

$$f = F(1) = \sum_{n=1}^{\infty} f_n \leqslant 1. \tag{2.4}$$

Ясно, что $1-f$ следует интерпретировать как *вероятность того, что $\mathscr{E}$ ни разу не происходит в продолженной до бесконечности последовательности испытаний*. Если $f=1$, то мы можем ввести случайную величину $\mathbf{T}$ с распределением

$$\mathbf{P}\{\mathbf{T}=n\} = f_n. \tag{2.5}$$

Обозначение (2.5) мы будем использовать и при $f<1$. В этой ситуации $\mathbf{T}$ *оказывается несобственной (или дефектной) случайной величиной, которая с вероятностью* $1-f$ *не принимает никакого численного значения*. (Нам будет удобно приписывать $\mathbf{T}$ значение ∞, и должно быть ясно, что при этом не потребуется вводить новые правила.)

Время ожидания события $\mathscr{E}$, т. е. число испытаний до первого появления $\mathscr{E}$ (включая это испытание), является случайной величиной с распределением (2.5); однако эта случайная величина естественно определяется только на пространстве бесконечных последовательностей $(E_{j_1}, E_{j_2}, \ldots)$.

По определению рекуррентных событий вероятность того, что $\mathscr{E}$ впервые произойдет при k-м испытании и *во второй раз* при n-м испытании, равна $f_k f_{n-k}$. Поэтому вероятность $f_n^{(2)}$ того, что $\mathscr{E}$ происходит во второй раз при n-м испытании, равна

$$f_n^{(2)} = f_1 f_{n-1} + f_2 f_{n-2} + \ldots + f_{n-1} f_1. \tag{2.6}$$

Правая часть—свертка $\{f_n\}$ с собой, и поэтому $\{f_n^{(2)}\}$ есть распределение вероятностей суммы двух независимых случайных величин, имеющих распределение (2.5). Вообще, если $f_n^{(r)}$—вероятность того, что r-е появление $\mathscr{E}$ происходит при n-м испытании, то

$$f_n^{(r)} = f_1 f_{n-1}^{(r-1)} + f_2 f_{n-2}^{(r-1)} + \ldots + f_{n-1} f_1^{(r-1)}. \tag{2.7}$$

Этот простой факт выражает следующая теорема.

Теорема. *Пусть $f_n^{(r)}$—вероятность того, что r-е появление $\mathscr{E}$ происходит при n-м испытании. Тогда $\{f_n^{(r)}\}$—распределение вероятностей суммы*

$$\mathbf{T}^{(r)} = \mathbf{T}_1 + \mathbf{T}_2 + \ldots + \mathbf{T}_r \tag{2.8}$$

r независимых случайных величин $\mathbf{T}_1, \ldots, \mathbf{T}_r$, имеющих распределение (2.5). *Иначе говоря, при фиксированном r последовательность $\{f_n^{(r)}\}$ имеет производящую функцию $F^r(s)$.*

Отсюда следует, в частности, что

$$\sum_{n=1}^{\infty} f_n^{(r)} = F^r(1) = f^r. \tag{2.9}$$

Иначе говоря, *вероятность того, что $\mathscr{E}$ произойдет по меньшей мере r раз, равна f^r* (это можно было предвидеть). Введем теперь следующее определение.

Определение 2. *Рекуррентное событие $\mathscr{E}$ будем называть возвратным*[1]), *если $f=1$, и невозвратным, если $f<1$.*

Для невозвратного события $\mathscr{E}$ вероятность f^r того, что оно произойдет по меньшей мере r раз, стремится к нулю, тогда как для возвратного $\mathscr{E}$ эта вероятность остается равной единице. Это можно описать словесно: *с вероятностью единица возвратное событие $\mathscr{E}$ происходит бесконечно много раз, а невозвратное событие $\mathscr{E}$ происходит лишь конечное число раз.* (Это утверждение является не только описательным, но и формально правильным при интерпретации в терминах пространства элементарных событий, состоящего из бесконечных последовательностей $E_{j_1}, E_{j_2}, \ldots$.)

Нам потребуется еще одно определение. При испытаниях Бернулли возвращение в начало (см. пример 1, а)) может произойти только при испытании с *четным* номером. В таком случае $f_{2n+1} = u_{2n+1} = 0$ и производящие функции $F(s)$ и $U(s)$ являются степенными рядами не по s, а по s^2. Аналогично, если в примере 1, в) число λ целое, то $\mathscr{E}$ может произойти при n-м испытании только при условии, что n делится на $\lambda+1$. В таких случаях мы будем называть $\mathscr{E}$ периодическим. По существу периодические рекуррентные события лишь обозначениями отличаются от непериодических, но в каждой теореме приходится специально упоминать об исключительном периодическом случае. Таким образом, периодические рекуррентные события доставляют много неприятностей, не внося ничего интересного.

Определение 3. *Рекуррентное событие $\mathscr{E}$ называется периодическим, если существует такое целое число $\lambda>1$, что $\mathscr{E}$ может происходить только при испытаниях с номерами $\lambda, 2\lambda, 3\lambda, \ldots$ (т. е. $u_n=0$, если n не делится на λ). Наибольшее λ, обладающее этим свойством, называется периодом $\mathscr{E}$.*

В заключение заметим, что на пространстве элементарных событий, состоящем из бесконечных последовательностей $E_{j_1}, E_{j_2}, \ldots$, число испытаний между $(r-1)$-м и r-м появлениями $\mathscr{E}$ является корректно определенной случайной величиной (может быть, несобственной) с тем же распределением, что $\mathbf{T}_r$. Иначе говоря, наши случайные величины $\mathbf{T}_r$ действительно соответствуют *про-*

[1]) В первом издании использовались термины «достоверное» (certain) и «недостоверное» (uncertain) событие; введенные здесь термины предпочтительнее в применениях к цепям Маркова,

межуткам времени между последовательными появлениями $\mathcal{E}$ *(временам возвращения).* Мы определили $\mathbf{T}_r$ аналитически, чтобы не использовать такие пространства элементарных событий, которые в этом томе не рассматриваются; будем надеяться, однако, что вероятностная подоплека показана во всей ее интуитивной простоте. Понятие рекуррентных событий предназначено для сведения весьма общей ситуации к суммам независимых случайных величин. Обратно, *любое распределение вероятностей* $\{f_n\}$, $n=1, 2, \ldots$, *можно использовать для определения рекуррентного события.* Это утверждение доказывает следующий пример.

Пример. *Самовосстанавливающиеся устройства.* Рассмотрим электрическую лампочку, плавкий предохранитель или иной элемент с конечным сроком службы. Как только этот элемент отказывает, он заменяется новым элементом того же вида, который рано или поздно заменяется третьим элементом, и т. д. Допустим, что срок службы — случайная величина, принимающая только значения, кратные единичному отрезку времени (году, дню или секунде). Тогда каждая единица времени представляет собой испытание с возможными исходами «замена» или «нет замены». Последовательные замены можно истолковывать как рекуррентные события. Если f_n — вероятность того, что новый элемент проработает ровно n единиц времени, то $\{f_n\}$ — распределение времени возвращения. Если срок службы с вероятностью 1 конечен, то $\sum f_n=1$ и рекуррентное событие является возвратным. Как правило, известно, что срок службы не может превосходить фиксированного числа m; в таком случае производящая функция $F(s)$ — многочлен степени не выше m. В приложениях нужно знать вероятность u_n того, что в момент n происходит замена. Значение u_n можно вычислить с помощью (3.1) (см. ниже). Мы получили здесь класс рекуррентных событий, определяемых произвольным распределением $\{f_n\}$. Случай $f<1$ не исключается: $1-f$ — это вероятность бесконечного срока службы нашего элемента. ▶

§ 3. ОСНОВНЫЕ СООТНОШЕНИЯ

Мы продолжаем пользоваться обозначениями (2.1)—(2.4) и собираемся изучить связь между $\{f_n\}$ и $\{u_n\}$. Вероятность того, что $\mathcal{E}$ происходит впервые при ν-м испытании, а затем снова при испытании с номером $n>\nu$, по определению равна $f_\nu u_{n-\nu}$. Вероятность того, что $\mathcal{E}$ происходит при n-м испытании впервые, есть $f_n=f_n u_0$. Поскольку эти случаи взаимно несовместны, мы имеем

$$u_n=f_1u_{n-1}+f_2u_{n-2}+\ldots+f_nu_0, \quad n\geqslant 1. \tag{3.1}$$

В выражении справа мы узнаем свертку $\{f_k\}*\{u_k\}$, производящая функция которой равна $F(s)\,U(s)$. В левой части мы обнаруживаем последовательность $\{u_n\}$ без члена u_0, так что ее производящая функция есть $U(s)-1$. Таким образом, $U(s)-1=F(s)U(s)$, и нами доказана следующая теорема.

Теорема 1. *Производящие функции* $\{u_n\}$ *и* $\{f_n\}$ *связаны соотношением*

$$U(s)=1/(1-F(s)). \tag{3.2}$$

Замечание. Правую часть (3.2) можно разложить в геометрический ряд $\sum F^r(s)$, сходящийся при $|s|<1$. Поскольку коэффициент $f_n^{(r)}$ при s^n в $F^r(s)$ равен вероятности r-го появления $\mathscr{E}$ при n-м испытании, соотношение (3.2) эквивалентно равенству

$$u_n=f_n^{(1)}+f_n^{(2)}+\ldots, \tag{3.3}$$

выражающему тот очевидный факт, что если $\mathscr{E}$ происходит при n-м испытании, то до момента n оно происходило $0, 1, 2, \ldots$ или $n-1$ раз. (Очевидно, $f_n^{(r)}=0$ при $r>n$.)

Теорема 2. *Для того чтобы событие* $\mathscr{E}$ *было невозвратным, необходимо и достаточно, чтобы величина*

$$u=\sum_{j=0}^{\infty} u_j \tag{3.4}$$

была конечной. В этом случае вероятность того, что $\mathscr{E}$ *происходит хотя бы раз, равна*

$$f=(u-1)/u. \tag{3.5}$$

Замечание. Мы можем интерпретировать u_j как математическое ожидание случайной величины, которая равна 1 или 0 в зависимости от того, происходит или не происходит $\mathscr{E}$ при j-м испытании. Следовательно, $u_1+u_2+\ldots+u_n$ есть *математическое ожидание числа появлений* $\mathscr{E}$ *в* n *испытаниях*, *а* $u-1$ можно интерпретировать как математическое ожидание числа появлений $\mathscr{E}$ в бесконечной последовательности испытаний.

Доказательство. Поскольку коэффициенты u_k неотрицательны, ясно, что $U(s)$ монотонно возрастает при $s\to 1$ и что при любом N

$$\sum_{n=0}^{N} u_n \leqslant \lim_{s\to 1} U(s) \leqslant \sum_{n=0}^{\infty} u_n = u.$$

Так как $U(s)\to(1-f)^{-1}$, когда $f<1$, и $U(s)\to\infty$, когда $f=1$, то теорема доказана. ▶

Следующая теорема чрезвычайно важна[1]). Ее доказательство

[1]) В частных случаях ее утверждение легко доказывается (см. задачу 1) и давно известно. Огромное количество работ было посвящено ослаблению ее условий, но все были уверены в том, что какие-то ограничения необходимы. В полной общности теорема 3 была доказана Эрдёшем, Феллером и Поллардом (Erdős P.,

элементарно, но мы отложим его до конца главы, поскольку оно не помогает понять вероятностную основу теоремы [1]).

Теорема 3. *Пусть событие $\mathscr{E}$ возвратно и непериодично; обозначим через μ математическое ожидание времени возвращения $\mathbf{T}_v$, т. е.*

$$\mu = \sum j f_j = F'(1) \tag{3.6}$$

(допускается случай $\mu = \infty$). Тогда

$$u_n \to \mu^{-1} \quad \text{при} \quad n \to \infty \tag{3.7}$$

($u_n \to 0$, если математическое ожидание времени возвращения бесконечно).

От условия непериодичности $\mathscr{E}$ легко избавиться. Действительно, если $\mathscr{E}$ имеет период λ, то ряд $\sum f_n s^n$ содержит только степени s^λ. Будем называть степенной ряд правильным, если это не так для любого целого $\lambda > 1$. Теорему 3 можно тогда переформулировать следующим образом: *если F — правильная вероятностная производящая функция и U определяется соотношением* (3.2), *то* $u_n \to 1/F'(1)$. Далее, если $\mathscr{E}$ имеет период λ, то $F(s^{1/\lambda})$ — правильная вероятностная производящая функция, и поэтому коэффициенты $U(s^{1/\lambda})$ сходятся к $\lambda/F'(1)$. Таким образом, нами доказана следующая теорема.

Теорема 4. *Если $\mathscr{E}$ возвратно и имеет период λ, то*

$$u_{n\lambda} \to \lambda/\mu \tag{3.8}$$

и $u_k = 0$ при любом k, не делящемся на λ.

§ 4. ПРИМЕРЫ

а) *Успехи в испытаниях Бернулли.* Рассмотрим банальный пример, когда $\mathscr{E}$ обозначает «успех» в последовательности испытаний Бернулли. Тогда $u_n = p$ при $n \geqslant 1$, откуда

$$U(s) = \frac{1-qs}{1-s}, \quad \text{и поэтому} \quad F(s) = \frac{ps}{1-qs} \tag{4.1}$$

Feller W., Pollard H., A property of power series with positive coefficients, Bull. Amer. Math. Soc., 55 (1949), 201—204). После выхода в свет первого издания Чжун Кайлай заметил, что теорему можно вывести из результатов Колмогорова, описывающих асимптотическое поведение цепей Маркова. Многие известные математики доказывали различные обобщения этой теоремы для разных классов распределений вероятностей. Эти исследования способствовали развитию методов современной теории вероятностей . В конце концов оказалось, что аналог теоремы 3 справедлив для любых распределений вероятностей. Элементарное (но не простое) доказательство см. в гл. XI, 9 тома 2.

[1]) Вероятностное доказательство этой теоремы (с явной оценкой скорости сходимости при $\mu < \infty$) можно найти в работе Калашников В. В. Равномерная оценка скорости сходимости в теореме восстановления для дискретного времени.— Теория вероятн. и ее примен., 1977, т. XXII, № 1, 399—403.— *Прим. перев.*

в силу (3.2). В этом частном случае теорема 2 попросту подтверждает очевидный факт: промежутки времени между последовательными успехами имеют геометрическое распределение с математическим ожиданием $1/p$.

б) *Возвращения в начало (пример* 1,а)). При k-м испытании суммарные числа успехов и неудач могут быть равны только тогда, когда $k=2n$ четно, и в этом случае вероятность возвращения в начало есть

$$u_{2n}=\binom{2n}{n}p^nq^n=\binom{-1/2}{n}(-4pq)^n. \tag{4.2}$$

Из разложения бинома (формула (8.7) гл. II) следует поэтому, что

$$U(s)=1/\sqrt{1-4pqs^2}, \tag{4.3}$$

а отсюда и из (3.2) — что

$$F(s)=1-\sqrt{1-4pqs^2}. \tag{4.4}$$

Повторное использование разложения бинома приводит к явному выражению для f_{2n}. (Явные выражения для u_{2n} и f_{2n} при $p=1/2$ выведены комбинаторными методами в гл. III, 2—3; производящие функции U и F найдены другими способами в гл. XI, 3. Следует отметить, что лишь предложенный выше метод не требует изобретательности.)

При $s=1$ квадратный корень в (4.4) равен $|p-q|$, и поэтому

$$f=1-|p-q|. \tag{4.5}$$

Таким образом, *возвращение в начало представляет собой рекуррентное событие с периодом* 2, *которое невозвратно при* $p\neq q$ *и возвратно в симметричном случае* $p=q$. Вероятность по меньшей мере r возвращений в начало равна f^r.

Когда $p=q=1/2$, время ожидания первого возвращения в начало является собственной[1]) случайной величиной, но $F'(1)=\infty$, и поэтому *среднее время возвращения* μ *бесконечно*. (Это следует также из теоремы 4 и из того, что $u_n\to 0$.) Бесконечность среднего времени возвращения означает, что случайные флуктуации в длительной реализации игры с бросанием монеты существенно отличаются от обычной модели, описываемой нормальным распределением. Весьма парадоксальный характер этих флуктуаций обсуждался в гл. III.

в) *Возвращение в начало по отрицательным значениям.* В примере 1, б) возвращение в начало удовлетворяло условию, состоящему в том, что ни одна из предшествующих частных сумм S_j не положительна. Распределение времени возвращения для этого ре-

[1]) Случайная величина называется *собственной* (proper), если она с вероятностью 1 принимает конечные значения (см. формулы (2.4), (2.5)).— *Прим. перев.*

куррентного события определяется равенством

$$f_{2n}^{-}=\mathbf{P}\{\mathbf{S}_{2n}=0,\ \mathbf{S}_1<0, \ldots, \mathbf{S}_{2n-1}<0\}, \tag{4.6}$$

и, конечно, $f_{2n-1}^{-}=0$. Кажется, что найти эти вероятности прямым рассуждением невозможно, однако их легко вычислить с помощью предыдущего примера. В самом деле, реализация последовательности $(\mathbf{X}_1, \ldots, \mathbf{X}_{2n})$, удовлетворяющая условию (4.6), содержит n положительных единиц и n отрицательных единиц и, значит, имеет ту же вероятность, что $(-\mathbf{X}_1, \ldots, -\mathbf{X}_{2n})$. Далее, *первое* возвращение в начало происходит либо по положительным, либо по отрицательным значениям, и мы заключаем, что эти две возможности имеют одинаковые вероятности. Следовательно, $f_{2n}^{-}=(1/2)f_{2n}$, где $\{f_n\}$ — распределение возвращений в начало, найденное в предыдущем примере. Поэтому производящая функция наших времен возвращения задается формулой

$$F^{-}(s)=1/2-(1/2)\sqrt{1-4pqs^2}, \tag{4.7}$$

и, значит,

$$U^{-}(s)=\frac{2}{1+\sqrt{1-4pqs^2}}=\frac{1-\sqrt{1-4pqs^2}}{2pqs^2}. \tag{4.8}$$

Событие $\mathscr{E}$ невозвратно, вероятность его наступления равна $1/2-(1/2)\,|p-q|$.

г) *Лестничные величины.* Первая положительная частная сумма может появиться при k-м испытании только тогда, когда $k=2n+1$ нечетно. Соответствующие вероятности мы запишем в виде

$$\varphi_{2n+1}=\mathbf{P}\{\mathbf{S}_1<0, \ldots, \mathbf{S}_{2n}=0,\ \mathbf{S}_{2n+1}=1\}. \tag{4.9}$$

Таким образом, $\{\varphi_k\}$ — распределение рекуррентного события из примера 1, г). Далее, из условия (4.9) следует, что $\mathbf{X}_{2n+1}=1$ и что при $2n$-м испытании происходит рекуррентное событие из предыдущего примера. Значит, $\varphi_{2n+1}=p\cdot u_{2n}^{-}$. Поэтому в очевидных обозначениях

$$\Phi(s)=psU^{-}(s)=(1-\sqrt{1-4pqs^2})/(2qs). \tag{4.10}$$

Это производящая функция времени первого достижения, найденная в гл. XI (формула (3.6)). Явное выражение для φ_{2n+1} можно вывести из (4.10) с помощью разложения бинома (гл. II, формула (8.7)). Оно совпадает с выражением для φ_{2n+1}, найденным комбинаторными методами в теореме 2 гл. III, 7.

д) *Счетчики Гейгера.* В примере 1, ж) счетчик остается свободным, если в момент времени 1 не происходит регистрации. В противном случае он запирается и вновь «освобождается» в момент времени $r+1$, если в этот момент не появляется частица; счетчик «освобождается» в момент времени $2r+1$, если частица появляется в момент времени $r+1$, но не появляется в момент

$2r+1$, и т. д. Производящей функцией времени возвращения является поэтому

$$qs+qps^{r+1}+qp^2s^{2r+1}+\ldots=qs/(1-ps^r). \tag{4.11}$$

(См. также задачи 7—9.)

е) *Простейшая задача теории очередей* (пример 1, з)). В этом случае продавец остается свободным, если в момент времени 1 не появляется покупатель. Если же покупатель появляется, то начинается так называемый «период занятости», который заканчивается в момент, когда продавец впервые освобождается. Производящая функция $\rho(s)$ периода занятости выводилась в примере гл. XII, 5, в) при помощи методов теории ветвящихся процессов. Нетрудно вывести, что в нашем случае производящая функция времени возвращения имеет вид $qs+ps\rho(s)$, что согласуется с формулой (5.7) гл. XII.

ж) *Ничьи в играх с бросанием нескольких монет.* В заключение приведем простой пример, демонстрирующий возможность получить некоторые выводы, не зная явного вида производящих функций. Пусть $r\geqslant 2$—произвольное целое число; рассмотрим последовательность одновременных независимых бросаний r монет. Пусть $\mathscr{E}$ обозначает рекуррентное событие, состоящее в том, что *все r монет находятся в одной и той же фазе* (т. е. что суммарные числа выпадения гербов одинаковы для всех r монет). Вероятность наступления этого события при n-м испытании есть

$$u_n=2^{-rn}\left[\binom{n}{0}^r+\binom{n}{1}^r+\ldots+\binom{n}{n}^r\right]. \tag{4.12}$$

В правой части мы узнаем члены биномиального распределения с $p=1/2$; применяя к последнему нормальное приближение, легко заключаем[1]), что при любом фиксированном r и $n\longrightarrow\infty$

$$u_n\sim[2/(\pi n)]^{r/2}\sum_j e^{-2rj^2/n} \tag{4.13}$$

---

[1]) Нормальное приближение $2^{-n}\binom{n}{k}\sim\left(\frac{2}{\pi n}\right)^{1/2}e^{-2(k-n/2)^2/n}$ справедливо, когда $n\longrightarrow\infty$ и $(k-n/2)^3/n^2\longrightarrow 0$ (см. гл. VII, 2). Из этого приближения следует, например, что

$$\sum_{k:\ (k-n/2)^{12}\leqslant n^7}2^{-rn}\binom{n}{k}^r\sim\sum_{j:\ j^{12}\leqslant n^7}\left(\frac{2}{\pi n}\right)^{r/2}e^{-2rj^2/n},\qquad n\longrightarrow\infty.$$

Для доказательства (4.13) остается воспользоваться соотношением (4.14) и соотношениями

$$\sum_{k:\ (k-n/2)^{12}>n^7}2^{-rn}\binom{n}{k}^r<n2^{-rn}\binom{n}{n/2+[n^{7/12}]}^r\sim n\left(\frac{2}{\pi n}\right)^{r/2}e^{-2n^{1/6}},$$

$$2\sqrt{\frac{r}{n}}\sum_{j:\ j^{12}>n^7}e^{-2rj^2/n}<2\int_{-\infty}^{-n^{7/12}+1}e^{-x^2/2}\,dx\longrightarrow 0,\qquad n\longrightarrow\infty.$$

— *Прим перев.*

(суммирование проводится по всем целым числам от $-n/2$ до $n/2$). Но по определению интеграла

$$2\sqrt{r/n}\sum_j e^{-xrj^2/n} \longrightarrow \int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}, \tag{4.14}$$

и отсюда мы выводим, что

$$u_n \sim (1/\sqrt{r})\,[2/(\pi n)]^{(r-1)/2}. \tag{4.15}$$

Следовательно, $\sum u_n$ расходится, когда $r \leqslant 3$, но сходится, когда $r \geqslant 4$. Значит, *$\mathscr{E}$ возвратно при $r \leqslant 3$ и невозвратно при $r \geqslant 4$*. Поскольку $u_n \to 0$, при $r \leqslant 3$ среднее время возвращения бесконечно. (Ср. с задачами 2 и 3.) ▶

§ 5. РЕКУРРЕНТНЫЕ СОБЫТИЯ С ЗАПАЗДЫВАНИЕМ. ОБЩАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Введем теперь незначительное обобщение понятия рекуррентного события, столь очевидное, что о нем не стоило бы и упоминать, если бы не было удобно иметь для него специальный термин и перечень основных уравнений.

Возможно, наилучшее неформальное описание рекуррентных событий с запаздыванием получится, если сказать, что они относятся к испытаниям, в которых мы «пропускаем начало и начинаем с середины». Время ожидания *первого* появления $\mathscr{E}$ имеет распределение $\{b_n\}$, отличающееся от распределения $\{f_n\}$ времени между последовательными появлениями $\mathscr{E}$. Теория переносится на этот случай без изменений, за исключением того, что испытания, следующие за каждым появлением $\mathscr{E}$, являются точными вероятностными копиями одной и той же последовательности случайных величин, которая не совпадает с исходной последовательностью.

Поскольку ситуация столь проста, мы воздержимся от формальностей и согласимся говорить о *рекуррентном событии $\mathscr{E}$ с запаздыванием, когда определение рекуррентных событий применимо лишь при игнорировании испытаний, предшествующих первому появлению $\mathscr{E}$; подразумевается, что время ожидания первого появления $\mathscr{E}$ есть случайная величина, не зависящая от последующих времен возвращения, но ее распределение $\{b_n\}$ может отличаться от общего распределения $\{f_n\}$ времен возвращения.*

Обозначим через v_n *вероятность появления $\mathscr{E}$ при n-м испытании.* При выводе выражения для v_n мы рассуждаем следующим образом. Допустим, что $\mathscr{E}$ происходит при испытании с номером $k<n$. По отношению к последующим испытаниям $\mathscr{E}$ оказывается обычным рекуррентным событием, и поэтому (условная) вероятность повторного появления $\mathscr{E}$ при n-м испытании равна u_{n-k}. Далее, если $\mathscr{E}$ происходит при n-м испытании, то либо это первое его появление, либо первое появление имело место при k-м испы-

тании с некоторым $k<n$. Суммируя по всем вариантам, получаем

$$v_n=b_n+b_{n-1}u_1+b_{n-2}u_2+\ldots+b_1u_{n-1}+b_0u_n. \qquad (5.1)$$

Таким образом, у нас есть возможность найти явное выражение для v_n. (Другое доказательство приводится в примере 10, а).) Соотношение (5.1) можно переписать в компактном виде уравнения свертки:

$$\{v_n\}=\{b_n\}*\{u_n\}. \qquad (5.2)$$

Это означает, что соответствующие производящие функции удовлетворяют тождеству

$$V(s)=B(s)\,U(s)=B(s)/(1-F(s)). \qquad (5.3)$$

Пример. а) В испытаниях Бернулли, рассмотренных в примерах 4, а)—4, г), событие $S_n=1$ является рекуррентным событием с запаздыванием. Время ожидания его первого появления имеет производящую функцию Φ, заданную формулой (4.10); промежутки времени между последовательными появлениями события $\{S_n=1\}$ имеют производящую функцию F возвращений в начало (см. (4.4)). Таким образом, в этом случае $V=\Phi/(1-F)$.

Легко показать, что асимптотическое поведение вероятностей v_n по существу такое же, как у u_n. Во избежание тривиальных оговорок предположим, что $\mathscr{E}$ непериодично [1]. Согласно § 3, в этом случае u_n стремится к конечному пределу и $\sum u_n<\infty$ тогда и только тогда, когда $\mathscr{E}$ невозвратно.

Теорема 1. *Если* $u_n\to\omega$, *то*

$$v_n\to b\omega, \quad \textit{где } b=\sum b_k=B(1). \qquad (5.4)$$

Если $\sum u_n=u<\infty$, *то*

$$\sum v_n=bu. \qquad (5.5)$$

В частности, $v_n\to\mu^{-1}$, *если* $\mathscr{E}$ *возвратно.*

Доказательство. Положим $r_k=b_{k+1}+b_{k+2}+\ldots$. Поскольку $u_n\leqslant 1$, из (5.1) следует, что при $n>k$

$$b_0u_n+\ldots+b_ku_{n-k}\leqslant v_n\leqslant b_0u_n+\ldots+b_ku_{n-k}+r_k. \qquad (5.6)$$

Выберем k столь большим, чтобы $r_k<\varepsilon$. Тогда для достаточно больших n левая часть (5.6) больше $b\omega-2\varepsilon$, а правая часть меньше $b\omega+2\varepsilon$. Тем самым (5.4) доказано. Утверждение (5.5) можно доказывать, либо суммируя (5.1) по n, либо полагая $s=1$ в (5.3). ►

Мы переходим теперь к предельной теореме, имеющей широкую область приложений. Допустим, что некоторая система имеет счет-

[1] Периодические рекуррентные события рассматриваются в теореме 2 § 10. Другое доказательство теоремы 1 приводится в примере 10, а).

ное множество возможных состояний E_0, $E_1, \ldots$ и что переходы из одного состояния в другое зависят от какого-то случайного механизма. Например, в простейшей модели очереди (пример 1,з)) мы говорим, что система находится в состоянии E_k, если в очереди находится k покупателей (вместе с покупателем, которого обслуживают). Чтобы задать состояние системы, включающей семнадцать продавцов, может потребоваться восемнадцать чисел, однако все мыслимые состояния могут быть упорядочены в последовательность E_0, $E_1, \ldots$. Нам не нужно думать о том, как это сделать наилучшим образом, потому что следующая теорема не дает конкретных методов оценки вероятностей. Она является чистой теоремой существования, показывающей, что стационарный режим существует в большинстве встречающихся на практике случаев. Это представляет теоретический интерес, но имеет также и практическое значение, поскольку, как правило, математическое исследование стационарного режима значительно проще, чем изучение зависящего от времени процесса.

Мы предполагаем, что для $n=1, 2, \ldots$ и для любого набора $(r_1, \ldots, r_n)$ определена вероятность того, что система в моменты времени $0, 1, \ldots, n-1$ проходит через состояния $E_{r_1}, \ldots, E_{r_n}$. Мы не делаем никаких предположений ни о взаимной зависимости этих событий, ни о вероятностях переходов из одного состояния в другое. Для простоты мы рассмотрим только *вероятности $p_n^{(r)}$ того, что в момент времени n система находится в состоянии* E_r. (Будет ясно, как теорема обобщается на пары состояний, тройки состояний и т. д.) Основное предположение состоит в том, что существует связанное с нашим процессом рекуррентное событие $\mathscr{E}$. Например, в модели очереди (пример 1, з)) таким рекуррентным событием является попадание в состояние E_0. Если бы в этом случае $\mathscr{E}$ было невозвратным, то была бы положительной вероятность того, что очередь никогда не окончится. Это значило бы, что раньше или позже мы столкнулись бы с бесконечной очередью, т. е. с очередью неограниченно возрастающего размера. Наша предельная теорема показывает, что такие системы на практике невозможны. Пример с очередью должен был пояснить роль условия возвратности события $\mathscr{E}$. (Условие непериодичности вводится только для того, чтобы избежать тривиальных оговорок.)

Теорема 2. *Допустим, что существует связанное с нашим процессом непериодическое возвратное рекуррентное событие $\mathscr{E}$ (возможно, с запаздыванием). Тогда при $n\to\infty$*

$$p_n^{(r)} \to p^{(r)}, \tag{5.7}$$

причем

$$\sum p^{(r)} = 1, \tag{5.8}$$

если среднее время возвращения μ конечно, и $p^r = 0$ в противном случае [1]).

Доказательство. Каждый раз, когда происходит $\mathscr{E}$, процесс начинается заново. Поэтому существует корректно определенная условная вероятность $g_n^{(r)}$ того, что если $\mathscr{E}$ происходит в некоторый момент времени, то состояние E_r появляется через n единиц времени, и это происходит *до* следующего наступления $\mathscr{E}$ (здесь $n = 0, 1, \ldots$). Для рекуррентных событий с запаздыванием нам потребуется также вероятность $\gamma_n^{(r)}$ того, что E_r появляется в момент времени n *до* первого наступления $\mathscr{E}$. (Очевидно, $\gamma_n^{(r)} = g_n^{(r)}$, если $\mathscr{E}$ — рекуррентное событие без запаздывания.)

Классифицируем теперь способы появления E_r в момент n по моментам последнего появления $\mathscr{E}$ до момента n. Прежде всего, $\mathscr{E}$ может еще не произойти. Вероятность этого равна $\gamma_n^{(r)}$. В противном случае существует такое $k \leqslant n$, что $\mathscr{E}$ происходит в момент времени k и не происходит между k и n. Вероятность этого равна $v_k g_{n-k}^{(r)}$. Суммируя по всем этим взаимно исключающим случаям, находим

$$p_n^{(r)} = \gamma_n^{(r)} + g_{n-1}^{(r)} v_1 + g_{n-2}^{(r)} v_2 + \ldots + g_0^{(r)} v_n. \tag{5.9}$$

(Мы пользуемся здесь обозначениями теоремы 1. Для событий с запаздыванием $v_0 = 0$, для событий без запаздывания $v_k = u_k$ и $\gamma_n^{(r)} = g_n^{(r)}$.)

Соотношение (5.9) аналогично (5.1) и отличается лишь наличием члена $\gamma_n^{(r)}$ в правой части. Очевидно, эта величина не больше вероятности того, что $\mathscr{E}$ не произойдет до момента n, и, поскольку $\mathscr{E}$ возвратно, $\gamma_n^{(r)} \to 0$ при $n \to \infty$. К остальным слагаемым мы можем применить теорему 1, заменив u_k на v_k и b_k на $g_k^{(r)}$. Так как $\mathscr{E}$ возвратно, то $v_n \to \mu^{-1}$, и поэтому

$$p_n^{(r)} \to \mu^{-1} \sum_{k=0}^{\infty} g_k^{(r)}. \tag{5.10}$$

Это доказывает существование пределов (5.7). Чтобы доказать, что их сумма равна 1, заметим, что в любой момент времени система находится в каком-то состоянии, и, значит,

$$\sum_{r=0}^{\infty} g_n^{(r)} = g_n \tag{5.11}$$

есть вероятность того, что время возвращения не меньше n, т. е.

$$g_n = f_n + f_{n+1} + \ldots$$

---

[1]) Последнее утверждение теоремы, вообще говоря, неверно (кстати, автор его и не доказывает). Контрпримером является последовательность испытаний Бернулли с $p = q = 1/2$ (процесс с двумя состояниями) и рекуррентное событие $\mathscr{E}$, состоящее в равенстве суммарных чисел успехов и неудач (см. пример 4, б)). — *Прим. перев.*

Таким образом,

$$\sum_{r=0}^{\infty} p^{(r)} = \mu^{-1} \sum_{n=0}^{\infty} g_n = 1 \tag{5.12}$$

в силу формулы (1.8) гл. XI. ►

(Предельную теорему из примера 10, б) можно рассматривать как частный случай только что доказанной теоремы.)

§ 6. ЧИСЛО ПОЯВЛЕНИЙ $\mathscr{E}$

До сих пор мы изучали свойства рекуррентного события $\mathscr{E}$, относящиеся к промежуткам времени между его последовательными появлениями. Часто оказывается предпочтительнее считать заданным число n испытаний, а в качестве основной случайной величины выбрать *число* $\mathbf{N}_n$ *появлений* $\mathscr{E}$ в первых n испытаниях. Мы исследуем теперь асимптотическое поведение распределения $\mathbf{N}_n$ при больших n. Для простоты будем считать, что $\mathscr{E}$ — рекуррентное событие без запаздывания.

Пусть, как и в (2.8), $\mathbf{T}^{(r)}$ обозначает число испытаний до r-го появления $\mathscr{E}$, включая это последнее испытание. Распределения вероятностей $\mathbf{T}^{(r)}$ и $\mathbf{N}_n$ связаны очевидным тождеством

$$\mathbf{P}\{\mathbf{N}_n \geqslant r\} = \mathbf{P}\{\mathbf{T}^{(r)} \leqslant n\}. \tag{6.1}$$

Начнем с простого случая, когда $\mathscr{E}$ возвратно и распределение $\{f_n\}$ его времени возвращения имеет конечные среднее μ и дисперсию σ^2. Так как $\mathbf{T}^{(r)}$ является суммой r независимых величин, то, согласно центральной предельной теореме из гл. X,1, для любого фиксированного x при $r \to \infty$

$$\mathbf{P}\{(\mathbf{T}^{(r)} - r\mu)/(\sigma\sqrt{r}) < x\} \to \mathfrak{N}(x), \tag{6.2}$$

где $\mathfrak{N}(x)$ — функция нормального распределения. Устремим теперь n и r к бесконечности так, чтобы

$$(n - r\mu)/(\sigma\sqrt{r}) \to x; \tag{6.3}$$

тогда из (6.1) и (6.2) будет следовать, что

$$\mathbf{P}\{\mathbf{N}_n \geqslant r\} \to \mathfrak{N}(x). \tag{6.4}$$

Чтобы записать это соотношение в более привычном виде, мы введем *нормированную случайную величину*

$$\mathbf{N}_n^* = (\mu\mathbf{N}_n - n)\sqrt{\mu/(\sigma^2 n)}. \tag{6.5}$$

Неравенство $\mathbf{N}_n \geqslant r$ эквивалентно неравенству

$$\mathbf{N}_n^* \geqslant [(r\mu - n)/(\sigma\sqrt{r})]\sqrt{r\mu/n} = -x\sqrt{r\mu/n}. \tag{6.6}$$

Разделив (6.3) на r, получим, что $n/r \to \mu$, и поэтому правая часть (6.6) стремится к $-x$. Так как $\mathfrak{N}(-x) = 1 - \mathfrak{N}(x)$, то

$$\mathbf{P}\{\mathbf{N}_n^* \geqslant -x\} \to \mathfrak{N}(x), \quad \textit{или} \quad \mathbf{P}\{\mathbf{N}_n^* < -x\} \to 1 - \mathfrak{N}(x), \qquad (6.7)$$

и нами доказана следующая теорема.

Теорема. (Нормальное приближение.) *Если рекуррентное событие $\mathscr{E}$ возвратно и его время возвращения имеет конечные среднее μ и дисперсию σ^2, то число $\mathbf{T}^{(r)}$ испытаний до r-го появления $\mathscr{E}$ и число $\mathbf{N}_n$ появлений $\mathscr{E}$ в первых n испытаниях имеют асимптотически нормальные распределения, указанные в* (6.2) *и* (6.7).

Заметим, что в (6.7) центральная предельная теорема применяется к последовательности *зависимых* случайных величин $\mathbf{N}_n$. Полезность нашей теоремы будет продемонстрирована в следующем параграфе на примере применений к сериям успехов.

Соотношения (6.7) делают правдоподобными формулы

$$\mathbf{E}(\mathbf{N}_n) \sim n/\mu, \quad \operatorname{Var}(\mathbf{N}_n) \sim n\sigma^2/\mu^3, \qquad (6.8)$$

где символ $\sim$ означает, что отношение левой части к правой стремится к 1. Чтобы доказать (6.8), мы заметим, что $\mathbf{N}_n$ есть сумма n таких (зависимых) случайных величин $\mathbf{Y}_k$, что $\mathbf{Y}_k$ равна единице или нулю в зависимости от того, происходит или не происходит $\mathscr{E}$ в k-м испытании. Тогда $\mathbf{E}(\mathbf{Y}_k) = u_k$ и

$$\mathbf{E}(\mathbf{N}_n) = u_1 + u_2 + \ldots + u_n. \qquad (6.9)$$

Поскольку $u_n \to \mu^{-1}$, отсюда следует первое соотношение (6.8). Второе доказывается аналогичными рассуждениями (см. задачу 20).

К сожалению, удивительно много времен возвращения, встречающихся в различных случайных процессах и в приложениях, имеют *бесконечное математическое ожидание*. В таких случаях нормальное приближение заменяется более общими предельными теоремами совершенно другого характера [1]) и случайные флуктуации обладают неожиданными свойствами. Например, интуиция подсказывает, что $\mathbf{E}(\mathbf{N}_n)$ должно расти линейно по n, «поскольку в среднем при удвоении числа испытаний число появлений $\mathscr{E}$ тоже должно удвоиться». Однако *это не так*. Если математическое ожидание времени возвращения бесконечно, то $u_n \to 0$, и поэтому $\mathbf{E}(\mathbf{N}_n)/n \to 0$ в силу (6.9). Это означает, что в длинной последовательности испытаний появления $\mathscr{E}$ становятся все реже и реже, и это возможно только за счет того, что некоторые времена возвращения фантастически велики. Следующие два примера показывают, насколько резко может быть выражено это явление.

Примеры. а) Если $\mathscr{E}$ — возвращение в начало в игре с бросанием монеты (пример 4, б) с $p = 1/2$), то $u_{2n} \sim 1/\sqrt{\pi n}$ и (6.9) яв-

[1]) Feller W., Fluctuation theory of recurrent events, Trans. Amer. Math. Soc., 67 (1949), 98—119.

ляется приближением к интегралу от $1/\sqrt{\pi x}$; отсюда следует, что $\mathbf{E}(\mathbf{N}_{2n}) \sim 2\sqrt{n/\pi}$. Поэтому *среднее* время возвращения в интервале от 0 до n растет примерно как $\sqrt{n}$. Любопытные следствия этого подробно обсуждались в гл. III.

б) Возвращаясь к примеру 4, ж), рассмотрим повторяющиеся бросания $r=3$ монет и обозначим через $\mathscr{E}$ событие, состоящее в том, что все три монеты находятся в одной и той же фазе. Мы видели, что $\mathscr{E}$ — возвратное рекуррентное событие и что $u_n \sim 2/(\sqrt{3}\pi n)$. Поэтому $\mathbf{E}(\mathbf{N}_n)$ *растет примерно как* $\log n$, и, значит, *среднее* времен возвращения до момента n имеет фантастическую величину порядка $n/\log n$. ▶

§ 7*). ПРИЛОЖЕНИЯ К ТЕОРИИ СЕРИЙ УСПЕХОВ

В дальнейшем r будет обозначать фиксированное положительное целое число, а $\mathscr{E}$ — появление серии из r успехов в последовательности испытаний Бернулли. Важно, чтобы длина серии определялась так же, как в примере 1, д), поскольку в противном случае серии не являются рекуррентными событиями и вычисления усложняются. Как в (2.1) и (2.2), u_n — *это вероятность появления* $\mathscr{E}$ *при* n*-м испытании, а* f_n — *вероятность того, что первая серия длиной* r *появляется при* n*-м испытании.*

Вероятность того, что r испытаний с номерами $n, n-1, n-2, \ldots, n-r+1$ закончатся успехами, очевидно, равна p^r. Тогда $\mathscr{E}$ происходит при одном из этих r испытаний; вероятность того, что $\mathscr{E}$ произойдет при испытании с номером $n-k$ ($k=0, 1, \ldots, r-1$), а следующие k испытаний приведут к k успехам, равна $u_{n-k}p^k$. Поскольку эти r возможностей исключают друг друга, мы получаем рекуррентное соотношение [1])

$$u_n + u_{n-1}p + \ldots + u_{n-r+1}p^{r-1} = p^r, \tag{7.1}$$

справедливое при $n \geqslant r$. Очевидно,

$$u_1 = u_2 = \ldots = u_{r-1} = 0, \quad u_0 = 1. \tag{7.2}$$

Умножая (7.1) на s^n и суммируя по $n = r, r+1, r+2, \ldots$, получаем в левой части

$$\{U(s) - 1\}(1 + ps + p^2s^2 + \ldots + p^{r-1}s^{r-1}), \tag{7.3}$$

а в правой $p^r(s^r + s^{r+1} + \ldots)$. Эти два ряда являются геометрическими, и мы находим, что

$$\{U(s) - 1\}(1 - (ps)^r)/(1 - ps) = p^r s^r/(1 - s), \tag{7.4}$$

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

[1]) Классический подход состоит в выводе рекуррентного соотношения для f_n. Этот метод сложнее и не применим, например, к сериям произвольного вида или к событиям *УУННУУ*, к которым наш метод применим без изменений (ср. с примером 8, в)).

или

$$U(s) = (1 - s + qp^r s^{r+1})/[(1-s)(1-p^r s^r)]. \tag{7.5}$$

Используя (3.2), находим теперь *производящую функцию времени возвращения*:

$$F(s) = \frac{p^r s^r (1-ps)}{1-s+qp^r s^{r+1}} = \frac{p^r s^r}{1-qs(1+ps+\ldots+p^{r-1}s^{r-1})}. \tag{7.6}$$

Равенство $F(1)=1$ показывает, что в удлиняющейся последовательности испытаний число серий любой длины с вероятностью 1 увеличивается неограниченно. Среднее время возвращения μ можно было бы получить непосредственно из (7.1), поскольку мы знаем, что $u_n \to \mu^{-1}$. Так как нам потребуется и дисперсия, удобнее вычислить производные $F(s)$. Лучше всего при этом дифференцировать, предварительно умножив обе части (7.6) на знаменатель. Несложные вычисления показывают, что *среднее и дисперсия времен возвращения для серий длиной r равны*

$$\mu = \frac{1-p^r}{qp^r}, \quad \sigma^2 = \frac{1}{(qp^r)^2} - \frac{2r+1}{qp^r} - \frac{p}{q^2} \tag{7.7}$$

соответственно. По теореме предыдущего параграфа при больших n *число* $\mathbf{N}_n$ *серий длиной r в n испытаниях имеет распределение, близкое к нормальному*, т. е. при фиксированных $\alpha < \beta$ вероятность того, что

$$n/\mu + \alpha\sigma\sqrt{n/\mu^3} < \mathbf{N}_n < n/\mu + \beta\sigma\sqrt{n/\mu^3}, \tag{7.8}$$

стремится к $\mathfrak{N}(\beta) - \mathfrak{N}(\alpha)$. Впервые это было доказано Мизесом при помощи весьма утомительных вычислений. В табл. 1 приведено несколько типичных значений математических ожиданий времен возвращения.

Таблица 1

Математические ожидания времен возвращения для серий успехов, когда производится по одному испытанию в секунду

Длина серии r	$p=0,6$	$p=0,5$ (монета)	$p=1/6$ (кость)
5	30,7 с	1 мин	2,6 ч
10	6,9 мин	34,1 мин	28,0 мес
15	1,5 ч	18,2 ч	18 098 лет
20	19 ч	24,3 сут	140,7 млн. лет

Метод разложения на простые дроби из гл. XI, 4 позволяет получить превосходные приближенные формулы. Второе представление в (7.6) показывает, что знаменатель имеет единственный *положительный корень* $s=x$. Для любого действительного или комплексного числа s с $|s|\leqslant x$ имеем

$$|qs(1+ps+\ldots+p^{r-1}s^{r-1})|\leqslant qx(1+px+\ldots+p^{r-1}x^{r-1})=1, \quad (7.9)$$

где знак равенства возможен только тогда, когда аргументы всех слагаемых в левой части одинаковы, т. е. когда $s=x$. Следовательно, x по абсолютной величине меньше любого другого корня знаменателя в (7.6). Поэтому мы можем применить формулы (4.5) и (4.9) гл. XI, положив в них $s_1=x$, $U(s)=p^rs^r(1-ps)$ и $V(s)=1-s+qp^r\times\times s^{r+1}$. Используя равенство $V(x)=0$, находим

$$f_n\sim\frac{(x-1)(1-px)}{(r+1-rx)q}\frac{1}{x^{n+1}}. \quad (7.10)$$

Вероятность отсутствия серий в n испытаниях есть $q_n=f_{n+1}+f_{n+2}+f_{n+3}+\ldots$, и, суммируя геометрическую прогрессию в (7.10), получаем

$$q_n\sim\frac{1-px}{(r+1-rx)q}\frac{1}{x^{n+1}}. \quad (7.11)$$

Таким образом, мы доказали, что *вероятность отсутствия серий успехов длиной r в n испытаниях удовлетворяет соотношению* (7.11). Табл. 2 показывает, что правая часть (7.11) дает удивительно хорошие приближения даже для очень малых n, и точность аппроксимации быстро увеличивается с ростом n. Это иллюстрирует эффективность метода производящих функций и разложения на простые дроби.

Таблица 2

Вероятность отсутствия серий успехов длиной $r=2$ в n испытаниях с $p=1/2$

n	q_n точное	q_n по (7.11)	Погрешность
2	0,75	0,76631	0,0163
3	0,625	0,61996	0,0050
4	0,500	0,50156	0,0016
5	0,40625	0,40577	0,0005

Численные оценки. Имея в виду интересы читателя-прикладника, мы покажем, что вычисления, связанные с разложением на простейшие дроби, часто оказываются проще, чем может показаться с первого взгляда, и что можно получить хорошие оценки для погрешности.

В связи с асимптотической формулой (7.11) возникают две задачи: во-первых, оценить вклад $r-1$ отброшенных корней и, во-вторых, получить оценки для главного корня x.

Первое представление в (7.6) показывает, что все корни знаменателя функции $F(s)$ удовлетворяют уравнению

$$s=1+qp^rs^{r+1}, \quad (7.12)$$

однако (7.12) имеет посторонний корень $s=p^{-1}$. При положительных s график функции $f(s)=1+qp^rs^{r+1}$ выпукл вниз; он пересекает биссектрису $y=s$ в x и в p^{-1}, а между x и p^{-1} лежит *ниже* биссектрисы. Далее, $f'(p^{-1})=(r+1)q$. Если эта величина больше 1, то график $f(s)$ пересекает биссектрису при $s=p^{-1}$ снизу, и, следовательно, $p^{-1}>x$. Мы будем предполагать для определенности, что

$$(r+1)q>1; \tag{7.13}$$

в этом случае $x<p^{-1}$ и $f(s)<s$ при $x<s<p^{-1}$. Значит, для всех таких комплексных чисел s, что $x<|s|<p^{-1}$, мы имеем $|f(s)|\leqslant f(|s|)<|s|$, так что нули s_k не могут лежать в кольце $x<|s|<p^{-1}$. Поскольку x выбирался как корень, наименьший по абсолютной величине, это означает, что

$$|s_k|\geqslant p^{-1}, \quad \text{когда } s_k\neq x. \tag{7.14}$$

Дифференцирование (7.12) показывает теперь, что все корни являются простыми.

Вклад в q_n каждого корня аналогичен вкладу (7.11) главного корня x, и поэтому $r-1$ членов, не учтенных в (7.11), имеют вид

$$A_k=\frac{ps_k-1}{rs_k-(r+1)}\,\frac{1}{qs_k^{n+1}}. \tag{7.15}$$

Нам потребуется оценка сверху для первой дроби в правой части. Чтобы получить ее, заметим, что для фиксированного $s>p^{-1}>(r+1)r^{-1}$

$$\left|\frac{pse^{i\theta}-1}{rse^{i\theta}-(r+1)}\right|\leqslant\frac{ps+1}{rs+r+1}; \tag{7.16}$$

действительно, величина в левой части, очевидно, принимает максимальное и минимальное значения при $\theta=0$ и $\theta=\pi$, и непосредственная подстановка показывает, что $\theta=0$ соответствует минимуму, а $\theta=\pi$ — максимуму. Тогда в силу (7.13) и (7.14)

$$|A_k|<\frac{2p^{n+1}}{(r+1+rp^{-1})\,q}<\frac{2p^{n+2}}{rq\,(1+p)}. \tag{7.17}$$

Мы приходим к выводу, что *погрешность в (7.11), возникшая из-за отбрасывания $r-1$ корней, отличных от x, по абсолютной величине меньше чем*

$$2(r-1)\,p^{n+2}/[rq\,(1+p)]. \tag{7.18}$$

Корень x легко вычислить при помощи (7.12) методом последовательных приближений, полагая $x_0=1$ и $x_{\nu+1}=f(x_\nu)$. Последовательность сходится к x монотонно, и каждый ее член дает оценку снизу для x, а любое значение s, для которого $s>f(s)$, дает оценку сверху. Нетрудно проверить, что

$$x=1+qp^r+(r+1)\,(qp^r)^2+\ldots\,. \tag{7.19}$$

§ 8*). СОБЫТИЯ БОЛЕЕ ОБЩЕГО ВИДА

Наш метод применим к более общим задачам, которые ранее казались значительно сложнее задач об обычных сериях.

Примеры. а) *Серии двух типов.* Пусть $\mathscr{E}$ означает событие *«либо серия успехов длины r, либо серия неудач длины ρ»* (см. пример 1, е)). Мы имеем дело с двумя рекуррентными событиями $\mathscr{E}_1$ и $\mathscr{E}_2$, где $\mathscr{E}_1$ — «серия успехов длины r», $\mathscr{E}_2$ — «серия неудач длины ρ», а $\mathscr{E}$ — событие «либо $\mathscr{E}_1$, либо $\mathscr{E}_2$». Событию $\mathscr{E}_1$ соответ-

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

ствует производящая функция (7.5), которая будет теперь обозначаться $U_1(s)$. Соответствующая производящая функция $U_2(s)$ для $\mathscr{E}_2$ получается из (7.5) перестановкой p и q и заменой r на ρ. Вероятность u_n того, что событие $\mathscr{E}$ произойдет при n-м испытании, есть сумма соответствующих вероятностей для $\mathscr{E}_1$ и $\mathscr{E}_2$, если исключить случай $u_0 = 1$. Следовательно,

$$U(s) = U_1(s) + U_2(s) - 1. \tag{8.1}$$

Производящая функция $F(s)$ времени возвращения для $\mathscr{E}$ есть снова $F(s) = 1 - U^{-1}(s)$, или

$$F(s) = \frac{(1-ps)\,p^r s^r\,(1-q^\rho s^\rho) + (1-qs)\,q^\rho s^\rho\,(1-p^r s^r)}{1-s+qp^r s^{r+1}+pq^\rho s^{\rho+1}-p^r q^\rho s^{r+\rho}}. \tag{8.2}$$

Математическое ожидание времени возвращения находится дифференцированием:

$$\mu = (1-p^r)(1-q^\rho)/(qp^r + pq^\rho - p^r q^\rho). \tag{8.3}$$

При $\rho \to \infty$ это выражение стремится к математическому ожиданию времени возвращения для серии успехов, которое вычисляется по формуле (7.7).

б) В гл. VIII, 1 мы нашли вероятность x того, что *серия успехов длины r появится раньше серии неудач длины* ρ. Определим рекуррентные события $\mathscr{E}_1$ и $\mathscr{E}_2$ так же, как в примере а). Пусть x_n — вероятность того, что $\mathscr{E}_1$ впервые происходит при n-м испытании и до этого ни разу не происходило $\mathscr{E}_2$, а f_n — вероятность того, что $\mathscr{E}_1$ впервые происходит при n-м испытании (без дополнительных условий на $\mathscr{E}_2$). Определим y_n и g_n так же, как x_n и f_n соответственно, но с перестановкой $\mathscr{E}_1$ и $\mathscr{E}_2$.

Производящую функцию для f_n дает формула (7.6), а $G(s)$ получается при перестановке p и q и замене r на ρ. Для x_n и y_n мы имеем очевидные рекуррентные соотношения

$$\begin{aligned} x_n &= f_n - (y_1 f_{n-1} + y_2 f_{n-2} + \ldots + y_{n-1} f_1), \\ y_n &= g_n - (x_1 g_{n-1} + x_2 g_{n-2} + \ldots + x_{n-1} g_1). \end{aligned} \tag{8.4}$$

Они имеют вид сверток, и поэтому для соответствующих производящих функций мы имеем

$$\begin{aligned} X(s) &= F(s) - Y(s)\,F(s), \\ Y(s) &= G(s) - X(s)\,G(s). \end{aligned} \tag{8.5}$$

Из этих двух линейных уравнений получаем

$$X(s) = \frac{F(s)\,(1-G(s))}{1-F(s)\,G(s)}, \quad Y(s) = \frac{G(s)\,(1-F(s))}{1-F(s)\,G(s)}. \tag{8.6}$$

Выражения для x_n и y_n можно найти, снова пользуясь методом разложения на простейшие дроби. При $s = 1$ имеем $X(1) = \sum x_n = x$, т. е. вероятности появления $\mathscr{E}_1$ раньше $\mathscr{E}_2$. Как числитель, так и знаменатель обращаются в нуль, и значение $X(1)$ вычисляется по правилу Лопиталя дифференцированием числителя и знаменателя: $X(1)=G'(1)/(F'(1)+G'(1))$. Используя равенства $F'(1)=(1-p^r)/(qp^r)$ и $G'(1)=(1-q^\rho)/(pq^\rho)$ (см. 7.7)), мы получаем $X(1)$ в том же виде, что и в формуле (1.3) гл. VIII.

в) Рассмотрим рекуррентное событие *УУННУУ*. Повторяя рассуждения § 7, легко находим, что

$$p^4q^2=u_n+p^2q^2u_{n-4}+p^3q^2u_{n-5}. \tag{8.7}$$

Поскольку известно, что $u_n\to\mu^{-1}$, мы получаем для среднего времени возвращения равенство $\mu=p^{-4}q^{-2}+p^{-2}+p^{-1}$. При $p=q=1/2$ находим, что $\mu=70$, тогда как среднее время возвращения для серии успехов длины 6 равно 126. Это показывает, что вопреки ожиданиям *при бросаниях монеты имеется существенное различие между сериями гербов и другими наборами той же длины.* ▶

§ 9. ОТСУТСТВИЕ ПАМЯТИ ДЛЯ ВРЕМЕН ОЖИДАНИЯ С ГЕОМЕТРИЧЕСКИМ РАСПРЕДЕЛЕНИЕМ

Геометрическое распределение времени ожидания обладает интересным и важным свойством, отличающим его от других распределений. Рассмотрим последовательность испытаний Бернулли и обозначим через $\mathbf{T}$ число испытаний до первого успеха (включая это испытание). Тогда $\mathbf{P}\{\mathbf{T}>k\}=q^k$. Пусть нам известно, что в первых m испытаниях успехов не было; время ожидания $\mathbf{T}^{(m)}$ от m-й неудачи до первого успеха имеет то же самое распределение $\{q^k\}$ и не зависит от числа предшествовавших неудач. Иначе говоря, вероятность того, что ожидание успеха продлится еще по крайней мере k единиц времени, всегда равно исходной вероятности того, что полное время ожидания превосходит k.

Если время жизни атома или срок службы элемента устройства имеет геометрическое распределение, то *старение* отсутствует: в любой момент своей жизни атом имеет одну и ту же вероятность распада при следующем испытании. Радиоактивные атомы на самом деле обладают этим свойством (однако в случае непрерывного времени роль геометрического распределения играет показательное).

Обратно, если известно, что некоторое явление характеризуется полным отсутствием памяти или старения, то распределение вероятностей его длительности должно быть геометрическим или показательным. Типичным примером является хорошо известный вид телефонных разговоров, часто рассматриваемый как модель бессвязности и полной зависимости от сиюминутных импульсов: окончание разговора — это мгновенное случайное событие, не связанное с предшествующей беседой. Напротив, если мы знаем, что в последние пять минут трамвая не было, то это увеличивает нашу надежду на его скорое появление. При бросании монеты вероятность того, что после второго испытания суммарные числа успехов и неудач будут равны, составляет 1/2. Однако если эти числа различны, то вероятность их совпадения после двух дополнительных испытаний равна 1/4. Это примеры последействия.

Для строгой формулировки утверждения предположим, что время ожидания T принимает значения 0, 1, 2, . . . с вероятностями p_0, p_1, p_2, . . . Пусть распределение T обладает следующим свойством: *условная вероятность окончания времени ожидания при k-м испытании (при условии, что оно не окончилось ранее) равна p_0 (вероятности окончания при первом испытании). Мы утверждаем, что тогда $p_k=(1-p_0)^k p_0$, так что* T *имеет геометрическое распределение.*

Для доказательства снова введем «хвосты»

$$q_k=p_{k+1}+p_{k+2}+p_{k+3}+\ldots=\mathbf{P}\{\mathrm{T}>k\}.$$

Наше условие состоит в том, что $\mathrm{T}>k-1$, а его вероятность есть q_{k-1}. Условная вероятность события $\mathrm{T}=k$ равна поэтому p_k/q_{k-1}, и по предположению при любом $k\geqslant 1$

$$p_k/q_{k-1}=p_0. \tag{9.1}$$

Так как $p_k=q_{k-1}-q_k$, то (9.1) приводится к виду

$$q_k/q_{k-1}=1-p_0. \tag{9.2}$$

Поскольку $q_0=p_1+p_2+\ldots=1-p_0$, отсюда следует, что $q_k=(1-p_0)^{k+1}$, и поэтому $p_k=q_{k-1}-q_k=(1-p_0)^k p_0$, как и утверждалось. ▶

В теории случайных процессов описанное выше отсутствие памяти связывается с *марковским свойством*; мы вернемся к этому вопросу в гл. XV, 13.

§ 10. ТЕОРИЯ ВОССТАНОВЛЕНИЯ

Область применения уравнений свертки, которые служат основой теории рекуррентных событий, значительно шире, чем это могло показаться при чтении предыдущих параграфов. Поэтому мы сформулируем заново и в более общем виде аналитическую часть полученных в них результатов и опишем как типичные вероятностные методы теории восстановления, так и их применения к изучению разнообразных ситуаций.

Мы начинаем с двух произвольных [1]) последовательностей действительных чисел f_1, f_2, . . . и b_0, b_1, . . . Новую последователь-

[1]) Мы полагаем $f_0=0$. Из (10.1) ясно, что случай $0<f_0<1$ приводит лишь к изменению обозначений, т. е. к замене f_k на $f_k/(1-f_0)$ и b_k на $b_k/(1-f_0)$.

ность $v_0, v_1, \ldots$ можно задать уравнениями свертки

$$v_n = b_n + f_1 v_{n-1} + f_2 v_{n-2} + \ldots + f_n v_0. \tag{10.1}$$

Они последовательно определяют $v_0, v_1, v_2, \ldots$, и поэтому v_n определяется однозначно в любом случае. Мы рассматриваем, однако, только последовательности, удовлетворяющие условиям [1])

$$f_n \geqslant 0, \quad f = \sum_{n=1}^{\infty} f_n < \infty; \quad b_n \geqslant 0, \quad b = \sum_{n=0}^{\infty} b_n < \infty. \tag{10.2}$$

В этом случае v_n неотрицательны и соответствующие производящие функции должны удовлетворять равенству

$$V(s) = B(s)/(1 - F(s)). \tag{10.3}$$

Производящие функции F и B сходятся по меньшей мере при $0 \leqslant s < 1$, и поэтому (10.3) определяет степенной ряд, который сходится, если $F(s) < 1$. Соотношения (10.1) и (10.3) полностью эквивалентны. В § 3 мы рассматривали частный случай, когда $B(s) = 1$ ($v_n = u_n$ при любом n). В § 5 рассмотрена общая ситуация при условии $f \leqslant 1$. Чтобы обеспечить возможность применения к теории популяций, мы теперь будем допускать, что $f > 1$; оказывается, этот случай легко сводится к стандартному случаю $f = 1$.

Мы будем говорить, что *последовательность* $\{f_n\}$ *имеет период* $\lambda > 1$, если $f_n = 0$, когда n не делится на λ, и λ — максимальное целое число с таким свойством. Это эквивалентно следующему утверждению: $F(s) = F_1(s^\lambda)$ — степенной ряд по s^λ, но не по $s^{r\lambda}$ при любом $r > 1$. Снова положим

$$\mu = \sum n f_n \leqslant \infty \tag{10.4}$$

и будем считать, что μ^{-1} нужно считать равным нулю, если $\mu = \infty$.

Теорема 1. (Теорема восстановления.) *Пусть выполняются условия* (10.2) *и* $\{f_n\}$ *непериодична. Тогда*

(i) *если* $f < 1$, *то* $v_n \to 0$ *и*

$$\sum_{n=0}^{\infty} v_n = b/(1 - f); \tag{10.5}$$

(ii) *если* $f = 1$, *то*

$$v_n \to b\mu^{-1}; \tag{10.6}$$

(iii) *если* $f > 1$, *то существует единственный положительный корень уравнения* $F(\xi) = 1$ *и*

$$\xi^n v_n \to B(\xi)/(\xi F'(\xi)). \tag{10.7}$$

---

[1]) Неотрицательность f_n существенна, а условие сходимости двух рядов введено только для удобства. Никаких общих выводов нельзя сделать, если $b = \infty$ и $f = \infty$. Соотношение (10.7) остается справедливым при $f = \infty$ (за исключением того, что в этом случае $F'(\xi)$ может быть равно бесконечности) и оказывается бессмысленным, если $b = \infty$ и $F'(\xi) = \infty$.

Очевидно, $\xi < 1$, и поэтому производная $F'(\xi)$ конечна; (10.7) показывает, что последовательность $\{v_n\}$ асимптотически ведет себя как геометрическая прогрессия со знаменателем $\xi^{-1} > 1$.

Доказательство. Утверждения (i) и (ii) были доказаны в § 5. Для доказательства (iii) достаточно применить (ii) к последовательностям $\{f_n\xi^n\}$, $\{b_n\xi^n\}$ и $\{v_n\xi^n\}$ с производящими функциями $F(\xi s)$, $B(\xi s)$ и $V(\xi s)$ соответственно. ►

Мы исключили периодические последовательности $\{f_n\}$, поскольку они не представляют особого интереса. Фактически они не дают ничего нового. В самом деле, если $\{b_n\}$ и $\{f_n\}$ имеют один и тот же период λ, то $B(s)$ и $F(s)$ — степенные ряды по s^λ, и поэтому то же верно для $V(s)$. К последовательностям $\{f_{n\lambda}\}$, $\{b_{n\lambda}\}$ и $\{v_{n\lambda}\}$ с производящими функциями $F(s^{1/\lambda})$, $B(s^{1/\lambda})$ и $V(s^{1/\lambda})$ применима теорема 1. В случае $F(1)=1$ отсюда следует, что $v_{n\lambda} \longrightarrow b\lambda/\mu$. Далее, произвольный степенной ряд B можно представить в виде линейной комбинации λ степенных рядов B_j, содержащих только степени s^λ:

$$B(s) = B_0(s) + sB_1(s) + \ldots + s^{\lambda-1}B_{\lambda-1}(s). \tag{10.8}$$

Подстановка этого разложения в (10.3) и применение только что сформулированного утверждения показывают, что справедлива следующая теорема.

Теорема 2. *Пусть выполняются условия* (10.2) *и* $\{f_n\}$ *имеет период* $\lambda > 1$. *Тогда*

(i) *если* $f < 1$, *то справедливо* (10.5);

(ii) *если* $f = 1$, *то для* $j = 0, 1, \ldots, \lambda-1$ *при* $n \longrightarrow \infty$

$$u_{n\lambda+j} \longrightarrow \lambda B_j\, 1)/\mu; \tag{10.9}$$

(iii) *если* $f > 1$, *то для* $j = 0, 1, \ldots, \lambda-1$ *при* $n \longrightarrow \infty$

$$\xi^{n\lambda} u_{n\lambda+j} \longrightarrow \lambda B_j(\xi)/(\xi\mu). \tag{10.10}$$

С помощью рассуждений, которые аналогичны проведенным для рекуррентных событий, можно показать, что некоторые вероятности, связанные с самыми разнообразными случайными процессами, удовлетворяют уравнению типа свертки (10.1). При этом многие важные предельные теоремы оказываются простыми следствиями теоремы 1. Такой подход, известный под названием *теории восстановления* [1]), теперь повсеместно вытеснил громоздкие старые методы. Полностью его преимущества раскрываются при применении к процессам с непрерывным временем, тем не менее первые два примера могут служить иллюстрацией. Другие примеры приводятся в задачах 8 и 9. Применение теоремы 1 к невероятностным предельным теоремам содержится в примере в). Последние два примера связаны с практическими приложениями.

Примеры. а) *Рекуррентные события с запаздыванием.* Выведем новым способом утверждение § 5 о рекуррентном событии $\mathscr{E}$ с запаздыванием в случае, когда $\{f_j\}$ — распределение времени

[1]) В оригинале renewal arguments. Многие авторы используют также термин «метод введения регенерирующего события». — *Прим. перев.*

возвращения, а $\{b_j\}$ — распределение момента первого появления $\mathscr{E}$. Пусть v_n означает вероятность того, что $\mathscr{E}$ происходит при n-м испытании. Покажем, что выполняется (10.1). Существуют две возможности появления $\mathscr{E}$ при n-м испытании. Появление $\mathscr{E}$ может быть первым, и вероятность этого равна b_n. В противном случае существует последнее появление $\mathscr{E}$ до n-го испытания, и поэтому существует такое число j $(1 \leqslant j < n)$, что $\mathscr{E}$ происходило при j-м испытании, а в *следующий* раз — при n-м. Вероятность этого равна $v_j f_{n-j}$. Указанные возможности исключают друг друга, и, значит,

$$v_n = b_n + v_1 f_{n-1} + v_2 f_{n-2} + \ldots + v_{n-1} f_1, \tag{10.11}$$

что совпадает с (10.1). Значит, производящая функция V вычисляется по формуле (10.3), которая согласуется с результатами § 5. (Хотя результаты совпадают по форме, ход рассуждений был разным: в § 5 перечисление происходило в соответствии с первым появлением $\mathscr{E}$, а сейчас использовалось последнее появление. Обе процедуры применяются и в других задачах и иногда приводят к формально различным уравнениям.)

б) *Вероятности попадания.* Рассмотрим последовательность испытаний с обычным (без запаздывания) возвратным рекуррентным событием $\mathscr{E}$. Пусть $\nu \geqslant 0$ — целое число. Предположим, что мы начинаем наблюдать процесс только после ν-го испытания и что нас интересует время до следующего появления $\mathscr{E}$. Более формально, для $r=1, 2, \ldots$ определим $w_\nu(r)$ как вероятность того, что *первое появление $\mathscr{E}$ после ν-го испытания* происходит при $(\nu+r)$-м испытании. Таким образом, $w_0(r)=f_r$ и $w_\nu(0)=0$. (Величины $w_\nu(r)$ называются вероятностями попадания в соответствии с их интерпретацией в теории случайных блужданий. При других обстоятельствах естественнее говорить о распределении остаточного времени ожидания, начинающегося с ν-го испытания. Ср. с примером гл. XV, 2, л).)

Чтобы найти эти вероятности, мы используем стандартный метод теории восстановления следующим образом. Событие $\mathscr{E}$ может произойти в самый первый раз при $(\nu+r)$-м испытании. Вероятность этого равна $f_{\nu+r}$. В противном случае существует такое число $k \leqslant \nu$, что $\mathscr{E}$ происходит впервые при k-м испытании. Поведение процесса после k-го испытания является вероятностной копией поведения всего процесса, но ν-е испытание становится теперь $(\nu-k)$-м. Вероятность нашего события равна поэтому $f_k w_{\nu-k}(r)$, и, значит, при любом $r>0$

$$w_\nu(r) = f_{\nu+r} + \sum_{k=1}^{\nu} f_k w_{\nu-k}(r). \tag{10.12}$$

Это уравнение имеет стандартный вид (10.1) с $b_n = f_{n+r}$. Нас интересует не производящая функция, а асимптотическое поведение вероятностей попадания при очень больших ν. Найти его позволяет теорема 1. Положим

$$\rho_k = f_{k+1} + f_{k+2} + \dots \tag{10.13}$$

и напомним (см. формулу (1.8) гл. XI), что математическое ожидание времени возвращения удовлетворяет равенству

$$\mu = \rho_1 + \rho_2 + \dots . \tag{10.14}$$

Если $\mathscr{E}$ непериодично, то из теоремы 1 мы выводим, что при $\nu \to \infty$

$$w_\nu(r) \to \begin{cases} \rho_r/\mu, & \textit{если } \mu < \infty, \\ 0, & \textit{если } \mu = \infty. \end{cases} \tag{10.15}$$

Этот результат чрезвычайно интересен. В случае когда среднее время возвращения конечно, (10.14) означает, что $\{\rho_r/\mu\}$ — распределение вероятностей, и, значит, мы имеем предельную теорему обычного типа. Однако *при* $\mu = \infty$ *вероятность того, что время ожидания превысит любое данное число* r, *стремится к* 1. Иначе говоря, наши времена ожидания ведут себя значительно хуже, чем времена возвращения. Это неожиданное явление имеет важные следствия, которые детально обсуждаются в томе 2. (См. также задачу 10.)

в) *Повторное осреднение.* Следующая задача имеет аналитический характер и рассматривалась в различных контекстах значительно более сложными методами. Допустим, что $f_1 + \dots + f_r = 1$ и $f_j \geqslant 0$. Для любых r чисел $v_1, \dots, v_r$ назовем *взвешенным средним* величину $f_1 v_r + \dots + f_r v_1$. Определим теперь бесконечную последовательность $v_1, v_2, \dots$, которая начинается с заданных r чисел, полагая v_n равным взвешенному среднему предыдущих r членов. Иначе говоря, для $n > r$

$$v_n = f_1 v_{n-1} + \dots + f_r v_{n-r}. \tag{10.16}$$

Так как последовательность $f_1, f_2, \dots$ обрывается на r-м члене, эти уравнения имеют вид (10.1). Определим теперь b_k так, чтобы (10.1) выполнялось при всех n. Для этого нужно положить $b_0 = v_0 = 0$ и

$$b_k = v_k - f_1 v_{k-1} + \dots - f_{k-1} v_1, \quad k \leqslant r. \tag{10.17}$$

(При $k > r$ по определению $b_k = 0$.) Без каких-либо вычислений из теоремы 1 следует, что при таком повторном осреднении v_n *стремится к конечному пределу*. Чтобы вычислить этот предел, мы должны найти $b = b_1 + \dots + b_r$. Используя обозначение (10.13) для «хвостов» $\sum f_k$, нетрудно вывести из (10.17) и (10.6), что

$$v_n \to (v_1 \rho_{r-1} + \dots + v_r \rho_0)/(f_1 + 2f_2 + \dots + rf_r). \tag{10.18}$$

Например, если $r = 3$ и вычисляются *средние арифметические*, то $f_1 = f_2 = f_3 = 1/3$ и

$$v_n \to (1/6)\ (v_1 + 2v_2 + 3v_3). \tag{10.19}$$

Легкость, с которой мы получили этот результат, не должна скрывать того факта, что задача является трудной, если ее рассматривать вне настоящего контекста. (Другой подход можно найти в задаче 15 гл. XV,14.)

г) *Самовосстанавливающиеся устройства.* Вернемся к примеру из § 2, в котором элемент устройства, устанавливаемый в момент n, имеет срок службы с распределением $\{f_k\}$. Когда срок службы оканчивается, элемент немедленно заменяется новым элементом того же типа; таким образом, последовательные замены образуют возвратное рекуррентное событие в последовательности зависимых испытаний (исходами которых являются решения о том, нужно или нет производить замену).

Допустим теперь, что элемент, установленный в момент времени 0, имеет возраст k, а не является новым. Это влияет только на время ожидания, т. е. $\mathscr{E}$ оказывается рекуррентным событием *с запаздыванием.* Чтобы получить распределение $\{b_n\}$ первого времени ожидания, заметим, что b_n есть (условное) математическое ожидание того, что элемент откажет в возрасте $n+k$ при условии, что он уже достиг возраста k. Поэтому при $k\geqslant 1$

$$b_n = f_{n+k}/r_k, \quad \text{где } r_k = f_{k+1} + f_{k+2} + \ldots \,. \tag{10.20}$$

На практике нас интересует обычно не отдельный элемент, а все устройство (например, сеть уличных фонарей в городе). Предположим поэтому, что *исходное устройство* (в момент времени 0) *состоит из N элементов, из которых β_k имеют возраст k* (причем $\sum\beta_k = N$). Каждый элемент порождает последовательность потомков, которым может потребоваться замена в момент n. *Математическое ожидание v_n числа всех замен в момент времени n, очевидно, удовлетворяет основному уравнению* (10.1) *с*

$$b_n = \sum \beta_k f_{n+k}/r_k. \tag{10.21}$$

Мы имеем здесь первый пример, в котором v_n — *математическое ожидание*, а не вероятность; нам известно только, что $v_n \leqslant N$.

Несложные вычисления показывают, что $b = \sum b_n = N$, и поэтому, согласно теореме 1, величины $v_n \to N/\mu$, если замены непериодичны. Это означает существование *устойчивого предельного распределения возраста.* Действительно, для того чтобы элемент в момент времени n имел возраст k, необходимо и достаточно, чтобы он был установлен в момент $n-k$ и работал до возраста k. Поэтому математическое ожидание числа таких элементов равно $v_{n-k}r_k$ и стремится к Nr_k/μ при $n\to\infty$. Иначе говоря, с течением времени *доля элементов устройства, имеющих возраст k, стремится к r_k/μ.* Таким образом, *предельное распределение возраста не зависит от начального распределения возраста* и определяется только вероятностями f_n. Аналогичное утверждение справедливо и при значительно более общих условиях. Численный пример приводится в табл. 3. При этом обнаруживается примечательный факт: стремление к пределу не является монотонным. (См. также задачи 16—18.)

Таблица 3

Изменение распределения возраста в устройстве, описанном в примере «г»

$k \backslash n$	0	1	2	3	4	5	6	7	∞
0	500	397	411,4	412	423,8	414,3	417,0	416,0	416,7
1	320	400	317,6	329,1	329,6	339,0	331,5	333,6	333,3
2	74	148	185	146,9	152,2	152,4	156,8	153,3	154,2
3	100	40	80	100	79,4	82,3	82,4	84,8	83,3
4	6	15	6	12	15	11,9	12,3	12,4	12,5

Столбцы дают распределение возраста в устройстве, состоящем из 1000 элементов, в моменты времени $n=0, 1, \ldots, 7$, а также предельное распределение. Предполагается, что распределение срока службы определяется вероятностями

$$f_1=0{,}20,\ f_2=0{,}43,\ f_3=0{,}17,\ f_4=0{,}17,\ f_5=0{,}03,$$

так что ни один элемент не может иметь возраст 5. Уравнение $1-F(s)=0$ имеет корни 1, $-5/3$, -5 и $\pm 2i$. Среднее время возвращения равно 2,40.

д) *Человеческие популяции.* В качестве примера, где $f=\sum f_n>1$, мы используем простейшую модель человеческой популяции. Она аналогична модели из предыдущего примера, однако размер популяции является теперь переменным, и рождения женщин играют роль замен. Новой особенностью является то, что женщина может иметь любое число дочерей, и поэтому ее потомство женского пола может исчезнуть, но может стать и многочисленным. Обозначим через f_n вероятность того, что новорожденная девочка доживет до возраста n и в этот момент у нее появится дочка. (Зависимостью от числа предыдущих детей и их возрастов мы пренебрегаем.) Тогда $f=\sum f_n$ — математическое ожидание числа дочерей, и поэтому в здоровой популяции $f>1$. Теорема 1 утверждает, что тогда размер популяции растет с примерно постоянной скоростью ξ и что распределение возраста в популяции стремится к пределу, как это описано в предыдущем примере [1]). Модель, по общему признанию, является грубой, но тем не менее представляет некоторый практический интерес: любопытную особенность предельного поведения ξ, конечно, нельзя было предугадать без соответствующего математического анализа.

[1]) Однако предельное распределение возраста вычисляется сложнее. Формулы для предельного распределения возраста частиц в ветвящихся процессах можно найти в книге Харрис Т. Е. Теория ветвящихся случайных процессов.— М.: Мир, 1966.— *Прим. перев.*

§ 11 *). ДОКАЗАТЕЛЬСТВО ОСНОВНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЫ

В § 3 мы опустили доказательство теоремы 3, которую мы теперь сформулируем следующим образом. *Пусть $f_1, f_2, \dots$ — такая последовательность неотрицательных чисел, что $\sum f_n = 1$ и 1 есть наибольший общий делитель тех n, для которых $f_n > 0$. Пусть $u_0 = 1$ и*

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \dots + f_n u_0, \quad n \geqslant 1. \tag{11.1}$$

Тогда

$$u_n \to \mu^{-1}, \quad \text{где} \quad \mu = \sum_{n=1}^{\infty} n f_n \tag{11.2}$$

(*μ^{-1} полагается равным нулю, когда $\mu = \infty$*).

Чтобы не прерывать изложение, мы предпошлем доказательству три известные леммы, которые часто используются вне рамок теории вероятностей.

Пусть A — множество всех целых чисел n, для которых $f_n > 0$; обозначим через A^+ множество всех возможных линейных комбинаций

$$p_1 a_1 + \dots + p_r a_r \tag{11.3}$$

чисел $a_1, \dots, a_r$ из A (коэффициенты p_j — положительные целые числа).

Лемма 1. *Существует такое число N, что A^+ содержит все целые числа $n > N$.*

Доказательство. Как было известно еще Евклиду, равенство единице наибольшего общего делителя чисел из A означает, что можно выбрать числа $a_1, \dots, a_r$ из A и (не обязательно положительные) целые числа c_j так, что

$$c_1 a_1 + \dots + c_r a_r = 1. \tag{11.4}$$

Положим $s = a_1 + \dots + a_r$. *Любое* целое число n может быть единственным образом представлено в виде $n = xs + y$, где x и y — целые и $0 \leqslant y < s$. Поэтому

$$n = \sum_{k=1}^{r} (x + c_k y)\, a_k, \tag{11.5}$$

и все коэффициенты положительны, если x в y раз больше наибольшего из чисел $|c_k|$. ▶

Лемма 2. (Принцип выбора.) *Допустим, что для любого целого $\nu > 0$ задана такая последовательность чисел $z_1^{(\nu)}, z_2^{(\nu)}, \dots$, что $0 \leqslant z_k^{(\nu)} \leqslant 1$. Тогда существует такая последовательность $\nu^{(1)}, \nu^{(2)}, \dots \to \infty$, что, когда ν пробегает ее, $z_k^{(\nu)}$ стремится к пределу при любом фиксированном k.*

*) Материал этого параграфа в дальнейшем не используется.

Доказательство[1]). Выберем возрастающую последовательность $\nu_1^{(1)}, \nu_2^{(1)}, \ldots$ так, чтобы, когда ν пробегало ее, $z_1^{(\nu)}$ стремилось бы к пределу z_1. Выберем из этой последовательности такую *подпоследовательность* $\nu_1^{(2)}, \nu_2^{(2)}, \ldots$, что, когда ν пробегает ее, $z_2^{(\nu)} \to z_2$. Продолжая таким же образом, мы построим для каждого n такую последовательность чисел $\nu_r^{(n)} \to \infty$, что, когда ν пробегает ее, $z_n^{(\nu)} \to z_n$, и любое $\nu_r^{(n)}$ является элементом предыдущей последовательности $\{\nu_r^{(n-1)}\}$. Положим теперь $\nu^{(r)} = \nu_r^{(r)}$. Пусть $r > n$. Все элементы $\nu^{(r)}$, кроме первых n, принадлежат последовательности $\nu_1^{(n)}, \nu_2^{(n)}, \ldots$, и поэтому $z_n^{(\nu)} \to z_n$, когда ν пробегает последовательность $\nu^{(1)}, \nu^{(2)}, \ldots$. ▶

Лемма 3. *Пусть* $\{w_n\}$ $(n = 0, \pm 1, \pm 2, \ldots)$ — *такая бесконечная в обе стороны последовательность, что* $0 \leq w_n \leq 1$ *и*

$$w_n = \sum_{k=1}^{\infty} f_k w_{n-k} \tag{11.6}$$

при любом n. *Если* $w_0 = 1$, *то* $w_n = 1$ *при всех* n.

Доказательство. Поскольку

$$w_0 = \sum_{k=1}^{\infty} f_k w_{-k} \leq \sum_{k=1}^{\infty} f_k = 1, \tag{11.7}$$

условие $w_0 = 1$ выполняется только тогда, когда эти два ряда почленно совпадают, а именно при любом k либо $f_k = 0$, либо $w_{-k} = 1$. Это означает, что $w_{-a} = 1$ для любого числа a из A. Но тогда рассуждения, проведенные при $n = 0$, применимы и к случаю $n = -a$, и мы заключаем, что $w_{-a-b} = 1$, если числа a и b принадлежат A. Продолжая по индукции, мы заключаем, что $w_{-m} = 1$ для любого числа m из A^+, и поэтому $w_{-m} = 1$ при любом $m > N$. Но это значит, что при $n = -N$ сумма в правой части (11.6) равна 1 и, таким образом, $w_{-N} = 1$. Полагая $n = -N + 1$, находим аналогично, что $w_{-N+1} = 1$, и, продолжая эти рассуждения, получаем по индукции, что $w_n = 1$ для всех n. ▶

Доказательство теоремы. Положим

$$\eta = \limsup_{n \to \infty} u_n. \tag{11.8}$$

В силу (11.1) очевидно, что $0 \leq \eta \leq 1$, и существует такая стремящаяся к бесконечности последовательность $r_1, r_2, \ldots$, что при $\nu \to \infty$

$$u_{r_\nu} \to \eta. \tag{11.9}$$

[1]) Это доказательство основано на так называемом *диагональном методе* Г. Кантора (1845—1918). Он давно стал стандартным, но во времена Кантора потрясал своей новизной.

Для каждого положительного целого числа ν введем бесконечную в обе стороны последовательность $\{u_n^{(\nu)}\}$ равенствами

$$u_n^{(\nu)}=\begin{cases} u_{r_\nu+n}, & \text{если } n\geqslant -r_\nu, \\ 0, & \text{если } n<-r_\nu. \end{cases} \tag{11.10}$$

Чтобы упростить изложение, лемма 2 была сформулирована для обычных последовательностей, однако она, очевидно, справедлива также и для бесконечных в обе стороны последовательностей. Поэтому можно найти возрастающую последовательность целых чисел $\nu_1, \nu_2, \ldots$, такую, что если индекс ν пробегает эту последовательность, то величина $u_n^{(\nu)}$ стремится к w_n для любого n. По построению $0\leqslant w_n\leqslant\eta$ и $w_0=\eta$. Далее, для любого ν и $n>-\nu$ определение (11.1) принимает вид

$$u_n^{(\nu)}=\sum_{k=1}^{\infty} f_k u_{n-k}^{(\nu)}, \tag{11.11}$$

и, переходя к пределу, мы получаем (11.6). По лемме 3 имеем $w_n=\eta$ при любом n.

Мы можем теперь завершить доказательство. Как и ранее, положим

$$\rho_k=f_{k+1}+f_{k+2}+\ldots, \tag{11.12}$$

так что $r_0=1$ и $\sum\rho_k=\mu$ (см. формулу (1.8) гл. XI). Суммируя соотношения (11.1) по $n=1, 2, \ldots, N$ и приводя подобные члены, мы получаем тождество

$$\rho_0 u_N+\rho_1 u_{N-1}+\ldots+\rho_N u_0=1. \tag{11.13}$$

Используем это соотношение последовательно для $N=\nu_1, \nu_2, \ldots$. Когда N пробегает эту последовательность, $u_{N-k}\to w_{-k}=\eta$ при любом k. Если $\sum\rho_k=\infty$, то $\eta=0$, и поэтому $u_n\to 0$, как и утверждалось. Если $\mu=\sum\rho_k<\infty$, то $\eta=\mu^{-1}$, и остается показать, что тогда $u_N\to\eta$ при любом способе стремления N к бесконечности. По определению верхнего предела мы имеем $u_{N-k}<\eta+\varepsilon$ при любом фиксированном k и достаточно большом N. Более того, $u_n\leqslant 1$ при всех n. Допустим теперь, что N стремится к бесконечности так, что $u_N\to\eta_0$. Из (11.13) легко вывести, что асимптотически

$$\rho_0\eta_0+(\rho_1+\ldots+\rho_r)(\eta+\varepsilon)+\rho_{r+1}+\rho_{r+2}+\ldots\geqslant 1, \tag{11.14}$$

и поэтому

$$\rho_0(\eta_0-\eta)+\mu(\eta+\varepsilon)\geqslant 1. \tag{11.15}$$

Но $\mu\eta=1$ и $\eta_0\leqslant\eta$ по определению η. Поскольку (11.15) выполняется для каждого $\varepsilon>0$, отсюда следует, что $\eta_0=\eta$, и, значит, $u_N\to\mu^{-1}$ при любом способе стремления N к бесконечности. ►

§ 12. ЗАДАЧИ

1. Допустим, что $F(s)$ — многочлен. Доказать для этого случая все теоремы § 3, используя метод разложения на простейшие дроби из гл. XI, 4.

2. Пусть подбрасываются r монет и рекуррентное событие $\mathscr{E}$ состоит в том, что для каждой из r монет суммарные числа появлений гербов и решеток одинаковы. Возвратно или невозвратно событие $\mathscr{E}$? Для минимального r, при котором $\mathscr{E}$ невозвратно, найти вероятность того, что $\mathscr{E}$ произойдет хотя бы один раз.

3. Пусть событие $\mathscr{E}$ состоит в том, что в последовательности независимых бросаний идеальной кости суммарные числа появлений единиц, двоек, ..., шестерок одинаковы. Показать, что $\mathscr{E}$ — невозвратное (периодическое) рекуррентное событие, и найти вероятность f того, что $\mathscr{E}$ произойдет хотя бы один раз.

4. Пусть в последовательности испытаний Бернулли событие $\mathscr{E}$ происходит, когда суммарное число успехов в λ раз превосходит суммарное число неудач; здесь λ — целое положительное число. (См. пример 1, в).) Показать, что $\mathscr{E}$ возвратно тогда и только тогда, когда $p/q=\lambda$, т. е. $p=\lambda/(\lambda+1)$. *Указание*. Воспользоваться нормальным приближением.

5. Мы говорим, что в последовательности испытаний Бернулли происходит событие $\mathscr{E}$, если суммарное число успехов вдвое превосходит суммарное число неудач и это отношение никогда ранее не превышало 2. Показать, что $\mathscr{E}$ невозвратно и периодично. Показать, далее, что производящая функция определяется кубическим уравнением $F(s)=qs(U(s)ps)^2$. (*Указание*. $U(s)ps$ — производящая функция времени ожидания первого момента, когда число успехов будет *больше* удвоенного числа неудач.)

6. Пусть $\mathbf{X}_j$ — независимые целочисленные одинаково распределенные случайные величины. Допустим, что они принимают как положительные, так и отрицательные значения. Доказать, что событие, определяемое условиями $\mathbf{S}_n=0$, $\mathbf{S}_1\leqslant 0$, ..., $\mathbf{S}_{n-1}\leqslant 0$, является рекуррентным и невозвратным.

7. *Счетчики Гейгера*. (См. примеры 1, ж) и 4, д).) Обозначим через $\mathbf{N}_n$ и $\mathbf{Z}_n$ число появлений $\mathscr{E}$ и число регистраций до момента времени n включительно. Установить связь между этими величинами и найти асимптотические формулы для $\mathbf{E}(\mathbf{Z}_n)$ и $\operatorname{Var}(\mathbf{Z}_n)$.

8. В *счетчиках Гейгера второго типа* каждая появляющаяся частица (независимо от того, регистрируется она или нет) запирает счетчик ровно на r единиц времени (т. е. на $r-1$ испытаний, следующих за моментом ее появления). Длительность мертвого времени, следующего за регистрацией, является поэтому случайной величиной. Найти ее производящую функцию G. Для рекуррентного события $\mathscr{E}$, состоящего в том, что счетчик «свободен», выразить производящую функцию F его времени возвращения через G. Наконец, найти среднее время возвращения.

9. *Более общий тип счетчиков Гейгера*. Пусть, как и в задаче 8, каждая новая частица полностью уничтожает влияние предыдущих, но теперь мы будем предполагать, что время, на которое частица запирает счетчик, является случайной величиной с заданной производящей функцией $B(s)$. (В предыдущей задаче $B(s)=s^r$.) Решить задачу 8 при этих более общих условиях.

10. Для рекуррентного события $\mathscr{E}$ с запаздыванием вероятности v_n не зависят от n только в случае, когда производящая функция момента первого появления $\mathscr{E}$ имеет вид $B(s)=(1-F(s))/[\mu(1-s)]$, т. е. когда $b_n=(f_{n+1}+f_{n+2}+\ldots)/\mu$. Как это связано с предельной теоремой для вероятностей попадания в примере 10, б)?

11. Найти приближенное значение вероятности того, что при 10 000 бросаний монеты число серий гербов длины 3 лежит между 700 и 730.

12. Пусть $\mathscr{E}$ означает появление *ГРГ* в последовательности бросаний монеты. Обозначим через r_n вероятность того, что $\mathscr{E}$ не произойдет в n испытаниях. Найти производящую функцию и при помощи разложения на простейшие дроби получить асимптотическое разложение.

13. Показать, что в примере 8, б) математическое ожидание времени до появ-

ления серии успехов длины r или серии неудач длины ρ равно

$$\mu_1\mu_2/(\mu_1+\mu_2).$$

где μ_1 и μ_2 — средние времена возвращения для серии успехов длины r и серии неудач длины ρ соответственно.

14. Возможные исходы A, B, C каждого испытания имеют вероятности α, β, γ $(\alpha+\beta+\gamma=1)$. Найти производящую функцию вероятностей того, что в n испытаниях отсутствуют серии длины r: а) из исходов A, б) из исходов A или B, в) любого вида.

15. *Продолжение*. Найти вероятность того, что первая A-серия длины r предшествует B-серии длины ρ и заканчивается при n-м испытании. (*Указание*. Производящая функция имеет тот же вид, что $X(s)$ в (8.6), причем F и G вычисляются по формулам (7.6) с заменой $p=1-q$ на α и β соответственно.)

16. *Самовосстанавливающиеся устройства*. В примере 10, г) найти предельное распределение возраста, предполагая, что распределение срока службы является геометрическим: $f_k=q^{k-1}p$.

17. *Продолжение*. Начальное *распределение возраста* $\{\beta_k\}$ *называется стационарным*, если оно не изменяется с течением времени. Показать (без вычислений), что это возможно только при $\beta_k=r_k/\mu$.

18. *Продолжение*. Обозначим через $w_k(n)$ математическое ожидание числа элементов, имеющих в момент времени n возраст k. Найти рекуррентные уравнения и вывести из них, что число элементов устройства остается постоянным. Показать далее, что математическое ожидание $w_0(n)$ удовлетворяет соотношению

$$w_0(n)=w_0(n-1)f_1/r_0+w_1(n-1)f_2/r_1+\ldots.$$

---

19. Пусть $\mathscr{E}$ — возвратное непериодическое рекуррентное событие. Допустим, что время возвращений имеет конечное среднее μ и дисперсию σ^2. Положим $q_n=f_{n+1}+f_{n+2}+\ldots$ и $r_n=q_{n+1}+q_{n+2}+\ldots$. Показать, что производящие функции $Q(s)$ и $R(s)$ сходятся при $s=1$. Доказать, что

$$u_0+\sum_{n=1}^{\infty}(u_n-1/\mu)s^n=R(s)/(\mu Q(s)) \tag{12.1}$$

и что поэтому

$$u_0+\sum_{n=1}^{\infty}(u_n-1/\mu)=(\sigma^2-\mu+\mu^2)/(2\mu^2). \tag{12.2}$$

20. Пусть $\mathscr{E}$ — возвратное рекуррентное событие и $\mathbf{N}_r$ — число появлений $\mathscr{E}$ в r испытаниях. Доказать, что

$$\mathbf{E}(\mathbf{N}_r^2)=u_1+\ldots+u_r+2\sum_{j=1}^{r-1}u_j(u_1+\ldots+u_{r-j}) \tag{12.3}$$

и что поэтому $\mathbf{E}(\mathbf{N}_r^2)$ является коэффициентом при s^r в

$$(F^2(s)+F(s))/[(1-s)(1-F(s))^2]. \tag{12.4}$$

(Заметить, что это можно сформулировать изящнее при помощи двойных производящих функций.)

21. Положим $q_{k,n}=\mathbf{P}\{\mathbf{N}_k=n\}$. Показать, что $q_{k,n}$ — это коэффициент при s^k в

$$F^n(s)(1-F(s))/(1-s). \tag{12.5}$$

Вывести отсюда, что $\mathbf{E}(\mathbf{N}_r)$ и $\mathbf{E}(\mathbf{N}_r^2)$ являются коэффициентами при s^r в

$$F(s)/[(1-s)(1-F(s))] \tag{12.6}$$

и в (12.4) соответственно.

22. Используя обозначения задачи 19, показать, что

$$\frac{F(s)}{(1-s)(1-F(s))} = -\frac{1}{1-s} + \frac{1}{\mu(1-s)^2} + \frac{R(s)}{\mu\{1-F(s)\}}. \tag{12.7}$$

Отсюда и из последней задачи вывести равенство

$$\mathrm{E}(\mathbf{N}_r) = r/\mu + (\sigma^2+\mu-\mu^2)/(2\mu^2) + \varepsilon_r, \tag{12.8}$$

где $\varepsilon_r \to 0$.

23. *Продолжение.* Используя аналогичные рассуждения, показать, что

$$\mathrm{E}(\mathbf{N}_r^2) = r^2/\mu^2 + [(2\sigma^2+\mu-\mu^2)/\mu^3]\, r + \alpha_r, \tag{12.9}$$

где $\alpha_r/r \to 0$. Поэтому

$$\mathrm{Var}(\mathbf{N}_r) \sim (\sigma^2/\mu^3)\, r. \tag{12.10}$$

(*Указание.* Представить разность выражений (12.4) и (12.7) в виде суммы трех дробей со знаменателями, кратными $(1-s)^k$, $k=1, 2, 3$.)

24. Пусть $q_{k,n}$ — вероятность того, что в k испытаниях Бернулли имеется ровно n серий успехов длины r. Используя решение задачи 21, показать, что производящая функция $Q_k(x) = \sum q_{k,n}x^n$ является коэффициентом при s^k в

$$(1-p^rs^r)/[1-s+qp^rs^{r+1}-(1-ps)\,p^rs^rx].$$

Показать далее, что наименьшим по абсолютной величине корнем знаменателя является $s_1 \approx 1+qp^r(1-x)$.

25. *Продолжение. Пуассоновское распределение длинных серий*[1]. Доказать, что если число k испытаний и длина r серии стремятся к бесконечности так, что $kqp^r \to \lambda$, то вероятность получить ровно n серий длины r стремится к $e^{-\lambda}\lambda^n/n!$. *Указание.* Используя предыдущую задачу, показать, что для производящей функции справедлива асимптотическая формула $\{1+qp^r(1-x)\}^{-k} \sim e^{-\lambda(1-x)}$. Использовать *теорему непрерывности* гл. XI, 6.

[1]) Эта теорема была доказана Мизесом, однако наш метод значительно проще.

ГЛАВА XIV

СЛУЧАЙНОЕ БЛУЖДАНИЕ И ЗАДАЧИ О РАЗОРЕНИИ

§ 1. ОБЩИЕ ПОНЯТИЯ

Первая часть этой главы посвящена испытаниям Бернулли, и для упрощения и оживления формулировок мы здесь снова воспользуемся образным языком пари и теории случайных блужданий.

Обратимся к уже известному нам игроку, который выигрывает или проигрывает доллар с вероятностями p и q соответственно. Пусть его начальный капитал равен z, и пусть он играет против соперника с начальным капиталом $a-z$, так что их суммарный капитал равен a. Игра продолжается до тех пор, пока капитал нашего игрока либо не уменьшится до нуля, либо не возрастет до a, т. е. до тех пор, пока один из двух играющих не разорится. Мы интересуемся вероятностью разорения нашего игрока и распределением вероятностей для продолжительности игры. Это *классическая задача о разорении.*

Физические приложения и аналогии подсказывают более гибкую интерпретацию — при помощи понятия блуждающей точки или «частицы» на оси x. Эта частица выходит из *начального положения* z и совершает через равные промежутки времени единичные скачки в положительном или отрицательном направлении в зависимости от того, чем закончилось соответствующее испытание — успехом или неудачей. Положение частицы после n шагов представляет капитал игрока по завершении n-го испытания. Испытания прекращаются, когда частица впервые достигает либо точки $x=0$, либо точки $x=a$, и мы говорим в этом случае, что частица совершает *случайное блуждание с поглощающими экранами в точках* $x=0$ *и* $x=a$. Это случайное блуждание *ограничено* возможными положениями $1, 2, \ldots$ $\ldots, a-1$; при отсутствии поглощающих экранов случайное блуждание называется *неограниченным*. Физики используют модель случайного блуждания в качестве грубого приближения для одномерной диффузии или броуновского движения, в котором физическая частица подвергается большому числу столкновений с молекулами, сообщающими ей случайное движение. Случай $p>q$ соответствует *сносу* вправо, когда толчки слева более вероятны; при $p=q=1/2$ случайное блуждание называется *симметричным*.

В предельном случае $a\to\infty$ мы получаем случайное блуждание на полубесконечной прямой: выходящая из $z>0$ частица совершает случайное блуждание до того момента, когда она впервые достигнет нуля. В этой формулировке мы узнаем *задачу о времени первого* *достижения*; она была решена элементарными методами в гл. III (по крайней мере для симметричного случая) и с использованием производящих функций в гл. XI,3. Мы встретимся с уже полученными ранее формулами, однако они будут выведены по-новому.

В этой главе мы воспользуемся методом *разностных уравнений*, которые являются аналогом дифференциальных уравнений теории диффузии. Эта аналогия естественным образом приводит к различным модификациям и обобщениям классической задачи о разорении; типичным и поучительным примером здесь является замена поглощающего экрана *отражающим* и *упругим* экранами. Чтобы описать отражающий экран, рассмотрим случайное блуждание в конечном интервале, определенное как и ранее, за тем исключением, что каждый раз, когда частица оказывается в точке $x=1$, она с вероятностью p перемещается в точку $x=2$ и с вероятностью q остается в точке $x=1$. В игровой терминологии это соответствует такому соглашению: когда игрок проигрывает свой последний доллар, этот доллар великодушно возвращается ему противником, так что игра может продолжаться. Физик представляет себе стенку, помещенную в точке $x=1/2$ оси x и обладающую тем свойством, что частица, движущаяся из точки $x=1$ в точку $x=0$, отражается от стенки и возвращается в точку $x=1$ вместо того, чтобы попасть в точку $x=0$.

Как поглощающий, так и отражающий экраны являются частными случаями так называемого упругого экрана. Мы определим *упругий экран в точке $x=0$ при помощи следующего правила: из точки $x=1$ частица с вероятностью p движется к точке $x=2$, с вероятностью δq останется в точке $x=1$ и с вероятностью $(1-\delta)q$ движется к точке $x=0$ и поглощается* (т. е. процесс прекращается). При $\delta=0$ мы имеем классическую задачу о разорении, или поглощающие экраны, при $\delta=1$ — отражающие экраны. Когда δ меняется от 0 до 1, мы получаем семейство промежуточных случаев. Чем больше δ, тем более вероятно, что процесс будет продолжаться, и процесс с двумя отражающими экранами никогда не сможет прекратиться.

Параграфы 2 и 3 посвящены элементарному обсуждению классической задачи о разорении и связанных с ней вопросов. Последующие три параграфа технически более сложны (и могут быть опущены); в § 4 и 5 мы выводим производящие функции и из них — явные выражения для распределения продолжительности игры и т. п. Параграф 6 содержит схему предельного перехода к уравнению диффузии (формальные решения этого уравнения являются предельными распределениями для случайного блуждания).

В § 7 изложение снова становится элементарным и посвящено *случайным блужданиям в двух и более измерениях*, где мы сталкиваемся с новыми явлениями. В § 8 рассматривается обобщение совершенно другого типа: одномерное случайное блуждание, при котором частица не обязательно перемещается единичными скачками, а может менять свое положение скачками произвольной величины, кратной единице. Такие обобщенные случайные блуждания привлекли широкий интерес в связи с развитой Вальдом теорией *последовательного анализа*.

Раздел задач содержит существенные дополнения к тексту и наброски альтернативных подходов. Можно надеяться, что сравнение используемых методов окажется весьма поучительным.

В заключение следует подчеркнуть, что каждое случайное блуждание представляет собой некоторую цепь Маркова, так что настоящая глава служит частично и введением в следующую, где некоторые задачи для случайных блужданий (например, задачи с упругими экранами) будут сформулированы иначе.

§ 2. КЛАССИЧЕСКАЯ ЗАДАЧА О РАЗОРЕНИИ

Рассмотрим задачу, поставленную в начале настоящей главы. Пусть q_z — вероятность окончательного разорения [1]) игрока, а p_z — вероятность того, что он в конце концов выиграет. В терминах теории случайного блуждания q_z и p_z суть вероятности того, что частица, выходящая из точки z, будет поглощена соответственно в 0 или в a. Мы покажем, что $p_z+q_z=1$, так что нам не нужно рассматривать возможность нескончаемой игры.

После первого испытания состояние игрока равно либо $z-1$, либо $z+1$, и поэтому

$$q_z=pq_{z+1}+qq_{z-1}, \tag{2.1}$$

если $1<z<a-1$. При $z=1$ первое же испытание может привести к разорению и (2.1) следует заменить на $q_1=pq_2+q$. Точно так же при $z=a-1$ первое испытание может закончиться победой, и поэтому $q_{a-1}=qq_{a-2}$. Чтобы объединить наши уравнения, положим

$$q_0=1, \quad q_a=0. \tag{2.2}$$

При этом соглашении вероятность q_z разорения удовлетворяет (2.1) для $z=1, 2, \ldots, a-1$.

Системы вида (2.1) известны под названием *разностных уравнений* [2]), а (2.2) представляют собой *граничные условия* на q_z. Мы выведем явное выражение для q_z при помощи *метода частных решений*, который будет использоваться и в более общих случаях.

Предположим сперва, что $p\neq q$. Легко проверить, что разностное уравнение (2.1) допускает два частных решения $q_z=1$ и $q_z=(q/p)^z$. Следовательно,

$$q_z=A+B\,(q/p)^z, \tag{2.3}$$

[1]) Строго говоря, вероятность разорения определена на пространстве элементарных исходов бесконечно продолжающихся игр, но мы можем работать и с пространством элементарных исходов для n испытаний. Вероятность разорения менее чем за n испытаний возрастает вместе с n и, следовательно, имеет предел. Мы назовем этот *предел* «вероятностью разорения». Все вероятности в этой главе можно интерпретировать таким образом, не обращаясь к бесконечным пространствам (ср. гл. VIII, 1).

[2]) Подробнее о разностных уравнениях см., например, Гельфонд А. О., Исчисление конечных разностей.— М.: Наука, 1967.— *Прим. перев.*

где A и B — произвольные константы, представляет собой формальное решение системы (2.1). Граничные условия (2.2) будут выполнены тогда и только тогда, когда A и B удовлетворяют двум линейным уравнениям $A+B=1$ и $A+B(q/p)^a=0$. Таким образом,

$$q_z = [(q/p)^a - (q/p)^z]/[(q/p)^a - 1] \tag{2.4}$$

формально является решением разностного уравнения (2.1), удовлетворяющим граничным условиям (2.2). Для того чтобы доказать, что (2.4) является искомой вероятностью разорения, осталось показать, что это решение единственно, т. е. что *все* решения уравнения (2.1) имеют вид (2.3). Если нам теперь дано произвольное решение уравнения (2.1), то можно выбрать две константы A и B так, что (2.3) будет совпадать с ним при $z=0$ и $z=1$. Подставляя в (2.1) последовательно $z=1, 2, 3, \ldots$, мы можем найти по этим двум значениям все остальные. Стало быть, два решения, совпадающие при $z=0$ и $z=1$, тождественно равны, и, следовательно, каждое решение имеет вид (2.3).

При $p=q=1/2$ наше рассуждение не проходит, ибо тогда (2.4) бессмысленно, так как частные решения $q_z=1$ и $q_z=(q/p)^z$ в этом случае совпадают. Однако при $p=q=1/2$ у нас есть второе решение $q_z=z$, и поэтому $q_z=A+Bz$ дает нам зависящее от двух констант решение уравнения (2.1). Для того чтобы были удовлетворены граничные условия (2.2), мы должны положить $A=1$ и $A+Ba=0$. Следовательно,

$$q_z = 1 - z/a. \tag{2.5}$$

(То же численное значение можно формально получить из (2.4), если, пользуясь правилом Лопиталя, найти предел этого выражения при $p \to 1/2$.)

Таким образом, мы доказали, что искомая *вероятность разорения игрока дается формулой* (2.4) *при* $p \neq q$ *и формулой* (2.5) *при* $p=q=1/2$. Вероятность p_z выигрыша нашим игроком всей игры равна вероятности разорения его противника и, стало быть, получается из наших формул при замене p, q и z на q, p и $a-z$ соответственно. Легко видеть, что $p_z+q_z=1$, как уже говорилось выше.

Мы можем переформулировать наш результат следующим образом. *Пусть игрок с начальным капиталом z играет против бесконечно богатого соперника, всегда готового продолжать игру, тогда как сам игрок имеет право прервать игру по своему желанию. Игрок, согласно принятой им стратегии, играет по тех пор, пока он либо не потеряет свой капитал, либо не увеличит его до a (с чистым выигрышем $a-z$). Тогда q_z равно вероятности его проигрыша, а $1-q_z$ — вероятности его выигрыша.*

При такой системе окончательный выигрыш или проигрыш игрока будет случайной величиной $\mathbf{G}$, принимающей значения $a-z$ и $-z$ с вероятностями $1-q_z$ и q_z соответственно. Ожидаемый выигрыш равен

$$\mathbf{E}(\mathbf{G}) = a(1 - q_z) - z. \tag{2.6}$$

Очевидно, $\mathbf{E}(\mathbf{G}) = 0$ тогда и только тогда, когда $p = q$. Это означает, что при описанной системе *«безобидная» игра остается безобидной, а «небезобидная» игра не может превратиться в безобидную.*

Из (2.5) мы видим, что в случае $p = q$ игрок с начальным капиталом $z = 999$ имеет вероятность 0,999 выиграть доллар до того, как он проиграет весь свой капитал. При $q = 0{,}6$, $p = 0{,}4$ игра весьма неблагоприятна, но вероятность (2.4) выиграть доллар до потери капитала все еще будет около 2/3. Вообще, игрок с относительно большим начальным капиталом z имеет приличные шансы выиграть небольшую сумму $a - z$ до своего разорения [1]).

(Удивительное следствие нашего результата см. в задаче 4.)

Изучим теперь влияние *изменения ставок*. Изменение единицы с доллара на полдоллара эквивалентно удвоению начальных капиталов. Соответствующая вероятность разорения q_z^* получается из (2.4) с заменой z на $2z$ и a на $2a$:

$$q_z^* = \frac{(q/p)^{2a} - (q/p)^{2z}}{(q/p)^{2a} - 1} = q_z \cdot \frac{(q/p)^a + (q/p)^z}{(q/p)^a + 1}. \tag{2.7}$$

При $q > p$ последняя дробь больше единицы и $q_z^* > q_z$. Мы переформулируем это заключение следующим образом: *если ставки удваиваются, а начальные капиталы остаются неизменными, то вероятность разорения уменьшается для игрока, у которого вероятность успеха $p < 1/2$, и возрастает для его противника (для которого игра выгодна* [2])).

Предположим, например, что Петр имеет 90 долларов, а Павел 10 и что $p = 0{,}45$, т. е. игра неблагоприятна для Петра. Если ставка в каждом испытании равна одному доллару, то табл. 1 показывает, что вероятность разорения Петра равна приблизительно 0,866. Если в этой же игре ставка равна 10 долларам, то вероятность разорения Петра падает ниже одной четвертой, а именно примерно до 0,210. Таким образом, влияние увеличения ставок выражено более резко, чем можно было бы ожидать.

---

[1]) Некий игрок из года в год ездил в Монте-Карло и всегда успешно покрывал расходы на свой отдых. Он твердо верил в магическую власть над случаем. В действительности же в его опыте нет ничего удивительного. Если предположить, что он начинал с суммой, в десять раз большей его окончательного выигрыша, то каждый год вероятность успеха составляла примерно 0,9. Вероятность непрерывающейся последовательности из десяти успехов равна приблизительно $(1 - 1/10)^{10} \approx e^{-1} \approx 0{,}37$. Таким образом, продолжительные успехи ни в коей мере не являются невероятными. Более того, в *одной* неудаче, по мнению игрока, конечно, были бы виноваты неосмотрительность или кратковременное недомогание.

[2]) Детальный анализ других возможных стратегий можно найти в (неэлементарной) книге Dubbins L. E., Savage L. J., How to gamble if you must (которая имеет более информативный подзаголовок: Inequalities for stochastic processes), New York, McGrow-Hill, 1965.

Таблица 1

К классической задаче о разорении игрока

p	q	z	a	Вероятность разорения	Вероятность успеха	Математическое ожидание выигрыша	Математическое ожидание продолжительности игры
0,5	0,5	9	10	0,1	0,9	0	9
0,5	0,5	90	100	0,1	0,9	0	900
0,5	0,5	900	1 000	0,1	0,9	0	90 000
0,5	0,5	950	1 000	0,05	0,95	0	47 500
0,5	0,5	8 000	10 000	0,2	0,8	0	16 000 000
0,45	0,55	9	10	0,210	0,790	−1,1	11
0,45	0,55	90	100	0,866	0,134	−76,6	765,6
0,45	0,55	99	100	0,182	0,818	−17,2	171,8
0,4	0,6	90	100	0,983	0,017	−88,3	441,3
0,4	0,6	99	100	0,333	0,667	−32,3	161,7

Начальный капитал равен z. Игра заканчивается разорением (проигрыш z) или обладанием капиталом a (выигрыш $a-z$).

Вообще, если при каждом испытании ставится k долларов, то вероятность разорения мы находим из (2.4), заменяя z на z/k и a на a/k; вероятность разорения убывает с ростом k. Поэтому в игре с постоянными ставками игрок будет минимизировать вероятность разорения, выбирая наибольшую ставку, совместимую с его намерением выиграть заранее фиксированную сумму. Эмпирическая обоснованность этого заключения подвергалась сомнению обычно теми людьми, которые утверждают, что каждое «невыгодное» пари неразумно. Если бы это было принято всерьез, то это означало бы конец страхового дела, ибо осторожный водитель, страхующийся на всякий случай, играет при этом, очевидно, в технически «невыгодную» игру. В действительности в теории вероятности не существует никакой теоремы, которая отбила бы у такого водителя охоту страховаться.

Предельный случай $a=\infty$ соответствует игре против бесконечно богатого соперника. Устремляя в (2.4) и (2.5) a к ∞, получаем

$$q_z=\begin{cases} 1, & \text{если } p\leqslant q, \\ (q/p)^z, & \text{если } p>q. \end{cases} \tag{2.8}$$

Мы интерпретируем q_z *как вероятность окончательного разорения игрока с начальным капиталом* z, *играющего против бесконечно* *богатого соперника* [1]). В терминологии теории случайных блужданий q_z есть вероятность того, что выходящая из $z>0$ частица когда-либо достигнет нуля. Более естественно перефразировать этот результат в следующем виде: *в случайном блуждании, начинающемся в нуле, вероятность когда-либо достичь положения $z>0$ равна* 1, *если $p\geqslant q$, и равна $(p/q)^z$ при $p<q$.*

§ 3. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ПРОДОЛЖИТЕЛЬНОСТИ ИГРЫ

Вероятностное распределение продолжительности игры будет выведено в следующих параграфах. Однако ее математическое ожидание может быть получено гораздо более простым методом, который применяется столь широко, что стоит объяснить его сейчас (хотя для этого придется несколько повториться).

Мы по-прежнему рассматриваем классическую задачу о разорении, сформулированную в начале этой главы. Будем считать известным тот факт, что продолжительность игры имеет конечное математическое ожидание D_z (это будет строго доказано в следующем параграфе).

Если первое испытание заканчивается успехом, то игра продолжается, как если бы начальным положением было $z+1$. Стало быть, условное математическое ожидание продолжительности игры при условии первого успеха равно $D_{z+1}+1$. Это соображение показывает, что ожидаемая продолжительность D_z удовлетворяет разностному уравнению

$$D_z=pD_{z+1}+qD_{z-1}+1,\quad 0<z<a, \tag{3.1}$$

с граничными условиями

$$D_0=0,\quad D_a=0. \tag{3.2}$$

Наличие члена 1 делает разностное уравнение (3.1) неоднородным. Если $p\neq q$, то $D_z=z/(q-p)$ будет формальным решением уравнения (3.1). Разность Δ_z любых двух решений (3.1) удовлетворяет однородным уравнениям $\Delta_z=p\Delta_{z+1}+q\Delta_{z-1}$, а мы уже знаем, что все решения этого уравнения имеют вид $A+B(q/p)^z$. Отсюда следует, что при $p\neq q$ все решения уравнения (3.1) представимы в виде

$$D_z=z/(q-p)+A+B(q/p)^z. \tag{3.3}$$

Для выполнения граничных условий (3.2) требуется, чтобы

$$A+B=0,\quad A+B(q/p)^a=-a/(q-p).$$

[1]) Легко видеть, что q_z представляет собой решение разностных уравнений (2.1), удовлетворяющее (теперь единственному) граничному условию $q_0=1$. При $p>q$ это решение не единственно. В действительности наш результат содержится в формуле (3.9) гл. XI и будет получен независимым образом (в усиленной форме) в § 4.

Разрешая эти равенства относительно A и B, находим, что

$$D_z = \frac{z}{q-p} - \frac{a}{q-p} \cdot \frac{1-(q/p)^z}{1-(q/p)^a}. \tag{3.4}$$

Этот прием опять-таки непригоден при $q=p=1/2$. В этом случае мы заменяем $z/(q-p)$ на $-z^2$, которое является теперь решением (3.1). Отсюда вытекает, что если $p=q=1/2$, то все решения (3.1) имеют вид $D_z=-z^2+A+Bz$. Искомое решение D_z, удовлетворяющее граничным условиям (3.2), равно

$$D_z=z(a-z). \tag{3.5}$$

Математическое ожидание продолжительности игры в классической задаче о разорении равно (3.4) *при* $p\neq q$ *и* (3.5) *при* $p=q=1/2$.

Следует отметить, что полученные нами значения для продолжительности игры значительно больше, чем мы обычно наивно ожидаем. Если два игрока, имеющие по 500 долларов каждый, бросают монету до тех пор, пока один из них не разорится, то средняя продолжительность игры равна 250 000 испытаниям. Если игрок имеет лишь один доллар, а его противник — 1000, то средняя продолжительность равна 1000 испытаниям. Дальнейшие примеры см. в табл. 1.

Как уже указывалось в конце предыдущего параграфа, мы можем перейти к пределу при $a\to\infty$ и рассмотреть игру против бесконечно богатого соперника. При $p>q$ игра может продолжаться бесконечно, и говорить о математическом ожидании ее продолжительности в этом случае бессмысленно. При $p<q$ мы получаем для математического ожидания продолжительности игры значение $z(q-p)^{-1}$, а при $p=q$ оно бесконечно. (Этот же результат был установлен в гл. XI,3 и будет независимо доказан в следующем параграфе.)

§ 4*). ПРОИЗВОДЯЩИЕ ФУНКЦИИ ДЛЯ ПРОДОЛЖИТЕЛЬНОСТИ ИГРЫ И ДЛЯ ВРЕМЕН ПЕРВОГО ДОСТИЖЕНИЯ

Мы воспользуемся методом производящих функций для изучения продолжительности игры в классической задаче о разорении, т. е. продолжительности ограниченного случайного блуждания с поглощающими экранами в точках $x=0$ и $x=a$. Начальным положением является z $(0<z<a)$. Пусть $u_{z,n}$ — вероятность того, что процесс окончится на n-м шаге на экране $x=0$ (разорение игрока при n-м испытании). После первого шага частица попадает или в точку $x=z+1$, или в точку $x=z-1$, и мы заключаем, что для $1<z<a-1$ и $n\geq 1$

$$u_{z,\,n+1} = pu_{z+1,\,n} + qu_{z-1,\,n}. \tag{4.1}$$

*) Этот параграф, как и связанный с ним § 5, может быть опущен при первом чтении.

Это — разностное уравнение, аналогичное (2.1), но зависящее уже от двух переменных z и n. Мы хотим (по аналогии с процедурой из § 2) определить граничные условия $u_{0,n}$, $u_{a,n}$ и $u_{z,0}$ так, чтобы (4.1) было также справедливо для $z=1$, $z=a-1$ и $n=0$. Для этой цели положим

$$u_{0,n}=u_{a,n}=0 \quad при \quad n\geqslant 1 \tag{4.2}$$

и

$$u_{0,0}=1, \quad u_{z,0}=0 \quad при \quad 0<z\leqslant a. \tag{4.3}$$

Тогда (4.1) справедливо для всех z при $0<z<a$ и всех $n\geqslant 0$.

Введем теперь производящие функции

$$U_z(s)=\sum_{n=0}^{\infty} u_{z,n}s^n. \tag{4.4}$$

Умножая (4.1) на s^{n+1} и суммируя по n, находим

$$U_z(s)=psU_{z+1}(s)+qsU_{z-1}(s), \quad 0<z<a; \tag{4.5}$$

граничные условия (4.2) и (4.3) приводят к равенствам

$$U_0(s)=1, \quad U_a(s)=0. \tag{4.6}$$

Система (4.5) состоит из разностных уравнений, аналогичных (2.1), а граничные условия (4.6) соответствуют (2.2). Новизна здесь заключается в том, что коэффициенты и неизвестные $U_z(s)$ зависят теперь от переменной s, однако для самого разностного уравнения s является просто произвольной постоянной. Если только нам удастся найти два частных решения системы (4.5), то мы снова сможем применять метод § 2. Естественно выяснить, существуют ли два решения $U_z(s)$ вида $U_z(s)=\lambda^z(s)$. Подставляя это выражение в (4.5), мы находим, что $\lambda(s)$ должно удовлетворять квадратному уравнению

$$\lambda(s)=ps\lambda^2(s)+qs, \tag{4.7}$$

имеющему два корня:

$$\lambda_1(s)=\frac{1+\sqrt{1-4pqs^2}}{2ps}, \quad \lambda_2(s)=\frac{1-\sqrt{1-4pqs^2}}{2ps} \tag{4.8}$$

(мы берем $0<s<1$ и положительный квадратный корень).

Теперь у нас есть два частных решения системы (4.5), и, как и в § 2, мы заключаем, что все ее решения имеют вид

$$U_z(s)=A(s)\lambda_1^z(s)+B(s)\lambda_2^z(s), \tag{4.9}$$

где $A(s)$ и $B(s)$ — произвольные функции. Чтобы удовлетворить граничным условиям (4.6), мы должны иметь $A(s)+B(s)=1$ и $A(s)\lambda_1^a(s)+B(s)\lambda_2^a(s)=0$, откуда

$$U_z(s)=\frac{\lambda_1^a(s)\lambda_2^z(s)-\lambda_1^z(s)\lambda_2^a(s)}{\lambda_1^a(s)-\lambda_2^a(s)}, \tag{4.10}$$

Пользуясь очевидным соотношением $\lambda_1(s)\lambda_2(s)=q/p$, это выражение можно упростить:

$$U_z(s)=\left(\frac{q}{p}\right)^z\frac{\lambda_1^{a-z}(s)-\lambda_2^{a-z}(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.11}$$

Это и есть *искомая производящая функция вероятности разорения (поглощения в точке $x=0$) при n-м испытании.* Тот же метод показывает, что производящая функция для вероятности поглощения в точке $x=a$ дается формулой

$$\frac{\lambda_1^z(s)-\lambda_2^z(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.12}$$

Производящая функция для *продолжительности игры* равна, конечно, сумме производящих функций (4.11) и (4.12).

Бесконечные интервалы и первые достижения

Проведенные выше рассуждения применимы равным образом и к случайным блужданиям на интервале $(0, \infty)$ с поглощающим экраном в точке $x=0$. Частица, выходящая из положения $z>0$, в конце концов поглощается в нуле, в противном случае блуждание продолжается вечно. Поглощение соответствует разорению игрока с начальным капиталом z, играющего против бесконечно богатого соперника. Производящая функция $U_z(s)$ вероятностей $u_{z,n}$ того, что поглощение происходит в точности при n-м испытании, вновь удовлетворяет разностным уравнениям (4.5) и, стало быть, имеет вид (4.9); однако если не выполняется условие $A(s)=0$, то это решение не ограничено на бесконечности. Единственное граничное условие теперь состоит в том, что $U_0(s)=1$, и поэтому $B(s)=1$, или

$$U_z(s)=\lambda_2^z(s) \tag{4.13}$$

(такой же результат можно получить, устремив в (4.11) a к ∞ и воспользовавшись тем, что $\lambda_1(s)\lambda_2(s)=q/p$).

Из (4.13) при $s=1$ следует, что поглощение неизбежно при $p\leqslant q$, а в противном случае имеет вероятность $(q/p)^z$. К такому же выводу мы пришли в § 2.

Рассматриваемое нами поглощение в нуле допускает и другую важную интерпретацию — как первое достижение в *неограниченном* случайном блуждании. Действительно, перенося начало координат в точку z, мы видим, что в начинающемся в нуле случайном блуждании на всей прямой $u_{z,n}$ *будет вероятностью того, что первое попадание в точку $-z<0$ произойдет при n-м испытании.* То, что соответствующая производящая функция (4.13) является z-й степенью λ_2, отражает тот очевидный факт, что время ожидания до первого прохождения через $-z$ является суммой z независимых времен ожидания между последовательными первыми прохождениями через $-1, -2, \ldots, -z$.

Явная формула для $u_{z,n}$ в частном случае $p=1/2$ была выведена элементарными методами в гл. III (формула (7.5)). Учитывая, что нужно совершить $(n+z)/2$ шагов влево и $(n-z)/2$ шагов вправо, легко заключить, что и в общем случае справедлива та же формула, за тем лишь исключением, что отдельные траектории имеют теперь вероятности $p^{(n-z)/2}q^{(n+z)/2}$, а не 2^{-n}. Таким образом,

$$u_{z,n}=\frac{z}{n}\binom{n}{(n+z)/2}p^{(n-z)/2}q^{(n+z)/2}, \tag{4.14}$$

где биномиальный коэффициент полагается равным нулю, если n и z — числа разной четности. (Относительно вывода этой формулы при помощи производящей функции см. конец § 3 гл. XI. Другой, внешне совершенно отличный вариант явной формулы содержится в задаче 13.)

§ 5*). ЯВНЫЕ ВЫРАЖЕНИЯ

Производящая функция U_z в (4.11) формально зависит от квадратного корня, однако в действительности она является рациональной функцией s. В самом деле, элементарное разложение бинома приводит знаменатель к виду

$$\lambda_1^a(s)-\lambda_2^a(s)=s^{-a}\sqrt{1-4pqs^2}P_a(s), \tag{5.1}$$

где P_a — многочлен четного порядка, равного $a-1$ при нечетном a и $a-2$ — при четном. Числитель имеет тот же вид, только вместо a стоит $a-z$. Таким образом, U_z является отношением двух многочленов, степени которых отличаются самое большее на 1. Следовательно, можно вывести явные выражения для вероятностей разорения при помощи описанного в гл. XI,4 метода разложения на простейшие дроби. Этот результат интересен из-за его связи с теорией диффузии, а его вывод является прекрасной демонстрацией техники, применяемой при практическом использовании простейших дробей.

Вычисления весьма упростятся, если мы воспользуемся вспомогательной переменной φ, определяемой соотношением

$$\cos\varphi=1/(2\sqrt{pq}\,s). \tag{5.2}$$

(Интервалу $0<s<1$ здесь соответствуют комплексные значения φ, однако это не оказывает никакого влияния на формальные выкладки.) Из (4.8) имеем

$$\lambda_1(s)=\sqrt{q/p}\,[\cos\varphi+i\sin\varphi]=\sqrt{q/p}\,e^{i\varphi}, \tag{5.3}$$

тогда как $\lambda_2(s)$ получается из этого выражения заменой i на $-i$. Поэтому

$$U_z(s)=(\sqrt{q/p})^z\sin[(a-z)\varphi]/\sin a\varphi. \tag{5.4}$$

*) См. примечание на с. 363.

Корни знаменателя $s_1, s_2, \ldots$ являются простыми, и, следовательно, существует разложение на простейшие дроби вида

$$(\sqrt{q/p})^z \frac{\sin(a-z)\varphi}{\sin a\varphi} = A + Bs + \frac{\rho_1}{s_1-s} + \ldots + \frac{\rho_{a-1}}{s_{a-1}-s}. \quad (5.5)$$

В принципе мы должны рассматривать только те корни s_ν, которые не являются в то же время и корнями числителя, но если s_ν — корень числителя, то $U_z(s)$ будет непрерывна при $s=s_\nu$, и, значит, $\rho_\nu=0$. Стало быть, такие взаимно сокращающиеся корни не дают никакого вклада в правую часть, и нет никакой необходимости рассматривать их отдельно.

Корни $s_1, \ldots, s_{a-1}$ соответствуют, очевидно, значениям $\varphi_\nu = \pi\nu/a$ при $\nu=1, \ldots, a-1$, так что

$$s_\nu = 1/[2\sqrt{pq}\cos(\pi\nu/a)]. \quad (5.6)$$

Это выражение теряет смысл, когда $\nu=a/2$ и a четно, но в этом случае φ_ν будет также корнем числителя и этот корень должен быть отброшен. Будучи собственным, соответствующий член в окончательном выражении пропадает.

Чтобы вычислить ρ_ν, умножим обе части (5.5) на $s_\nu - s$ и устремим s к s_ν. Вспоминая, что $\sin a\varphi_\nu = 0$ и $\cos a\varphi_\nu = 1$, получаем

$$\rho_\nu = (\sqrt{q/p})^z \sin z\varphi_\nu \cdot \lim_{s\to s_\nu}[(s-s_\nu)/\sin a\varphi].$$

Последний предел определяется при помощи правила Лопиталя и дифференцирования (5.2), что дает

$$\rho_\nu = a^{-1}\cdot 2\sqrt{pq}\,(\sqrt{q/p})^z \sin z\varphi_\nu \cdot \sin\varphi_\nu \cdot s_\nu^2.$$

Из разложения правой части (5.5) в геометрический ряд мы получаем для $n>1$

$$u_{z,n} = \sum_{\nu=1}^{a-1}\rho_\nu s_\nu^{-n-1} = a^{-1} 2\sqrt{pq}\,(\sqrt{q/p})^z \sum_{\nu=1}^{a-1} s_\nu^{-n+1}\cdot\sin\varphi_\nu\cdot\sin z\varphi_\nu$$

и окончательно

$$u_{z,n} = a^{-1}2^n p^{(n-z)/2} q^{(n+z)/2} \sum_{\nu=1}^{a-1}\cos^{n-1}(\pi\nu/a)\sin(\pi\nu/a)\sin(\pi z\nu/a). \quad (5.7)$$

Итак, мы вывели явную формулу для *вероятности разорения при n-м испытании*. Она восходит к Лагранжу и многократно выводилась классиками [1]), однако продолжает переоткрываться и в совре-

[1]) Элементарный вывод, основанный на тригонометрической интерполяции, см. в работах Эллиса (Ellis R. E., Cambridge Math. J., 4 (1844); Ellis R. E., Collected works, Cambridge and London, 1863).

менной литературе. Интересно, что метод отражения приводит к другому явному выражению для $u_{z,n}$ — через биномиальные коэффициенты (задача 21). Другой метод вывода (5.7) будет описан в гл. XVI,3.

Переходя к пределу при $a \to \infty$, получаем вероятность того, что в игре против бесконечно богатого соперника игрок с начальным капиталом z будет разорен при n-м испытании. (См. задачу 13.)

Из вида суммы в (5.7) ясно, что члены, соответствующие $\nu=k$ и $\nu=a-k$, равны по абсолютной величине; они имеют один и тот же знак, когда n и z — числа одинаковой четности, и взаимно уничтожаются в противном случае. Следовательно, *при $n-z$ нечетном $u_{z,n}=0$, а при $n-z$ четном и $n>1$*

$$u_{z,n}=a^{-1}2^{n+1}p^{(n-z)/2}q^{(n+z)/2}\sum_{\nu<a/2}\cos^{n-1}(\pi\nu/a)\sin(\pi\nu/a)\sin(\pi z\nu/a), \tag{5.8}$$

причем суммирование ведется по всем положительным целым числам ν, меньшим $a/2$. Представление в таком виде более естественно, чем (5.7), поскольку коэффициенты $\cos(\pi\nu/a)$ образуют убывающую последовательность, и поэтому для больших n существенным является только первый член.

§ 6. СВЯЗЬ С ДИФФУЗИОННЫМИ ПРОЦЕССАМИ

Этот параграф посвящен неформальному обсуждению случайных блужданий, в которых длина δ отдельных шагов мала, но шаги столь быстро следуют друг за другом, что движение представляется практически непрерывным. Переход к пределу приводит к винеровскому процессу (броуновскому движению) и другим диффузионным процессам. Такая тесная связь между этими процессами и случайными блужданиями очень способствует пониманию обоих этих объектов [1]). Задачу здесь можно сформулировать и в математических, и в физических терминах.

Лучше всего начать с *неограниченного случайного блуждания*, начинающегося в нуле. В нем n-й шаг приводит частицу в положение $\mathbf{S}_n$, где $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ — сумма n независимых случайных величин, каждая из которых принимает значения $+1$ и -1 с вероятностями p и q соответственно. Поэтому

$$\mathbf{E}(\mathbf{S}_n)=(p-q)n, \quad \mathrm{Var}(\mathbf{S}_n)=4pqn. \tag{6.1}$$

На рис. 4 в гл. III,6 представлены первые 10 000 шагов такого случайного блуждания с $p=q=1/2$; чтобы уместить его график на

---

[1]) Этот подход был плодотворным и исторически. Он был сполна использован (хотя и с эвристических позиций) Л. Башелье, работа которого вдохновила А. Н. Колмогорова на развитие формальных оснований теории марковских процессов. См., в частности, Bachelier L., Calcul des probabilités, Paris, Gauthier-Villars, 1912.

странице книги, пришлось выбрать подходящие масштабы для обеих осей. Теперь мы пойдем дальше и представим себе запечатлевший случайное блуждание кинофильм. Предположим, что он должен длиться 1000 с (почти 17 минут). Для того чтобы это соответствовало миллиону шагов, необходимо, чтобы случайное блуждание происходило со скоростью один шаг в миллисекунду, и это фиксирует шкалу времени. Какую же величину шага должны мы выбрать для блуждания, чтобы быть вполне уверенными, что запись его поместится на экране заданной высоты? Итак, мы рассматриваем вместо $\mathbf{S}_n$ величины $\delta\mathbf{S}_n$, где δ означает длину отдельного шага. Теперь

$$\mathrm{E}(\delta\mathbf{S}_n) = (p-q)\delta n, \quad \mathrm{Var}(\delta\mathbf{S}_n) = 4pq\delta^2 n, \tag{6.2}$$

и из центральной предельной теоремы ясно, что наш фильм будет возможен только в том случае, если при $n = 1\,000\,000$ обе величины [1]) (6.2) будут меньше высоты экрана. (Здесь подразумевается, что мы пользуемся фиксированной единицей измерения, скажем дюймами или футами, как для экрана, так и для величины отдельных шагов.) Но если $p \neq q$ и δn сравнимо с высотой экрана, то $\delta^2 n$ будет неотличимо от нуля и фильм покажет нам линейное движение без видимых случайных флуктуаций. Характер случайного блуждания может быть выявлен лишь тогда, когда $\delta^2 n$ имеет умеренную положительную величину, а это возможно только при $p-q$, сравнимом с δ.

Если бы вопрос был чисто математическим, мы заключили бы, что желаемое графическое представление невозможно, если только не выполняется условие $p=q$, однако с физической точки зрения ситуация выглядит совершенно иначе. В броуновском движении мы наблюдаем двигающиеся случайным образом взвешенные в жидкости частицы, и естественно возникает вопрос: можно ли интерпретировать это движение как результат огромного числа столкновений с более мелкими частицами в жидкости? Конечно, предполагать, что столкновения происходят через равные интервалы времени и что каждое столкновение вызывает перемещение, в точности равное $\pm\delta$,— это сверхупрощение. Во всяком случае, для начала мы будем считать, что столкновения управляются испытаниями Бернулли, и посмотрим, совместимо ли с этой картиной наблюдаемое движение частиц. Из реальных наблюдений мы находим среднее смещение c и дисперсию D за единичный интервал времени. Обозначим через r (неизвестное) число столкновений за единицу времени. Тогда мы приближенно должны иметь

$$(p-q)\delta r = c, \quad 4pq\delta^2 r = D. \tag{6.3}$$

В моделирующем эксперименте не наблюдалось бы никаких случайных флуктуаций, если бы оба условия (6.3) не выполнялись при

[1]) Точнее, первая из них и корень квадратный из второй.— *Прим. перев.*

$D>0$. Вполне мыслим эксперимент с $p=0{,}6$ и $\delta r=1$, однако в нем дисперсия была бы столь малой, что движение казалось бы детерминированным: группа частиц, сначала близких друг к другу, так и оставалась бы вместе, как если бы была твердым телом.

По существу те же рассуждения применимы и ко многим другим явлениям в физике, экономике, теории обучения, эволюционной теории и т. д., когда медленные флуктуации состояния системы интерпретируются как результат огромного числа последовательных малых изменений, вызванных случайными воздействиями. Простая модель случайного блуждания не представляется реалистичной ни в одном конкретном случае, но, к счастью, ситуация здесь подобна той, что имеет место в центральной предельной теореме. Характер отдельных изменений не будет играть никакой роли уже при удивительно слабых условиях, ибо наблюдаемый эффект будет зависеть лишь от их математических ожиданий и дисперсий [1]). При таких обстоятельствах естественно принять в качестве универсального прототипа простую модель случайного блуждания.

Итак, в качестве подготовки к более глубокому изучению различных стохастических процессов естественно рассмотреть случайные блуждания, в которых длина δ отдельных шагов мала, число r шагов за единицу времени велико, а $p-q$ мало, причем соотношение между ними таково, что справедливо (6.3) (где c и $D>0$ — заданные постоянные). Понятия «мало» и «велико» строго не определены и в практических применениях должны варьироваться [2]).

Аналитически проблема формулируется следующим образом. Каждому выбору δ, r и p соответствует конкретное случайное блуждание. *Спрашивается, что произойдет в пределе, когда* $\delta\to 0$, $r\to\infty$, *и* $p\to 1/2$ *так, что*

$$(p-q)\delta r\to c,\quad 4pq\delta^2 r\to D. \tag{6.4}$$

Тут возможны два подхода. Если мы располагаем явными выражениями для нужных нам вероятностей, то можем непосредственно перейти к пределу. Мы продемонстрируем этот метод, поскольку он проливает новый свет на нормальное приближение и предельные теоремы, выведенные в гл. III. Однако возможности этого метода ограничены, ибо он не годится для обобщений. Более

---

[1]) Это так называемый принцип инвариантности, впервые установленный М. Донскером. Подробное изложение этого вопроса см. в книге Биллингсли П. Сходимость вероятностных мер.— М.: Наука, 1977.— *Прим. перев.*

[2]) Число молекулярных толчков в единицу времени находится за пределом воображения. В противоположность этому в эволюционной теории рассматриваются небольшие изменения от одного поколения к другому, и разделяющее два поколения время по обыденным стандартам отнюдь не является малым. Число рассматриваемых поколений также не является фантастическим, но может составлять многие тысячи. Здесь дело в том, что этот процесс протекает так, что изменения кажутся практически непрерывными, и диффузионная модель с непрерывным временем предпочтительнее модели случайного блуждания.

плодотворным будет начать с разностных уравнений, описывающих случайные блуждания, и вывести из них предельные дифференциальные уравнения. Оказывается, что эти дифференциальные уравнения описывают вполне определенные стохастические процессы, зависящие от непрерывного временно́го параметра. То же справедливо и относительно различных очевидных обобщений этих дифференциальных уравнений, так что второй метод приводит к важному общему классу диффузионных процессов.

Чтобы описать прямой метод в простейшем случае, будем по-прежнему обозначать через $\{\mathbf{S}_n\}$ обычное случайное блуждание с единичными шагами и положим

$$v_{k,\,n} = \mathbf{P}\,\{\mathbf{S}_n = k\}. \tag{6.5}$$

В нашем ускоренном блуждании n-й шаг произойдет в момент времени n/r, а положением частицы в этот момент будет $\mathbf{S}_n\delta = k\delta$. Мы интересуемся вероятностью найти частицу в данный момент t в окрестности заданной точки x и поэтому должны изучать асимптотическое поведение $v_{k,n}$, когда $k \to \infty$ и $n \to \infty$ таким образом, что $n/r \to t$ и $k\delta \to x$. Событие $\{\mathbf{S}_n = k\}$ требует, чтобы n и k были числами одинаковой четности, и происходит, когда ровно $(n+k)/2$ из первых n шагов направлены вправо. Стало быть, из теоремы Муавра — Лапласа мы заключаем, что при переходе к пределу

$$v_{k,\,n} \sim \frac{1}{\sqrt{2\pi npq}}\, e^{-[(n+k)/2 - np]^2/(2npq)} = \frac{1}{\sqrt{2\pi npq}}\, e^{-[k - n(p-q)]^2/(8npq)} \sim$$
$$\sim \frac{2\delta}{\sqrt{2\pi Dt}}\, e^{-(x-ct)^2/(2Dt)}, \tag{6.6}$$

где знак $\sim$ означает, что отношение величин стремится к единице. Поскольку $v_{k,\,n}$ есть вероятность того, что $\mathbf{S}_n\delta$ находится между $k\delta$ и $(k+2)\delta$, а этот интервал имеет длину 2δ, мы можем сказать, что отношение $v_{k,n}/(2\delta)$ локально измеряет вероятность на единицу длины, т. е. является *плотностью* вероятности. Из второго соотношения (6.6) вытекает, что отношение $v_{k,\,n}/(2\delta)$ стремится к

$$v(t,\ x) = \left(1/\sqrt{2\pi Dt}\right) e^{-(x-ct)^2/(2Dt)}. \tag{6.7}$$

Отсюда следует, что суммы вероятностей $v_{k,\,n}$ можно аппроксимировать интегралами от $v(t,\ x)$, и наш результат может быть переформулирован следующим образом: при выбранном нами переходе к пределу

$$\mathbf{P}\,\{\alpha < \mathbf{S}_n\delta < \beta\} \to \left(1/\sqrt{2\pi Dt}\right) \int\limits_{\alpha}^{\beta} e^{-(x-ct)^2/(2Dt)}\, dx. \tag{6.8}$$

Входящий сюда интеграл можно выразить через функцию нормального распределения $\mathfrak{N}$, и (6.8) в действительности является лишь вариантом записи предельной теоремы Муавра — Лапласа для биномиального распределения.

Более интересен подход, основанный на соответствующих разностных уравнениях. Если рассмотреть положение частицы при n-м и $(n+1)$-м испытаниях, то будет очевидно, что вероятности $v_{k,n}$ удовлетворяют разностным уравнениям

$$v_{k,\,n+1} = pv_{k-1,\,n} + qv_{k+1,\,n}. \tag{6.9}$$

Умножая это равенство на 2δ, мы видим (как следует из нашего предыдущего результата), что $v(t,\ x)$ должно быть *приближенным* решением разностного уравнения

$$v(t+r^{-1},\ x) = pv(t,\ x-\delta) + qv(t,\ x+\delta). \tag{6.10}$$

Поскольку v имеет непрерывные производные, мы можем разложить члены в этом равенстве в ряды Тейлора. Пользуясь приближением первого порядка в левой части и второго порядка — в правой, мы получаем (после сокращения главных членов)

$$\frac{\partial v(t,\ x)}{\partial t} = (q-p)\,\delta r\,\frac{\partial v(t,\ x)}{\partial x} + \frac{1}{2}\,\delta^2 r\,\frac{\partial^2 v(t,\ x)}{\partial x^2} + \ldots\ . \tag{6.11}$$

При переходе к пределу опущенные нами члены устремятся к нулю и (6.11) превратится в

$$\frac{\partial v(t,\ x)}{\partial t} = -c\,\frac{\partial v(t,\ x)}{\partial x} + \frac{1}{2}\,D\,\frac{\partial^2 v(t,\ x)}{\partial x^2}. \tag{6.12}$$

Это частный вид *уравнения диффузии*, известный также под названием *уравнения Фоккера — Планка*. Наши выкладки были исключительно формальными и эвристическими, однако не будет неожиданным, что функция v из (6.7) действительно удовлетворяет дифференциальному уравнению (6.12). Более того, можно показать, что (6.7) представляет собой единственное решение уравнения диффузии, обладающее очевидными свойствами, обусловленными вероятностной интерпретацией.

Уравнение диффузии (6.12) можно обобщить, допустив, что коэффициенты c и D зависят от x и t. Кроме того, у него есть очевидные аналоги в многомерном случае, и все эти обобщения могут быть выведены непосредственно из общих вероятностных постулатов. Этот вопрос будет обсуждаться в гл. X тома 2; здесь же мы должны довольствоваться краткими и эвристическими указаниями на связь между случайными блужданиями и общей теорией диффузии.

В качестве второго примера мы рассмотрим вероятности разорения $u_{z,\,n}$, обсуждавшиеся в предыдущих двух параграфах. Отвечающие им разностные уравнения (4.1) отличаются от (6.9) тем, что коэффициенты p и q в них меняются местами[1]). Формальные выкладки, указанные в (6.11), приводят теперь к уравнениям диффузии, получающимся из (6.12) при замене $-c$ на c. При нашей предельной процедуре мы переходим от вероятностей $u_{z,n}$ к удовлетворяющей модифицированному уравнению диффузии функции $u(t, \xi)$, которая по вероятностному смыслу подобна $u_{z,n}$: в диффузионном процессе, начинающемся в точке $\xi>0$, вероятность того, что частица попадет в нуль *до* момента достижения точки $\alpha>\xi$ и что это событие произойдет в интервале времени $t_1<t<t_2$, дается интегралом от $u(t, \xi)$ по этому интервалу.

Формальные вычисления выглядят следующим образом. Для $u_{z,n}$ у нас есть явное выражение (5.8). Поскольку z и n должны быть числами одинаковой четности, то $u_{z,n}$ соответствует интервалу между n/r и $(n+2)/r$, и мы должны вычислить предел отношения $u_{z,n}r/2$ при $r\to\infty$ и $\delta\to 0$ в соответствии с (6.4). Длина интервала a и начальное положение z должны быть подобраны так, чтобы получить в пределе α и ξ. Таким образом, $z\sim\xi/\delta$ и $a\sim\alpha/\delta$. Теперь легко получить пределы для отдельных множителей в (5.8).

Из первого соотношения (6.4) мы получаем, что $2p\sim 1+c\delta/D$ и $2q\sim 1-c\delta/D$; из второго соотношения (6.4) видим, что $\delta^2 r\to D$. Стало быть,

$$(4pq)^{n/2}(q/p)^{z/2}\sim(1-c^2\delta^2/D^2)^{rt/2}(1-2c\delta/D)^{\xi/(2\delta)}\sim e^{-c^2t/(2D)}\cdot e^{-c\xi/D}. \tag{6.13}$$

Аналогично для фиксированного ν

$$\cos^n(\nu\pi\delta/\alpha)\sim[1-\nu^2\pi^2\delta^2/(2\alpha^2)]^{tr}\sim e^{-\nu^2\pi^2Dt/(2\alpha^2)}. \tag{6.14}$$

Наконец, $\sin\nu\pi\delta/\alpha\sim\nu\pi\delta/\alpha$. Подстановка полученных выражений в (5.8) формально приводит к результату

$$u(t,\xi)=\pi D\alpha^{-2}e^{-(ct+2\xi)c/(2D)}\sum_{\nu=1}^{\infty}\nu e^{-\nu^2\pi^2Dt/(2\alpha^2)}\sin(\pi\xi\nu/\alpha). \tag{6.15}$$

(Поскольку этот ряд сходится равномерно, нетрудно показать законность наших формальных выкладок.) В физической теории диффузии формула (6.15) называется *формулой Фюрта для первого прохождения*. (Относительно предельного случая $\alpha=\infty$ см. задачу 14. Другую форму (6.15) см. в задаче 22.)

---

[1]) Причина этого состоит в том, что в $u_{z,n}$ переменная z означает *начальное* положение, тогда как вероятность $v_{k,n}$ относится к положению в текущий момент времени. В терминологии, которая будет введена в томе 2, вероятности, зависящие от начального положения, удовлетворяют *обратным* уравнениям, а остальные — *прямым* уравнениям (или уравнениям Фоккера — Планка). В физике последние иногда называются *уравнениями неразрывности*. С такой же ситуацией мы столкнемся в гл. XVII.

§ 7*). СЛУЧАЙНЫЕ БЛУЖДАНИЯ НА ПЛОСКОСТИ И В ПРОСТРАНСТВЕ

В двумерном случайном блуждании частица передвигается единичными шагами в четырех направлениях, параллельных осям x и y. Для выходящей из начала координат частицы возможными положениями будут все точки плоскости с целочисленными координатами. Каждое положение имеет четыре *соседних*. Аналогично в трех измерениях каждое положение имеет шесть соседних. Случайное блуждание определяется выбором соответствующих четырех или шести вероятностей. Для простоты мы будем рассматривать только *симметричный* случай, в котором все направления равновероятны. Сложность задач здесь намного выше, чем в одномерном случае, ибо теперь области, в которых происходит движение частицы, могут иметь произвольную форму и роль экранов играют границы сложного вида.

Мы начнем с интересной теоремы, принадлежащей Пойа [1]).

Теорема. *В одномерном и двумерном симметричных случайных блужданиях частица с вероятностью единица рано или поздно (а поэтому и бесконечное число раз) возвратится в свое начальное положение. Однако в трехмерном случае вероятность этого меньше единицы и равна примерно* 0,35. (Среднее число возвращений равно тогда $0{,}65\sum k(0{,}35)^k = 0{,}35/0{,}65 \approx 0{,}53$.)

Прежде чем доказывать эту теорему, приведем две другие ее формулировки, также принадлежащие Пойа. Прежде всего, почти очевидно, что из этой теоремы вытекает, что *в одномерном и в двумерном случаях с вероятностью* 1 *частица бесконечное число раз пройдет через каждое возможное положение*, однако в трехмерном случае это неверно. Таким образом, для двух измерений в известном смысле справедливо утверждение «все дороги ведут в Рим».

С другой стороны, рассмотрим *две* частицы, совершающие независимые симметричные случайные блуждания, причем перемещения их происходят одновременно. Встретятся ли они когда-нибудь? Чтобы упростить изложение, мы определим *расстояние* между двумя возможными положениями как наименьшее число шагов, ведущих из одного положения в другое. (Это расстояние равно сумме абсолютных величин разностей координат.) Если две частицы продвигаются на один шаг каждая, то расстояние между ними либо останется тем же, либо изменится на две единицы, и поэтому расстояние между частицами будет либо всегда четным, либо всегда нечетным. Во втором случае наши две частицы никогда не смогут занять одно и то же положение. В первом случае легко видеть, что веро-

*) Этот параграф посвящен специальному вопросу и может быть опущен при первом чтении.

[1]) Polya G., Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt in Strassennetz, Mathematische Annalen, 84 (1921), 149—160. Численное значение 0,35 было найдено в работе Мак-Кри и Уиппла (McCrea W. H., Whipple F. J. W., Random paths in two and three dimensions, Proceedings of the Royal Society of Edinburgh, 60 (1940), 281—298).

ятность их встречи на n-м шаге равна вероятности того, что первая частица за $2n$ шагов достигнет начального положения второй частицы.

Следовательно, наша теорема утверждает, что в двумерном (но не в трехмерном) случае две частицы наверняка бесконечное число раз будут занимать одно и то же положение. Если начальное расстояние между двумя частицами нечетно, то аналогичное рассуждение показывает, что они будут бесконечно много раз занимать соседние положения. Если назвать это встречей, то наша теорема утверждает, что *в одномерном и двумерном случаях две частицы с достоверностью встретятся бесконечное число раз, однако в трехмерном случае они с положительной вероятностью никогда не встретятся.*

Доказательство. Для одномерного случая теорема была доказана в примере гл. XIII, 4, б) при помощи метода рекуррентных событий. Доказательство для двумерного и трехмерного случаев проводится примерно так же. Пусть u_n — вероятность того, что n-е испытание приведет частицу в начальное положение. Согласно теореме 2 гл. XIII,3, нам надо доказать, что в двумерном случае ряд $\sum\limits_{n\geqslant 1} u_n$ расходится, тогда как в трехмерном случае $\sum\limits_{n\geqslant 1} u_n \approx 0,53$. В двумерном случае возвращение в начальное положение возможно только тогда, когда количества шагов в положительных направлениях осей x и y равны соответственно количествам шагов в отрицательных направлениях этих осей. Следовательно, $u_n=0$, если n нечетно, и (мы пользуемся полиномиальным распределением; см. формулу (9.2) гл. VI)

$$u_{2n}=\frac{1}{4^{2n}}\sum_{k=0}^{n}\frac{(2n)!}{k!\,k!\,(n-k)!\,(n-k)!}=\frac{1}{4^{2n}}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}^2. \qquad (7.1)$$

Согласно формуле (12.11) гл. II, правая часть равна $4^{-2n}\binom{2n}{n}^2$. Формула Стирлинга показывает теперь, что u_{2n} является величиной порядка $1/n$, так что ряд $\sum u_{2n}$, как и утверждалось, расходится.

Для трехмерного случая аналогично находим

$$u_{2n}=6^{-2n}\sum_{j,k}\frac{(2n)!}{j!\,j!\,k!\,k!\,(n-j-k)!\,(n-j-k)!}, \qquad (7.2)$$

где суммирование проводится по всем j и k, таким, что $j+k\leqslant n$. Легко убедиться, что

$$u_{2n}=\frac{1}{2^{2n}}\binom{2n}{n}\sum_{j,k}\left\{\frac{1}{3^n}\,\frac{n!}{j!\,k!\,(n-j-k)!}\right\}^2. \qquad (7.3)$$

Внутри скобок стоят члены триномиального распределения, и мы знаем, что в сумме они составляют единицу. Следовательно, сумма их квадратов не превосходит максимального из членов в скобках, а последнему соответствуют значения j и k, равные примерно $n/3$. Формула Стирлинга показывает, что этот максимум имеет порядок величины n^{-1}, и поэтому u_{2n} имеет величину $1/\sqrt{n^3}$, так что ряд $\sum u_{2n}$, как и утверждалось, сходится. ►

Мы завершаем этот параграф еще одной задачей, которая обобщает понятие *поглощающего экрана*. Рассмотрим двумерный случай, в котором вместо интервала $0 \leqslant x \leqslant a$ у нас имеется плоская область D, т. е. набор точек с целочисленными координатами. Каждая точка имеет четыре соседних, однако у некоторых точек из D одна или более соседних точек лежат вне D. Такие точки образуют границу области D, а все остальные ее точки называются внутренними.

В одномерном случае границу образуют два экрана, и наша задача состояла в нахождении вероятности того, что, выходя из z, частица достигнет граничной точки $x=0$ до попадания в точку $x=a$. Теперь по аналогии мы интересуемся вероятностью того, что частица достигнет определенного участка границы до того момента, как она попадет в какую-либо граничную точку, не принадлежащую этому участку. Это означает, что мы разбиваем все граничные точки на два множества B' и B''. Если (x, y) является внутренней точкой, то мы ищем вероятность того, что выходящая из (x, y) частица достигнет какой-нибудь точки из B' ранее, чем любой точки из B''. В частности, если B' состоит из одной-единственной точки, то $u(x, y)$ будет вероятностью того, что частица рано или поздно будет поглощена в этой точке.

Пусть (x, y) — некоторая внутренняя точка. Первый шаг переводит частицу из (x, y) в одну из четырех соседних ей точек $(x \pm 1, y)$, $(x, y \pm 1)$, и если все четыре из них суть внутренние точки, то, очевидно,

$$u(x, y) = (1/4)[u(x+1, y) + u(x-1, y) + u(x, y+1) + u(x, y-1)]. \quad (7.4)$$

Это и есть разностное уравнение, которое появляется здесь вместо (2.1) (при $p=q=1/2$). Если $(x+1, y)$ является граничной точкой, то ее вклад $u(x+1, y)$ следует заменить на 1 и 0 в соответствии с тем, принадлежит ли $(x+1, y)$ множеству B' или B''. *Следовательно*, (7.4) *будет справедливо для всех внутренних точек, если мы будем считать, что* $u(\xi, \eta)=1$ *для граничной точки* (ξ, η) *из* B' *и* $u(\xi, \eta)=0$ *для граничной точки из* B''. *Это соглашение играет здесь роль граничных условий* (2.2).

Итак, мы получили систему линейных уравнений (7.4) для неизвестных $u(x, y)$; каждой внутренней точке соответствуют в ней одно неизвестное и одно уравнение. Эта система неоднородна, поскольку имеется хотя бы одна граничная точка (ξ, η) из B', привносящая в правую часть вклад 1/4. Если область D конечна, то уравнений у нас столько же, сколько и неизвестных, и известно, что такая система имеет единственное решение тогда и только тогда, когда соответствующая однородная система (с $u(\xi, \eta)=0$ для всех граничных точек) не имеет ненулевых решений. Но $u(x, y)$ является средним арифметическим четырех соседних значений $u(x\pm1, y)$, $u(x, y\pm1)$ и не может быть больше (а также меньше) всех этих значений. Иначе говоря, во внутренних точках $u(x, y)$ не имеет ни максимума, ни минимума в строгом смысле, и наибольшее и наименьшее ее значения достигаются в граничных точках. Следовательно, если все граничные значения равны нулю, то таковы же значения $u(x, y)$ и во всех внутренних точках, что доказывает существование и единственность решения (7.4). Поскольку граничные значения равны 0 или 1, все значения $u(x, y)$ лежат между 0 и 1, как и требуется для вероятностей. Как мы увидим из общей теоремы для бесконечных цепей Маркова, эти утверждения справедливы и для случая бесконечных областей [1]).

§ 8*). ОБОБЩЕННОЕ ОДНОМЕРНОЕ СЛУЧАЙНОЕ БЛУЖДАНИЕ (ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ)

Вернемся теперь к одномерному случаю, отказавшись от ограничения, что частица движется единичными шагами. Вместо этого *частица на каждом шаге будет иметь вероятность p_k перейти из точки x в точку $x+k$*, где целое k может быть нулем, положительным или отрицательным числом. Мы будем рассматривать следующую *задачу о разорении. Частица выходит из точки z, такой, что $0<z<a$; мы ищем вероятность u_z того, что частица достигает какой-либо точки $x\leqslant0$ раньше, чем любой из точек $x\geqslant a$.* Иначе говоря, после n-го испытания частица будет находиться в точке $x=z+\mathbf{X}_1+\mathbf{X}_2+\ldots\mathbf{X}_n$ на оси x, где $\{\mathbf{X}_k\}$ — взаимно независимые случайные величины с общим распределением $\{p_\nu\}$; процесс останавливается, когда впервые будет выполнено либо неравенство $\mathbf{X}_1+\ldots+\mathbf{X}_n\leqslant-z$, либо неравенство $\mathbf{X}_1+\ldots+\mathbf{X}_n\geqslant a-z$.

Эта задача привлекла широкий интерес в связи с *последовательным анализом*. Там $\mathbf{X}_k$ представляют некоторые характеристики выборок или наблюдений. Измерения производятся до тех пор, пока сумма $\mathbf{X}_1+\ldots+\mathbf{X}_k$ не выйдет за пределы двух заранее определенных границ (наших $-z$ и $a-z$). В первом случае эта процедура приводит (если следовать специальной терминологии) к тому, что гипотеза отклоняется, а во втором случае — к тому, что она принимается [2]).

---

[1]) Явные решения известны лишь для немногих случаев и имеют, как правило, очень сложный вид. Решения для случаев прямоугольных областей, бесконечных полос и т. д. можно найти в указанной в предыдущем примечании статье Мак-Кри и Уиппла.

*) Материал этого параграфа в дальнейшем не используется.

[2]) Общая теория последовательных статистических процедур была развита Абрахамом Вальдом во время второй мировой войны в связи с важными практическими задачами. Ее современное изложение можно найти в ряде курсов математической статистики. Описанная в примере схема Бартки относится к 1943 г. и яв-

Пример. а) В качестве иллюстрации возьмем предложенную Бартки многовыборочную схему контроля. Чтобы проверить партию изделий, производятся выборки объема N, которые и подвергаются экспертизе. Предполагается, что эти выборки стохастически независимы и что число дефектных изделий в каждой из них имеет одно и то же биномиальное распределение. Допускается наличие одного дефектного изделия на выборку, и мы положим $\mathbf{X}_k+1$ равным числу дефектных элементов k-й выборки. Тогда для $k\geqslant 0$

$$p_k=\binom{N}{k+1}p^{k+1}q^{N-k-1},$$

и $p_{-1}=q^N$, $p_x=0$ для $x<-1$.

Процедурное правило состоит в следующем. Извлекается первая выборка, и если она не содержит ни одного дефектного изделия, то принимается вся партия; если число дефектных образцов в ней превосходит a, то вся партия бракуется. В любом из этих случаев процесс прекращается. Однако если число z дефектных изделий лежит в интервале $1\leqslant z\leqslant a$, то процедура извлечения выборок продолжается описанным образом до тех пор, пока сумма величин $\mathbf{X}_k$ остается между 1 и a. Рано или поздно она либо обратится в нуль, и тогда партия принимается, либо станет $\geqslant a$, и тогда партия бракуется. ►

Предположим, не теряя общности, что шаги возможны как в положительном, так и в отрицательном направлениях. В противном случае мы имели бы либо $u_z=0$, либо $u_z=1$ для всех z. Вероятность разорения на *первом* шаге, очевидно, равна

$$r_z=p_{-z}+p_{-z-1}+p_{-z-2}+\dots \tag{8.1}$$

(величина, которая может быть равна нулю). Случайное блуждание продолжается только в том случае, когда частица перемещается в положение x, $0<x<a$; вероятность скачка из z в x равна p_{x-z}, и вероятность последующего разорения составляет тогда u_x. Поэтому

$$u_z=\sum_{x=1}^{a-1}u_xp_{x-z}+r_z. \tag{8.2}$$

Мы снова имеем здесь $a-1$ линейных уравнений для $a-1$ неизвестных u_z. Эта система неоднородна, поскольку по крайней мере при $z=1$ вероятность r_1 отлична от нуля (ибо возможны шаги в отрицательном направлении). Чтобы доказать, что линейная система (8.2) имеет единственное решение, мы должны доказать, что соот-

---

ляется, по-видимому, самой первой последовательной процедурой, предложенной в литературе. [Подробнее об этом см. книги Вальд А. Последовательный анализ.— М.: Физматгиз, 1960; Ширяев А. Н. Статистический последовательный анализ.— 2-е изд., перераб.— М.: Наука, 1976.— *Перев.*]

ветствующая однородная система

$$u_z = \sum_{x=1}^{a-1} u_x p_{x-z} \tag{8.3}$$

не имеет ненулевых решений. Предположим (для сокращения числа появляющихся в доказательстве индексов), что $p_{-1} \neq 0$ (однако те же рассуждения применимы и при иных положительных членах с отрицательным индексом). Допустим теперь, что u_z удовлетворяет (8.3), и обозначим через M максимум значений u_z. Пусть $u_r = M$. Поскольку сумма всех коэффициентов p_{x-z} в (8.3) равна единице, то равенство это возможно при $z=r$ только тогда, когда те из u_x, которые действительно входят в правую часть (с положительными коэффициентами), равны M и если коэффициенты при них в сумме дают единицу. Следовательно, $u_{r-1}=M$, и, рассуждая таким же образом, получаем, что $u_{r-2}=u_{r-3}=\ldots=u_1=M$. Однако при $z=1$ сумма коэффициентов p_{x-r} в (8.3) меньше единицы, так что M должно быть равно нулю.

Отсюда следует, что (8.2) имеет единственное решение, так что наша задача вполне определена. Снова упростим запись, введя граничные условия

$$u_x = \begin{cases} 1, & \text{если } x \leqslant 0, \\ 0, & \text{если } x \geqslant a. \end{cases} \tag{8.4}$$

Тогда (8.2) можно переписать в виде

$$u_z = \sum u_x p_{x-z}, \tag{8.5}$$

и суммирование теперь ведется уже по всем x (слагаемые при $x \geqslant a$ не вносят никакого вклада в силу второго условия из (8.4); в силу первого условия слагаемые для $x \leqslant 0$ дают в сумме r_z).

При больших a непосредственное решение $a-1$ линейного уравнения громоздко, и лучше воспользоваться *методом частных решений*, аналогичным процедуре § 2. Он хорошо работает, когда распределение вероятностей $\{p_k\}$ содержит сравнительно немного положительных членов. Предположим, что отличны от нуля только p_k при $-\nu \leqslant k \leqslant \mu$, так что наибольшие возможные скачки в положительном и отрицательном направлениях суть μ и ν соответственно. *Характеристическое уравнение*

$$\sum p_k \sigma^k = 1 \tag{8.6}$$

эквивалентно алгебраическому уравнению степени $\nu+\mu$. Если σ — корень (8.6), то $u_z = \sigma^z$ будет формальным решением (8.5) для всех z, однако это решение не удовлетворяет граничным условиям (8.4). Если (8.6) имеет $\mu+\nu$ различных корней $\sigma_1, \sigma_2, \ldots$, то линейная комбинация

$$u_z = \sum A_k \sigma_k^z \tag{8.7}$$

снова будет формальным решением (8.5) при всех z, и мы должны подобрать постоянные A_k так, чтобы удовлетворялись граничные условия. Далее, при $0<z<a$ в (8.5) входят лишь значения x из интервала $-\nu+1\leqslant x\leqslant a+\mu-1$. Поэтому достаточно удовлетворить граничным условиям (8.4) только для $x=0, -1, -2, \ldots, -\nu+1$ и $x=a, a+1, \ldots, a+\mu-1$, так что всего мы имеем $\mu+\nu$ условий. Если σ_k — двойной корень (8.6), то мы теряем одну постоянную, но в этом случае легко видеть, что еще одним формальным решением будет $u_z=z\sigma_k^z$. В любом случае $\mu+\nu$ граничных условий определяют $\mu+\nu$ произвольных постоянных.

Пример. б) Предположим, что каждый отдельный шаг переводит частицу в одно из четырех ближайших положений и что $p_{-2}=p_{-1}=p_1=p_2=1/4$. Характеристическое уравнение (8.6) здесь имеет вид $\sigma^{-2}+\sigma^{-1}+\sigma+\sigma^2=4$. Положим $t=\sigma+\sigma^{-1}$: при такой подстановке наше уравнение превращается в $t^2+t=6$, корни которого суть $t=2$ и $t=-3$. Разрешая $t=\sigma+\sigma^{-1}$ относительно σ, находим четыре корня

$$\sigma_1=\sigma_2=1, \quad \sigma_3=\frac{-3+\sqrt{5}}{2}=\sigma_4^{-1}, \quad \sigma_4=\frac{-3-\sqrt{5}}{2}=\sigma_3^{-1}. \tag{8.8}$$

Поскольку σ_1 — двойной корень, общим решением уравнения (8.5) в нашей случае является

$$u_z=A_1+A_2z+A_3\sigma_3^z+A_4\sigma_4^z. \tag{8.9}$$

Граничные условия $u_0=u_{-1}=1$ и $u_a=u_{a+1}=0$ приводят к четырем линейным уравнениям для коэффициентов A_j и к окончательному решению

$$u_z=1-\frac{z}{a}+\frac{(2z-a)(\sigma_3^a-\sigma_4^a)-a(\sigma_3^{2z-a}-\sigma_4^{2z-a})}{a\{(a+2)(\sigma_3^a-\sigma_4^a)-a(\sigma_3^{a+2}-\sigma_4^{a+2})\}}. \tag{8.10}$$

►

Численные приближения. Найти все корни уравнения (8.6) обычно бывает весьма затруднительно, однако для них можно получить вполне удовлетворительные приближения удивительно простым способом. Рассмотрим сперва случай, когда распределение вероятностей $\{p_k\}$ имеет нулевое среднее. Тогда характеристическое уравнение (8.6) имеет двойной корень в точке $\sigma=1$ и $A+Bz$ является формальным решением (8.5). Конечно, двух постоянных A и B недостаточно для того, чтобы удовлетворялись $\mu+\nu$ граничных условий (8.4). Однако если мы определим A и B так, чтобы $A+Bz$ обращалось в нуль при $z=a+\mu-1$ и было равно 1 при $z=0$, то $A+Bx\geqslant 1$ при $x\leqslant 0$ и $A+Bx\geqslant 0$ при $a\leqslant x<a+\mu$, так что $A+Bz$ будет удовлетворять граничным условиям (8.4), в которых знак $=$ заменен на $\geqslant$. Следовательно, разность $A+Bz-u_z$ будет формальным решением (8.5) с неотрицательными граничными значениями, и, стало быть, $A+Bz-u_z\geqslant 0$. Подобным же образом мы можем получить и нижнюю границу для u_z, определяя A и B так, чтобы $A+Bz$ обращалось в нуль при $z=a$ и в единицу при $z=-\nu+1$. Следовательно,

$$\frac{a-z}{a+\nu-1}\leqslant u_z\leqslant\frac{a+\mu-z-1}{a+\mu-1}. \tag{8.11}$$

Эта оценка очень точна, когда a велико по сравнению с $\mu+\nu$. [Конечно, приближение $u_z\approx(1-z/a)$ лучше, однако оно не дает точных границ для u_z.]

Рассмотрим теперь общий случай, когда среднее распределения $\{p_k\}$ отлично от нуля. Тогда характеристическое уравнение (8.6) имеет в точке $\sigma=1$ простой корень. Левая часть (8.6) стремится к ∞ при $\sigma\to 0$ и при $\sigma\to\infty$. При положительных σ кривая $y=\sum p_k\sigma^k$ непрерывна и выпукла вниз, и поскольку она пересекает прямую $y=1$ при $\sigma=1$, то она пересекает ее еще ровно в одной точке. Стало быть, характеристическое уравнение (8.6) имеет ровно два положительных корня, 1 и σ_1. Отсюда, как мы уже знаем, следует, что $A+B\sigma_1^z$ является формальным решением уравнения (8.5), и мы можем применить наши предыдущие рассуждения к этому решению вместо $A+Bz$. Мы находим, что в этом случае

$$\frac{\sigma_1^a-\sigma_1^z}{\sigma_1^a-\sigma_1^{-\nu+1}}\leqslant u_z\leqslant\frac{\sigma_1^{a+\mu-1}-\sigma_1^z}{\sigma_1^{a+\mu-1}-1}, \tag{8.12}$$

так что справедлива следующая теорема.

Теорема. *Решение нашей задачи о разорении удовлетворяет неравенствам* (8.11), *если* $\{p_k\}$ *имеет нулевое среднее, и неравенствам* (8.12) *в противном случае. Здесь* σ_1 *есть единственный отличный от нуля положительный корень* (8.6), *а* μ *и* $-\nu$ *равны соответственно наибольшему и наименьшему из индексов* k, *при которых* $p_k\neq 0$.

Пусть $m=\sum kp_k$ — *математическое ожидание выигрыша* в единичном испытании (или математическое ожидание длины одного шага). Из (8.6) легко видеть, что $\sigma_1>1$ при $m<0$ и $\sigma_1<1$ при $m>0$. Полагая $a\to\infty$, мы заключаем из нашей теоремы, что *в игре против бесконечно богатого соперника вероятность окончательного разорения равна единице тогда и только тогда, когда* $m\leqslant 0$.

Продолжительность игры можно исследовать аналогичными методами (ср. задачу 9).

§ 9. ЗАДАЧИ

Замечание. Задачи 1—4 относятся только к § 2 и не требуют никаких вычислений.

1. Найти вероятность того, что в случайном блуждании, начинающемся в нуле, точка $x=a>0$ будет достигаться раньше точки $x=-b<0$.

2. Доказать, что (в обозначениях § 2)

а) в случайном блуждании, начинающемся в нуле, вероятность достижения точки $x=a>0$ до возвращения в нуль равна $p(1-q_1)$;

б) в случайном блуждании, начинающемся в точке $x=a>0$, вероятность достижения нуля до возвращения в начальную точку равна qq_{a-1}.

3. Вывести из предыдущей задачи, что если $q\geqslant p$, то в начинающемся в нуле случайном блуждании число попаданий в точку $x=a>0$ до первого возвращения в нуль имеет геометрическое распределение со знаменателем $1-qq_{a-1}$. (Почему необходимо условие $q\geqslant p$?)

4. Используя результаты, полученные в двух предыдущих задачах, доказать следующую теорему [1]). *Математическое ожидание числа попаданий в точку* $x=a>0$ *до первого возвращения в нуль равно* $(p/q)^a$ *при* $p<q$ *и равно* 1 *при* $p=q$.

5. Рассмотреть задачу о разорении из § 2, 3 для случая видоизмененного

[1]) Поистине удивительные следствия этого результата проявляются лучше всего применительно к безобидным играм. Симметричная монета подбрасывается до тех пор, пока впервые не сравняются суммарные числа гербов и решеток. Игрок получает один цент каждый раз, когда суммарное число гербов будет превосходить суммарное число решеток на m. *«Справедливая плата за вступление в игру» равна* 1 *центу независимо от* m.

Иное (элементарное) доказательство см. в задачах 1 и 2 гл. XII, 10 тома 2.

случайного блуждания, в котором частица либо совершает единичный шаг вправо или влево, либо остается на месте с вероятностями α, β и γ соответственно ($\alpha+\beta+\gamma=1$). (В игровой терминологии это означает, что партия может окончиться вничью.)

6. Рассмотреть задачу о разорении из § 2, 3 для случая, когда в нуле помещен *упругий* экран (определение см. в § 1). Разностные уравнения для вероятности разорения (поглощения в нуле) и для средней продолжительности игры будут такими же, но уже с другими граничными условиями.

7. Частица при каждом шаге перемещается на две единицы направо или на одну единицу налево с вероятностями p и q соответственно ($p+q=1$). Найти вероятность q_z того, что, выходя из точки $z>0$, частица когда-нибудь достигнет нуля. (Это задача о разорении при игре с бесконечно богатым противником.)

Указание. Аналог уравнения (2.1) приводит к кубическому уравнению, имеющему частное решение $q_z=1$ и два частных решения вида λ^z, где λ удовлетворяет квадратному уравнению.

8. *Продолжение* [1]). Показать, что q_1 равно вероятности того, что в последовательности испытаний Бернулли суммарное число неудач когда-нибудь будет вдвое больше суммарного числа успехов.

(При $p=q$ эта вероятность равна $(\sqrt{5}-1)/2$.)

9. В задаче § 8 для обобщенного случайного блуждания по аналогии с (8.1) положим $\rho_z=p_{a-z}+p_{a+1-z}+p_{a+2-z}+\ldots$ и обозначим через $d_{z,n}$ вероятность того, что игра продлится ровно n шагов. Показать, что при $n\geqslant 1$

$$d_{z,\,n+1}=\sum_{x=1}^{a-1} d_{x,\,n}p_{x-z},$$

где $d_{z,1}=r_z+\rho_z$. Исходя из этого, доказать, что производящая функция $d_z(\sigma)=\sum d_{z,n}\sigma^n$ является решением системы линейных уравнений

$$\sigma^{-1}d_z(\sigma)-\sum_{x=1}^{a-1} d_x(\sigma)\,p_{x-z}=r_z+\rho_z.$$

Дифференцируя, вывести отсюда, что средняя продолжительность e_z игры будет решением системы уравнений

$$e_z-\sum_{x=1}^{a-1} e_xp_{x-z}=1.$$

10. Пусть $w_{z,n}(x)$ — вероятность того, что в случайном блуждании с начальным положением z и *поглощающими* экранами в точках $x=0$ и $x=a$ n-й шаг приводит частицу в точку x. Найти разностные уравнения и граничные условия, которым удовлетворяет $w_{z,n}(x)$.

11. *Продолжение.* Видоизменить граничные условия на случай двух *отражающих* экранов (т. е. упругих экранов с $\delta=1$).

12. Пусть при симметричном случайном блуждании ($p=q$) частица может находиться в точках $1, 2, \ldots, a-1$. В нуле расположен поглощающий экран, а на другом конце интервала — отражающий. Найти производящую функцию для времени ожидания поглощения.

---

[1]) Эта задача была сформулирована Д. Дж. Ньюменом. То, что ее решение является простым следствием предыдущей задачи, заметил У. Во. Читатель может попробовать воспользоваться аналогичным подходом для более общих задач, в которых множитель 2 заменяется каким-то другим рациональным числом. Решение, основанное на иных идеях, предложил Дж. С. Фрэйм; см. Frame J. S., Solution to problem 4864, Amer. Math. Monthly, 67 (1960), 700—702.

13. *Другой вариант записи вероятностей первого достижения.* В явной формуле (5.7) для вероятностей разорения устремим a к ∞. Показать, что в результате получится

$$u_{z,n}=2^n p^{(n-z)/2} q^{(n+z)/2} \int_0^1 \cos^{n-1} \pi x \cdot \sin \pi x \cdot \sin \pi x z \cdot dx.$$

Следовательно, эта формула должна быть эквивалентна (4.14). Проверить это, показав, что удовлетворяются соответствующие разностные уравнения и граничные условия.

14. *Продолжение. Моменты первого достижения в диффузии.* Показать, что при описанном в § 6 предельном переходе из последней формулы получается выражение

$$(z/\sqrt{2\pi D t^3})\, e^{-(z+ct)^2/(2Dt)}$$

для плотности вероятности для времени ожидания поглощения в нуле в диффузии, начинающейся из точки $z>0$. При $p=q$ этот результат эквивалентен предельной теореме 3 из гл. III, 7.

Замечание. В следующих ниже задачах $v_{x,n}$ означает вероятность (6.1) того, что в начинающемся в нуле неограниченном случайном блуждании частица на n-м шаге попадает в положение x. Принцип отражения (гл. III, 1) приводит к иной интерпретации.

15. *Метод отражения*[1]. Пусть $p=q=1/2$. Обозначим через $u_{z,n}(x)$ вероятность того, что в случайном блуждании на $(0, \infty)$ с поглощающим экраном в нуле и начальным положением $z>0$ n-й шаг приведет частицу в положение $x>0$. Показать, что $u_{z,n}(x)=v_{x-z,n}-v_{x+z,n}$. (*Указание.* Показать, что разностное уравнение, соответствующее (4.1), и подходящие граничные условия в этом случае удовлетворяются.)

16. *Продолжение.* Доказать, что если в точке $x=0$ помещен *отражающий* экран, то

$$u_{z,n}(x)=v_{x-z,n}+v_{x+z-1,n}.$$

17. *Продолжение.* Доказать, что если случайное блуждание ограничено интервалом $(0, a)$ и оба экрана являются *поглощающими*, то

$$u_{z,n}(x)=\sum_k \{v_{x-z-2ka,n}-v_{x+z-2ka,n}\}, \tag{9.1}$$

где суммирование производится по всем k, положительным и отрицательным (лишь конечное число слагаемых отлично от нуля). Доказать, что если оба экрана являются *отражающими*, то уравнение (9.1) остается в силе при замене *минуса* на *плюс* и $x+z$ на $x+z-1$.

18. *Распределение максимумов.* Пусть $\mathbf{M}_n$ — максимальная абсцисса частицы за первые n шагов симметричного неограниченного случайного блуждания,

---

[1]) Задачи 15—17 являются примерами применения *метода отражения*. Член $v_{x-z,n}$ соответствует частице в неограниченном случайном блуждании, а $v_{x+z,n}$ — ее «отражению». В (9.1) используются отраженные точки, выходящие из различных положений, полученных повторными отражениями от обеих границ. В задачах 20 и 21 при помощи производящих функций мы получаем общий результат для несимметричного случайного блуждания. В теории дифференциальных уравнений метод отражений всегда связывают с именем Кельвина. В вероятностной литературе эквивалентный принцип приписывается Д. Андрэ. (См. примечание к лемме в гл. III, 1).

начинающегося в нуле. Используя результат задачи 15, показать, что

$$\mathbf{P}\{\mathbf{M}_n = z\} = v_{z,\,n} + v_{z+1,\,n}. \tag{9.2}$$

19. Пусть $V_x(s) = \sum v_{x,\,n} s^n$ (см. замечание перед задачей 15). Доказать, что $V_x(s) = V_0(s)\,\lambda_2^{-x}(s)$ при $x \leqslant 0$ и $V_x(s) = V_0(s)\,\lambda_1^{-x}(s)$ при $x \geqslant 0$, где $\lambda_1(s)$ и $\lambda_2(s)$ определены формулами (4.8). Более того, $V_0(s) = (1 - 4pqs^2)^{-1/2}$.

Замечание. Эти соотношения следуют *непосредственно* из того факта, что $\lambda_1(s)$ и $\lambda_2(s)$ являются производящими функциями времен первого достижения, как это было установлено в конце § 4.

20. Пусть $u_{z,\,n}(x)$ — вероятность того, что в случайном блуждании в $(0, \infty)$ с поглощающим экраном в нуле и начальным положением z n-й шаг приведет частицу в положение x, и пусть

$$U_z(s;\,x) = \sum_{n=0}^{\infty} u_{z,\,n}(x)\,s^n. \tag{9.3}$$

Используя результат задачи 19, показать, что $U_z(s;x) = V_{x-z}(s) - \lambda_2^z(s) V_x(s)$. Получить отсюда, что

$$u_{z,\,n}(x) = v_{x-z,\,n} - (q/p)^z \cdot v_{x+z,\,n}. \tag{9.4}$$

Сравнить с результатом задачи 15 и вывести (9.4) из последнего при помощи комбинаторных методов.

21. *Другая формула для вероятности разорения* (5.7). Разложив (4.11) в геометрический ряд, доказать, что

$$u_{z,\,n} = \sum_{k=0}^{\infty} (p/q)^{ka} w_{2ka+z,\,n} - \sum_{k=1}^{\infty} (p/q)^{ka-z} w_{2ka-z,\,n},$$

где $w_{z,\,n}$ — вероятность (4.14) первого достижения.

22. Показать, что если к выражению для $u_{z,\,n}$, приведенному в предыдущей задаче, применить предельный переход из § 6, то плотность вероятности для времени поглощения будет равна [1])

$$\left(1/\sqrt{2\pi D t^3}\right) e^{-(c t + 2\xi)c/(2D)} \sum_{k=-\infty}^{\infty} (\xi + 2k\alpha) e^{-(\xi + 2k\alpha)^2/(2Dt)}.$$

(*Указание.* Использовать нормальное приближение для биномиального распределения.)

23. *Метод восстановления для задачи о разорении* [2]). Пусть в случайном блуждании с двумя поглощающими экранами $u_{z,\,n}$ и $u^*_{z,\,n}$ — вероятности поглощения на левом и на правом экранах соответственно. Посредством надлежащей интерпретации доказать справедливость двух следующих уравнений:

$$V_{-z}(s) = U_z(s)\,V_0(s) + U_z^*(s)\,V_{-a}(s),$$
$$V_{a-z}(s) = U_z(s)\,V_a(s) + U_z^*(s)\,V_0(s).$$

Вывести (4.11), разрешив эту систему относительно $U_z(s)$.

24. Пусть $u_{z,\,n}(x)$ — вероятность того, что выходящая из точки z частица на n-м шаге попадет в точку x, не коснувшись предварительно поглощающих экранов. Показать, что для соответствующей производящей функции $U_z(s;x) =$

[1]) Совпадение этой новой формулы с предельной формой (6.15) — известный факт теории тэта-функций. См. формулу (5.8) гл. XIX тома 2.

[2]) Задачи 23—25 содержат новый и независимый вывод основных результатов, относящихся к одномерным случайным блужданиям.

$=\sum u_{z,\,n}(x)\,s^n$ мы имеем (используя обозначения задачи 23) равенство

$$U_z(s;x)=V_{x-z}(s)-U_z(s)\,V_x(s)-U_z^*(s)\,V_{x-a}(s).$$

(Здесь не требуется никаких вычислений.)

25. *Продолжение.* Доказать, что производящую функцию $U_z(s;x)$ из предыдущей задачи можно получить в виде $U_z(s;x)=V_{x-z}(s)-A\lambda_1^z(s)-B\lambda_2^z(s)$, где постоянные определяются так, чтобы при $z=0$ и $z=a$ выполнялись граничные условия $U_z(s;x)=0$. (В случае *отражающих экранов* граничные условия имеют вид $U_0(s;x)=U_1(s;x)$ и $U_a(s;x)=U_{a-1}(s;x)$.)

26. Доказать формулу

$$v_{x,\,n}=(2\pi)^{-1}2^n\,p^{(n+x)/2}q^{(n-x)/2}\int_{-\pi}^{\pi}\cos^n t\cdot\cos tx\cdot dt,$$

показав, что удовлетворяется соответствующее разностное уравнение. Вывести отсюда, что

$$V_x(s)=(2\pi)^{-1}(p/q)^{x/2}\int_{-\pi}^{\pi}\frac{\cos tx}{1-2\sqrt{pq}\cdot s\cdot\cos t}\,dt.$$

27. Показать, что в трехмерном симметричном случайном блуждании частица с вероятностью единица бесконечное число раз пересечет любую фиксированную прямую $x=m$, $y=n$. (*Указание.* Ср. с задачей 5.)

28. В двумерном симметричном случайном блуждании, начинающемся в нуле, вероятность того, что n-й шаг приведет частицу в точку (x,y), равна

$$(2\pi)^{-2}2^{-n}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(\cos\alpha+\cos\beta)^n\cdot\cos x\alpha\cdot\cos y\beta\cdot d\alpha\,d\beta.$$

Проверить эту формулу и найти ее аналог для трехмерного случая. (*Указание.* Проверить, что это выражение удовлетворяет соответствующему разностному уравнению.)

29. В двумерном симметричном случайном блуждании обозначим через $\mathbf{D}_n^2=x^2+y^2$ квадрат расстояния частицы от начала координат в момент времени n. Доказать, что $\mathbf{E}(\mathbf{D}_n^2)=n$. (*Указание.* Вычислить $\mathbf{E}(\mathbf{D}_{n-1}^2-\mathbf{D}_n^2)$.)

30. Доказать, что в симметричном случайном блуждании в d-мерном пространстве частица с вероятностью 1 будет бесконечное число раз возвращаться в положения, которые уже были ею заняты ранее. (*Указание.* При каждом шаге вероятность попадания в новое положение не превышает $(2d-1)/(2d)$.)

31. Показать, что описанный в § 8 метод применим также к производящей функции $U_z(s)$ для времени ожидания разорения.

§ 1. ОПРЕДЕЛЕНИЕ

До сих пор мы занимались главным образом независимыми испытаниями, которые можно описать следующим образом. Задано множество возможных исходов $E_1, E_2, \ldots$ (в конечном или бесконечном числе), и каждому из них соотнесена некоторая вероятность p_k; вероятности последовательностей исходов определяются по правилу умножения: $\mathbf{P}\{(E_{j_0}, E_{j_1}, \ldots, E_{j_n})\}=p_{j_0}p_{j_1}\ldots p_{j_n}$. В теории цепей Маркова мы рассматриваем простейшее обобщение этой схемы, которое состоит в том, что для любого испытания допускается зависимость его от непосредственно предшествующего ему испытания (и только от него). С исходом E_k не связана более фиксированная вероятность p_k, но зато каждой паре (E_j, E_k) теперь соответствует *условная вероятность* p_{jk}; при условии, что E_j появился в некотором испытании, вероятность появления E_k в следующем испытании равна p_{jk}. Помимо p_{jk} нам должны быть заданы вероятности a_k исходов E_k в *начальном* испытании. Чтобы p_{jk} имели приписанный им смысл, вероятности последовательностей исходов, соответствующих двум, трем или четырем испытаниям, должны быть определены равенствами

$$\mathbf{P}\{(E_j, E_k)\}=a_jp_{jk}, \quad \mathbf{P}\{(E_j, E_k, E_r)\}=a_jp_{jk}p_{kr},$$
$$\mathbf{P}\{(E_j, E_k, E_r, E_s)\}=a_jp_{jk}p_{kr}p_{rs},$$

и вообще

$$\mathbf{P}\{(E_{j_0}, E_{j_1}, \ldots, E_{j_n})\}=a_{j_0}p_{j_0j_1}p_{j_1j_2}\ldots p_{j_{n-2}j_{n-1}}p_{j_{n-1}j_n}. \quad (1.1)$$

Здесь начальному испытанию присвоен номер нуль, так что испытание номер один является вторым. (Это соглашение удобно и — без специального на то указания — уже использовалось в предыдущей главе.)

Некоторые процессы, рассматривавшиеся нами в предыдущих главах, являются цепями Маркова, однако в конкретных случаях часто бывает лучше воспользоваться другими обозначениями и способами описания. Основные результаты настоящей главы касаются существования некоторых пределов и равновесных распределений; эти результаты, конечно же, не зависят от обозначений и применимы ко всем цепям Маркова.

Примеры. а) *Случайные блуждания.* Случайное блуждание на прямой является цепью Маркова, однако в этом случае возможные положения естественно представить в виде бесконечной в обе стороны последовательности ..., -2, -1, 0, 1, 2, ... При таком порядке переходы будут возможны только между соседними положениями, т. е. $p_{jk}=0$, если $k \neq j \pm 1$. Чтобы воспользоваться нашими нынешними обозначениями, нам пришлось бы расположить целые числа в простую последовательность, скажем 0, 1, -1, 2, -2, . . . , и это привело бы к громоздким формулам для вероятностей p_{jk}. То же замечание справедливо и в отношении случайных блужданий в пространствах высшей размерности: для практических вычислений лучше обозначать точки значениями их координат, а для теоретических целей можно пользоваться символикой настоящей главы.

б) *Ветвящиеся процессы.* Вместо того чтобы говорить, что n-е испытание окончилось исходом E_k, в гл. XII, 3 мы говорили, что в n-м поколении имеется k частиц. Мы рассматривали там обычную цепь Маркова, у которой переходная вероятность p_{jk} равна коэффициенту при s^k в разложении j-й степени $p^j(s)$ заданной производящей функции.

в) *Урновые модели.* Очевидно, что некоторые урновые модели из гл. V, 2 представляют собой цепи Маркова. И обратно, как мы сейчас увидим, каждая цепь Маркова эквивалентна некоторой урновой модели.

Пусть каждый из имеющихся у состояний цепи индексов представлен отдельной урной и каждая урна содержит шары с метками E_1, E_2, ... Содержимое урн остается фиксированным, но меняется от урны к урне; вероятность извлечь из j-й урны шар с меткой E_k равна p_{jk}. В *начальном*, или нулевом, испытании в соответствии с распределением вероятностей $\{a_i\}$ выбирается урна. Из этой урны случайным образом извлекается шар, и если он помечен E_j, то следующее извлечение производится из j-й урны и т. д. Очевидно, что при такой процедуре вероятность выборки $(E_{j_0}, \ldots, E_{j_n})$ дается формулой (1.1). Мы видим, что понятие цепи Маркова не является более общим, чем понятие урновой модели, однако новая символика окажется более практичной и более интуитивно ясной. ►

Если a_k — вероятность появления E_k в начальном (или нулевом) испытании, то мы должны иметь $a_k \geqslant 0$ и $\sum a_k = 1$. Более того, после появления E_j непременно должно произойти одно из E_k, и поэтому необходимо, чтобы для всех j и k

$$p_{j1}+p_{j2}+\ldots=1, \qquad p_{jk} \geqslant 0. \tag{1.2}$$

Теперь мы покажем, что для любых чисел a_k и p_{jk}, удовлетворяющих этим условиям, формула (1.1) является допустимым определением вероятностей в пространстве элементарных событий, соответствующем $n+1$ испытаниям. Поскольку числа, определяемые по (1.1), неотрицательны, нам нужно доказать только то, что в сумме они составляют единицу. Фиксируем первые j_0, j_1, . . ., j_{n-1} и сложим числа (1.1) при всех возможных j_n. Используя (1.2) при $j=j_{n-1}$, мы немедленно убеждаемся, что эта сумма равна $a_{j_0}p_{j_0j_1}\ldots p_{j_{n-2}j_{n-1}}$. Таким образом, сумма всех чисел (1.1) не зависит от n, и, поскольку $\sum a_{j_0}=1$, сумма эта равна единице для всех n.

Определение (1.1) формально зависит от числа испытаний, однако наше рассуждение доказывает взаимную согласованность определений (1.1) при всех n. Например, чтобы получить вероятность события «первые два испытания окончились исходами (E_j, E_k)», мы должны фиксировать $j_0=j$ и $j_1=k$ и сложить вероятности (1.1) при всех возможных $j_2, j_3, \ldots, j_n$. Только что мы показали, что эта сумма равна a_jp_{jk} и, таким образом, не зависит от n. Это означает, что обычно нет необходимости говорить о полном числе испытаний; событие $(E_{j_0}, \ldots, E_{j_r})$ имеет одну и ту же вероятность во всех пространствах элементарных событий для более чем r испытаний. В связи с независимыми испытаниями неоднократно отмечалось, что с математической точки зрения наиболее удовлетворительным будет ввести только одно пространство элементарных событий — неограниченных последовательностей испытаний — и рассматривать результат конечного числа испытаний как начальный отрезок бесконечной последовательности. Это утверждение справедливо также и для цепей Маркова. К сожалению, пространства бесконечных последовательностей испытаний выводят изложение за пределы теории дискретных вероятностей, которой мы ограничились в этом томе.

Суммируя сказанное выше, получаем в качестве отправного пункта следующее определение.

Определение. *Последовательность испытаний с возможными исходами* $E_1, E_2, \ldots$ *называется цепью Маркова* [1]*, если вероятности последовательностей исходов определяются формулой* (1.1) *через распределение вероятностей* $\{a_k\}$ *для* E_k *в начальном (или нулевом) испытании и через фиксированные условные вероятности* p_{jk} *появления* E_k *при условии, что в предыдущем испытании появился* E_j.

Для приложений цепей Маркова удобнее несколько видоизмененная терминология. Возможные исходы E_k обычно называются возможными *состояниями системы*; вместо того чтобы говорить, что n-е испытание окончилось появлением E_k, говорят, что *n-й шаг приводит к состоянию E_k* или что система попадает в E_k на n-м шаге. Наконец, p_{jk} называется *вероятностью перехода из E_j в E_k*. Как обычно, мы считаем, что испытания происходят через равные ин-

---

[1]) Это нестандартная терминология. Здесь мы рассматриваем только частный класс цепей Маркова, и, строго говоря, здесь и в последующих параграфах термин «цепь Маркова» должен всякий раз уточняться добавлением слов «со стационарными переходными вероятностями». На самом деле общий тип цепей Маркова рассматривается редко. Он будет определен в § 13, где марковское свойство будет обсуждаться в связи с общими стохастическими процессами. Там же читатель найдет примеры зависимых испытаний, не образующих цепи Маркова.

тервалы времени, так что номер шага служит временным параметром.

Вероятности перехода p_{jk} будут расположены в *матрицу переходных вероятностей*

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{bmatrix}, \tag{1.3}$$

где первый индекс означает номер строки, а второй — номер столбца. Ясно, что P — квадратная матрица с неотрицательными элементами и единичными суммами по строкам. Такая матрица (конечная или бесконечная) называется *стохастической матрицей. Любая стохастическая матрица может служить матрицей переходных вероятностей; вместе с нашим начальным распределением* $\{a_k\}$ *она полностью определяет цепь Маркова с состояниями* E_1, E_2, ...

В некоторых частных случаях бывает удобно нумеровать состояния, начиная с 0, а не с 1. Тогда к матрице P следует добавить нулевые строку и столбец.

Историческое замечание. Многие задачи, рассматривавшиеся в классической литературе при помощи урновых моделей, представляются теперь при помощи цепей Маркова частного вида, однако первоначальные методы исследования были совершенно другими. Более того, ряд урновых моделей носят совсем иной характер, ибо в них имеет место эффект последействия, и эта существенная разница в должной мере не была понята. В действительности эта путаница сохранялась долгое время после первооткрывающей работы Маркова. А. А. Марков заложил основания теории конечных цепей Маркова, однако конкретные приложения ограничивались главным образом тасованием карт и лингвистическими задачами. Теоретическое изучение проводилось обычно алгебраическими методами, близкими к описанным в следующей главе. Основы этого подхода изложены в монографии М. Фреше [1]).

Теория цепей с бесконечным числом состояний была введена А. Н. Колмогоровым [2]). Этот новый подход, изложенный в первом издании настоящей книги, сделал эту теорию доступной более широкому кругу специалистов и привлек внимание к множеству ее возможных приложений. С тех пор цепи Маркова стали стандартным элементом теории вероятностей и знакомым инструментом во многих приложениях. Более современные теоретические разработки упоминаются в замечаниях в § 11 и 12.

---

[1]) Fréchet M., Recherches théoriques modernes sur le calcul des probabilités, t. 2 (Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles), Paris, 1938.

[2]) Kolmogoroff A., Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen, Математический сборник, н. с., 1 (1936), 607—610. Эта статья не содержит доказательств. Полное изложение см. в работе Колмогоров А. Н., Цепи Маркова со счетным числом состояний, Бюлл. МГУ (А), 1 : 3 (1937), 1—16.

§ 2. ПОЯСНИТЕЛЬНЫЕ ПРИМЕРЫ

(Приложения к классической задаче о тасовании карт см. в § 10.)

а) Когда у цепи есть только два возможных состояния E_1 и E_2, матрица переходных вероятностей с необходимостью имеет вид

$$P=\begin{bmatrix}1-p & p\\ \alpha & 1-\alpha\end{bmatrix}.$$

Подобная цепь могла бы быть реализована в следующем мысленном эксперименте. Частица движется вдоль оси x таким образом, что абсолютная величина ее скорости остается постоянной, но направление движения может меняться на противоположное. Говорят, что система находится в состоянии E_1, если частица движется направо, и в состоянии E_2, если она движется налево. Тогда p — вероятность поворота, когда частица движется направо, а α — вероятность поворота при движении налево. (Подробный анализ этой цепи см. в примере гл. XVI, 2,а).)

б) *Случайное блуждание с поглощающими экранами.* Пусть возможными состояниями будут $E_0, E_1, \ldots, E_\rho$; рассмотрим матрицу переходных вероятностей

$$P=\begin{bmatrix}1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0\\ q & 0 & p & 0 & \ldots & 0 & 0 & 0\\ 0 & q & 0 & p & \ldots & 0 & 0 & 0\\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \\ 0 & 0 & 0 & 0 & \ldots & q & 0 & p\\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1\end{bmatrix}.$$

Из каждого «внутреннего» состояния $E_1, \ldots, E_{\rho-1}$ возможны переходы в правое и левое соседние состояния (с вероятностями $p_{i,\,i+1}=p$ и $p_{i,\,i-1}=q$). Однако ни из E_0, ни из E_ρ невозможны переходы в какое-либо иное состояние; система может переходить из одного состояния в другое, но коль скоро будет достигнуто E_0 или E_ρ, система останется неизменной навсегда. Ясно, что эта цепь Маркова только терминологией отличается от модели случайного блуждания с поглощающими экранами в точках 0 и ρ, обсуждавшейся в предыдущей главе. Там случайное блуждание начиналось в фиксированной точке z рассматриваемого интервала. В терминологии цепей Маркова это равносильно выбору такого начального распределения, что $a_z=1$ (и, следовательно, $a_x=0$ при $x\neq z$). Случайно выбранному начальному состоянию соответствует начальное распределение $a_k=1/(\rho+1)$.

в) *Отражающие экраны.* Интересный вариант предыдущего примера представляет собой цепь с возможными состояниями

$E_1, \ldots, E_\rho$ и переходными вероятностями

$$P=\begin{bmatrix} q & p & 0 & 0 & \ldots & 0 & 0 & 0 \\ q & 0 & p & 0 & \ldots & 0 & 0 & 0 \\ 0 & q & 0 & p & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & q & 0 & p \\ 0 & 0 & 0 & 0 & \ldots & 0 & q & p \end{bmatrix}.$$

Эту цепь можно интерпретировать на языке азартных игр, рассматривая двух игроков, ведущих игру с единичными ставками и с соглашением, что каждый раз, когда один из игроков проигрывает свой последний доллар, тот немедленно возвращается ему его противником, так что игра может продолжаться бесконечно. Мы предполагаем, что игроки вместе имеют $\rho+1$ долларов, и говорим, что система находится в состоянии E_k, если их капиталы равны k и $\rho-k+1$ соответственно. Тогда переходные вероятности даются нашей матрицей P. В терминологии, введенной в гл. XIV, 1, наша цепь представляет собой случайное блуждание с отражающими экранами в точках $1/2$ и $\rho+1/2$. Аналогичным образом можно рассматривать случайные блуждания с упругими экранами. Подробный анализ цепи для случая отражающих экранов читатель найдет в гл. XVI, 3 (см. также пример 7, в)).

г) *Циклические случайные блуждания.* Пусть возможными состояниями снова будут $E_1, E_2, \ldots, E_\rho$, но теперь мы упорядочим их циклическим образом, так что для E_ρ соседними состояниями будут $E_{\rho-1}$ и E_1. Если, как и ранее, система всегда переходит либо в правое, либо в левое соседнее состояние, то строки матрицы P будут такими же, как в примере б), за тем исключением, что первая строка будет $(0, p, 0, 0, \ldots, 0, q)$, а последняя $(p, 0, 0, 0, \ldots, 0, q, 0)$.

В более общем случае мы можем допустить переходы между любыми двумя состояниями. Пусть $q_0, q_1, \ldots, q_{\rho-1}$ — соответственно вероятности остаться на месте или передвинуться на $1, 2, \ldots, \rho-1$ единиц вправо (причем переход на k единиц вправо — то же самое, что переход на $\rho-k$ единиц влево). Тогда P будет циклической матрицей

$$P=\begin{bmatrix} q_0 & q_1 & q_2 & \ldots & q_{\rho-2} & q_{\rho-1} \\ q_{\rho-1} & q_0 & q_1 & \ldots & q_{\rho-3} & q_{\rho-2} \\ q_{\rho-2} & q_{\rho-1} & q_0 & \ldots & q_{\rho-4} & q_{\rho-3} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ q_1 & q_2 & q_3 & \ldots & q_{\rho-1} & q_0 \end{bmatrix}.$$

Анализ этой цепи см. в примере гл. XVI, 2, г).

д) *Модель Эренфестов для диффузии.* Снова рассмотрим цепь с $\rho+1$ состоянием $E_0, E_1, \ldots, E_\rho$, переходы в которой возможны только в правое и левое соседние состояния; на этот раз мы положим $p_{j,\,j+1}=1-j/\rho$ и $p_{j,\,j-1}=j/\rho$, так что

$$P=\begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ \rho^{-1} & 0 & 1-\rho^{-1} & 0 & \ldots & 0 & 0 \\ 0 & 2\rho^{-1} & 0 & 1-2\rho^{-1} & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & 0 & \rho^{-1} \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Эта цепь имеет две интересные физические интерпретации. При обсуждении различных задач о возвращении в статистической механике П. и Т. Эренфесты [1]) описали мысленный урновый эксперимент, в котором ρ молекул распределены по двум сосудам A и B. При каждом испытании случайным образом выбирается одна молекула и перемещается из своего сосуда в другой. Состояние системы определяется числом молекул в A. Предположим, что в некоторый момент времени в сосуде A находится ровно k молекул. При следующем испытании система переходит в E_{k-1} или E_{k+1} в зависимости от того, выбрана ли молекула из A или из B; соответствующие вероятности суть k/ρ и $(\rho-k)/\rho$, и, стало быть, наша цепь действительно описывает эксперимент Эренфестов. Однако нашу цепь можно интерпретировать и как *диффузию при наличии центральной силы*, т. е. как случайное блуждание, в котором вероятность шага вправо меняется вместе с состоянием. Из $x=j$ переход частицы вправо (влево) будет более вероятен, если $j<\rho/2$ ($j>\rho/2$); это означает, что частица имеет тенденцию двигаться к точке $x=\rho/2$, что соответствует упругой силе притяжения, возрастающей прямо пропорционально расстоянию. (Модель Эренфестов была описана в примере гл. V, 2,в); см. также пример 7, г) и задачу 12.)

е) *Модель Бернулли — Лапласа для диффузии* [2]). Модель, похожая на модель Эренфестов, была предложена Д. Бернулли в качестве вероятностного аналога течения двух несжимаемых жид-

---

[1]) Ehrenfest P., Ehrenfest T., Über zwei bekannte Einwände gegen das Boltzmannsche *H*-Theorem, Physikalische Zeitschrift, 8 (1907), 311—314; Ming Chen Wang, Uhlenbeck G. E., On the theory of the Brownian motion II, Reviews of Modern Physics, 17 (1945), 323—342. Более полное обсуждение см. Kac M., Random walk and the theory of Brownian motion, Amer. Math. Monthly, 54 (1947), 369—391. Эти авторы не упоминали цепи Маркова, однако Кац использовал методы, близко связанные с описываемыми в следующей главе. См. также Friedman B., A simple urn model, Communications on Pure and Applied Mathematics, 2 (1949), 59—70.

[2]) В виде урновой модели эту задачу рассматривал Даниил Бернулли в 1769 г., критиковал Мальфатти в 1782 г. и анализировал Лаплас в 1812 г. См. Todhunter I., A history of the mathematical theory of probability, Cambridge, 1865.

костей между двумя сосудами. На этот раз мы имеем всего 2ρ частиц, среди которых ρ черных и ρ белых. Поскольку предполагается, что частицы представляют несжимаемые жидкости, их плотности не должны меняться, так что число ρ частиц в каждой урне остается постоянным. Мы говорим, что система находится в состоянии E_k ($k=0, 1, \ldots, \rho$), если первая урна содержит k белых частиц. (Это означает, что она содержит $\rho-k$ черных частиц, тогда как вторая урна содержит $\rho-k$ белых и k черных частиц.) При каждом испытании из каждой урны выбирается по одной частице, и эти две частицы меняются местами. Переходные вероятности даются тогда формулами

$$p_{j,\,j-1}=\left(\frac{j}{\rho}\right)^2, \quad p_{j,\,j+1}=\left(\frac{\rho-j}{\rho}\right)^2, \quad p_{jj}=2\,\frac{j(\rho-j)}{\rho^2} \tag{2.1}$$

и $p_{jk}=0$, если $|j-k|>1$ (здесь $j=0, \ldots, \rho$). (О стационарном распределении см. в примере 7, д); обобщение этой модели см. в задаче 10.)

ж) *Случайные размещения шаров*. Рассмотрим последовательность независимых испытаний, в каждом из которых шар случайным образом помещается в один из заданных ρ ящиков (или урн). Мы скажем, что система находится в состоянии E_k, если заняты ровно k ящиков. Это определяет цепь Маркова с набором состояний $E_0, \ldots, E_\rho$ и переходными вероятностями, такими, что

$$p_{jj}=j/\rho, \qquad p_{j,\,j+1}=(\rho-j)/\rho, \tag{2.2}$$

и, конечно, $p_{jk}=0$ для всех других комбинаций j и k. Если сначала все ящики были пустыми, то распределение $\{a_k\}$ определяется равенствами $a_0=1$ и $a_k=0$ при $k>0$. (Дальнейший анализ этой цепи см. в примере гл. XVI, 2, д). Случайные размещения шаров рассматривались нами с других точек зрения в гл. II, 5 и гл. IV, 2.)

з) *Пример из клеточной генетики* [1]). Цепь Маркова с набором состояний $E_0, \ldots, E_N$ и переходными вероятностями

$$p_{jk}=\binom{2j}{k}\binom{2N-2j}{N-k}\binom{2N}{N}^{-1} \tag{2.3}$$

появляется в биологической задаче, которую упрощенно можно сформулировать следующим образом. Каждая клетка некоторого организма содержит N частиц, причем одни из них относятся к типу A, а другие — к типу B. Говорят, что клетка находится в состоянии E_j, если она содержит ровно j частиц типа A. Дочерние клетки образуются в результате клеточного деления, но перед делением каждая из частиц реплицируется (т. е. удваивается); дочерняя клетка

[1]) Schensted I. V., Model of subnuclear segregation in the macronucleus of ciliates, The Amer. Naturalist, 92 (1958), 161—170. Этот автор существенно использует методы гл. XVI, но не упоминает цепи Маркова. Наша формулировка задачи математически эквивалентна рассматривавшейся в этой работе, но биологически сверхупрощена.

наследует N частиц, выбранных случайным образом из $2j$ частиц типа A и $2N-2j$ частиц типа B, имевшихся в родительской клетке. Вероятность того, что дочерняя клетка будет находится в состоянии E_k, дается тогда гипергеометрическим распределением (2.3).

В примере 8, б) будет показано, что *после достаточно большого числа поколений вся популяция будет состоять (и останется состоящей) из клеток, находящихся в одном из чистых состояний E_0 или E_N*; вероятности того, что отдельная линия потомков придет к тому или иному чистому состоянию, суть соответственно $1-j/N$ и j/N (E_j есть начальное состояние).

и) *Примеры из популяционной генетики* [1]. Рассмотрим последовательные поколения в популяции (такой, как растения на кукурузном поле), величина которой сохраняется постоянной в результате отбора N особей в каждом поколении. Отдельный ген, который может относиться к типу A или a, имеет в популяции $2N$ представителей; если в n-м поколении A встречается j раз, то a встречается $2N-j$ раз. В этом случае мы говорим, что популяция находится в состоянии E_j ($0 \leqslant j \leqslant 2N$). Если предположить, что скрещивание происходит случайным образом, то состав следующего поколения определяется $2N$ испытаниями Бернулли, в которых вероятность появления гена A будет равна $j/(2N)$. Стало быть, мы имеем цепь Маркова с

$$p_{jk} = \binom{2N}{k} \left(\frac{j}{2N}\right)^k \left(1-\frac{j}{2N}\right)^{2N-k}. \qquad (2.4)$$

В состояниях E_0 и E_{2N} все гены относятся к одному типу, и уход из этих состояний невозможен. (Такие состояния называются гомозиготными.) В примере 8, б) будет показано, что *в конце концов вся популяция окажется в одном из гомозиготных состояний E_0 или E_{2N}*. Если популяция находится вначале в состоянии E_j, то соответствующие вероятности суть $1-j/(2N)$ и $j/(2N)$.

Эту модель можно модифицировать с тем, чтобы учесть возможные мутации и селективные преимущества генов.

к) *Задача о скрещивании.* В так называемом братско-сестринском скрещивании скрещиваются две особи, и среди их прямых потомков случайным образом выбираются две особи разного пола. Они вновь скрещиваются, и процесс этот продолжается бесконечно. Имея три генотипа AA, Aa, aa для каждого родителя, мы должны различать шесть комбинаций родителей, которые мы пометим следующим образом: $E_1 = AA\times AA$, $E_2 = AA\times Aa$, $E_3 = Aa\times Aa$, $E_4 = Aa\times aa$, $E_5 = aa\times aa$, $E_6 = AA\times aa$. Используя правила из гл. V. 5, легко видеть, что матрица переходных вероятностей имеет в этом слу-

[1] Эту задачу исследовали различными методами Р. Э. Фишер и С. Райт. Формулировка в терминах цепей Маркова дается в статье Malécot G., Sur un problème de probabilités en chaîne que pose la génétique, Comptes rendus de l'Académie des Sciences, 219 (1944), 379—381.

чае вид

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\ 1/16 & 1/4 & 1/4 & 1/4 & 1/16 & 1/8 \\ 0 & 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

(Обсуждение этой цепи продолжается в задаче 4; подробный ее анализ дан в примере гл. XVI, 4, б).)

л) *Рекуррентные события и остаточные времена ожидания.* Мы неоднократно будем использовать цепь с набором состояний E_0, $E_1, \ldots$ и переходными вероятностями

$$P = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 & \ldots \\ 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 1 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \ldots \\ \cdot & \cdot & \cdot & \cdot & \ldots \end{bmatrix};$$

вероятности f_k произвольны, за тем лишь исключением, что в сумме они должны давать единицу. Чтобы наглядно представить себе этот процесс, предположим, что начальным состоянием будет E_0. Если первый шаг приводит в E_{k-1}, то система обязана пройти последовательно через состояния E_{k-2}, E_{k-3}, ... и на k-м шаге вернуться в E_0, откуда процесс начнется сначала. Таким образом, последовательные возвращения в E_0 представляют из себя возвратное рекуррентное событие $\mathcal{E}$ с распределением $\{f_k\}$ для времен возвращения. Состояние системы в любой момент времени определяется временем ожидания *следующего* прохождения через E_0.

В большинстве конкретных реализаций рекуррентных событий время ожидания следующего осуществления события зависит от будущего, и, значит, наша цепь Маркова не имеет никакого практического значения. Однако эта цепь имеет смысл в том случае, когда можно представить себе, что одновременно с каждым осуществлением события $\mathcal{E}$ производится случайный эксперимент, исход которого определяет величину следующего времени ожидания. Такие ситуации встречаются на практике, хотя они и составляют скорее исключение, чем правило. Например, в теории самовосстанавливающихся устройств (пример гл. XIII, 10, г)) иногда предполагается, что срок службы вновь установленного элемента зависит от выбора этого элемента, но вполне определен, коль скоро выбор уже сделан. С другой стороны, в теории массового обслуживания (в очередях к продавцу или на телефонных линиях) последовательные моменты начала обслуживания отдельных клиентов обычно соответствуют рекуррентным событиям. Предположим теперь, что имеется много типов клиентов, и для каждого из этих типов требуется обслуживание известной продолжительности. Тогда время ожидания между двумя последовательными моментами начала обслуживания определяется единственным образом с того момента, когда начинается обслуживание соответствующего клиента. (См. пример 7, ж).)

м) *Другая цепь, связанная с рекуррентными событиями.* Рассмотрим цепь с набором возможных состояний E_0, E_1, ... и переходными вероятностями

$$P=\begin{bmatrix} q_1 & p_1 & 0 & 0 & 0 & \dots \\ q_2 & 0 & p_2 & 0 & 0 & \dots \\ q_3 & 0 & 0 & p_3 & 0 & \dots \\ q_4 & 0 & 0 & 0 & p_4 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \end{bmatrix},$$

где $p_k+q_k=1$. Для наглядности мы можем интерпретировать состояние E_k как представляющее «возраст» системы. По достижении системой возраста k процесс старения с вероятностью p_{k+1} продолжается, а с вероятностью q_{k+1} система «омолаживается», и процесс начинается заново — с нулевого возраста. Последовательные прохождения через состояние E_0 здесь снова представляют рекуррентное событие, и вероятность того, что время возвращения равно k, дается произведением $p_1p_2\dots p_{k-1}q_k$. Можно подобрать $\{p_k\}$ так, чтобы получить заданное распределение $\{f_k\}$ для времен возвращения; достаточно положить $q_1=f_1$, затем $q_2=f_2/p_1$ и т. д. В общем виде

$$p_k=\frac{1-f_1-\dots-f_k}{1-f_1-\dots-f_{k-1}}. \tag{2.5}$$

Таким образом, произвольное рекуррентное событие $\mathscr{E}$ с распределением $\{f_k\}$ времен возвращения соответствует цепи Маркова с матрицей P, определяемой вероятностями (2.5). После n-го испытания система окажется в состоянии E_k тогда и только тогда, когда последним испытанием, при котором произошло событие $\mathscr{E}$, было испытание с номером $n-k$ (здесь $k=0, 1, \dots$). Номер этого состояния часто называется «затраченным временем ожидания». (Обсуждение продолжается в примерах 5, б), 7, е) и 8, д).)

н) *Серии успехов.* В качестве частного случая предыдущего примера рассмотрим последовательность испытаний Бернулли и условимся, что при n-м испытании система будет в состоянии E_k, если последняя *неудача* наблюдалась при испытании с номером $n-k$. Здесь $k=0, 1, \dots$, и считается, что нулевое испытание привело к неудаче. Иначе говоря, индекс k равен длине непрерывшейся последовательности успехов, оканчивающейся n-м испытанием. Переходные вероятности здесь те же, что и в предыдущем примере с $p_k=p$ и $q_k=q$ при всех k.

§ 3. ВЕРОЯТНОСТИ ПЕРЕХОДА ЗА НЕСКОЛЬКО ШАГОВ

Мы обозначим через $p_{jk}^{(n)}$ *вероятность перехода из* E_j *в* E_k *ровно за* n *шагов*. Иначе говоря, $p_{jk}^{(n)}$ есть условная вероятность попадания в E_k на n-м шаге при условии, что начальным состоянием было E_j; она равна сумме вероятностей всех путей $E_jE_{j_1}\ldots$ $\ldots E_{j_{n-1}}E_k$ длины n, начинающихся в E_j и оканчивающихся в E_k. В частности, $p_{jk}^{(1)}=p_{jk}$ и

$$p_{jk}^{(2)}=\sum_{\nu} p_{j\nu}p_{\nu k}. \tag{3.1}$$

По индукции мы получаем *общую рекуррентную формулу*

$$p_{jk}^{(n+1)}=\sum_{\nu} p_{j\nu}p_{\nu k}^{(n)}; \tag{3.2}$$

дальнейшая индукция по m приводит к *основному тождеству*

$$p_{jk}^{(m+n)}=\sum_{\nu} p_{j\nu}^{(m)}p_{\nu k}^{(n)} \tag{3.3}$$

(которое является частным случаем уравнения Колмогорова — Чэпмена). Оно отражает тот простой факт, что первые m шагов приводят из E_j в некоторое промежуточное состояние E_ν и что вероятность последующего перехода из E_ν в E_k не зависит от того, каким образом было достигнуто E_ν[1]).

Так же как и в случае p_{jk}, образовавших матрицу P, *мы расположим* $p_{jk}^{(n)}$ *в матрицу, которую обозначим* P^n. Тогда (3.2) утверждает, что для того, чтобы получить элемент $p_{jk}^{(n+1)}$ матрицы P^{n+1}, мы должны умножить элементы j-й строки P на соответствующие элементы k-го столбца P^n и сложить все полученные произведения. Эта операция называется умножением матриц P и P^n и выражается символически равенством $P^{n+1}=PP^n$. Данное определение позволяет назвать P^n n-й степенью P; уравнение (3.3) выражает известный закон $P^{m+n}=P^mP^n$.

Для того чтобы (3.3) было справедливо для всех $n\geqslant 0$, мы определим $p_{jk}^{(0)}$, положив $p_{jj}^{(0)}=1$ и $p_{jk}^{(0)}=0$ при $j\neq k$, что вполне естественно.

Примеры. а) *Независимые испытания.* Обычно бывает трудно получить явные выражения для вероятностей перехода за несколь-

[1]) Последнее свойство является характерным для марковских процессов, которые будут определены в § 13. Долгое время предполагалось, что (3.3) можно использовать для определения цепей Маркова, однако это неожиданно оказалось неверным (см. пример 13,е)),

ко шагов, однако, к счастью, они не представляют особого интереса. Как важное, хотя и тривиальное исключение, мы отметим частный случай независимых испытаний. Этот случай имеет место тогда, когда все строки P тождественно совпадают с данным распределением вероятностей, и ясно без вычислений, что отсюда следует равенство $P^n=P$ при всех n.

б) *Серии успехов.* В примере 2, н) легко видеть (либо из рекуррентной формулы (3.2), либо из самого определения процесса), что

$$p_{jk}^{(n)}=\begin{cases} qp^k & \text{при } k=0,\ 1,\ \ldots,\ n-1, \\ p^n & \text{при } k=n+j, \\ 0 & \text{в остальных случаях.} \end{cases}$$

В этом случае ясно, что P^n сходится к матрице, такой, что все элементы в ее столбце с номером k равны qp^k. ►

Безусловные вероятности

Пусть снова a_j означает вероятность состояния E_j в начальном (нулевом) испытании. Тогда (безусловная) вероятность попадания в E_k на n-м шаге равна

$$a_k^{(n)}=\sum_j a_j p_{jk}^{(n)}. \tag{3.4}$$

Обычно мы считаем, что процесс начинается из фиксированного состояния E_i, т. е. полагаем $a_i=1$. В этом случае $a_k^{(n)}=p_{ik}^{(n)}$.

Интуитивно мы чувствуем, что влияние начального состояния должно постепенно ослабевать, так что при больших n распределение (3.4) должно быть почти независимым от начального распределения $\{a_j\}$. Так оно и будет, если (как в последнем примере) $p_{jk}^{(n)}$ сходится к не зависящему от j пределу, т. е. если P^n сходится к матрице с одинаковыми строками. Мы увидим, что обычно это действительно так, хотя нам и придется еще принимать в расчет досадные исключения, обусловленные периодичностью.

§ 4. ЗАМЫКАНИЯ И ЗАМКНУТЫЕ МНОЖЕСТВА

Будем говорить, что E_k *достижимо из* E_j, *если существует такое* $n\geqslant 0$, *что* $p_{jk}^{(n)}>0$ (т. е. если имеется положительная вероятность попасть в E_k из E_j, включая случай $E_k=E_j$). Например, в неограниченном случайном блуждании каждое состояние достижимо из любого другого состояния, однако из поглощающего экрана не достижимо никакое другое состояние.

Определение. *Множество состояний C называется замкнутым, если никакое состояние вне C не достижимо ни из одного состояния E_j из C. Для произвольного множества C замыканием C называется наименьшее замкнутое множество, содержащее C.*

Если одно состояние E_k образует замкнутое множество, то оно называется поглощающим.

Цепь Маркова называется неприводимой, если не существует замкнутых множеств, отличных от множества всех состояний.

Ясно, что C замкнуто тогда и только тогда, когда $p_{jk}=0$ при всех j и k, таких, что E_j принадлежит C, а E_k не принадлежит C, ибо в этом случае мы видим из (3.2), что $p_{jk}^{(n)}=0$ при всех n. Таким образом, имеет место очевидная теорема.

Теорема. *Если вычеркнуть в матрицах P^n все строки и столбцы, соответствующие состояниям, не принадлежащим замкнутому множеству C, то останутся стохастические матрицы, для которых вновь будут справедливы фундаментальные соотношения* (3.2) *и* (3.3).

Это означает, что мы имеем цепь Маркова, определенную на C, и эту *подцепь можно изучать независимо от всех остальных состояний.*

Состояние E_k является поглощающим тогда и только тогда, когда $p_{kk}=1$; в этом случае матрица из последней теоремы сводится к одному-единственному элементу. Вообще, ясно, что все состояния E_k, достижимые из данного состояния E_j, образуют замкнутое множество. (Поскольку замыкание E_j не может быть меньше этого множества, то оно с ним совпадает.) Неприводимая цепь не содержит собственных замкнутых подмножеств, так что мы имеем следующий простой, но полезный критерий.

Критерий. *Цепь неприводима тогда и только тогда, когда каждое ее состояние достижимо из любого другого состояния.*

Примеры. а) Для того чтобы найти все замкнутые множества, достаточно знать, какие p_{jk} равны нулю и какие положительны. В соответствии с этим мы воспользуемся символом $*$ для обозначения положительных элементов и рассмотрим типичную матрицу, скажем

$$P=\begin{bmatrix} 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\ 0 & * & * & 0 & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \\ * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \end{bmatrix}.$$

Занумеруем состояния числами от 1 до 9. В пятой строке $*$ стоит только на пятом месте, и, стало быть, $p_{55}=1$: состояние E_5 является *поглощающим*. Третья и восьмая строки содержат лишь по одному положительному элементу каждая, и ясно, что E_3 и E_8 образуют *замкнутое* множество. Из E_1 возможны переходы в E_4 и E_9, а из них — только в E_1, E_4, E_9. Следовательно, три состояния E_1, E_4, E_9 образуют другое *замкнутое* множество.

Из E_2 возможны непосредственные переходы в него самого и в E_3, E_5 и E_8. Пара (E_3, E_8) образует замкнутое множество, тогда как E_5 есть поглощающее состояние; поэтому замыкание E_2 состоит из множества E_2, E_3, E_5, E_8. Легко видеть, что замыкания оставшихся состояний E_6 и E_7 состоят из всех девяти состояний.

Внешний вид нашей матрицы и процедуру выделения замкнутых множеств можно упростить, перенумеровав состояния в следующем порядке:

$$E_5E_3E_8E_1E_4E_9E_2E_6E_7.$$

Тогда замкнутые множества будут содержать лишь соседние состояния и структура цепи будет понятна с первого взгляда на новую матрицу.

б) В матрице из примера 2, к) состояния E_1 и E_5 являются поглощающими, и других замкнутых множеств не существует.

в) В генетическом примере 2, и) состояния E_0 и E_{2N} являются поглощающими. При $0<j<2N$ замыкание E_j содержит все состояния. В примере 2, з) состояния E_0 и E_N являются поглощающими.

Рассмотрим цепь с набором состояний $E_1, \ldots, E_\rho$, такую, что $E_1, \ldots, E_r$ образуют замкнутое множество $(r<\rho)$. Тогда $(r\times r)$-подматрица, стоящая в левом верхнем углу P, тоже будет стохастической, и мы можем представить P в виде блочной матрицы

$$P=\begin{bmatrix} Q & 0 \\ U & V \end{bmatrix}. \tag{4.1}$$

Матрица в верхнем правом углу имеет r строк и $\rho-r$ столбцов и состоит только из нулевых элементов. Аналогично U означает матрицу с $\rho-r$ строками и r столбцами, тогда как V есть квадратная матрица. Мы будем пользоваться символическим блочным разбиением (4.1) и в том случае, когда замкнутое множество C и его дополнение C' содержат бесконечно много состояний; это разбиение просто указывает на группировку состояний и на тот факт, что $p_{jk}=0$, если E_j принадлежит C, а E_k — его дополнению C'. Из рекуррентной формулы (3.2) очевидно, что вероятности перехода за несколько шагов допускают аналогичное разбиение:

$$P^n=\begin{bmatrix} Q^n & 0 \\ U_n & V^n \end{bmatrix}. \tag{4.2}$$

Сейчас мы не интересуемся видом элементов матрицы U_n, стоящей в левом нижнем углу. Главное здесь в том, что из (4.2) становится ясны три очевидных, но важных факта. Во-первых, $p_{jk}^{(n)}=0$, когда $E_j \in C$, но $E_k \in C'$. Во-вторых, появление степени Q^n показывает, что когда оба состояния E_j и E_k принадлежат C, то переходные вероятности $p_{jk}^{(n)}$ получаются из рекуррентной формулы (3.2) при суммировании, ограниченном только состояниями из замкнутого множества C. Наконец, наличие V^n указывает, что последнее утверждение остается справедливым при замене C на его дополнение C'. Следовательно, можно будет упростить дальнейшее изучение цепей Маркова, рассматривая отдельно состояния, принадлежащие замкнутому множеству C, и состояния, принадлежащие дополнению C'.

Заметим, что мы не предполагали, что Q неприводима. Если C разбивается на несколько замкнутых подмножеств, то Q допускает дальнейшее разбиение. Существуют цепи с бесконечным числом замкнутых подмножеств.

Пример. г) Как уже упоминалось выше, случайное блуждание на плоскости представляет собой специальную цепь Маркова, хотя упорядочивание состояний в простую последовательность было бы неудобным для практических целей. Предположим теперь, что мы видоизменяем случайное блуждание, считая, что по достижении оси x частица будет продолжать случайное блуждание вдоль по этой оси, уже не покидая ее. Тогда точки оси x образуют бесконечное замкнутое множество. С другой стороны, если мы условимся, что по достижении оси x частица навсегда остается в точке первого попадания на эту ось, то каждая точка оси x превратится в поглощающее состояние. ►

§ 5. КЛАССИФИКАЦИЯ СОСТОЯНИЙ

В процессе с начальным состоянием E_j последовательные возвращения в E_j представляют собой рекуррентное событие, тогда как последовательные прохождения через какое-либо иное состояние будут представлять собой уже рекуррентное событие с запаздыванием (определенное в гл. XIII,5). Стало быть, теория цепей Маркова сводится к одновременному изучению многих рекуррентных событий. Общая теория рекуррентных событий применима к ней без каких-либо модификаций, однако, чтобы избежать чрезмерного количества ссылок на гл. XIII, мы заново сформулируем основные определения. Таким образом, настоящая глава будет по существу самостоятельной и независимой от гл. XIII, за тем лишь исключением, что трудное доказательство сходимости (5.8) не будет повторено во всей полноте.

Состояния цепи Маркова будут классифицированы независимым образом с двух точек зрения. Классификация по свойствам возвратности и невозвратности является фундаментальной, тогда как классификация по свойствам периодичности и непериодичности касается технических деталей. Периодичность связана с некоторыми неудобствами, ибо требует постоянных тривиальных замечаний; начинающий читатель должен концентрировать свое внимание на цепях, не имеющих периодических состояний. Все определения в этом параграфе связаны только с матрицей переходных вероятностей и не зависят от начального распределения $\{a_j\}$.

Определение 1. *Состояние E_j имеет период $t > 1$, если $p_{jj}^{(n)} = 0$, когда n не является кратным t, и t — наибольшее целое число, обладающее этим свойством. Состояние E_j является непериодическим, если такого $t > 1$ не существует*[1].

Для изучения периодического состояния E_j достаточно рассмотреть нашу цепь при испытаниях с номерами t, $2t$, $3t$, ... Таким образом мы получим новую цепь Маркова с переходными вероятностями $p_{ik}^{(t)}$, и в этой новой цепи E_j будет уже непериодическим состоянием. При помощи этого приема результаты для непериодических состояний могут быть перенесены на периодический случай. Детали мы обсудим в § 9 и (за исключением следующего примера) сосредоточим теперь наше внимание на непериодических цепях.

Пример. а) В неограниченном случайном блуждании все состояния имеют период 2. В случайном блуждании с поглощающими экранами в точках 0 и ρ (пример 2, б)) внутренние состояния имеют период 2, однако поглощающие состояния E_0 и E_ρ будут, конечно же, непериодическими. Если хотя бы один из экранов сделать отражающим (пример 2, в)), то все состояния станут непериодическими. ▶

Обозначения. *На всем протяжении этой главы $f_{jk}^{(n)}$ означает вероятность того, что в начинающемся из E_j процессе первое попадание в E_k произойдет на n-м шаге. Мы положим $f_{jk}^{(0)} = 0$ и*

$$f_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)}, \tag{5.1}$$

$$\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}. \tag{5.2}$$

Очевидно, f_{jk} есть вероятность того, что, выходя из E_j, система когда-нибудь пройдет через E_k. Поэтому $f_{jk} \leqslant 1$. Когда $f_{jk} = 1$, последовательность $\{f_{jk}^{(n)}\}$ есть собственное распределение вероятностей, и мы будем называть его *распределением времени первого достижения E_k*. В частности, $\{f_{jj}^{(n)}\}$ будет распределением для *времен возвращения в E_j*. Определение (5.2) имеет смысл

[1]) Состояние E_j, возвращение в которое невозможно (т. е. для которого $p_{jj}^{(n)} = 0$ при всех $n > 0$), будет считаться непериодическим.

только тогда, когда $f_{jj}=1$, т. е. когда возвращение в E_j достоверно. В этом случае $\mu_j \leqslant \infty$ будет *средним временем возвращения в* E_j.

Для наших целей не потребуется действительного вычисления вероятностей $f_{jk}^{(n)}$, однако для ясности мы укажем, как можно определить $f_{jk}^{(n)}$ (при помощи стандартных для теории восстановления рассуждений). Если первое достижение E_k осуществляется при ν-м испытании ($1 \leqslant \nu \leqslant n-1$), то (условная) вероятность попадания в E_k при n-м испытании будет равна $p_{kk}^{(n-\nu)}$. Вспоминая наше соглашение $p_{kk}^{(0)}=1$, мы заключаем, что

$$p_{jk}^{(n)} = \sum_{\nu=1}^{n} f_{jk}^{(\nu)} p_{kk}^{(n-\nu)}. \tag{5.3}$$

Полагая последовательно $n=1, 2, \ldots$, мы получим рекуррентно $f_{jk}^{(1)}, f_{jk}^{(2)}, \ldots$. И обратно, если для пары j, k известны $f_{jk}^{(n)}$, то (5.3) определяет все переходные вероятности $p_{jk}^{(n)}$.

Для любого состояния E_j прежде всего встает вопрос о том, является ли достоверным возвращение в него. Если оно достоверно, то возникает вопрос: конечно или нет среднее время возвращения μ_j? Следующее определение согласуется с терминологией гл. XIII.

Определение 2. *Состояние E_j возвратно, если $f_{jj}=1$, и невозвратно, если $f_{jj}<1$.*

Возвратное состояние E_j называется нулевым, если для него среднее время возвращения $\mu_j=\infty$.

Это определение применимо и к периодическим состояниям. Все возвратные состояния оно подразделяет на нулевые и ненулевые состояния. Последние представляют особый интерес, и, поскольку обычно мы фокусируем наше внимание на непериодических состояниях, нам будет удобно воспользоваться термином «эргодическое состояние» для непериодических возвратных ненулевых состояний[1]). Итак, мы вводим следующее определение.

Определение 3. *Непериодическое возвратное состояние E_j, у которого $\mu_j<\infty$, называется эргодическим.*

В следующей теореме приведены выраженные через переходные вероятности $p_{jj}^{(n)}$ условия того, что E_j является состоянием того или иного типа. Содержащиеся в ней критерии весьма важны, хотя обычно слишком трудны для практического их использования. Бо-

---

[1]) К сожалению, эта терминология не является общепринятой. В терминологии Колмогорова невозвратные состояния называются *«несущественными»*, однако в этой главе мы намереваемся показать, что именно невозвратные состояния часто представляют теоретический и практический интерес. (Это мнение поддерживается и современной теорией потенциала.) Эргодические состояния иногда называются *«положительными»*, термин «эргодическое» иногда используется вместо нашего «возвратное».

лее удобные критерии можно найти в § 7 и 8, однако, к сожалению, простого универсального критерия не существует.

Теорема. (i) *E_j является невозвратным состоянием тогда и только тогда, когда*

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty. \tag{5.4}$$

В этом случае при всех i

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty. \tag{5.5}$$

(ii) *E_j является (возвратным) нулевым состоянием тогда и только тогда, когда*

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty, \text{ но } p_{jj}^{(n)} \to 0 \tag{5.6}$$

при $n \to \infty$. В этом случае при всех i

$$p_{ij}^{(n)} \to 0. \tag{5.7}$$

(iii) *Непериодическое (возвратное) состояние E_j является эргодическим тогда и только тогда, когда $\mu_j < \infty$. В этом случае при $n \to \infty$*

$$p_{ij}^{(n)} \to f_{ij}\mu_j^{-1}. \tag{5.8}$$

Следствие. *Если E_j — непериодическое состояние, то $p_{ij}^{(n)}$ стремится либо к нулю, либо к пределу в (5.8).*

Доказательство. Утверждение (5.4) содержится в теореме 2 гл. XIII, 3. Утверждение (5.5) является непосредственным следствием (5.4) и (5.3), однако оно содержится также в теореме 1 гл. XIII, 5.

Для непериодического возвратного состояния E_j теорема 3 гл. XIII, 3 утверждает, что $p_{jj}^{(n)} \to \mu_j^{-1}$, где правую часть надо считать равной нулю, если $\mu_j = \infty$. Утверждения (5.7) и (5.8) непосредственно следуют из этого соотношения и (5.3) или же из теоремы 1 гл. XIII, 5.

Пусть E_j возвратно и $\mu_j = \infty$. По теореме 4 гл. XIII, 3 в этом случае $p_{jj}^{(n)} \to 0$, откуда снова следует (5.7). ▶

Примеры. б) Рассмотрим состояние E_0 в цепи из примера 2, м). Своеобразная структура матрицы переходных вероятностей показывает, что первое возвращение в E_0 при n-м испытании возможно лишь при наличии последовательности переходов

$$E_0 \to E_1 \to E_2 \to \ldots \to E_{n-1} \to E_0,$$

и поэтому при $n > 1$

$$f_{00}^{(n)} = p_1 p_2 \ldots p_{n-1} q_n, \tag{5.9}$$

а $f_{00}^{(1)}=q_1$. В частном случае, когда p_k определяются формулой (2.5), это сводится к $f_{00}^{(n)}=f_n$. Поэтому E_0 невозвратно, если $\sum f_n<1$. Для возвратного E_0 среднее время возвращения μ_0 совпадает со средним распределения $\{f_n\}$. Наконец, если E_0 имеет период t, то $f_n=0$, за исключением тех n, которые кратны t. Короче говоря, как и следовало ожидать, при любых обстоятельствах E_0 имеет тот же тип, что и рекуррентное событие $\mathscr{E}$, связанное с нашей цепью Маркова.

в) Если в примере 4, а) система покинет состояние E_2, то возвращение в это состояние станет невозможным, и поэтому E_2 невозвратно. Небольшое усовершенствование этого рассуждения показывает, что состояния E_6 и E_7 также невозвратны. Из теоремы 4 § 6 следует, что все остальные состояния цепи эргодичны.

§ 6. НЕПРИВОДИМЫЕ ЦЕПИ. РАЗЛОЖЕНИЯ

Для краткости мы будем говорить, что *два состояния однотипны*, если они одинаково характеризуются во всех классификациях, введенных в предыдущем параграфе. Другими словами, два однотипных состояния либо имеют один и тот же период, либо оба непериодичны; или оба невозвратны, или оба возвратны; в последнем случае средние времена возвращения у обоих либо бесконечны, либо конечны.

Полезность нашей классификации обусловлена в значительной степени тем фактом, что для всех практических целей всегда можно ограничиться рассмотрением состояний какого-то одного типа. Следующая теорема показывает, что для неприводимых цепей это справедливо всегда.

Теорема 1. *В неприводимой цепи все состояния однотипны.*

Доказательство. Пусть E_j и E_k — два произвольных состояния неприводимой цепи. В силу критерия из § 4 каждое состояние достижимо из любого другого состояния, и поэтому существуют такие целые числа r и s, что $p_{jk}^{(r)}=\alpha>0$ и $p_{kj}^{(s)}=\beta>0$. Очевидно,

$$p_{jj}^{(n+r+s)}\geqslant p_{jk}^{(r)}p_{kk}^{(n)}p_{kj}^{(s)}=\alpha\beta p_{kk}^{(n)}. \tag{6.1}$$

Здесь j, k, r и s фиксированы, тогда как n произвольно. Если E_j невозвратно, то левая часть (6.1) является членом сходящегося ряда, и, стало быть, то же справедливо и для $p_{kk}^{(n)}$. Более того, если $p_{jj}^{(n)}\to 0$, то и $p_{kk}^{(n)}\to 0$. Аналогичные утверждения будут иметь место и в случае, когда j и k поменяются ролями, и поэтому либо E_j и E_k оба невозвратны, либо оба возвратны; если одно из них является нулевым, то таким же будет и другое состояние.

Предположим, наконец, что E_j имеет период t. При $n=0$ правая часть (6.1) положительна, и, следовательно, $r+s$ кратно t. Но тогда левая часть (6.1) при n, не кратных t, будет обращаться в нуль, и поэтому E_k имеет период, который кратен t. Меняя j и k ролями, мы видим, что эти состояния имеют одинаковый период. ►

Важность теоремы 1 отчасти выявляет следующая теорема.

Теорема 2. *Для любого возвратного состояния E_j существует единственное неприводимое замкнутое множество C, содержащее E_j и такое, что для каждой пары E_i, E_k состояний из C*

$$f_{ik}=1 \quad и \quad f_{ki}=1. \tag{6.2}$$

Иначе говоря: выходя из произвольного состояния E_i из C, система с достоверностью пройдет через каждое состояние из множества C; по определению замыкания выход из C невозможен.

Доказательство. Пусть E_k — достижимое из E_j состояние. Тогда очевидно, что в E_k можно попасть, не возвращаясь предварительно в E_j; вероятность этого события мы обозначим через α. Но, попав в E_k, мы никогда не вернемся в E_j с вероятностью $1-f_{kj}$. Стало быть, вероятность того, что выходящая из E_j система никогда не вернется в E_j, не меньше $\alpha(1-f_{kj})$. Однако для возвратного E_j вероятность невозвращения равна нулю, и поэтому $f_{kj}=1$ для любого E_k, достижимого из E_j.

Обозначим через C совокупность всех достижимых из E_j состояний. Если E_i и E_k принадлежат C, то, как мы видели, E_j достижимо из E_k, и, следовательно, E_i также достижимо из E_k. Таким образом, каждое состояние из C достижимо из любого другого состояния из C, и поэтому в силу критерия из § 4 C неприводимо. Отсюда следует, что все состояния из C возвратны, и поэтому роль E_j в первой части рассуждения может играть любое E_i из этого множества. Это означает, что $f_{ki}=1$ для всех E_k из C, и тем самым справедливо (6.2). ►

Из этой теоремы вытекает, что замыкание возвратного состояния неприводимо. Для невозвратного состояния это не обязательно так.

Пример. Предположим, что $p_{jk}=0$ при $k\leqslant j$, но $p_{j,\,j+1}>0$. Здесь имеют место только переходы в состояния с бо́льшими номерами, и поэтому невозможно возвращение ни в одно состояние. Каждое E_j невозвратно, и замыкание E_j состоит из состояний E_j, E_{j+1}, E_{j+2}, ..., однако и само это замыкание содержит замкнутое подмножество, получающееся при вычеркивании E_j. Отсюда следует, что неприводимых множеств у этой цепи не существует. ►

Из последней теоремы вытекает, в частности, что из возвратного состояния не достигается ни одно невозвратное состояние. Если же цепь содержит оба типа состояний, то это означает, что матрица P допускает символическое разбиение вида (4.1), где матрица Q соответствует возвратным состояниям. Излишне говорить, что матрица Q вновь может быть разложимой. Но каждое возвратное состояние принадлежит единственному *неприводимому* подмножеству, и между этими подмножествами невозможны никакие переходы. Сказанное резюмирует следующая теорема.

Теорема 3. *Состояния цепи Маркова единственным образом могут быть разбиты на неперекрывающиеся множества $T, C_1, C_2, \ldots$, такие, что*

(i) *T состоит из всех невозвратных состояний;*

(ii) *если E_j принадлежит C_ν, то $f_{jk}=1$ для всех E_k из C_ν, тогда как $f_{jk}=0$ для всех E_k, не принадлежащих C_ν.*

Отсюда вытекает, что C_ν неприводимо и содержит только однотипные возвратные состояния. Приведенный выше пример показывает, что все состояния могут быть невозвратными, а пример 4, г) доказывает, что множеств C_ν может быть бесконечно много.

Следующую теорему мы выведем как простое следствие из теоремы 2, однако она может быть доказана и другими простыми способами (см. задачи 18—20).

Теорема 4. *В конечной цепи не существует нулевых состояний, и все состояния ее не могут быть невозвратными.*

Доказательство. Для каждой строки матрицы P^n сумма элементов равна единице, и так как строки эти состоят из фиксированного числа элементов, то невозможно, чтобы $p_{jk}^{(n)} \to 0$ для всех j, k. Поэтому не все состояния невозвратны. Но возвратное состояние принадлежит некоторому неприводимому множеству C. Все состояния из C однотипны. Поэтому тот факт, что C содержит возвратное состояние и хотя бы одно ненулевое состояние, и будет означать, что в C нет ни одного нулевого состояния. ▶

§ 7. ИНВАРИАНТНЫЕ РАСПРЕДЕЛЕНИЯ

Поскольку каждое возвратное состояние принадлежит некоторому неприводимому множеству, асимптотическое поведение которого можно изучать независимо от остальных состояний, мы сейчас займемся исключительно неприводимыми цепями. Все состояния такой цепи однотипны, и мы начнем с простейшего случая, а именно с цепей с конечными средними временами возвращения μ_j. Чтобы избежать тривиальных переформулировок, мы отложим обсуждение периодических цепей до § 9. Иначе говоря, мы рассмотрим теперь цепи, состояния которых непериодичны и возвратны с конечными временами возвращения. Такие цепи называются эргодическими (определение 5.3).

Теорема. *В неприводимой эргодической цепи существуют не зависящие от начального состояния j пределы*

$$u_k = \lim_{n\to\infty} p_{jk}^{(n)}. \tag{7.1}$$

Более того, $u_k > 0$,

$$\sum u_k = 1, \tag{7.2}$$

и [1])

$$u_j = \sum_i u_i p_{ij}. \tag{7.3}$$

Обратно, предположим, что цепь неприводима и непериодична и что существуют числа $u_k \geqslant 0$, *удовлетворяющие* (7.2) *и* (7.3). *Тогда цепь эргодична, и для* u_k *справедливы соотношения* (7.1) *и*

$$u_k = 1/\mu_k, \tag{7.4}$$

где μ_k *— среднее время возвращения в* E_k.

Доказательство. (i) Предположим, что цепь неприводима и эргодична, и определим u_k по формуле (7.4). Теорема 2 § 6 гарантирует, что $f_{ij}=1$ для каждой пары состояний, и поэтому утверждение (7.1) сводится к (5.8). Далее,

$$p_{ik}^{(n+1)} = \sum_j p_{ij}^{(n)} p_{jk}. \tag{7.5}$$

При $n \to \infty$ левая часть сходится к u_k, тогда как общий член справа стремится к $u_j p_{jk}$. Взяв справа лишь конечное число членов, мы заключаем, что

$$u_k \geqslant \sum_j u_j p_{jk}. \tag{7.6}$$

Для фиксированных i и n левые части в (7.5) дают в сумме единицу, и поэтому

$$s = \sum u_k \leqslant 1. \tag{7.7}$$

Суммируя по k в (7.6), мы получаем соотношение $s \geqslant s$, знак неравенства в котором невозможен. Следовательно, в (7.6) при всех k справедливо равенство, и тем самым первая часть теоремы доказана.

(ii) Допустим теперь, что $u_k \geqslant 0$ и выполняются равенства (7.2) и (7.3). По индукции

$$u_k = \sum_i u_i p_{ik}^{(n)} \tag{7.8}$$

для всех $n > 1$. Поскольку цепь предполагается неприводимой, все состояния однотипны. Если бы они были невозвратными или нулевыми, то правая часть (7.8) стремилась бы к 0 при $n \to \infty$, а этого при всех k быть не может, ибо u_k в сумме составляют единицу. Поскольку случай периодических цепей мы исключили, то это означает, что цепь эргодична, так что применима первая часть теоремы.

---

[1]) Если представить $\{u_j\}$ как вектор-строку, то (7.3) можно будет записать в матричной форме: $u = uP$.

Поэтому, устремляя n к ∞, получаем

$$u_k = \sum_i u_i \mu_k^{-1}. \tag{7.9}$$

Следовательно, распределение вероятностей $\{u_k\}$ пропорционально распределению $\{\mu_k^{-1}\}$, и поэтому, как утверждалось, $u_k = \mu_k^{-1}$. ►

Чтобы оценить все значение этой теоремы, рассмотрим эволюцию процесса от начального распределения $\{a_j\}$. Вероятность попадания при n-м шаге в состояние E_k равна

$$a_k^{(n)} = \sum_j a_j p_{jk}^{(n)} \tag{7.10}$$

(см. (3.4)). Поэтому в силу (7.1) при $n \to \infty$

$$a_k^{(n)} \to u_k. \tag{7.11}$$

Иначе говоря, каким бы ни было начальное распределение, вероятность состояния E_k стремится к u_k. С другой стороны, если начальное распределение равно $\{u_k\}$ (т. е. если $a_k = u_k$), то из (7.3) вытекает, что $a_k^{(1)} = u_k$ и — по индукции — что $a_k^{(n)} = u_k$ при всех n. Таким образом, удовлетворяющее (7.3) начальное распределение «увековечивается» на все времена. По этой причине оно называется инвариантным.

Определение. *Распределение вероятностей $\{u_k\}$, удовлетворяющее условию* (7.3), *называется инвариантным или стационарным (для данной цепи Маркова).*

Теперь главную часть предыдущей теоремы можно переформулировать следующим образом.

Неприводимая непериодичная цепь Маркова обладает инвариантным распределением вероятностей $\{u_k\}$ тогда и только тогда, когда она эргодична. В этом случае $u_k > 0$ при всех k и безусловные вероятности $a_k^{(n)}$ стремятся к u_k независимо от начального распределения.

Физический смысл стационарности станет понятным, если мы представим себе большое число одновременно протекающих процессов. Конкретнее, рассмотрим N частиц, независимо совершающих случайные блуждания одного типа. Среднее число частиц, находящихся на n-м шаге в состоянии E_k, равно величине $Na_k^{(n)}$, которая стремится к Nu_k при $n \to \infty$. После достаточно длительного промежутка времени распределение станет приблизительно инвариантным, и физик в такой ситуации сказал бы, что он наблюдает частицы в равновесии. Поэтому распределение $\{u_k\}$ называется также *равновесным распределением*. К сожалению, этот термин отвлекает внимание от того важного обстоятельства, что относится он к так называемому *макроскопическому равновесию*, т. е.

равновесию, поддерживаемому большим числом переходов в противоположных направлениях. Отдельная же частица не проявляет никакой тенденции к равновесию, и для этого индивидуального процесса никаких следствий из нашей предельной теоремы не проистекает.

Типичным в этом отношении является симметричное случайное блуждание, рассматривавшееся в гл. III. Если большое число частиц независимо совершают такие случайные блуждания, начавшиеся в нуле, то в любой момент времени приблизительно половина из них будет справа, а другая половина слева от нуля. Однако это не означает, что большинство частиц проводят половину своего времени на положительной стороне. Напротив, законы арксинуса показывают, что большинство частиц проводят непропорционально большую часть своего времени на одной стороне от нуля, и в этом смысле *большинство не является представителем ансамбля*. Этот случай является крайним в том смысле, что средние времена возвращения в нем бесконечны. В эргодических цепях случайные флуктуации более умеренны, но практически они будут носить тот же характер, если времена возвращения будут иметь очень большие (или бесконечные) дисперсии. Должным образом осознав статистическую природу «тенденции к равновесию», можно было бы избежать многих пространных рассуждений и ложных заключений.

В предыдущей теореме мы предполагали цепь неприводимой и непериодичной, и теперь уместно спросить, в какой степени существенны эти предположения. Внимательное чтение доказательства показывает, что на самом деле мы доказали больше, чем утверждалось в теореме. В частности, мы по ходу дела получили следующий критерий, применимый к произвольным цепям (включая периодические и приводимые цепи).

Критерий. *Если цепь обладает инвариантным распределением вероятностей $\{u_k\}$, то $u_k=0$ для каждого E_k, являющегося либо невозвратным, либо возвратным нулевым состоянием.*

Иначе говоря, из $u_k>0$ вытекает, что E_k возвратно и имеет конечное среднее время возвращения, но может быть и периодическим.

Доказательство. Мы видели, что из стационарности распределения $\{u_k\}$ следует соотношение (7.8). Если E_k — невозвратное или нулевое состояние, то $p_{jk}^{(n)} \to 0$ при всех j, и поэтому, как и утверждалось, $u_k=0$. ▶

Что касается периодических цепей, то, забегая вперед, упомянем результат, доказанный в § 9 и заключающийся в том, что *для каждой неприводимой цепи, состояния которой имеют конечные средние времена возвращения, существует единственное инвариантное распределение вероятностей* $\{u_k\}$. Случай периодических цепей был исключен в настоящем параграфе лишь потому, что простые предельные соотношения (7.1) и (7.11) принимают для него менее привлекательный вид, который только отвлекает от существа дела, в действительности на него не влияя.

Примеры. а) Цепи с несколькими неприводимыми компонентами могут допускать несколько стационарных решений. Банальный, но типичный пример этого — случайное блуждание с двумя поглощающими состояниями E_0 и E_ρ (см. пример 2, б)). Для него каждое распределение вероятностей вида $(\alpha, 0, 0, \ldots, 0, 1-\alpha)$, приписывающее положительные веса только состояниям E_0 и E_ρ, является стационарным.

б) По данной матрице переходных вероятностей не всегда легко определить, существует ли инвариантное распределение $\{u_k\}$. Замечательное исключение составляет случай, когда

$$p_{jk}=0 \quad \textit{при} \quad |k-j|>1, \tag{7.12}$$

т. е. когда все ненулевые элементы матрицы лежат на главной диагонали или на непосредственно примыкающей к ней прямой. Если состояния перенумерованы, начиная с 0, то соотношения (7.3) здесь примут вид

$$\begin{aligned} u_0 &= p_{00}u_0+p_{10}u_1, \\ u_1 &= p_{01}u_0+p_{11}u_1+p_{21}u_2 \end{aligned} \tag{7.13}$$

и т. д. Чтобы избежать тривиальных оговорок, предположим, что $p_{j,\,j+1}>0$ и $p_{j,\,j-1}>0$ при всех $j>0$; относительно диагональных элементов p_{jj} не предполагается ничего. Уравнения (7.13) могут быть последовательно разрешены относительно $u_1, u_2, \ldots$. Вспоминая, что для каждой строки матрицы P сумма элементов равна единице, мы находим

$$u_1=\frac{p_{01}}{p_{10}}u_0, \quad u_2=\frac{p_{01}p_{12}}{p_{10}p_{21}}u_0, \quad u_3=\frac{p_{01}p_{12}p_{23}}{p_{10}p_{21}p_{32}}u_0 \tag{7.14}$$

и т. д. Получаемая в результате (конечная или бесконечная) последовательность $u_0, u_1, \ldots$ представляет собой единственное решение (7.13). Чтобы превратить ее в распределение вероятностей, надо выбрать нормирующий множитель u_0 так, чтобы $\sum u_k=1$. Такой выбор возможен тогда и только тогда, когда

$$\sum \frac{p_{01}p_{12}p_{23}\cdots p_{k-1,\,k}}{p_{10}p_{21}p_{32}\cdots p_{k,\,k-1}}<\infty. \tag{7.15}$$

Таким образом, это есть необходимое и достаточное условие для существования инвариантного распределения вероятностей; если же последнее существует, то оно с необходимостью единственно. (Если (7.15) не выполняется, то (7.14) является так называемой инвариантной мерой; см. § 11.)

В примере 8,г) мы выведем аналогичный критерий для проверки возвратности состояний. Следующие три примера иллюстрируют применимость нашего критерия.

в) *Отражающие экраны.* Пример 2, в) (с $\rho \leqslant \infty$) является частным случаем предыдущего примера с $p_{j,\, j+1} = p$ при всех $j < \rho$ и $p_{j,\, j-1} = q$ при всех $j > 1$. Когда число состояний конечно, существует инвариантное распределение с u_k, пропорциональными $(p/q)^k$. Если же число состояний бесконечно, то для сходимости (7.15) требуется, чтобы $p < q$, и в этом случае $u_k = (1 - p/q)(p/q)^k$. Из общей теории случайных блужданий ясно, что при $p > q$ состояния цепи будут невозвратными, а при $p = q$ — возвратными нулевыми. Это будет вытекать также из критерия в примере 8,г).

г) *Модель Эренфестов для диффузии.* Для матрицы из примера 2,д) решение (7.14) сводится к

$$u_k = \binom{\rho}{k} u_0, \quad k = 0, \ \ldots, \ \rho. \tag{7.16}$$

Биномиальные коэффициенты являются членами разложения бинома Ньютона $(1+1)^\rho$, и поэтому, чтобы получить распределение вероятностей, мы должны положить $u_0 = 2^{-\rho}$. Эта цепь имеет период 2, состояния имеют конечные средние времена возвращения, и *инвариантным является биномиальное распределение с $p = 1/2$*.

Данный результат можно интерпретировать следующим образом: каково бы ни было начальное число молекул в первом сосуде, через достаточно долгое время вероятность обнаружения в нем k молекул будет примерно такой же, как если бы ρ молекул были распределены по сосудам случайным образом, причем для каждой молекулы вероятность находиться в первом сосуде равнялась бы 1/2. Это типичный пример того, как наши результаты приобретают физический смысл.

При больших ρ нормальное приближение для биномиального распределения показывает, что, коль скоро приблизительно установится предельное распределение, мы практически наверняка найдем примерно половину молекул в каждом из сосудов. Для физика число $\rho = 10^6$, конечно, является малым. Но даже при $\rho = 10^6$ молекул вероятность обнаружить более 505 000 молекул в одном сосуде (флуктуация плотности примерно на 1 процент) имеет величину порядка 10^{-23}. При $\rho = 10^8$ флуктуация плотности на одну тысячную будет иметь такую же пренебрежимо малую вероятность. Верно, что система будет изредка попадать в весьма невероятные состояния, однако времена возвращения для них фантастически велики по сравнению с временами возвращения в состояния, близкие к равновесию. Физическая необратимость проявляется в том факте, что когда бы система ни оказалась в состоянии, далеком от равновесия, для нее гораздо правдоподобнее изменения в сторону равновесия, чем в противоположном направлении.

д) *Модель Бернулли — Лапласа для диффузии.* Для матрицы с элементами (2.1) получаем из (7.14), что

$$u_k = \binom{\rho}{k}^2 u_0, \quad k = 0, \ \ldots, \ \rho. \tag{7.17}$$

Сумма квадратов биномиальных коэффициентов равна $\binom{2\rho}{\rho}$ (см. формулу (12.11) гл. II), и, следовательно,

$$u_k = \binom{\rho}{k}^2 \binom{2\rho}{\rho}^{-1} \tag{7.18}$$

представляет собой *инвариантное распределение*. Это гипергеометрическое распределение (см. гл. II, 6). Таким образом, в состоянии равновесия распределение частиц по цвету в каждой урне такое же, как если бы содержащиеся в ней ρ частиц были выбраны случайным образом из набора из ρ черных и ρ белых частиц.

е) В примере 2, м) соотношения для инвариантного распределения вероятностей примут вид

$$u_k = p_k u_{k-1}, \quad k = 1, 2, \ldots, \tag{7.19 а}$$

$$u_0 = q_1 u_0 + q_2 u_1 + q_3 u_2 + \ldots . \tag{7.19 б}$$

Из (7.19а) получаем

$$u_k = p_1 \ldots p_k u_0, \tag{7.20}$$

и теперь легко видеть, что первые k членов в правой части (7.19б) в сумме составляют $u_0 - u_k$. Поэтому (7.19б) автоматически выполняется, если $u_k \to 0$, и *инвариантное распределение вероятностей существует тогда и только тогда, когда*

$$\sum_k p_1 p_2 \ldots p_k < \infty. \tag{7.21}$$

(См. также примеры 8, д) и 11, в).)

ж) *Рекуррентные события*. В примере 2,л) условия для инвариантного распределения вероятностей сводятся к

$$u_k = u_{k+1} + f_{k+1} u_0, \quad k = 0, 1, \ldots . \tag{7.22}$$

Суммируя по $k = 0, 1, \ldots$, получаем

$$u_n = r_n u_0, \quad \textit{где } r_n = f_{n+1} + f_{n+2} + \ldots . \tag{7.23}$$

Но $r_0 + r_1 + \ldots = \mu$ равно математическому ожиданию распределения $\{f_k\}$. Итак, *инвариантное распределение вероятностей дается формулой $u_n = r_n/\mu$, если $\mu < \infty$*; при $\mu = \infty$ такого распределения не существует.

Напомним, что наша цепь Маркова связана с рекуррентным событием $\mathscr{E}$ с распределением $\{f_k\}$ времени возвращения. В частном случае $p_k = r_k/r_{k-1}$ цепь из предыдущего примера связана с тем же рекуррентным событием $\mathscr{E}$, и в этом случае (7.20) и (7.23) эквивалентны. Следовательно, *инвариантные распределения здесь совпадают*. На языке теории массового обслуживания следовало бы сказать, что *распределения затраченного времени ожидания и оста-*

точного времени ожидания сходятся к одному и тому же распределению, а именно к $\{r_n/\mu\}$.

Основные предельные теоремы для цепей Маркова мы вывели из теории рекуррентных событий. Теперь мы видим, что и, наоборот, рекуррентные события могут рассматриваться как специальные цепи Маркова. (См. также пример 11,г).)

з) *Дважды стохастические матрицы*. Стохастическая матрица P называется дважды стохастической, если не только суммы по строкам, но и суммы по столбцам равны единице. Если такая цепь содержит лишь конечное число a состояний, то инвариантным распределением будет $u_k=a^{-1}$. Это означает, что при макроскопическом равновесии все состояния равновероятны. ▶

§ 8. НЕВОЗВРАТНЫЕ СОСТОЯНИЯ

Мы видели в § 6, что возвратные состояния любой цепи Маркова могут быть разбиты на неперекрывающиеся замкнутые неприводимые множества $C_1, C_2, \ldots$. В общем случае существует также непустой класс T невозвратных состояний. Когда система выходит из невозвратного состояния, возникают две возможности: либо система в конце концов попадает в одно из замкнутых множеств C_ν и останется в нем навсегда, либо система останется навсегда во множестве T невозвратных состояний. Наша главная задача состоит в определении соответствующих вероятностей. Решение ее даст нам критерий, позволяющий решить, каким является данное состояние — возвратным или невозвратным.

Примеры. а) *Мартингалы*. Цепь называется мартингалом, если для каждого j математическое ожидание распределения вероятностей $\{p_{jk}\}$ равно j, т. е. если [1])

$$\sum_k p_{jk}k = j. \tag{8.1}$$

Рассмотрим конечную цепь с набором состояний $E_0, \ldots, E_a$. Полагая в (8.1) $j=0$ и $j=a$, мы видим, что $p_{00}=p_{aa}=1$, и поэтому E_0 и E_a суть поглощающие состояния. Чтобы избежать тривиальных оговорок, предположим, что цепь не содержит других замкнутых множеств. Отсюда следует, что внутренние состояния $E_1, \ldots, E_{a-1}$ невозвратны, и поэтому процесс рано или поздно закончится либо в E_0, либо в E_a. Из (8.1) выводим по индукции, что при всех n

$$\sum_{k=0}^{a} p_{ik}^{(n)}k = i. \tag{8.2}$$

[1]) Очевидно, если отождествить E_j с целым числом j, то такая цепь будет мартингалом в смысле обычного определения (см. том 2). — *Прим. перев.*

Но $p_{ik}^{(n)} \to 0$ для каждого невозвратного состояния E_k, и поэтому из (8.2) вытекает, что при всех $i > 0$

$$p_{ia}^{(n)} \to i/a. \tag{8.3}$$

Иначе говоря, если процесс начинается в E_i, то *вероятности окончательного поглощения в E_0 и E_a равны $1-i/a$ и i/a соответственно.*

б) *Частные случаи.* Рассмотренные в примерах из генетики 2, з) и 2, и) цепи имеют вид, обсуждавшийся в предыдущем примере, с $a=N$ и $a=2N$ соответственно. Стало быть, при условии, что начальным состоянием является E_i, вероятность окончательного прекращения процесса в E_0 равна $1-i/a$.

в) Рассмотрим цепь с набором состояний $E_0, E_1, \ldots$, такую, что E_0 — поглощающее состояние, тогда как из других состояний E_j возможны переходы только в правое соседнее состояние E_{j+1} и в E_0. Для $j \geqslant 1$ положим

$$p_{j0}=\varepsilon_j, \quad p_{j,\,j+1}=1-\varepsilon_j, \tag{8.4}$$

где $\varepsilon_j > 0$. При начальном состоянии E_j вероятность непоглощения в E_0 в первых n испытаниях равна

$$(1-\varepsilon_j)(1-\varepsilon_{j+1})\ldots(1-\varepsilon_{j+n-1}). \tag{8.5}$$

Это произведение убывает с ростом n и, следовательно, стремится к некоторому пределу λ_j. Отсюда мы выводим, что вероятность окончательного поглощения равна $1-\lambda_j$, тогда как с вероятностью λ_j система навсегда останется в множестве невозвратных состояний. Для положительности λ_j необходимо и достаточно, чтобы $\sum \varepsilon_k < \infty$. ►

Основой для изучения невозвратных состояний является подматрица матрицы P, получающаяся вычеркиванием всех строк и столбцов, соответствующих возвратным состояниям, и содержащая лишь те элементы p_{jk}, для которых и E_j, и E_k невозвратны. Суммы элементов строк этой подматрицы уже не равны единице, и здесь удобно ввести следующее определение.

Определение. *Квадратная матрица Q с элементами q_{ik} называется субстохастической, если $q_{ik} \geqslant 0$ и суммы элементов каждой строки $\leqslant 1$.*

Каждая стохастическая матрица является субстохастической в смысле этого определения, и, наоборот, каждая субстохастическая матрица может быть дополнена до стохастической добавлением поглощающего состояния E_0. (Иначе говоря, мы добавляем сверху строку 1, 0, 0, ..., а слева — столбец, элементы которого равны дефектам строк [1]) матрицы Q.) Поэтому очевидно, что все сказанное

[1]) Дефектом строки называется разность между единицей и суммой элементов этой строки.— *Прим. перев.*

относительно стохастических матриц применимо без существенных изменений и к субстохастическим матрицам. В частности, рекуррентное соотношение (3.2) определяет n-ю степень Q^n как матрицу с элементами

$$q_{ik}^{(n+1)}=\sum_{\nu} q_{i\nu}q_{\nu k}^{(n)}. \tag{8.6}$$

Обозначим через $\sigma_i^{(n)}$ сумму элементов i-й строки Q^n. Тогда для $n\geqslant 1$

$$\sigma_i^{(n+1)}=\sum_{\nu} q_{i\nu}\sigma_{\nu}^{(n)}, \tag{8.7}$$

и это соотношение остается справедливым и для $n=0$, если положить $\sigma_{\nu}^{(0)}=1$ при всех ν. Тот факт, что Q является субстохастической, означает, что $\sigma_i^{(1)}\leqslant\sigma_i^{(0)}$, и из (8.7) выводим по индукции, что $\sigma_i^{(n+1)}\leqslant\sigma_i^{(n)}$. Стало быть, для фиксированного i последовательность $\{\sigma_i^{(n)}\}$ монотонно убывает к пределу $\sigma_i\geqslant 0$, и ясно, что

$$\sigma_i=\sum_{\nu} q_{i\nu}\sigma_{\nu}. \tag{8.8}$$

Вся теория невозвратных состояний основывается на решениях этой системы уравнений. В некоторых случаях ненулевых решений у (8.8) не существует (т. е. $\sigma_i=0$ для всех i). В других случаях может существовать бесконечно много линейно независимых решений, т. е. различных последовательностей чисел x_i, удовлетворяющих системе

$$x_i=\sum_{\nu} q_{i\nu}x_{\nu}. \tag{8.9}$$

Первая наша задача состоит в том, чтобы охарактеризовать частное решение $\{\sigma_i\}$. Мы заинтересованы только в тех решениях $\{x_i\}$, для которых $0\leqslant x_i\leqslant 1$ при всех i. Это можно переписать в виде $0\leqslant x_i\leqslant\sigma_i^{(0)}$; сравнивая (8.9) и (8.7), мы видим, что по индукции $x_i\leqslant\sigma_i^{(n)}$ при всех n, и поэтому

$$0\leqslant x_i\leqslant 1 \text{ влечет } x_i\leqslant\sigma_i\leqslant 1. \tag{8.10}$$

Решение $\{\sigma_i\}$ мы будем называть *максимальным*, однако следует помнить, что во многих случаях $\sigma_i=0$ при всех i. Наш результат резюмирует следующая лемма.

Лемма. *Для субстохастической матрицы Q линейная система (8.9) обладает максимальным решением $\{\sigma_i\}$ со свойством (8.10). Эти σ_i являются пределами сумм элементов строк матриц Q^n.*

Отождествим теперь Q с подматрицей матрицы P, составленной только из тех элементов p_{jk}, для которых E_j и E_k невозвратны.

Линейную систему (8.9) можно записать в виде

$$x_i = \sum_T p_{i\nu} x_\nu, \quad E_i \in T, \tag{8.11}$$

где суммирование распространяется только на те ν, при которых E принадлежит множеству T невозвратных состояний. При таком отождествлении $\sigma_i^{(n)}$ будет вероятностью того, что за первые n испытаний с начальным состоянием E_i не произойдет ни одного перехода в какое-либо возвратное состояние. Следовательно, предел σ_i равен вероятности того, что ни один такой переход не произойдет никогда. Таким образом, доказана следующая теорема.

Теорема 1. *Вероятности x_i того, что система с начальным состоянием E_i навсегда останется среди невозвратных состояний, даются максимальным решением системы* (8.11).

Такие же рассуждения дают нам следующий критерий.

Критерий. *В неприводимой*[1]) *цепи Маркова с набором состояний E_0, $E_1, \ldots$ состояние E_0 является возвратным тогда и только тогда, когда линейная система*

$$x_i = \sum_{\nu=1}^{\infty} p_{i\nu} x_\nu, \quad i \geqslant 1, \tag{8.12}$$

не допускает никаких решений, удовлетворяющих $0 < x_i \leqslant 1$, за исключением $x_i = 0$ при всех i.

Доказательство. Отождествим матрицу Q из леммы с подматрицей матрицы P, получающейся вычеркиванием строки и столбца, соответствующих E_0. Рассуждения, использовавшиеся нами в теореме 1, показывают, что σ_i есть вероятность того, что (при начальном состоянии E_i) система навсегда останется среди состояний E_1, $E_2, \ldots$. Но если E_0 возвратно, то вероятность f_{i0} достижения E_0 равна 1, и, следовательно, $\sigma_i = 0$ при всех i. ▶

Примеры. г) Как и в примере 7, б), рассмотрим цепь с набором состояний E_0, $E_1, \ldots$, такую, что

$$p_{jk} = 0, \text{ когда } |k - j| > 1. \tag{8.13}$$

Чтобы избежать тривиальных оговорок, мы предположим, что $p_{j,\,j+1} \neq 0$ и $p_{j,\,j-1} \neq 0$. Эта цепь неприводима, потому что каждое состояние достижимо из любого другого состояния. Поэтому все состояния однотипны, и нам достаточно проверить характер E_0.

---

[1]) Неприводимость предполагается только для того, чтобы избежать усложнения обозначений. Никакого ограничения она здесь не представляет, поскольку нам достаточно рассмотреть замыкание E_0. Критерий применим также к периодическим цепям.

Уравнения (8.12) сводятся к рекуррентной системе

$$x_1=p_{11}x_1+p_{12}x_2,\qquad p_{j,j-1}(x_j-x_{j-1})=p_{j,j+1}(x_{j+1}-x_j),\quad j\geqslant 2. \tag{8.14}$$

Поэтому

$$x_j-x_{j+1}=\frac{p_{21}p_{32}\cdots p_{j,j-1}}{p_{23}p_{34}\cdots p_{j,j+1}}(x_1-x_2). \tag{8.15}$$

Так как $p_{10}>0$, то $x_2-x_1>0$, и поэтому ограниченное неотрицательное решение $\{x_j\}$ существует тогда и только тогда, когда

$$\sum\frac{p_{21}\cdots p_{j,j-1}}{p_{23}\cdots p_{j,j+1}}<\infty. \tag{8.16}$$

Наша цепь возвратна тогда и только тогда, когда этот ряд расходится. В частном случае случайных блужданий мы имеем $p_{j,j+1}=p$ и $p_{j,j-1}=q$ при всех $j>1$, и мы вновь видим, что все состояния возвратны тогда и только тогда, когда $p\leqslant q$.

(Цепь из этого примера можно интерпретировать как случайное блуждание на прямой с вероятностями, меняющимися от точки к точке.)

д) Для матрицы из примера 2, м) уравнения (8.12) сводятся к

$$x_j=p_{j+1}x_{j+1}, \tag{8.17}$$

и здесь ограниченное положительное решение существует тогда и только тогда, когда бесконечное произведение $p_1p_2\ldots$ сходится. Если наша цепь связана с рекуррентным событием $\mathcal{E}$, то p_k даются формулой (2.5), и это произведение сходится тогда и только тогда, когда $\sum f_j<\infty$. Таким образом (как и можно было бы предвидеть), и цепь, и событие $\mathcal{E}$ либо оба возвратны, либо оба невозвратны. ►

Чтобы ответить на последний вопрос, поставленный в начале этого параграфа, снова обозначим через T класс невозвратных состояний; пусть C — произвольное *замкнутое* множество возвратных состояний. (Мы не требуем, чтобы C было неприводимым.) Обозначим через y_i *вероятность окончательного поглощения в* C при условии, что начальным состоянием было E_i. Мы намереваемся показать, что y_i удовлетворяют системе неоднородных уравнений

$$y_i=\sum_T p_{i\nu}y_\nu+\sum_C p_{i\nu},\quad E_i\in T, \tag{8.18}$$

где суммы берутся по тем ν, для которых $E_\nu\in T$ и $E_\nu\in C$ соответственно. Система (8.18) может допускать несколько независимых решений, однако следующее доказательство покажет, что среди них существует *минимальное* решение, определяемое очевидным образом по аналогии с (8.10).

Теорема 2. *Вероятности y_i окончательного поглощения замкнутым возвратным множеством C даются минимальным неотрицательным решением системы* (8.18).

Доказательство. Обозначим через $y_i^{(n)}$ вероятности поглощения множеством C на n-м шаге или до него. Тогда ясно, что при $n\geq 1$

$$y_i^{(n+1)} = \sum_T p_{i\nu} y_\nu^{(n)} + \sum_C p_{i\nu}, \tag{8.19}$$

и если мы положим $y_\nu^{(0)} = 0$ для всех ν, то (8.19) будет справедливо и при $n=0$. Для фиксированного i последовательность $\{y_i^{(n)}\}$ не убывает, но остается ограниченной единицей. Пределы этих последовательностей, очевидно, удовлетворяют (8.18). Обратно, если $\{y_i\}$ — произвольное неотрицательное решение системы (8.18), то $y_i \geqslant y_i^{(1)}$ поскольку вторая сумма в (8.18) равна $y_i^{(1)}$. По индукции $y_i \geqslant y_i^{(n)}$ при всех n, так что пределы $y_i^{(n)}$ действительно являются минимальным решением. ▶

Иллюстрацию к этой теореме см. в примере в).

§ 9*). ПЕРИОДИЧЕСКИЕ ЦЕПИ

Ни трудностей, ни неожиданных новых свойств в случае периодических цепей не появляется. Эти цепи были исключены из формулировки главной теоремы в § 7 лишь потому, что они представляют вторичный интерес, а описание их оказывается непропорционально многословным. Содержащееся в настоящем параграфе обсуждение этого случая проводится скорее ради полноты изложения, чем из-за представляемого им интереса.

Простейшим примером цепи с периодом 3 является цепь с тремя состояниями, в которой возможны только переходы $E_1 \to E_2 \to E_3 \to E_1$. В этом случае

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad P^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Мы покажем теперь, что этот пример во многих отношениях типичен.

Рассмотрим неприводимую цепь с конечным или бесконечным числом состояний $E_1, E_2, \ldots$. По теореме 6.1 все состояния имеют один и тот же период t (мы предполагаем, что $t>1$). Поскольку в неприводимой цепи каждое состояние достижимо из любого другого состояния, то для каждого состояния E_k существуют два целых числа a и b, такие, что $p_{1k}^{(a)} > 0$ и $p_{k1}^{(b)} > 0$. Но $p_{11}^{(a+b)} \geqslant p_{1k}^{(a)} p_{k1}^{(b)}$,

*) Результаты этого параграфа не будут использоваться в дальнейшем.

и поэтому сумма $a+b$ должна быть кратной периоду t. Фиксируя b, мы заключаем, что каждое целое a, для которого $p_{1k}^{(a)}>0$, имеет вид $\alpha+\nu t$, где α—фиксированное целое, $0\leqslant\alpha<t$. Целое число α является характеристикой состояния E_k, и поэтому множество всех состояний может быть разбито на t непересекающихся классов $G_0, \ldots, G_{t-1}$ так, что

$$\text{если } E_k\in G_\alpha, \quad \text{то } p_{1k}^{(n)}=0 \quad \text{при } n\neq\alpha+\nu t. \tag{9.1}$$

Мы будем считать, что классы $G_0, \ldots, G_{t-1}$ упорядочены циклическим образом, так что G_{t-1} является левым соседом G_0.

Теперь очевидно, что за *один шаг* переход возможен только в состояние из правого соседнего класса, и, следовательно, каждый путь из t шагов всегда приводит в состояние из исходного класса. Отсюда вытекает, что в цепи Маркова с переходной матрицей P^t каждый класс G_α будет замкнутым множеством[1]). Это множество будет неприводимо, потому что в исходной цепи каждое состояние достижимо из любого другого, и внутри одного класса необходимое для этого число шагов с необходимостью должно делиться на t. Таким образом, нами доказана следующая теорема.

Теорема. *В неприводимой цепи с периодом t состояния могут быть разбиты на t непересекающихся классов $G_0, \ldots, G_{t-1}$, таких, что справедливо* (9.1) *и переходы за один шаг всегда приводят в состояние из правого соседнего класса (в частности, из G_{t-1} в G_0). В цепи с матрицей P^t каждый класс G_α соответствует неприводимому замкнутому множеству.*

Теперь, используя эту теорему, легко описать асимптотическое поведение переходных вероятностей $p_{jk}^{(n)}$. Мы знаем, что $p_{jk}^{(n)}\to 0$, если E_k является либо невозвратным, либо возвратным нулевым состоянием, а также что все состояния однотипны (§ 6). Стало быть, нам надо рассмотреть лишь случай, когда каждое состояние E_k имеет конечное среднее время возвращения μ_k. В цепи с матрицей P^t состояние E_k имеет среднее время возвращения μ_k/t, и по отношению к этой цепи каждый класс G_α эргодичен. Поэтому, если E_j принадлежит G_α, то

$$\lim_{n\to\infty} p_{jk}^{(nt)}=\begin{cases} t/\mu_k, & \text{если } E_k\in G_\alpha,\\ 0 & \text{в противном случае,}\end{cases} \tag{9.2}$$

---

[1]) При $t=3$ имеется три класса, и в символической блочной записи, введенной в § 4, матрица P принимает вид

$$\begin{pmatrix} 0 & A & 0\\ 0 & 0 & B\\ C & 0 & 0\end{pmatrix},$$

где A—матрица вероятностей переходов из G_0 в G_1, и т. д.

и веса t/μ_k определяют распределение вероятностей на состояниях из класса G_α (см. теорему § 7). Поскольку всего у нас t таких классов, то *числа* $u_k = 1/\mu_k$ *определяют распределение вероятностей на множестве всех состояний*, как это было и в случае непериодических цепей. Мы покажем, что это распределение инвариантно. Для этой цели нам нужны соотношения, соответствующие (9.2), когда показатель степени не делится на период t.

Мы отправимся от фундаментального соотношения

$$p_{jk}^{(nt+\beta)} = \sum_\nu p_{j\nu}^{(\beta)} p_{\nu k}^{(nt)}. \tag{9.3}$$

Множитель $p_{j\nu}^{(\beta)}$ обращается в нуль, за исключением тех случаев когда E_ν принадлежит $G_{\alpha+\beta}$. (Если $\alpha+\beta \geqslant t$, то под $G_{\alpha+\beta}$ следует понимать $G_{\alpha+\beta-t}$.) В этих случаях $p_{\nu k}^{(nt)}$ обращается в нуль, если E_k не лежит в $G_{\alpha+\beta}$, и, следовательно, для фиксированных β и E_j из G_α

$$\lim_{n\to\infty} p_{jk}^{(nt+\beta)} = \begin{cases} t/\mu_k, & \textit{если } E_k \in G_{\alpha+\beta}, \\ 0 & \textit{в противном случае}. \end{cases} \tag{9.4}$$

Перепишем теперь (9.3) в виде

$$p_{ik}^{(nt+1)} = \sum_\nu p_{i\nu}^{(nt)} p_{\nu k}. \tag{9.5}$$

Рассмотрим произвольное состояние E_k; пусть G_ρ — тот класс, которому оно принадлежит. Тогда $p_{\nu k} = 0$, если $E_\nu \notin G_{\rho-1}$, и поэтому обе части равенства (9.5) обращаются в нуль, если $E_i \notin G_{\rho-1}$. В этом случае $p_{ik}^{(nt+1)} \to t u_k$, откуда

$$u_k = \sum_\nu u_\nu p_{\nu k}. \tag{9.6}$$

Поскольку E_k — произвольное состояние, то тем самым мы доказали, что *распределение вероятностей* $\{u_k\}$ *инвариантно*.

§ 10. ПРИМЕНЕНИЕ К ТАСОВАНИЮ КАРТ

Колоду из N карт, занумерованных числами 1, 2, . . ., N, можно упорядочить $N!$ различными способами, и каждый из них представляет возможное состояние этой системы. Каждая отдельная операция тасования колоды осуществляет переход от имеющегося состояния в некоторое другое. Например, «снимание» заменит порядок $(1, 2, \ldots, N)$ на один из N циклически эквивалентных порядков $(r, r+1, \ldots, N, 1, 2, \ldots, r-1)$. Та же операция, примененная к обратному порядку $(N, N-1, \ldots, 1)$, приведет к состоянию $(N-r+1, N-r, \ldots, 1, N, N-1, \ldots, N-r+2)$. Иначе говоря, каждую отдельную операцию тасования колоды мы представляем себе как преобразование $E_j \to E_k$. Если повторять *в точности* одну и ту же операцию, то система пройдет (начиная с данного состояния E_j) через вполне определенную последовательность состояний, и в результате после конечного числа шагов восстановится первоначальный порядок. С этих пор эта последовательность состояний будет периодически повторяться. Для большинства операций этот период будет довольно мал, и *ни в одном случае* с помощью такой процедуры не могут быть получены все состояния [1]). Например, точная «врезка» сменила бы в колоде из $2m$ карт порядок $(1, \ldots, 2m)$ на $(1, m+1, 2, m+2, \ldots, m, 2m)$. В колоде из 6 карт начальный порядок восстанавливает четыр[illegible]ехкратное применение этой операции. При десяти картах начальный порядок по[illegible]вится вновь после шести операций, так что повторное применение этой операции к колоде из десяти карт может привести только к шести из $10! = 3\,628\,800$ возможных порядков расположения карт.

На практике игрок может пожелать варьировать опера[illegible] сования, и, конечно, разного рода отклонения будут вы[illegible]цию та[illegible] случая. Предположим, что мы можем учесть при[illegible]ваны игрой влияние случайных отклонений, считая, что [illegible]ычки игрока и операция имеет определенную вероятност[illegible] каждая отдельная Ни в каких допущениях относительно [illegible] (возможно, нулевую). роятностей мы не нуждаемся, но [illegible]численных значений этих ве- игрока не зависят от прошл[illegible] будем предполагать, что действия Отсюда вытекает, что [illegible]го и что он не знает порядка карт [2]). [illegible] испыт[illegible] последовательные операции соответствуют независим[illegible] [illegible]аниям с фиксированными вероятностями; для самой колоды [illegible] карт мы тогда имеем цепь Маркова.

Покаже[illegible]м теперь, что матрица переходных вероятностей P является дважды стохастической (пример 7, з)). Действительно, если операция изменяет состояние (порядок карт) E_j на E_k, то существует другое состояние E_r, которое этой операцией изменяется на E_j. Это означает, что элементы j-го столбца матрицы P совпадают с элементами ее j-й строки с той лишь разницей, что располагаются они там в другом порядке. Поэтому все суммы по столбцам равны единице.

Отсюда следует, что ни одно из состояний не может быть невозвратным. *Если цепь неприводима и непериодична, то в пределе все состояния станут равновероятными.* Иначе говоря, годится *любой* способ тасования, если только он приводит к неприводимой и непериодичной цепи. Можно с уверенностью полагать, что обычно так оно и есть. Предположим, однако, что колода содержит четное число карт и наша процедура состоит в разделении их на две равные

---

[1]) В терминологии теории групп это равносильно утверждению, что группа перестановок не является циклической и поэтому не может порождаться одной операцией.

[2]) Это предположение соответствует обычной ситуации при игре в бридж. Легко придумать более сложную технику тасования, при которой операции будут зависеть от предшествующих действий и окончательный результат не будет цепью Маркова (см. пример 13, д)).

части и в тасовании последних отдельно друг от друга любым методом. Если эти две части складываются затем вместе в их первоначальном порядке, то полученная цепь Маркова будет приводимой (поскольку не из каждого состояния достижимо любое другое состояние). Если же эти части складываются в обратном порядке, то цепь будет иметь период 2. Теоретически могут возникнуть обе эти возможности, но практически они едва ли осуществимы, поскольку из-за влияния случая операции не могут выполняться абсолютно точно.

Итак, вполне можно ожидать, что продолжительное тасование приведет к полной «случайности» и уничтожит все следы первоначального порядка. Следует отметить, однако, что необходимое для этой цели число операций чрезвычайно велико [1]).

§ 11 *). ИНВАРИАНТНЫЕ МЕРЫ. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ДЛЯ ОТНОШЕНИЙ

В этом параграфе мы рассмотрим неприводимую цепь с возвратными нулевыми состояниями. Наша главная цель — вывести аналоги для результатов, полученных в § 7 для цепей, в которых состояния имеют конечные средние времена возвращения. Характерным свойством таких цепей является существование инвариантного (или стационарного) распределения вероятностей, определяемого соотношением

$$u_k = \sum_\nu u_\nu p_{\nu k}. \tag{11.1}$$

Мы знаем, что в случае, когда средние времена возвращения бесконечны, такое инвариантное распределение вероятностей не существует, но мы покажем, что линейная система (11.1) все же допускает положительное решение $\{u_k\}$, такое, что $\sum u_k = \infty$. Такое решение $\{u_k\}$ называется *инвариантной (или стационарной) мерой*. Если цепь неприводима и возвратна, то инвариантная мера единственна с точностью до произвольной нормирующей постоянной.

Примеры. а) Предположим, что матрица переходных вероятностей P дважды стохастическая, т. е. суммы и по столбцам, и по стро-

[1]) По поводу анализа невероятно плохих результатов тасования в записях экспериментов по экстрасенсорному восприятию см. Feller W., Statistical aspects of ESP, Journal of Parapsychology, 4 (1940), 271—298. Гринвуд и Стюарт в своей поразительной статье (Greenwood J. A., Stuart C. E., A review of Dr. Feller's critique, там же, 299—319) пытаются показать, что эти результаты обусловлены случаем, однако и их вычисления, и их эксперименты имеют определенный оттенок сверхъестественности.

*) Этот и следующий параграфы посвящены вопросам, играющим важную роль в современных исследованиях, однако полученные в них результаты не будут использоваться в настоящей книге.

кам равны единице. Тогда справедливо (11.1) при $u_k=1$ для всех k. Чтобы выразить этот факт, говорят, что *равномерная мера инвариантна*.

б) *Случайные блуждания*. Интересным частным случаем цепи из предыдущего примера является неограниченное случайное блуждание на прямой. Занумеруем состояния в их естественном порядке от $-\infty$ до ∞. Это не позволит нам представить переходные вероятности в обычной форме матрицы, однако необходимые изменения обозначения очевидны. Если переходы в правое и левое соседние состояния имеют вероятности p и q соответственно, то система (11.1) принимает вид

$$u_k = pu_{k-1} + qu_{k+1}, \quad -\infty < k < \infty.$$

Состояния возвратны только при $p=q=1/2$, и в этом случае $u_k=1$ является единственным положительным решением. Это решение сохраняется и при $p \neq q$, однако оно уже не будет единственным; вторым неотрицательным решением будет $u_k=(p/q)^k$. Этот пример доказывает, что инвариантная мера может существовать и у невозвратных цепей, но уже не обязана быть единственной. Мы вернемся к этому интересному вопросу в следующем параграфе.

Инвариантную меру $\{u_j\}$ можно интерпретировать на интуитивном уровне, если рассмотреть одновременно бесконечно много процессов с одной и той же матрицей P переходных вероятностей. Для каждого j определим случайную величину $\mathbf{N}_j$, имеющую распределение Пуассона со средним u_j, и рассмотрим $\mathbf{N}_j$ независимых процессов, начинающихся из E_j. Мы делаем это для всех состояний одновременно, предполагая, что все процессы взаимно независимы. Нетрудно показать, что в каждый данный момент времени в любом заданном состоянии E_k с вероятностью единица может быть лишь конечное число процессов. Стало быть, число процессов, находящихся на n-м шаге в состоянии E_k, является случайной величиной $\mathbf{X}_k^{(n)}$, и *инвариантность* $\{u_k\}$ *влечет равенство* $\mathbf{E}\{\mathbf{X}_k^{(n)}\}=u_k$ *для всех* n (см. задачу 29).

в) В примере 7, е) мы обнаружили, что инвариантное распределение вероятностей существует только тогда, когда сходится ряд (7.21). В случае когда он расходится, (7.20) все еще будет инвариантной мерой, если $u_k \to 0$, а это равносильно тому, что $p_1 p_2 \ldots p_k \to 0$. Инвариантной меры не существует, когда произведение $p_1 \ldots p_k$ остается отделенным от 0, например когда $p_k = 1-(k+1)^{-2}$. В этом случае цепь невозвратна.

г) В примере 7, ж) соотношения (7.23) определяют инвариантную меру даже тогда, когда $\mu=\infty$. ▶

В эргодических цепях вероятности $p_{jk}^{(n)}$ стремятся к значению u_k инвариантного распределения вероятностей. Для возвратных нулевых цепей мы докажем слабый вариант этого результата, а именно что для всех E_α и E_β при $N \to \infty$

$$\left(\sum_{n=0}^{N} p_{\alpha i}^{(n)}\right)\left(\sum_{n=0}^{N} p_{\beta j}^{(n)}\right)^{-1} \to u_i/u_j. \tag{11.2}$$

Суммы в левой части представляют собой математические ожидания числа прохождений через E_i и E_j за первые N испытаний. Грубо говоря, (11.2) утверждает, что эти математические ожидания асимптотически независимы от начальных состояний E_α и E_β и находятся в той же пропорции, что и соответствующие значения инвариантных мер. Таким образом, характерные факты здесь те же, что и в случае эргодических цепей, хотя ситуация и более запутана. С другой стороны, периодические цепи не требуют теперь специального рассмотрения. (В действительности (11.2) справедливо для *всех* [неприводимых.— *Перев.*] возвратных цепей. Для эргодической цепи числитель в левой части равен $\sim Nu_i$.)

Соотношения вида (11.2) называются *предельными теоремами для отношений*. Мы выведем (11.2) из более сильного результата, который до недавнего времени рассматривался как более сложное уточнение. Наши доказательства основываются на рассмотрении только тех путей, которые не проходят через некоторое выделенное состояние E_r. Следуя Чжун Кайлаю, мы назовем это запрещенное состояние *табу*, а вероятности перехода в него — *табу-вероятностями*.

Определение. *Пусть E_r — произвольное фиксированное состояние. Для $E_k \neq E_r$ и $n \geqslant 1$ мы обозначим через ${}_rp_{jk}^{(n)}$ вероятность того, что начинающийся из E_j процесс попадет на n-м шаге в состояние E_k, не заходя по пути в E_r.*

Здесь E_j может совпадать с E_r. Мы распространим это определение естественным образом на случаи $E_k = E_r$ и $n = 0$, полагая

$${}_rp_{jr}^{(n)} = 0, \quad n \geqslant 1, \tag{11.3}$$

и

$${}_rp_{jk}^{(0)} = \begin{cases} 1, & \textit{если } E_j = E_k, \\ 0 & \textit{в противном случае.} \end{cases} \tag{11.4}$$

В аналитической записи мы имеем для $n \geqslant 0$ и $E_k \neq E_r$

$${}_rp_{jk}^{(n+1)} = \sum_{\nu} {}_rp_{j\nu}^{(n)} p_{\nu k}. \tag{11.5}$$

В действительности при $n=0$ сумма в правой части сводится к одному члену, а именно к p_{jk}. При $n \geqslant 1$ соответствующий $\nu = r$ член обращается в нуль в силу (11.3), так что (11.5) эквивалентно первоначальному определению.

Введение табу на состояние E_r равносильно рассмотрению исходного марковского процесса лишь до первого достижения состояния E_r. В неприводимой возвратной цепи состояние E_r достигается с вероятностью единица из любого начального состояния E_j. Отсюда следует, что в цепи с табу E_r последовательные прохождения через начальное состояние E_j составляют *невозвратное* рекуррентное событие, а прохождения через любое другое состояние $E_k \neq E_r$ составляют невозвратное рекуррентное событие с запаздыванием. Поэтому в силу основной теоремы 2 гл. XIII, 3 для $E_k \neq E_r$

$$\sum_{n=0}^{\infty} {}_r p_{jk}^{(n)} = {}_r\pi_{jk} < \infty. \tag{11.6}$$

При $E_k = E_r$ слагаемые с $n \geqslant 1$ обращаются в нуль и сумма сводится к 1 при $j=r$ и к 0 при $j \neq r$.

Теперь мы в состоянии доказать существование инвариантной меры, т. е. чисел u_k, удовлетворяющих (11.1). В доказательстве теоремы 2 это использоваться не будет.

Теорема 1. *Если цепь неприводима и возвратна, то последовательность чисел*

$$u_k = {}_r\pi_{rk} \tag{11.7}$$

представляет собой инвариантную меру; более того, $u_k > 0$ для всех k, а $u_r = 1$.

Обратно, если $u_k \geqslant 0$ при всех k и справедливо (11.1), *то существует постоянная λ, такая, что $u_k = \lambda \cdot {}_r\pi_{rk}$.*

Состояние E_r здесь произвольно, однако из единственности вытекает, что получаемые при разных r последовательности $\{u_k\}$ отличаются лишь множителями пропорциональности. Заметим, что теорема и ее доказательство охватывают также случай цепей с конечными средними временами возвращения.

Доказательство. Если $k \neq r$, то, суммируя (11.5) при $j=r$ по $n=0, 1, \ldots$, получаем

$${}_r\pi_{rk} = \sum_{\nu} {}_r\pi_{r\nu} p_{\nu k}, \tag{11.8}$$

так что числа (11.5) удовлетворяют уравнениям (11.1) хотя бы при $k \neq r$. При $j=k=r$ ясно, что

$$\sum_{\nu} {}_r p_{r\nu}^{(n)} p_{\nu r} = f_{rr}^{(n+1)} \tag{11.9}$$

есть вероятность того, что (в исходной цепи) первое возвращение в E_r произойдет на $(n+1)$-м шаге. Поскольку цепь неприводима и возвратна, в сумме эти вероятности составляют единицу. Поэтому, суммируя (11.9) по $n=0, 1, \ldots$, мы получаем

$$\sum_{\nu} {}_r\pi_{r\nu} p_{\nu r} = 1. \tag{11.10}$$

Но по определению ${}_r\pi_{rr}=1$, так что (11.8) справедливо и при $k=r$. Следовательно, (11.7) действительно представляет собой инвариантную меру.

Рассмотрим, далее, произвольную неотрицательную инвариантную меру $\{u_k\}$. Из определения (11.1) ясно, что если $u_k=0$ при некотором k, то $u_\nu=0$ для всех ν, таких, что $p_{\nu k}>0$. По индукции отсюда следует, что $u_\nu=0$ при каждом ν, таком, что E_k достижимо из E_ν. Поскольку цепь неприводима, отсюда вытекает, что $u_\nu=0$ для всех ν. Поэтому инвариантная мера строго положительна (или тождественно равна нулю).

Стало быть, в обратном утверждении теоремы мы можем считать, что заданная инвариантная мера нормирована условием $u_r=1$ для некоторого заранее выбранного r. Тогда

$$u_k=p_{kr}+\sum_{j\neq r} u_j p_{jk}. \tag{11.11}$$

Предположим, что $k\neq r$. Подставим под знаком суммы вместо чисел u_j соответствующие им выражения из соотношения (11.1) и в полученной двойной сумме снова выделим содержащий u_r член. В результате имеем

$$u_k=p_{rk}+{}_rp_{rk}^{(2)}+\sum_{\nu\neq r} u_\nu\cdot {}_rp_{\nu k}^{(2)}. \tag{11.12}$$

Продолжая в том же духе, мы получим для каждого n

$$u_k=p_{rk}+{}_rp_{rk}^{(2)}+\ldots+{}_rp_{rk}^{(n)}+\sum_{\nu\neq r} u_\nu\cdot {}_rp_{\nu k}^{(n)}. \tag{11.13}$$

Устремляя n к ∞, мы видим, что $u_k\geqslant {}_r\pi_{rk}$. Отсюда следует, что $\{u_k-{}_r\pi_{rk}\}$ определяет [неотрицательную.— *Перев.*] инвариантную меру, обращающуюся в нуль при $k=r$. Но такая мера равна нулю тождественно, и тем самым справедливо (11.7). ►

Вскоре мы увидим, что уточнением предельной теоремы для отношений является следующая теорема.

Теорема 2. *В неприводимой возвратной цепи для всех* N

$$0\leqslant \sum_{n=0}^{N} p_{kk}^{(n)}-\sum_{n=0}^{N} p_{\alpha k}^{(n)}\leqslant {}_\alpha\pi_{kk} \tag{11.14}$$

и

$$-1\leqslant (1/{}_j\pi_{ij})\sum_{n=0}^{N} p_{ii}^{(n)}-(1/{}_i\pi_{ji})\sum_{n=0}^{N} p_{jj}^{(n)}\leqslant 1. \tag{11.15}$$

Доказательство (11.14). Рассмотрим первое попадание в E_k; ясно, что при $\alpha\neq k$

$$p_{\alpha k}^{(n)}=\sum_{\nu=1}^{n} f_{\alpha k}^{(\nu)} p_{kk}^{(n-\nu)}. \tag{11.16}$$

(Это то же самое, что (5.3).) Суммируя по n, получаем

$$\sum_{n=0}^{N} p_{\alpha k}^{(n)} \leqslant \sum_{n=0}^{N} p_{kk}^{(n)} \cdot \sum_{\nu=1}^{\infty} f_{\alpha k}^{(\nu)} = \sum_{n=0}^{N} p_{kk}^{(n)}, \tag{11.17}$$

что доказывает первое неравенство в (11.14).

Заметим теперь, что при выходе из E_k возвращение в E_k [на n-м шаге.— *Перев.*] произойдет либо без промежуточного прохождения через E_α, либо первое попадание в E_α произойдет на ν-м шаге, $1 \leqslant \nu \leqslant n$. Это означает, что

$$p_{kk}^{(n)} = {}_\alpha p_{kk}^{(n)} + \sum_{\nu=1}^{n} f_{k\alpha}^{(\nu)} p_{\alpha k}^{(n-\nu)}. \tag{11.18}$$

Суммирование по n приводит ко второму неравенству в (11.14).

Доказательство (11.15). В силу очевидной симметрии i и j достаточно доказать второе неравенство. Мы отправимся от тождества

$$p_{ii}^{(n)} = {}_j p_{ii}^{(n)} + \sum_{\nu=1}^{n-1} p_{ij}^{(n-\nu)} \cdot {}_j p_{ji}^{(\nu)}, \tag{11.19}$$

которое выражает тот факт, что возвращение из E_i в E_i происходит либо без промежуточного прохождения через E_j, либо *последнее* попадание в E_j имеет место на $(n-\nu)$-м шаге и следующие ν шагов приводят из E_j в E_i без возвращения в E_j. Суммируя по n, в силу (11.14) получаем

$$\sum_{n=0}^{N} p_{ii}^{(n)} \leqslant {}_j\pi_{ii} + {}_j\pi_{ji} \sum_{n=0}^{N} p_{ij}^{(n)} \leqslant {}_j\pi_{ii} + {}_j\pi_{ji} \sum_{n=0}^{N} p_{jj}^{(n)}. \tag{11.20}$$

Чтобы привести это неравенство к симметричному виду (11.15), достаточно заметить, что

$${}_j\pi_{ji} = {}_j\pi_{ii} / {}_i\pi_{jj}. \tag{11.21}$$

Действительно, по аналогии с (11.16)

$${}_j\pi_{ji} = {}_j f_{ji} \cdot {}_j\pi_{ii}, \tag{11.22}$$

где ${}_j f_{ji}$ — вероятность достижения E_i из E_j без предварительного возвращения в E_j. Альтернативой для этого события является возвращение в E_j до достижения E_i, и, следовательно,

$${}_j f_{ji} = 1 - {}_i f_{jj} = 1/{}_i\pi_{jj}. \tag{11.23}$$

(Последнее уравнение является основным тождеством для невозвратного рекуррентного события, состоящего в возвращении в E_j без промежуточного прохождения через E_i.) Подставляя (11.23) в (11.22), получаем утверждение (11.21), и доказательство теоремы этим завершается.

►

Из соотношения (11.21) вытекает интересное следствие.

Следствие 1. *Если* $\{u_k\}$ — *инвариантная мера, то*

$$ {}_j\pi_{ii}/{}_i\pi_{jj}=u_i/u_j. \tag{11.24}$$

Доказательство. Инвариантная мера определяется с точностью до постоянного множителя, и поэтому правая часть в (11.24) определена единственным образом. Стало быть, мы можем предположить, что $\{u_k\}$ есть инвариантная мера, определенная в (11.7), когда запрещенное состояние E_r отождествляется с E_j. Но тогда $u_j=1$ и ${}_j\pi_{ji}=u_i$, и поэтому (11.21) сводится к (11.24). ►

Следствие 2. (Предельная теорема для отношений.) *Для неприводимой возвратной цепи справедливо предельное соотношение* (11.2).

Доказательство. Суммы из теоремы 2 стремятся к ∞ при $N\to\infty$. Стало быть, отношение двух сумм в (11.14) стремится к единице, и поэтому достаточно доказать (11.2) для частного случая $\alpha=i$ и $\beta=j$. Однако при таком выборе (11.2) является непосредственным следствием (11.15) и (11.24). ►

Существование инвариантной меры для возвратной цепи впервые доказал Дерман[1]). Существование предела в (11.2) установил Дёблин[2]). Табу-вероятности в качестве мощного инструмента теории цепей Маркова ввел Чжун Кайлай[3]); более детально они изучаются в его фундаментальном труде[4]). Ограниченность частных сумм $\sum\limits_{n=0}^{N}\left(p_{kk}^{(n)}-p_{ii}^{(n)}\right)$ доказал Орей, рассматривавший также вопрос о сходимости[5]).

§ 12 *). ОБРАЩЕННЫЕ ЦЕПИ. ГРАНИЦЫ

При изучении эволюции некоторой системы мы обычно интересуемся вероятностями возможных будущих событий, однако иногда бывает необходимо изучать и прошлое. В специальном случае цепи Маркова мы можем поинтересоваться, чему равна (условная) вероятность того, что в определенный момент в прошлом система

---

[1]) Derman C., A solution to a set of fundamental equations in Markov chains, Proc. Amer. Math. Soc., 5 (1954), 332—334.

[2]) Doeblin W., Sur deux problémes de A. Kolmogoroff concernant les chaines dénombrables, Buil. Soc. Math. de France, 66 (1938), 1—11.

[3]) Chung K. L., Contribution to the theory of Markov chains, I, Journ. Research, Nat. Bureau of Standards, 50 (1953), 203—208; II, Trans. Amer. Math. Soc., 76 (1954), 397—419.

[4]) Chung K. L., Markov chains with stationary transition probabilities, Berlin, Springer, 1960. [Имеется перевод: Чжун Кай-лай. Однородные цепи Маркова.— М.: Мир, 1964.]

[5]) Orey S., Sums arising in the theory of Markov chains, Proc. Amer. Math. Soc., 12 (1961), 847—856.

*) См. примечание на с. 423.

была в состоянии E_i при условии, что в настоящем ее состоянием является E_j.

Рассмотрим сперва цепь, имеющую строго положительное инвариантное распределение вероятностей $\{u_k\}$, т. е. предположим, что $u_k>0$ и $\sum u_k=1$, где

$$u_k=\sum_\nu u_\nu p_{\nu k}. \tag{12.1}$$

(Напомним, что по теореме из § 7 инвариантное распределение вероятностей в неприводимой цепи автоматически будет строго положительным.)

Если начальное распределение совпадает с $\{u_k\}$, то в любой момент времени вероятность того, что система находится в состоянии E_i, будет равна u_i. При условии, что это событие в состоянии E_i осуществилось, условная вероятность того, что n единиц времени назад система была в состоянии E_j, равна

$$q_{ij}^{(n)}=u_j p_{ji}^{(n)}/u_i. \tag{12.2}$$

При $n=1$ получаем

$$q_{ij}=u_j p_{ji}/u_i. \tag{12.3}$$

В силу (12.1) ясно, что q_{ij} являются элементами *стохастической* матрицы Q. Более того, вероятности $q_{ij}^{(n)}$ являются просто элементами n-й степени Q^n этой матрицы (иначе говоря, $q_{ij}^{(n)}$ можно вычислить по q_{ij} точно так же, как $p_{ji}^{(n)}$ вычисляются по p_{ji}). Теперь очевидно, что *изучение прошлой эволюции нашей цепи Маркова сводится к изучению цепи Маркова с переходными вероятностями* q_{ij}. Безусловные вероятности для новой цепи совпадают, конечно же, с инвариантным распределением вероятностей $\{u_k\}$. Вероятности q_{ij} называются *обращенными вероятностями* (по отношению к исходной цепи), и процедура, приводящая от одной цепи к другой, называется *обращением времени*. В частном случае, когда $q_{ij}=p_{ij}$, говорят, что цепь *обратима*; вероятностные соотношения для такой цепи симметричны по времени.

Мы знаем, что неприводимая цепь обладает инвариантным распределением вероятностей только тогда, когда ее состояния имеют конечные средние времена возвращения. Если состояния цепи возвратные нулевые, то существует инвариантная мера, единственная с точностью до постоянного множителя. Для невозвратных цепей возможны все варианты: некоторые цепи не имеют инвариантных мер, другие имеют их бесконечно много (см. примеры 11,б) и 11,в)). При этих обстоятельствах заслуживает внимания тот факт, что преобразование (12.3) *определяет стохастическую матрицу Q, если $\{u_k\}$ является строго положительной инвариантной мерой.* Степени матрицы Q даются формулой (12.2). В этом смысле *каждая строго положительная инвариантная мера определяет обращенную цепь Маркова.* К сожалению, новые переходные вероятности q_{ij}

нельзя интерпретировать непосредственно как условные вероятности для исходного процесса [1]).

Из (12.3) ясно, что $\{u_j\}$ является инвариантной мерой и для обращенной цепи. Более того, из (12.2) следует, что ряды $\sum_n q_{ij}^{(n)}$ и $\sum_n p_{ji}^{(n)}$ оба сходятся или оба расходятся. Отсюда следует, что *состояния обеих цепей имеют один тип*: если одна из цепей невозвратна, либо возвратна, то такова же и другая.

Примеры. а) Инвариантное распределение вероятностей, соответствующее *модели Эренфестов* (пример 2, д)), было найдено в (7.16). Простое вычисление показывает, что *модель Эренфестов обратима* в том смысле, что $q_{ij}=p_{ij}$.

б) В примере 11,б) мы нашли инвариантные меры, соответствующие случайному блужданию на прямой, в котором переходы в правую и левую соседние точки имеют вероятности p и q соответственно. Если мы выберем меру с $u_k=1$ при $k=0, \pm 1, \pm 2, \ldots$, то получим $q_{ij}=p_{ji}$ и тем самым придем к случайному блужданию, в котором p и q поменялись ролями. С другой стороны, инвариантная мера с $u_k=(p/q)^k$ приводит к обращенному блужданию, совпадающему с исходным.

в) В примерах 2, л) и 2, м) мы ввели две цепи Маркова, связанные с рекуррентным событием $\mathcal{E}$. В примере 7,ж) мы видели, что для возвратного $\mathcal{E}$ с конечным средним временем возвращения μ эти две цепи имеют одно и то же инвариантное распределение вероятностей, определенное в (7.23). При $\mu=\infty$ эти соотношения определяют общую для обеих цепей инвариантную меру (см. примеры 11,в) и 11,г)). Простое вычисление показывает теперь, что *эти две цепи получаются одна из другой при обращении времени*. Это и не удивительно, поскольку цепь из примера 2,л) связана со временем ожидания следующего осуществления события $\mathcal{E}$, тогда как пример 2,м) относится ко времени, прошедшему с момента последнего осуществления этого события.

Рассмотрим теперь произвольную неприводимую невозвратную цепь с инвариантной мерой $\{u_k\}$. Определяющие инвариантную меру уравнения (12.1) могут допускать и другие решения, и вопрос о единственности для этой системы тесно связан с вопросом о единственности для сопряженной системы линейных уравнений [2])

$$\xi_i=\sum_\nu p_{i\nu}\xi_\nu, \tag{12.4}$$

---

[1]) Для конкретной интерпретации q_{ij} необходимо рассмотреть бесконечно много одновременно протекающих процессов, как было указано в примере 11, б).

[2]) Если через ξ обозначить вектор-столбец с компонентами ξ_i, то система (12.4) будет эквивалентна векторно-матричному уравнению $\xi=P\xi$. Система (12.1) соответствует уравнению $u=uP$, где u — вектор-строка.

сыгравшей важную роль в § 8. Эта система допускает тривиальное решение $\xi_i = c$ при всех i. Каждое неотрицательное решение автоматически будет строго положительным. (В самом деле, из $\xi_i = 0$ вытекало бы, что $\xi_\nu = 0$ при всех ν, таких, что $p_{i\nu} > 0$. Это, в свою очередь, означало бы, что $\xi_\nu = 0$ при $p_{i\nu}^{(2)} > 0$, и вообще $\xi_\nu = 0$ для каждого достижимого из E_i состояния E_ν. Поэтому $\xi_\nu = 0$ для всех ν, ибо цепь неприводима.) Если $\{\xi_i\}$ — непостоянное решение (12.4), то (12.3) показывает, что

$$v_i = u_i \xi_i \tag{12.5}$$

определяет инвариантную меру для обращенной цепи с матрицей Q. И обратно, если обозначить такую меру через $\{v_i\}$, то (12.5) будет определять положительное решение системы (12.4). Иначе говоря, *положительные решения* (12.4) *состоят во взаимно однозначном соответствии с инвариантными мерами для обращенной цепи* [1]) *с матрицей* Q.

В современной теории цепей Маркова и потенциалов положительные решения $\{\xi_i\}$ и $\{u_k\}$ используются для определения границ. Описание того, как это делается, выходит за рамки нашей книги, однако следующие примеры могут дать некоторое понятие о том, что называется *границей-выходом* (exit boundary).

Примеры. а) Рассмотрим случайное блуждание на бесконечной прямой, такое, что из состояния $j \neq 0$ частица с вероятностью p перемещается на единичный шаг в направлении от начала координат и с вероятностью q — на единичный шаг к началу координат. Из нуля частица перемещается с равными вероятностями в $+1$ или в -1. Мы предположим, что $p > q$.

В этой цепи Маркова состояния занумерованы числами от $-\infty$ до $+\infty$ и уравнения (12.4) принимают вид

$$\begin{aligned} &\xi_i = p\xi_{i+1} + q\xi_{i-1}, && i > 0, \\ &\xi_0 = (1/2)\,\xi_1 + (1/2)\,\xi_{-1}, && \\ &\xi_i = q\xi_{i+1} + p\xi_{i-1}, && i < 0. \end{aligned} \tag{12.6}$$

---

[1]) Для неприводимой возвратной цепи инвариантная мера единственна с точностью до постоянного множителя. Поскольку цепи с матрицами P и Q однотипны, то нами доказана следующая теорема.

Теорема. *Для неприводимой возвратной цепи единственным неотрицательным решением системы* (12.4) *является* $\xi_i = \text{const}$.

Это можно доказать также почти дословным повторением последней части доказательства теоремы 1 § 11. В самом деле, по индукции мы находим, что для произвольных i, r и n

$$\xi_i = \left[f_{ir}^{(1)} + \ldots + f_{ir}^{(n)}\right] \xi_r + \sum_\nu {}_r p_{i\nu}^{(n)} \xi_\nu.$$

Для возвратной цепи выражение внутри скобок стремится к 1, тогда как ряд сходится к 0. Следовательно, $\xi_i = \xi_r$, как и утверждалось.

Положим

$$\eta_i=\begin{cases}1-(1/2)(q/p)^i & \text{при } i\geqslant 0,\\ (1/2)(q/p)^{-i} & \text{при } i\leqslant 0.\end{cases}\tag{12.7}$$

Легко видеть, что $\xi_i=\eta_i$ и $\xi_i=1-\eta_i$ определяют два[1]) нетривиальных решения системы (12.6). Отсюда следует, что наша цепь невозвратна, и поэтому положение частицы с необходимостью устремится либо к $+\infty$, либо к $-\infty$.

К этому заключению можно прийти и непосредственно из теории случайных блужданий. В самом деле, из гл. XIV, 2 мы знаем, что если частица выходит из положения $i>0$, то вероятность того, что она когда-либо достигнет нуля, равна $(q/p)^i$. В силу симметрии выходящая из нуля частица имеет равные вероятности сноса в $+\infty$ и в $-\infty$, и поэтому вероятность окончательного сноса в частицы в $-\infty$ равна $(1/2(q/p)^i$. Мы заключаем, что η_i *есть вероятность того, что выходящую из произвольного положения i частицу в конце концов снесет в* $+\infty$. Снос в $-\infty$ имеет вероятность $1-\eta_i$. В современной теории эта ситуация была бы описана посредством введения «точек границы-выхода» $+\infty$ и $-\infty$.

б) Простота предыдущего примера несколько обманчива, и поэтому может быть полезно иметь пример *границы, состоящей из бесконечного числа точек*. Рассмотрим для этой цели следующее случайное блуждание в плоскости x, y. Координата x совершает обычное случайное блуждание, в котором шаги $+1$ и -1 имеют вероятности p и $q<p$. Координата y остается фиксированной, за исключением тех моментов, когда координата x обращается в нуль, и в эти моменты координата y уменьшается на 1. Точнее когда $j\neq 0$, возможны только переходы $(j, k)\to(j+1, k)$ и $(j-1, k)$, и они имеют вероятности p и $q<p$ соответственно. Из положения $(0, k)$ частица с вероятностью p перемещается в $(1, k-1)$ и с вероятностью q — в $(-1, k-1)$.

Из теории случайных блужданий мы знаем, что координата x обязана стремиться к $+\infty$ и что (с вероятностью единица) она лишь конечное число раз пройдет через 0. Отсюда следует, что (исключая событие нулевой вероятности) координата y изменится лишь конечное число раз. Это означает, что после конечного числа изменений координаты y частица навсегда останется на прямой $y=r$. В этом смысле существует бесконечно много «путей бегства в бесконечность», и для каждого начального положения (j, k) мы можем вычислить[2]) вероятность $\xi_{j,k}^{(r)}$ того, что частица в конце концов

---

[1]) Наиболее общее решение системы (12.6) дается формулой $\xi_i=A+B\eta_i$, где A и B — произвольные постоянные. Действительно, эти постоянные могут быть выбраны так, чтобы получить заданные значения ξ_1 и ξ_{-1}, а из (12.6) очевидно, что значения ξ_1 и ξ_{-1} определяют все ξ_i единственным образом.

[2]) Явное выражение для $\xi_{j,k}^{(r)}$ может быть получено из относящихся к одномерным случайным блужданиям результатов гл. XIV,2. Для начального положения $i\leqslant 0$ вероятность того, что частица побывает в нуле ровно $\rho>0$

обосуется на прямой $y=r$. Легко видеть, что для фиксированного r вероятности $\xi_{i,k}^{(r)}$ представляют собой решение системы, соответствующей (12.4), и что наиболее общее решение такой системы является линейной комбинацией этих частных решений. Более того, частное решение $\xi_{i,k}^{(r)}$ характеризуется тем интуитивно очевидным «граничным условием», что $\xi_{i,k}^{(r)} \to 0$ при $j \to \infty$, за исключением случая $k=r$, в котором $\xi_{i,r}^{(r)} \to 1$. ►

Эти примеры типичны в следующем смысле. Для данной неприводимой невозвратной цепи Маркова всегда можно определить «границу», такую, что с вероятностью единица состояние системы стремится к некоторой точке этой границы. Если нам дано множество Γ на границе, то мы можем поинтересоваться, какова вероятность η_i того, что, выходя из начального состояния E_i, система устремится к точке из Γ. Назовем $\{\eta_i\}$ *вероятностями поглощения в* Γ. Оказывается, что такие вероятности поглощения всегда являются решениями линейной системы (12.4), и, наоборот, все ограниченные решения (12.4) представляют собой линейные комбинации вероятностей поглощения. Более того, вероятности $\{\eta_i\}$ поглощения в Γ даются единственным решением системы (12.4), которое имеет граничное значение 1 на Γ и граничное значение 0 на дополнении Γ до границы. Мы теперь можем образовать новую стохастическую матрицу $\hat{P}$ с элементами

$$\hat{p}_{ik} = p_{ik}\eta_k/\eta_i. \tag{12.8}$$

Это условная вероятность перехода из E_i в E_k при условии, что система в конечном счете устремится к одной из точек Γ. Марковский процесс с матрицей $\hat{P}$ может быть описан как условный процесс, получающийся из исходного процесса при условии окончательного поглощения в Γ. Поскольку будущая эволюция никогда не может быть известна заранее, такой переход на первый взгляд кажется бессмысленным. Тем не менее он является мощным аналитическим инструментом и даже имеет реальный смысл для процессов, которые протекают уже очень длительное время.

Граница может быть определена также и для матрицы Q, получаемой при обращении времени. Стало быть, в общем случае существуют две различные границы, соответствующие данной цепи. Они

раз, равна $(2q)^{\rho-1}(p-q)$; при $i \geqslant 0$ эта вероятность равна $(q/p)^i(2q)^{\rho-1}(p-q)$. Вероятность никогда не попасть в нуль равна 0 при $i \leqslant 0$ и $1-(q/p)^i$ при $i \geqslant 0$. Отсюда легко получить, что при $i \leqslant 0$

$$\xi_{i,k}^{(r)} = (2q)^{k-r-1}(p-q), \quad k > r,$$

тогда как при $i > 0$

$$\xi_{i,k}^{(r)} = (q/p)^i(2q)^{k-r-1}(p-q), \quad k > r,$$
$$\xi_{i,r}^{(r)} = 1-(q/p)^i$$

и, конечно же, $\xi_{i,r}^{(k)} = 0$ при $k < r$, называются соответственно *гра*[illegible]*ница-вход* (entrance boundary) и *граница-выход*. Грубо говоря, [illegible] первая из них [illegible] прошлому, а вторая — к [illegible] относится к удаленному [illegible] удаленному будущему.

Обращенные [illegible] системы (12[illegible] цепи впервые рассмотрел А. Н. Колмогоров [1]). Роль решений вход [illegible] была подчеркнута в предыдущих изданиях этой книги. Границу-[illegible] границу-выход ввел В. Феллер [2]). Предложенное им построение удовлетворительно, когда существует лишь конечное число граничных точек, а в общем случае, как указал Дж. Л. Дуб [3]), проще принять построение, введенное в теории гармонических функций Р. С. Мартином. Величины (12.8) были введены В. Феллером [2]); аналогичное преобразование в теории классических гармонических функций было определено примерно в то же время М. Брело [4]).

§ 13. ОБЩИЙ МАРКОВСКИЙ ПРОЦЕСС

В приложениях обычно бывает удобно описывать цепи Маркова в терминах случайных величин. Это может быть сделано с помощью простой замены в предыдущих параграфах символа E_k на целое число k. Состояние системы в момент времени n будет тогда случайной величиной $\mathbf{X}^{(n)}$, принимающей значение k с вероятностью $a_k^{(n)}$; совместное распределение $\mathbf{X}^{(n)}$ и $\mathbf{X}^{(n+1)}$ дается формулой $\mathbf{P}\{\mathbf{X}^{(n)}=j, \mathbf{X}^{(n+1)}=k\}=a_j^{(n)}p_{jk}$, а совместное распределение $(\mathbf{X}^{(0)}, \ldots, \mathbf{X}^{(n)})$ — формулой (1.1). Можно также (иногда это предпочтительнее) приписывать состоянию E_k численное значение e_k, отличное от k. При такой записи цепь Маркова становится специальным стохастическим процессом [5]), или, иначе говоря, последовательностью (зависимых) случайных величин [6]) $(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \ldots)$. Верхний индекс n играет роль времени. В гл. XVII мы получим некоторое представление о более общих стохастических процессах, в которых временной параметр может меняться непрерывно. Термин «марковский процесс» применяется к весьма обширному и важному классу стохастических процессов (как с дискретным, так и с непрерывным временем). Даже в дискретном случае существуют более общие марковские процессы, чем те простые цепи, которые изучались нами до сих пор. Поэтому будет полезно дать определение марковского свойства, указать спе-

---

[1]) Kolmogoroff A., Zur Theorie der Markoffschen Ketten, Mathematische Annalen, 112 (1935), 155—160.

[2]) Feller W., Boundaries induced by positive matrices, Trans. Amer. Math. Soc., 83 (1956), 19—54.

[3]) Doob J. L., Discrete potential theory and boundaries, J. Math. Mechanics, 8 (1959), 433—458.

[4]) Brelot M., Le problème de Dirichlet. Axiomatique et frontière de Martin, J. Math. Pures Appl., 35 (1956), 297—335.

[5]) Термины «стохастический процесс» и «случайный процесс» являются синонимами и охватывают практически всю теорию вероятностей от бросания монеты до гармонического анализа. На практике термин «стохастический процесс» используется главным образом тогда, когда рассматривается случай непрерывного времени.

[6]) Эта формулировка относится к бесконечному произведению пространств, но в действительности мы имеем дело лишь с совместными распределениями конечных наборов случайных величин.

циальное условие, характеризующее наши цепи Маркова, и, наконец, привести несколько примеров немарковских процессов.

По идее марковский процесс является вероятностным аналогом процессов классической механики, где будущая эволюция полностью определяется состоянием в настоящий момент и не зависит от того, каким образом это состояние было достигнуто. Эти процессы существенно отличаются от процессов с последействием (или с наследственностью [1])), таких, которые встречаются, например, в теории пластичности, где вся предыдущая история системы влияет на ее будущее. В стохастических процессах будущее не определяется единственным образом, однако у нас по крайней мере имеются вероятностные соотношения, позволяющие нам делать предсказания. Для изучавшихся в этой главе цепей Маркова ясно, что вероятностные соотношения, относящиеся к будущему, зависят лишь от состояния в настоящий момент, но не от того, каким образом это состояние возникло в прошлом. Иначе говоря, если две независимые системы с одинаковыми переходными вероятностями оказываются в одном и том же состоянии, то все вероятности, относящиеся к их будущей эволюции, совпадают. Это довольно неопределенное описание формализуется следующим определением.

Определение. *Последовательность дискретных случайных величин является марковским процессом, если для каждого конечного набора целых чисел* $n_1<n_2<\ldots<n_r<n$ *соответствующее совместное распределение* $(\mathbf{X}^{(n_1)}, \mathbf{X}^{(n_2)}, \ldots, \mathbf{X}^{(n_r)}, \mathbf{X}^{(n)})$ *определено таким образом, что условная вероятность соотношения* $\mathbf{X}^{(n)}=x$ *при условии* $\mathbf{X}^{(n_1)}=x_1, \ldots, \mathbf{X}^{(n_r)}=x_r$ *совпадает с условной вероятностью события* $\mathbf{X}^{(n)}=x$ *при единственном условии* $\mathbf{X}^{(n_r)}=x_r$. *Здесь* $x_1, \ldots, x_r$ *суть произвольные числа, для которых упомянутые условия имеют положительные вероятности.*

Проще говоря, в определении сказано, что если дано настоящее состояние x_r, то никакие дополнительные сведения относительно состояний системы в прошлом не могут изменить (условную) вероятность состояния x в некоторый будущий момент времени.

Изучавшиеся до сих пор в этой главе цепи Маркова являются, очевидно, марковскими процессами, однако они обладают тем дополнительным свойством, что для них *переходные вероятности* $p_{jk}=\mathbf{P}\{\mathbf{X}^{(m+1)}=k\,|\,\mathbf{X}^{(m)}=j\}$ *не зависят от* m. Более общие переходные вероятности

$$p_{jk}^{(n-m)}=\mathbf{P}\{\mathbf{X}^{(n)}=k\,|\,\mathbf{X}^{(m)}=j\}, \quad m<n, \tag{13.1}$$

зависят только от разности $n-m$. Такие переходные вероятности называются *стационарными (или однородными по времени)*. Для общей целочисленной цепи Маркова правая часть (13.1) зависит от m и n. Мы обозначим ее через $p_{jk}(m, n)$, так что $p_{jk}(n, n+1)$ опреде-

[1]) В оригинале hereditary processes.— *Прим. перев.*

ляют вероятности перехода за один шаг. Для вероятности пути $(j_0, j_1, \ldots, j_n)$ мы получаем теперь вместо (1.1) выражение

$$a_{j_0}^{(0)} p_{j_0 j_1}(0, 1)\, p_{j_1 j_2}(1, 2) \ldots p_{j_{n-1} j_n}(n-1, n). \qquad (13.2)$$

Соответствующим обобщением (3.3) будет, очевидно, тождество

$$p_{jk}(m, n) = \sum_{\nu} p_{j\nu}(m, r)\, p_{\nu k}(r, n), \qquad (13.3)$$

справедливое для всех r, таких, что $m<r<n$. Это тождество следует непосредственно из определения марковского процесса, а также из (13.2); оно называется *уравнением Колмогорова — Чэпмена*. (Переходные вероятности $p_{jk}(m, n)$ определены и для немарковских дискретных процессов, однако для них множитель $p_{\nu k}(r, n)$ в (13.3) должен быть заменен выражением, зависящим не только от ν и k, но и от j.)

Изучавшиеся в этой главе цепи Маркова представляют общий однородный по времени дискретный марковский процесс. Мы не будем останавливаться на неоднородных марковских процессах. Следующие примеры могут быть полезными для понимания марковского свойства и проиллюстрируют ситуации, в которых уравнение Колмогорова — Чэпмена (13.3) не имеет места.

Примеры немарковских процессов.

а) *Урновая схема Пойа* (пример гл. V,2, в)). Пусть $\mathbf{X}^{(n)}$ равно 1 или 0 в зависимости от того, какой шар будет вынут при n-м извлечении — черный или красный. Последовательность $\{\mathbf{X}^{(n)}\}$ *не является* марковским процессом. Например,

$$\mathbf{P}\{\mathbf{X}^{(3)}=1 \mid \mathbf{X}^{(2)}=1\} = (b+c)/(b+r+c),$$

но

$$\mathbf{P}\{\mathbf{X}^{(3)}=1 \mid \mathbf{X}^{(2)}=1,\ \mathbf{X}^{(1)}=1\} = (b+2c)/(b+r+2c)$$

(см. задачи 19 и 20 гл. V). С другой стороны, если $\mathbf{Y}^{(n)}$ — число черных шаров в урне в момент времени n, то $\{\mathbf{Y}^{(n)}\}$ является обычной цепью Маркова с постоянными переходными вероятностями.

б) *Суммы высших порядков*. Пусть $\mathbf{Y}_0, \mathbf{Y}_1, \ldots$ — взаимно независимые случайные величины, и пусть $\mathbf{S}_n = \mathbf{Y}_0 + \ldots + \mathbf{Y}_n$. Разность $\mathbf{S}_n - \mathbf{S}_m$ (при $m<n$) зависит лишь от $\mathbf{Y}_{m+1}, \ldots, \mathbf{Y}_n$, и поэтому легко видеть, что последовательность $\{\mathbf{S}_n\}$ является марковским процессом. Определим теперь новую последовательность случайных величин $\mathbf{U}_n$, полагая

$$\mathbf{U}_n = \mathbf{S}_0 + \mathbf{S}_1 + \ldots + \mathbf{S}_n = \mathbf{Y}_n + 2\mathbf{Y}_{n-1} + 3\mathbf{Y}_{n-2} + \ldots + (n-1)\mathbf{Y}_0.$$

Последовательность $\{\mathbf{U}_n\}$ образует стохастический процесс, вероятностные соотношения для которого могут быть в принципе выражены через распределения величин $\mathbf{Y}_k$. Но этот процесс в общем случае уже не является марковским, поскольку нет никакой причины, по которой, например, $\mathbf{P}\{\mathbf{U}_n=0 \mid \mathbf{U}_{n-1}=a\}$ должна была бы равняться $\mathbf{P}\{\mathbf{U}_n=0 \mid \mathbf{U}_{n-1}=a,\ \mathbf{U}_{n-2}=b\}$; зная $\mathbf{U}_{n-1}$ и $\mathbf{U}_{n-2}$, можно сделать лучшие предсказания, чем зная одно лишь $\mathbf{U}_{n-1}$.

В случае непрерывного временно́го параметра суммирование здесь заменяется интегрированием. В теории диффузии $\mathbf{Y}_n$ играют роль ускорений; тогда $\mathbf{S}_n$ — скорости, а $\mathbf{U}_n$ — положения. Если могут быть измерены только положения, то мы вынуждены будем изучать немарковский процесс, хотя он и определяется косвенным образом через марковский процесс.

в) *Скользящие средние.* Пусть снова $\{\mathbf{Y}_n\}$ — последовательность взаимно независимых случайных величин. Скользящие средние порядка r определяются как $\mathbf{X}^{(n)}=(\mathbf{Y}_n+\mathbf{Y}_{n+1}+\ldots+\mathbf{Y}_{n+r-1})/r$. Легко видеть, что $\{\mathbf{X}^{(n)}\}$ не является марковским процессом. Процессы этого типа встречаются во многих приложениях (см. задачу 25).

г) *Задача об уличном движении.* Чтобы получить эмпирический пример немарковского процесса, Р. Фюрт [1]) провел наблюдения над числом пешеходов на определенном участке улицы. Идеализированная математическая модель этого процесса может быть получена следующим образом. Для простоты мы предположим, что все пешеходы имеют одну и ту же скорость v, и рассмотрим только пешеходов, движущихся в одном направлении. Разобьем ось x на отрезки $I_1, I_2, \ldots$ фиксированной длины d и будем регулярно наблюдать за расположением пешеходов в моменты, отстоящие друг от друга на d/v единиц времени. Определим случайную величину $\mathbf{Y}_k$ как число пешеходов, находившихся первоначально в I_k. При n-м наблюдении эти пешеходы будут обнаружены в I_{k-n}, тогда как интервал I_k будет содержать $\mathbf{Y}_{k+n}$ пешеходов. Стало быть, полное число пешеходов в интервале $0<x<Nd$ будет тогда равно $\mathbf{X}^{(n)}=\mathbf{Y}_{n+1}+\ldots+\mathbf{X}_{n+N}$, и поэтому наш процесс является по существу процессом скользящего среднего. Простейшая модель для случайных величин $\mathbf{Y}_k$ представляется испытаниями Бернулли. В пределе при $d\to 0$ они приводят к непрерывной модели, в которой биномиальное распределение заменяется распределением Пуассона.

д) *Суперпозиция марковских процессов (сложное тасование).* Существует много технических устройств (таких, как группы селекторов в телефонных коммутаторах, счетчики, фильтры), действие которых может быть описано как суперпозиция двух марковских процессов, результат которой уже не является марковским. Некоторое представление о таких механизмах можно получить, изучая следующий метод тасования карт.

Пусть помимо основной колоды из N карт у нас есть такая же вспомогательная колода, и эта вспомогательная колода тасуется обычным образом. Если карты в ней расположатся в порядке $(a_1, a_2, \ldots, a_N)$, то мы переставим карты в основной колоде так, чтобы

---

[1]) Fürth R., Schwankungserscheinungen in der Physik, Sammlung Vieweg, Braunschweig, 1920, 17ff. Результаты наблюдений появились в Physikalische Zeitschrift, 19 (1918), 20 (1919).

первая, вторая, . . ., N-я карта переместились на места с номерами a_1, a_2, . . ., a_N. Таким образом, тасование вспомогательной колоды косвенно определяет последовательные порядки карт в основной колоде. Эти последние образуют *стохастический процесс немарковского типа*. Для доказательства этого достаточно показать, что два последовательных порядка карт в основной колоде дают в общем случае больше информации о будущем, чем один последний порядок. Мы покажем это для простого частного случая.

Пусть $N=4$; предположим, что вспомогательная колода находится вначале в порядке (2431). Предположим, кроме того, что операция тасования всегда представляет собой только «снимание», т. е. порядок (a_1, a_2, a_3, a_4) меняется на один из трех следующих: (a_2, a_3, a_4, a_1), (a_3, a_4, a_1, a_2), (a_4, a_1, a_2, a_3); каждой из этих трех возможностей мы припишем вероятность 1/3. При этих соглашениях вспомогательная колода в любой момент времени будет иметь один из четырех порядков (2431), (4312), (3124), (1243). С другой стороны, уже непродолжительное экспериментирование покажет, что основная колода постепенно пройдет через все 24 возможных порядка и что каждый из них будет появляться в комбинации с каждым из четырех возможных порядков вспомогательной колоды. Это означает, что порядок (1234) основной колоды появится бесконечное число раз и что за ним всегда будет следовать один из четырех порядков (4132), (3421), (2314), (1243). Но вспомогательная колода не может оставаться в одном и том же порядке, и, следовательно, основная колода не может дважды подряд подвергнуться одной и той же перестановке. Стало быть, если при испытаниях с номерами $n-1$ и n имели место соответственно порядки карт (1234) и (1243), то в следующем испытании порядок (1234) будет невозможен. Таким образом, два последовательных наблюдения дают больше информации, чем одно-единственное.

е) *Немарковский процесс, удовлетворяющий уравнению Колмогорова—Чэпмена.* Тождество (3.3) было выведено из того предположения, что переход из E_ν в E_k не зависит от того способа, каким было достигнуто состояние E_ν. Поэтому сначала представлялось интуитивно ясным, что ни один немарковский процесс не должен удовлетворять этому тождеству; это предположение, казалось, подтверждалось тем фактом, что вероятности перехода за n шагов для такого процесса должны удовлетворять целому множеству любопытных тождеств. Тем не менее оказалось, что существуют исключения (хотя бы в теории). Действительно, в гл. IX,1 мы встретились с бесконечной последовательностью попарно независимых одинаково распределенных случайных величин, принимающих значения 1, 2 и 3 с вероятностью 1/3 каждое. Таким образом, мы имели процесс с возможными состояниями 1, 2, 3, такой, что $p_{jk}=1/3$ для всех комбинаций j и k. Стало быть, тождество (3.3) выполнялось тривиальным образом при $p_{jk}^{(n)}=1/3$. Тем не менее этот процесс не является марковским. Чтобы убедиться в этом, предположим, что первый шаг приводит систему в состояние 2. Тогда переход в 3 на следующем шаге будет возможен тогда и только тогда, когда начальным состоянием было 1. Таким образом, переходы, следующие за первым шагом, зависят не только от настоящего, но и от начального состояния. (Модификации различного вида можно найти в замечании в гл. IX,1 и в примечании к нему.)

§ 14. ЗАДАЧИ

1. Рассмотрим последовательность испытаний Бернулли и скажем, что в момент времени n наблюдается состояние E_1, если испытания с номерами $n-1$ и n привели к результату $УУ$. Аналогично E_2, E_3, E_4 означают исходы $УН$, $НУ$, $НН$. Найти матрицу P и все ее степени. Обобщить схему.

2. Классифицировать состояния четырех цепей, матрицы P которых имеют приведенные ниже строки. В каждом случае найти $P^{(2)}$ и асимптотическое поведение $p_{jk}^{(n)}$:

а) (0, 1/2, 1/2), (1/2, 0, 1/2), (1/2, 1/2, 0);

б) (0, 0, 0, 1), (0, 0, 0, 1), (1/2, 1/2, 0, 0), (0, 0, 1, 0);

в) (1/2, 0, 1/2, 0, 0), (1/4, 1/2, 1/4, 0, 0), (1/2, 0, 1/2, 0, 0), (0, 0, 0, 1/2, 1/2), (0, 0, 0, 1/2, 1/2);

г) (0, 1/2, 1/2, 0, 0, 0), (0, 0, 0, 1/3, 1/3, 1/3), (0, 0, 0, 1/3, 1/3, 1/3), (1, 0, 0, 0, 0, 0), (1, 0, 0, 0, 0, 0), (1, 0, 0, 0, 0, 0).

3. Рассмотрим бросания правильной кости и условимся говорить, что в момент n система находится в состоянии E_j, если j — наибольшее из чисел, выпавших в первых n бросаниях. Найти матрицу P^n и убедиться в том, что (3.3) выполняется.

4. В примере 2,к) найти вероятности (поглощения) x_k и y_k того, что, выходя из E_k, система закончит эволюцию соответственно в E_1 или в E_5 ($k=2$, 3, 4, 6). (Решить эту задачу, исходя из основных определений и не обращаясь к § 8.)

5. Рассмотреть пример гл. 1, 5,б) как цепь Маркова. Вычислить вероятность выигрыша для каждого игрока.

6. Пусть E_0 — поглощающее состояние (т. е. $p_{00}=1$). Пусть $p_{jj}=p$ и $p_{j,j-1}=q$ при $j>0$, где $p+q=1$. Найти вероятность $f_{j0}^{(n)}$ того, что поглощение в E_0 произойдет в точности на n-м шаге. Найти также математическое ожидание этого распределения.

7. Первая строка матрицы P равна v_0, v_1, ... При $j>0$ имеем (как и в предыдущей задаче) $p_{jj}=p$ и $p_{j,j-1}=q$. Найти распределение времени возвращения в E_0.

8. Пусть $p_{j,j+2}=v_j$ и $p_{j0}=1-v_j$ при $j=0$, 1, ... Выяснить характер состояний.

9. *Два отражающих экрана.* Цепь с набором состояний 1, 2, ..., ρ имеет матрицу, первая и последняя строки которой суть (q, p, 0, ..., 0) и (0, ..., 0, q, p). Во всех остальных строках $p_{k,k+1}=p$, $p_{k,k-1}=q$. Найти стационарное распределение. Может ли эта цепь быть периодической?

10. Обобщить *модель Бернулли — Лапласа для диффузии* (пример 2,е)), предполагая, что имеются $b\geqslant\rho$ черных и $w=2\rho-b$ белых частиц. Число частиц в каждой урне остается равным ρ.

11. Цепь с набором состояний E_0, E_1, ... имеет переходные вероятности

$$p_{jk}=e^{-\lambda}\sum_{\nu=0}^{j}\binom{j}{\nu}p^{\nu}q^{j-\nu}\frac{\lambda^{k-\nu}}{(k-\nu)!},$$

где слагаемые с $\nu>k$ следует считать равными нулю. Показать, что

$$p_{jk}^{(n)}\to e^{-\lambda/q}(\lambda/q)^k/k!.$$

Замечание. Эта цепь встречается в статистической механике[1]) и может быть интерпретирована следующим образом. Состояние системы определяется числом

[1]) Chandrasekhar S. Stochastic problems in physics and astronomy, Rew. of Modern Physics, 15(1943), 1—89; p. 45. [Имеется перевод: Чандрасекар С. Стохастические процессы в физике и астрономии.— М.: ИЛ, 1947, с. 1—126, в частности с. 82.]

частиц в некоторой области пространства. В течение каждого единичного интервала времени каждая частица может с вероятностью q покинуть эту область, причем все частицы стохастически независимы. Кроме того, в эту область могут попадать новые частицы, и вероятность появления [за единицу времени.— *Перев.*] r новых частиц дается выражением Пуассона $e^{-\lambda}\lambda^r/r!$. Тогда стационарным распределением будет распределение Пуассона с параметром λ/q.

12. *Модель Эренфестов.* Пусть в примере 2,д) в первом сосуде сначала было j молекул, и пусть $\mathbf{X}^{(n)}=2k-\rho$, если на n-м шаге система находилась в состоянии k (так что $\mathbf{X}^{(n)}$ есть разность числа молекул в двух наших сосудах). Пусть $e_n=\mathbf{E}(\mathbf{X}^{(n)})$. Доказать, что $e_{n+1}=(\rho-2)e_n/\rho$, откуда $e_n=(1-2/\rho)^n(2j-\rho)$. (Заметим, что $e_n\to 0$ при $n\to\infty$.)

13. Рассмотреть модель из задачи о счетчике (пример гл. XIII, 1, ж)) как цепь Маркова.

14. *Случайное блуждание на плоскости с отражающими экранами.* Рассмотрим симметричное случайное блуждание в ограниченной области плоскости. Ее граница является отражающей в том смысле, что каждый раз, когда при неограниченном случайном блуждании частица должна была бы покинуть эту область, она бывает вынуждена вернуться в предыдущее положение. Показать, что если каждая точка области достижима из любой другой точки, то существует стационарное распределение, и что $u_k=1/a$, где a — число положений в области. (Если область не ограничена, то состояния являются возвратными нулевыми и $u_k=1$ определяет инвариантную меру.)

15. *Повторное осреднение.* Пусть $\{x_1, x_2, \ldots\}$ — ограниченная последовательность чисел, и пусть P — матрица эргодической цепи. Доказать, что $\sum\limits_j p_{ij}^{(n)}x_j\to\sum u_jx_j$. Показать, что процедура повторного осреднения из примера гл. XIII, 10,в) является частным случаем этой задачи.

16. В теории *массового обслуживания* встречается матрица переходных вероятностей

$$\begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \end{bmatrix},$$

где $\{p_k\}$ — некоторое распределение вероятностей. Используя производящие функции, выяснить характер состояний. Найти производящую функцию стационарного распределения, если оно существует.

17. *Время ожидания поглощения.* Пусть $\mathbf{Y}_j$ — момент времени, когда выходящая из невозвратного состояния E_j система *впервые* попадает в возвратное состояние. Предполагая, что вероятность остаться навсегда в множестве невозвратных состояний равна нулю, доказать, что $d_j=\mathbf{E}(\mathbf{Y}_j)$ однозначно определяется как решение системы линейных уравнений

$$d_j=\sum_T p_{j\nu}d_\nu+1,$$

где суммирование проводится по всем ν, таким, что E_ν невозвратно. Однако d_ν не обязательно конечны.

18. Если число состояний $a<\infty$ и если E_k достижимо из E_j, то оно достижимо не более чем за $a-1$ шагов ($j\neq k$).

19. Пусть цепь содержит a состояний, и пусть E_j — возвратное состояние. Существует число $q<1$, такое, что при $n\geqslant a$ вероятность того, что время возвращения в E_j превосходит n, меньше чем q^n. (*Указание.* Использовать результат задачи 18.)

20. В конечной цепи состояние E_j невозвратно тогда и только тогда, когда существует E_k, такое, что E_k достижимо из E_j, а E_j недостижимо из E_k. (Для бесконечных цепей это неверно, как показывают случайные блуждания.)

21. Неприводимая цепь, для которой положителен хотя бы *один* диагональный элемент p_{jj}, не может быть периодической.

22. Конечная неприводимая цепь непериодична тогда и только тогда, когда существует n, такое, что $p_{jk}^{(n)} > 0$ при всех j и k.

23. Пусть a состояний $(x_1, \ldots, x_a)$, некоторой цепи дают решение системы линейных уравнений $x_j = \sum p_{j\nu} x_\nu$. Доказать, что а) если $x_j \leqslant 1$ при всех j, то состояния, для которых $x_r = 1$, образуют замкнутое множество; б) если E_j и E_k принадлежат одному неприводимому множеству, то $x_j = x_k$; в) в неприводимой цепи решение $\{x_j\}$ сводится к константе. *Указание.* Рассмотреть сужение этих уравнений на замкнутом множестве.

24. *Продолжение.* Если $(x_1, \ldots, x_a)$ есть (комплекснозначное) решение системы $x_j = s \sum p_{j\nu} x_\nu$ при $|s| = 1$ и $s \neq 1$, то существует целое число $t > 1$, такое, что $s^t = 1$. Если цепь неприводима, то наименьшее такое целое число является периодом цепи.

Указание. Предположить без потери общности, что $x_1 = 1 \geqslant |x_\nu|$. Рассмотреть последовательно состояния, которые достигаются за 1, 2, ... шагов.

25. *Скользящие средние.* Пусть $\{Y_k\}$ — последовательность взаимно независимых случайных величин, каждая из которых принимает значения ± 1 с вероятностями 1/2, и пусть $X^{(n)} = (Y_n + Y_{n+1})/2$. Найти переходные вероятности

$$p_{jk}(m, n) = P\{X^{(n)} = k \mid X^{(m)} = j\},$$

где $m < n$ и $j, k = -1, 0, 1$. Вывести отсюда, что $\{X^{(n)}\}$ не является марковским процессом и что (13.3) не выполняется.

26. Рассмотрим последовательность испытаний Бернулли и скажем, что в момент времени n наблюдается состояние E_1, если испытания с номерами $n-1$ и n привели к успеху; в противном случае система находится в состоянии E_2. Найти вероятность перехода за n шагов и установить немарковский характер этого процесса.

Замечание. Этот процесс получается из цепи задачи 1 при объединении трех состояний в одно. Такая *группировка* применима к любой цепи Маркова и нарушает марковское свойство. Процессы такого типа изучал Харрис[1]).

27. *Смесь цепей Маркова.* Пусть заданы две цепи Маркова с одним и тем же числом состояний и матрицами P_1 и P_2. Новый процесс определяется некоторым начальным распределением и матрицей вероятностей перехода за n шагов $(1/2)P_1^n + (1/2)P_2^n$. Установить немарковский характер этого процесса и его связь с урновыми моделями гл. V, 2.

28. Пусть N — случайная величина, имеющая распределение Пуассона со средним λ. Рассмотрим N независимых марковских процессов, начинающихся из E_0 и имеющих одну и ту же матрицу P. Обозначим через $Z_k^{(n)}$ число этих процессов, находящихся после n шагов в состоянии E_k. Показать, что $Z_k^{(n)}$ имеет распределение Пуассона со средним $\lambda p_{0k}^{(n)}$.

Указание. Использовать результат примера гл. XII, 1, б).

29. Используя результат предыдущей задачи, показать, что случайная величина $X_k^{(n)}$ из примера 11, б) имеет распределение Пуассона со средним $\sum_j u_j p_{jk}^{(n)} = u_k$.

[1]) Harris T. E., On chains of infinite order, Pacific Journal of Mathematics, 5 (1955), Supplement 1, 707—724.

ГЛАВА XVI*)

АЛГЕБРАИЧЕСКАЯ ТРАКТОВКА КОНЕЧНЫХ ЦЕПЕЙ МАРКОВА

В этой главе мы рассматриваем цепь Маркова с конечным числом состояний $E_1, \ldots, E_\rho$ и заданной матрицей переходных вероятностей p_{jk}. Наша главная цель состоит в выводе явных формул для вероятностей $p_{jk}^{(n)}$ перехода за n шагов. Нам не потребуются результаты предыдущей главы, за исключением общих понятий и обозначений из § 3.

Мы воспользуемся методом производящих функций и получим желаемые результаты при помощи разложения на простые дроби, описанного в гл. XI, 4. Наши результаты можно также получить непосредственно из теории приведения матриц к каноническому виду (которая, в свою очередь, может быть выведена из наших результатов). Кроме того, для *конечных* цепей из результатов настоящей главы следуют эргодические свойства, доказанные в гл. XV. Однако для простоты мы несколько ограничим общность и будем игнорировать исключительные случаи, которые усложняют общую теорию и едва ли встречаются в практических примерах.

Общий метод намечен в основных чертах в § 1 и иллюстрируется в § 2 и 3. В § 4 особое внимание уделяется невозвратным состояниям и вероятностям поглощения. В § 5 наша теория применяется для нахождения дисперсий времен возвращения в состояния E_j.

§ 1. ОБЩАЯ ТЕОРИЯ

Введем для фиксированных j и k производящую функцию[1])

$$P_{jk}(s) = \sum_{n=0}^{\infty} p_{jk}^{(n)} s^n. \tag{1.1}$$

Умножая на sp_{ij} и суммируя по $j = 1, \ldots, \rho$, получаем

$$s \sum_{j=1}^{\rho} p_{ij} P_{jk}(s) = P_{ik}(s) - p_{ik}^{(0)}. \tag{1.2}$$

*) Эта глава посвящена специальному вопросу и может быть опущена. [Подробное изложение близких вопросов можно найти в книге Романовский В. И. Дискретные цепи Маркова.— М.— Л.: Гостехиздат, 1949.— *Перев.*]

[1]) Напомним, что $p_{jk}^{(0)}$ равняется 0 при $j \neq k$ и 1 при $j = k$. (Такие $p_{jk}^{(0)}$ известны под названием символов Кронекера.)

Это означает, что при фиксированных k и s величины $z_j = P_{jk}(s)$ удовлетворяют системе линейных уравнений вида

$$z_i - s \sum_{j=1}^{\rho} p_{ij} z_j = b_i. \tag{1.3}$$

Решения z_j системы (1.3), очевидно, являются рациональными функциями от s с общим знаменателем $D(s)$ — детерминантом этой системы. Для согласования со стандартными обозначениями из линейной алгебры положим $s = t^{-1}$. Тогда $t^{\rho} D(t^{-1})$ есть многочлен степени ρ (называемый характеристическим многочленом матрицы P переходных вероятностей p_{jk}). Его корни $t_1, \ldots, t_{\rho}$ называются *характеристическими числами* (или собственными значениями) матрицы P.

Введем теперь *упрощающие предположения о том, что характеристические числа просты* (отличны друг от друга) *и отличны от нуля*[1]. Это представляет некоторое ограничение общности, однако наша теория будет все же охватывать большинство представляющих интерес случаев.

Как уже говорилось, для фиксированных k и ρ величины $P_{jk}(s)$ суть рациональные функции от s с общим знаменателем $D(s)$. Корни многочлена $D(s)$ равны обратным величинам не обращающихся в нуль характеристических чисел t_{ν}. Поэтому из результатов гл. XI,4 следует, что существуют постоянные $b_{jk}^{(\nu)}$, такие, что[2]

$$P_{jk}(s) = b_{jk}^{(1)}/(1 - st_1) + \ldots + b_{jk}^{(\rho)}/(1 - st_{\rho}). \tag{1.4}$$

Разлагая дроби в геометрические ряды, мы получаем эквивалентные соотношения

$$p_{jk}^{(n)} = b_{jk}^{(1)} t_1^n + \ldots + b_{jk}^{(\rho)} t_{\rho}^n, \tag{1.5}$$

справедливые для всех целых $n \geqslant 0$. Покажем теперь, что коэффициенты $b_{jk}^{(\nu)}$ однозначно определяются как решения некоторых систем линейных уравнений. Величину $p_{ik}^{(n+1)}$ можно получить, заменяя в (1.5)[3] n на $n+1$, но ее можно получить также, умножая (1.5) на p_{ij} и суммируя по $j = 1, \ldots, \rho$. Приравнивая эти два выражения, мы получаем тождество вида

$$C_1 t_1^n + \ldots + C_{\rho} t_{\rho}^n = 0, \tag{1.6}$$

---

[1]) Условие $t_r \neq 0$ вскоре будет отброшено. В примере 4, б) проводится численный анализ цепи с кратными корнями.

[2]) Теоретически мы должны были бы опустить те корни t_r, которые совпадают с корнями числителя. Однако для таких корней мы положим $b_{jk}^{(\nu)} = 0$, и поэтому (1.4) и (1.5) остаются верными при любых обстоятельствах.

[3]) При $j = i$. — *Прим. перев.*

справедливое для всех n. Это, очевидно, невозможно, если все коэффициенты не обращаются в нуль, и мы заключаем, что

$$\sum_{j=1}^{\rho} p_{ij} b_{jk}^{(\nu)} = t_\nu b_{ik}^{(\nu)} \tag{1.7}$$

для всех комбинаций i, k и ν. Умножая (1.5) на p_{kr} и суммируя по k, аналогичным образом получаем

$$\sum_{k=1}^{\rho} b_{jk}^{(\nu)} p_{kr} = t_\nu b_{jr}^{(\nu)}. \tag{1.8}$$

Рассмотрим $(\rho \times \rho)$-матрицу $b^{(\nu)}$ с элементами $b_{ik}^{(\nu)}$. В силу соотношений[1]) (1.7) ее k-й столбец представляет решение системы ρ линейных уравнений

$$\sum_{j=1}^{\rho} p_{ij} x_j - t x_i = 0 \tag{1.9}$$

при $t = t_\nu$; аналогично в силу (1.8) ее j-я строка удовлетворяет системе

$$\sum_{k=1}^{\rho} y_k p_{kr} - t y_r = 0 \tag{1.10}$$

при $t = t_\nu$. Система (1.10) получается из (1.9), когда строки и столбцы меняются местами, и поэтому детерминанты этих систем одинаковы. Детерминант системы (1.9) обращается в нуль только тогда, когда t совпадает с одним из характеристических значений $t_1, \ldots, t_\rho$. Иначе говоря, системы (1.9) и (1.10) допускают нетривиальные решения тогда и только тогда, когда $t = t_\nu$ для некоторого ν. Обозначим пару соответствующих решений через $(x_1^{(\nu)}, \ldots, x_\rho^{(\nu)})$ и $(y_1^{(\nu)}, \ldots, y_\rho^{(\nu)})$. Они определены с точностью до постоянного множителя, так что

$$b_{jk}^{(\nu)} = c^{(\nu)} x_j^{(\nu)} y_k^{(\nu)}, \tag{1.11}$$

где $c^{(\nu)}$ — постоянная (не зависящая от j и k). Чтобы найти эту неизвестную постоянную, заметим, что из (1.9) по индукции вытекает, что

$$\sum_{j=1}^{\rho} p_{ij}^{(n)} x_j = t^n x_i \tag{1.12}$$

при всех n. Воспользуемся этим соотношением при $t = t_\lambda$, где λ — произвольное целое число, лежащее между 1 и ρ. Подставляя

---

1) Системы (1.7) и (1.8) можно записать в компактной векторно-матричной форме $Pb^{(\nu)} = t_\nu b^{(\nu)}$ и $b^{(\nu)} P = t_\nu b^{(\nu)}$.

вместо $p_{ij}^{(n)}$ выражения (1.5), находим, что

$$t_\lambda^n x_i = t_1^n c^{(1)} x_i^{(1)} \sum_{k=1}^{\rho} y_k^{(1)} x_k^{(\lambda)} + \ldots + t_\rho^n c^{(\rho)} x_i^{(\rho)} \sum_{k=1}^{\rho} y_k^{(\rho)} x_k^{(\lambda)}. \qquad (1.13)$$

Это тождество вида (1.6), которое может выполняться лишь тогда, когда все коэффициенты обращаются в нуль. Приравнивая коэффициенты при t_λ^n в обеих частях, получаем окончательно[1])

$$c^{(\lambda)} \sum_{k=1}^{\rho} y_k^{(\lambda)} x_k^{(\lambda)} = 1. \qquad (1.14)$$

Это соотношение определяет коэффициент $b_{jk}^{(\lambda)}$ в (1.11). Правда, $x_j^{(\lambda)}$ и $y_k^{(\lambda)}$ определяются только с точностью до постоянного множителя, но замена $x_j^{(\lambda)}$ на $Ax_j^{(\lambda)}$ и $y_k^{(\lambda)}$ на $By_k^{(\lambda)}$ меняет $c^{(\lambda)}$ на $c^{(\lambda)}/(AB)$, и коэффициент $b_{jk}^{(\lambda)}$ остается неизменным.

Резюмируем этот результат следующим образом. Системы линейных уравнений (1.9) и (1.10) допускают нетривиальные решения не более чем для ρ различных значений t (одних и тех же для обеих систем). Предположим, что существует ровно ρ таких значений $t_1, \ldots, t_\rho$ и что все они отличны от 0. *Выберем для каждого t_λ ненулевое решение $(x_1^{(\lambda)}, \ldots, x_\rho^{(\lambda)})$ системы* (1.9) *и ненулевое решение $(y_1^{(\lambda)}, \ldots, y_\rho^{(\lambda)})$ системы* (1.10). *Тогда при $c^{(\lambda)}$, определяемых в* (1.14), *имеем для* $n = 0, 1, \ldots$

$$p_{jk}^{(n)} = \sum_{\lambda=1}^{\rho} c^{(\lambda)} x_j^{(\lambda)} y_k^{(\lambda)} t_\lambda^n. \qquad (1.15)$$

Таким образом, мы нашли явное выражение для всех переходных вероятностей[2]).

Предположение о том, что все характеристические числа различны, удовлетворяется в большинстве практических случаев, за исключением случая разложимых цепей, который требует лишь небольших изменений в рассуждениях (см. § 4). Однако нередко одним из характеристических чисел оказывается нуль. Положим в этом случае $t_\rho = 0$. Новизна состоит здесь в том, что детерминант $D(s)$

[1]) Обращение в нуль остальных коэффициентов влечет равенство $\sum_{k=1}^{\rho} y_k^{(\lambda)} x_k^{(\nu)} = 0$ при $\lambda \neq \nu$.

[2]) В векторно-матричной форме окончательная формула (1.15) становится более изящной. Пусть $X^{(\lambda)}$ — вектор-*столбец* (или $(\rho \times 1)$-матрица) с элементами $x_j^{(\lambda)}$, а $Y^{(\lambda)}$ — вектор-*строка* (или $(1 \times \rho)$-матрица) с элементами $y_k^{(\lambda)}$. Тогда (1.15) принимает вид

$$P^n = \sum_{\lambda=1}^{\rho} c^{(\lambda)} X^{(\lambda)} Y^{(\lambda)} t_\lambda^n,$$

а $c^{(\lambda)}$ определяется скалярным уравнением $c^{(\lambda)} Y^{(\lambda)} X^{(\lambda)} = 1$.

системы (1.3) имеет только $\rho-1$ корней $t_1^{-1}, \ldots, t_{\rho-1}^{-1}$, и поэтому производящая функция $P_{jk}(s)$ является отношением двух многочленов степени $\rho-1$. Для разложения на простейшие дроби требуется, чтобы степень числителя была меньше степени знаменателя, и, чтобы добиться этого, мы должны сперва вычесть из $P_{jk}(s)$ подходящую постоянную. Таким образом мы получим для $P_{jk}(s)$ разложение на простые дроби, отличающееся от (1.4) тем, что последний член в нем заменяется постоянной. Из (1.15) ясно, что такая замена влияет на правую часть этого равенства лишь при $n=0$. Иначе говоря, *явное представление* (1.15) *для* $p_{jk}^{(n)}$ *остается справедливым при* $n \geqslant 1$, *даже если* $t_\rho = 0$ (при условии, что корни $t_1, \ldots, t_{\rho-1}$ различны и отличны от нуля).

Левая часть (1.15) может оставаться ограниченной при всех n только тогда, когда $|t_\lambda| \leqslant 1$ для всех λ. При $t=1$ уравнения (1.9) имеют решение $x_j = 1$, так что одно характеристическое число равно 1. Не ограничивая общности, мы можем положить $t_1 = 1$. Если цепь непериодична, то для всех других характеристических чисел мы имеем $|t_\lambda| < 1$, и из (1.15) видно, что при $n \to \infty$

$$p_{jk}^{(n)} \to c^{(1)} y_k^{(1)}. \tag{1.16}$$

Иначе говоря, *инвариантное распределение вероятностей можно характеризовать как решение системы* (1.10) *при* $t=1$.

2. ПРИМЕРЫ

а) Рассмотрим сначала цепь, имеющую только два состояния. Матрица переходных вероятностей принимает простой вид

$$P = \begin{pmatrix} 1-p & p \\ \alpha & 1-\alpha \end{pmatrix},$$

где $0 < p < 1$ и $0 < \alpha < 1$. Вычисления здесь тривиальны, поскольку рассматриваются лишь системы из двух уравнений. Характеристические числа суть $t_1 = 1$ и $t_2 = 1 - \alpha - p$. Явное представление (1.15) для $p_{jk}^{(n)}$ можно записать в матричной форме:

$$P^n = \frac{1}{\alpha+p}\begin{pmatrix} \alpha & p \\ \alpha & p \end{pmatrix} + \frac{(1-\alpha-p)^n}{\alpha+p}\begin{pmatrix} p & -p \\ -\alpha & \alpha \end{pmatrix}$$

(где множители, общие для всех четырех элементов матриц, вынесены как коэффициенты при матрицах). Эта формула справедлива для $n \geqslant 0$.

б) Пусть

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.1}$$

(это матрица из задачи 2, б гл. XV, 14). Система (1.9) сводится к

$$x_4 = tx_1, \quad x_4 = tx_2, \quad (1/2)(x_1 + x_2) = tx_3, \quad x_3 = tx_4. \qquad (2.2)$$

Значению $t=0$ здесь соответствует решение $(1, -1, 0, 0)$, однако мы видели, что для явного представления $p_{jk}^{(n)}$ при $n \geqslant 1$ корень 0 не требуется. Стандартная процедура последовательного исключения переменных показывает, что остальные корни удовлетворяют кубическому уравнению $t^3=1$. Если для сокращения записи положить

$$\theta = e^{2\pi i/3} = \cos(2\pi/3) + i \sin(2\pi/3) \qquad (2.3)$$

(где $i^2=-1$), то этими тремя корнями будут $t_1=1$, $t_2=\theta$ и $t_3=\theta^2$ (или, что то же самое, $t_3=\theta^{-1}$). Теперь мы должны решить при этих значениях t системы (1.9) и (1.10). Поскольку постоянный множитель остается произвольным, мы можем положить $x_1^{(\nu)}=y_1^{(\nu)}=1$. Тогда решения в окончательном явном представлении будут совпадать соответственно с первыми столбцами и первыми строками трех матриц:

$$P^n = \frac{1}{6}\begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix} + \frac{\theta^n}{6}\begin{bmatrix} 1 & 1 & 2\theta & 2\theta^2 \\ 1 & 1 & 2\theta & 2\theta^2 \\ \theta^2 & \theta^2 & 2 & 2\theta \\ \theta & \theta & 2\theta^2 & 2 \end{bmatrix} +$$

$$+\frac{\theta^{2n}}{6}\begin{bmatrix} 1 & 1 & 2\theta^2 & 2\theta \\ 1 & 1 & 2\theta^2 & 2\theta \\ \theta & \theta & 2 & 2\theta^2 \\ \theta^2 & \theta^2 & 2\theta & 2 \end{bmatrix}. \qquad (2.4)$$

Поскольку мы отбросили корень $t=0$, эта формула справедлива только при $n \geqslant 1$.

Из (2.4) очевидно, что наша цепь имеет период 3. Чтобы найти асимптотическое поведение P^n, заметим, что $1+\theta+\theta^2=0$. Используя этот факт, легко проверить, что при $n \to \infty$ по числам вида $n=3k$ строки матрицы P^n будут стремиться к $(1/2, 1/2, 0, 0)$. Для $n=3k+1$ и $n=3k+2$ соответствующие пределы суть $(0, 0, 0, 1)$ и $(0, 0, 1, 0)$. Отсюда следует, что инвариантное распределение вероятностей дается вектором $(1/6, 1/6, 1/3, 1/3)$.

в) Пусть $p+q=1$ и

$$P = \begin{bmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{bmatrix}. \qquad (2.5)$$

Эта цепь представляет собой частный случай цепи из следующего примера, но из-за своей простоты рассматривается отдельно. Легко видеть, что система (1.9) сводится к двум линейным уравнениям для двух неизвестных x_1+x_3 и x_2+x_4, и, следовательно, четыре характеристических числа здесь суть

$$t_1=1,\ t_2=-1,\ t_3=i(q-p),\ t_4=-i(q-p). \tag{2.6}$$

Соответствующими решениями будут $(1, 1, 1, 1)$, $(-1, 1, -1, 1)$, $(-i, -1, i, 1)$ и $(i, -1, -i, 1)$. (Заметим, что они имеют вид $(\theta, \theta^2, \theta^3, \theta^4)$, где θ — корень четвертой степени из единицы.) Система (1.10) отличается от (1.9) лишь тем, что p и q в ней меняются местами, и поэтому без дальнейших вычислений мы получаем

$$p_{jk}^{(n)}=(1/4)\{1+(q-p)^n i^{j-k-n}\}\{1+(-1)^{k+j-n}\}. \tag{2.7}$$

г) В *общем циклическом случайном блуждании* из примера гл. XV, 2, г) первая строка матрицы P имеет вид $q_0, \ldots, q_{\rho-1}$, а остальные строки получаются из нее циклическими перестановками. В предыдущем примере было показано, что в частном случае $\rho=4$ $x_j^{(\nu)}$ и $y_k^{(\nu)}$ могут быть представлены как степени корней четвертой степени из единицы. Поэтому естественно попытаться найти здесь аналогичное представление через корни ρ-й степени из единицы, а именно через величины

$$\theta=e^{2i\pi/\rho}. \tag{2.8}$$

Все корни ρ-й степени из единицы суть $1, \theta, \theta^2, \ldots, \theta^{\rho-1}$. Положим для $r=0, \ldots, \rho-1$

$$t_r=\sum_{\nu=0}^{\rho-1} q_\nu\theta^{\nu r}. \tag{2.9}$$

Легко проверить, что при $t=t_r$ системы (1.9) и (1.10) имеют решения

$$x_j^{(r)}=\theta^r, \quad y_k^{(r)}=\theta^{-rk} \tag{2.10}$$

и что для соответствующих коэффициентов $c^{(r)}$ мы во всех случаях имеем $c^{(r)}=1/\rho$. Таким образом, окончательно[1])

$$p_{jk}^{(n)}=\rho^{-1}\sum_{r=0}^{\rho-1}\theta^{r(j-k)}t_r^n. \tag{2.11}$$

[1]) При $n=0$ правая часть (2.11) определена только тогда, когда ни одно из t_r не обращается в нуль. В действительности мы доказали (2.11) для $n\geqslant 1$, предполагая, что все корни t_r различны, а это отнюдь не всегда верно в настоящей ситуации. Например, если $q_k=\rho^{-1}$ при всех k, то $t_0=1$, а $t_1=\ldots=t_{\rho-1}=0$. Но даже в этом крайнем случае (2.11) остается справедливым при всех j, k и $n\geqslant 1$. К счастью, (2.11) нетрудно проверить непосредственной индукцией по n. В частности, для $n=1$ множитель при q_ν

д) *Задача о размещении.* Пример гл. XV, 2, ж) показывает, что классическая задача о размещении может изучаться методом цепей Маркова. Система находится в состоянии j, если имеется j занятых и $\rho-j$ пустых ящиков. Если это состояние является начальным и по ящикам случайным образом размещается еще n дополнительных шаров, то $p_{jk}^{(n)}$ есть вероятность того, что после этого будет k занятых и $\rho-k$ пустых ящиков (так что $p_{jk}^{(n)}=0$ при $k<j$). Для $j=0$ эта вероятность получается по формуле (11.7) гл. II. Теперь мы выведем формулу для $p_{jk}^{(n)}$, обобщив таким образом результаты гл. II.

Поскольку $p_{jj}=j/\rho$ и $p_{j,j+1}=(\rho-j)/\rho$, система (1.9) сводится к

$$(\rho t-j)\,x_j=(\rho-j)\,x_{j+1}. \tag{2.12}$$

При $t=1$ отсюда вытекает, что $x_j=1$ для всех j. Когда $t\neq 1$, то с необходимостью $x_\rho=0$, и поэтому существует такой индекс r, что $x_{r+1}=0$, но $x_r\neq 0$; из (2.12) следует, что $\rho t=r$. Стало быть, характеристические числа здесь суть

$$t_r=r/\rho,\quad r=1,\ \ldots,\ \rho. \tag{2.13}$$

Соответствующие решения (2.12) даются формулой

$$x_j^{(r)}=\binom{r}{j}\binom{\rho}{j}^{-1}, \tag{2.14}$$

так что $x_j^{(r)}=0$ при $j>r$. Для $t=t_r$ система (1.10) сводится к

$$(r-j)\,y_j^{(r)}=(\rho-j+1)\,y_{j-1}^{(r)} \tag{2.15}$$

и имеет решение

$$y_j^{(r)}=\binom{\rho-r}{j-r}(-1)^{j-r}, \tag{2.16}$$

где, конечно же, $y_j^{(r)}=0$ при $j<r$. Поскольку $x_j^{(r)}=0$ при $j>r$ и $y_j^{(r)}=0$ при $j<r$, мы имеем

$$1/c^{(r)}=x_r^{(r)}y_r^{(r)}=\binom{\rho}{r},$$

и, следовательно,

$$p_{jk}^{(n)}=\sum_{r=j}^{k}\left(\frac{r}{\rho}\right)^n\binom{\rho}{r}\binom{r}{j}\binom{\rho-r}{k-r}(-1)^{k-r}\binom{\rho}{j}^{-1}. \tag{2.17}$$

в (2.11) сводится к

$$\rho^{-1}\sum_{r=0}^{\rho-1}\theta^{r(j-k+\nu)}.$$

Сумма здесь равна нулю, за исключением случаев $j-k+\nu=0$ и $j-k+\nu=\rho$, а в этих случаях каждое слагаемое равно единице. Следовательно, $p_{jk}^{(1)}$ сводится к q_{k-j}, если $k\geqslant j$, и к $q_{\rho+k-j}$, если $k<j$, а это и есть заданная матрица (p_{jk}).

Если выразить биномиальные коэффициенты через факториалы, то эта формула упростится:

$$p_{jk}^{(n)} = \binom{\rho-j}{\rho-k} \sum_{\nu=0}^{k-j} \left(\frac{\nu+j}{\rho}\right)^n (-1)^{k-j-\nu} \binom{k-j}{\nu} \tag{2.18}$$

и $p_{jk}^{(n)} = 0$ при $k < j$. ▶

(Численную иллюстрацию см. в примере 4, б).)

§ 3. СЛУЧАЙНОЕ БЛУЖДАНИЕ С ОТРАЖАЮЩИМИ ЭКРАНАМИ

Теперь применение цепей Маркова будет продемонстрировано на примере полного обсуждения случайного блуждания с состояниями 1, 2, ..., ρ и двумя отражающими экранами[1]). Матрица P для него приведена в примере гл. XV, 2, в). Для $2 \leqslant k \leqslant \rho - 1$ имеем $p_{k,\,k+1} = p$ и $p_{k,\,k-1} = q$; первая и последняя строки суть $(q, p, 0, \ldots, 0)$, $(0, \ldots, 0, q, p)$.

Для удобства сравнения с выводами гл. XIV мы теперь откажемся от переменной $t = s^{-1}$ и запишем характеристические числа в виде s_r^{-1} (вместо t_r); будет удобно пронумеровать их от 0 до $\rho - 1$. Для переменной s линейная система (1.9) запишется в виде

$$\begin{aligned} &x_1 = s\,(qx_1 + px_2), \\ &x_j = s\,(qx_{j-1} + px_{j+1}), \quad j = 2, 3, \ldots, \rho - 1, \\ &x_\rho = s\,(qx_{\rho-1} + px_\rho). \end{aligned} \tag{3.1}$$

Эта система допускает решение $x_j = 1$, соответствующее корню $s = 1$. Чтобы найти все остальные решения, применим метод частных решений (которым мы уже пользовались для решения аналогичных уравнений в гл. XIV, 4). Среднее уравнение в (3.1) удовлетворяется при $x_j = \lambda^j$, если λ — корень квадратного уравнения $\lambda = qs + \lambda^2 ps$. Два корня этого уравнения суть

$$\lambda_1(s) = \frac{1 + \sqrt{1 - 4pqs^2}}{2ps}, \quad \lambda_2(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}, \tag{3.2}$$

и поэтому общим решением среднего из уравнений (3.1) является

$$x_j = A(s)\,\lambda_1^j(s) + B(s)\,\lambda_2^j(s), \tag{3.3}$$

где $A(s)$ и $B(s)$ произвольны. Первое и последнее из уравнений (3.1) будут иметь решение (3.3) тогда и только тогда, когда $x_0 = x_1$ и $x_\rho = x_{\rho+1}$. Для этого требуется, чтобы $A(s)$ и $B(s)$ удов-

[1]) Часть последующего изложения является повторением теории из гл. XIV. Там приведенное в тексте квадратное уравнение встречается под номером (4.7); выражения (3.2) для $\lambda_1(s)$ и $\lambda_2(s)$ даются в (4.8), а общее решение (3.3) появляется в виде (4.9). Два эти метода связаны друг с другом, однако во многих случаях детали вычислений оказываются совершенно различными.

летворяли условиям

$$A(s)\{1-\lambda_1(s)\}+B(s)\{1-\lambda_2(s)\}=0,$$
$$A(s)\lambda_1^{\rho}(s)\{1-\lambda_1(s)\}+B(s)\lambda_2^{\rho}(s)\{1-\lambda_2(s)\}=0. \tag{3.4}$$

И обратно, если эти два уравнения справедливы для некоторого значения s, то (3.3) представляет собой решение линейной системы (3.1), и это решение тождественно обращается в нуль только тогда, когда $\lambda_1(s)=\lambda_2(s)$. Стало быть, наша задача состоит в том, чтобы найти те значения s, для которых

$$\lambda_1^{\rho}(s)=\lambda_2^{\rho}(s), \quad \text{но} \quad \lambda_1(s)\neq\lambda_2(s). \tag{3.5}$$

Поскольку $\lambda_1(s)\lambda_2(s)=q/p$, то из первого соотношения вытекает, что $\lambda_1(s)\sqrt{p/q}$ должно быть корнем (2ρ)-й степени из единицы, т. е. мы должны иметь

$$\lambda_1(s)=\sqrt{q/p}\,e^{i\pi r/\rho}, \tag{3.6}$$

где r — целое число, $0\leqslant r<2\rho$. Из определения (3.2) легко видеть, что (3.6) имеет место только при $s=s_r$, где

$$s_r^{-1}=2\sqrt{pq}\cos\pi r/\rho. \tag{3.7}$$

Для значения $s=s_0$ второе условие (3.5) нарушается; более того, $s_r=s_{2\rho-r}$, и поэтому ρ различных характеристических чисел даются формулой (3.7) при $r=0, 1, \ldots, \rho-1$.

Решая (3.4) при $s=s_r$ и подставляя результат в (3.3), мы получаем

$$x_j^{(r)}=(q/p)^{j/2}\sin(\pi rj/\rho)-(q/p)^{(j+1)/2}\sin[\pi r(j-1)/\rho] \tag{3.8}$$

для $r=1, \ldots, \rho-1$, тогда как для $r=0$

$$x_j^{(0)}=1. \tag{3.9}$$

Сопряженная система (1.10) сводится к

$$\begin{aligned} y_1&=sq(y_1+y_2),\\ y_k&=s(py_{k-1}+qy_{k+1}), \quad k=2, \ldots, \rho-1,\\ y_\rho&=sp(y_{\rho-1}+y_\rho). \end{aligned} \tag{3.10}$$

Среднее уравнение здесь такое же, как в (3.1), но с переставленными p и q, и поэтому его общее решение получается из (3.3) с переставленными p и q. Первое и последнее уравнения могут быть удовлетворены, если $s=s_r$, и простые вычисления показывают, что для $r=1, 2, \ldots, \rho-1$ решением (3.10) является

$$y_k^{(r)}=(p/q)^{k/2}\sin(\pi rk/\rho)-(p/q)^{(k-1)/2}\sin[\pi r(k-1)/\rho]. \tag{3.11}$$

Для $s_0=1$ мы аналогично получаем

$$y_k^{(0)}=(p/q)^k. \tag{3.12}$$

Остается найти коэффициенты $c^{(r)}$, определяемые соотношениями

$$c^{(r)}\sum_{k=0}^{\rho-1} x_k^{(r)} y_k^{(r)} = 1. \tag{3.13}$$

При $r=0$ k-й член этой суммы равняется $(p/q)^k$, и поэтому

$$c^{(0)} = (q/p)\cdot[(p/q)-1]/[(p/q)^\rho - 1], \tag{3.14}$$

за исключением случая $p=q$, когда $c_0 = 1/\rho$. При $r \geqslant 1$ элементарные, хотя и трудоемкие, вычисления[1]) приводят к выражению

$$c^{(r)} = (2p/\rho)\,\{1-2\sqrt{pq}\cos(\pi r/\rho)\}^{-1}. \tag{3.15}$$

Поэтому общее представление (1.15) для переходных вероятностей высших порядков приводит к окончательному результату[2])

$$p_{jk}^{(n)} = \frac{(p/q)-1}{(p/q)^\rho - 1}\left(\frac{p}{q}\right)^{k-1} + \frac{2p}{\rho}\sum_{r=1}^{\rho-1} \frac{x_j^{(r)} y_k^{(r)}\,[2\sqrt{pq}\cos(\pi r/\rho)]^n}{1-2\sqrt{pq}\cos(\pi r/\rho)}, \tag{3.16}$$

где $x_j^{(r)}$ и $y_k^{(r)}$ определены формулами (3.8) и (3.11). При $p=q$ первый член в правой части следует считать равным $1/\rho$.

§ 4. НЕВОЗВРАТНЫЕ СОСТОЯНИЯ; ВЕРОЯТНОСТИ ПОГЛОЩЕНИЯ

Теорема § 1 была выведена в предположении, что корни t_1, t_2, ... различны. Наличие кратных корней не требует существенных изменений, однако мы обсудим лишь особо важный частный случай. Корень $t_1=1$ является кратным, когда цепь содержит две или несколько замкнутых подцепей; эта ситуация часто встречается в задачах, связанных с вероятностями поглощения. Легко приспособить метод § 1 к этому случаю. Для краткости и ясности мы объясним эту процедуру на примерах, в которых обнаружатся главные черты общего случая.

---

[1]) Эти вычисления значительно упростятся, если воспользоваться комплексной записью и тождеством $\sin v = [e^{iv} - e^{-iv}]/(2i)$. Сумма в (3.13) сводится к линейной комбинации (с комплексными коэффициентами) сумм вида

$$\sum_{j=0}^{\rho-1} e^{2j\pi i m/\rho},$$

где $m=0$ или $m=\pm 1$. В первом случае эта сумма равна ρ, во втором — нулю, откуда (3.15) следует тривиальным образом.

[2]) Аналогичные формулы в случае одного отражающего и одного поглощающего экрана см. в статье Каца (Kac M., Random walk and the theory of Brownian motion, Amer. Math. Monthly, 54 (1947), 369—391). Определение отражающего экрана модифицировано там таким образом, что частица может достичь 0; когда это происходит, следующий шаг переводит ее в 1. Явные формулы оказываются в этом случае более сложными. Статья Каца содержит также формулы для $p_{jk}^{(n)}$ в модели Эренфестов (пример гл. XV, 2, д)).

Примеры. а) Рассмотрим матрицу переходных вероятностей

$$P=\begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/5 & 4/5 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 1/4 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}. \tag{4.1}$$

Ясно, что E_1 и E_2 образуют замкнутое множество (т. е. из них невозможен переход ни в одно из остальных четырех состояний; ср. гл. XV, 4). Аналогично E_3 и E_4 образуют другое замкнутое множество. Наконец, E_5 и E_6 суть невозвратные состояния. После конечного числа шагов система перейдет в одно из двух замкнутых множеств и останется там навсегда.

Матрица P имеет вид блочной матрицы

$$P=\begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ U & V & T \end{bmatrix}, \tag{4.2}$$

где каждая буква обозначает (2×2)-матрицу, а каждый нуль — матрицу из четырех нулей. Например, A имеет строки $(1/3, 2/3)$ и $(2/3, 1/3)$; это матрица переходных вероятностей, соответствующая цепи, образованной двумя состояниями E_1 и E_2. Эту матрицу можно изучать отдельно, а степени A^n могут быть получены из примера 2, а) при $p=\alpha=2/3$. Когда степени P^2, P^3, ... будут вычислены, окажется, что на первые две строки остальные четыре никак не влияют. Точнее, P^n имеет вид

$$P^n=\begin{bmatrix} A^n & 0 & 0 \\ 0 & B^n & 0 \\ U_n & V_n & T^n \end{bmatrix}, \tag{4.3}$$

где A^n, B^n, T^n суть n-е степени A, B и T соответственно и могут быть вычислены¹) методом § 1 (ср. пример 2, а), где выполнены все вычисления). Вместо шести уравнений с шестью неизвестными перед нами теперь только системы из двух уравнений с двумя неизвестными каждая.

Следует отметить, что матрицы U_n и V_n в (4.3) не являются степенями U и V и не могут быть получены тем же простым способом, что и A^n, B^n и T^n. Однако при вычислении P^2, P^3, ...

¹) Суммы по строкам в T не равны единице, так что T не является стохастической матрицей (это субстохастическая матрица в смысле определения гл. XV, 8). Метод § 1 применим без изменений, за исключением того, что $t=1$ не является более корнем (так что $T^n \to 0$).

третий и четвертый столбцы никак не влияют на остальные четыре столбца. Иначе говоря, вычеркнув в P^n соответствующие E_3 и E_4 строки и столбцы, мы получим матрицу

$$\begin{pmatrix} A^n & 0 \\ U_n & T^n \end{pmatrix}, \tag{4.4}$$

которая является n-й степенью соответствующей подматрицы матрицы P, т. е. матрицы

$$\begin{pmatrix} A & 0 \\ U & T \end{pmatrix} = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 \\ 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}. \tag{4.5}$$

Стало быть, матрица (4.4) может быть вычислена методом § 1, который в настоящем случае значительно упрощается. Матрица V_n может быть получена аналогичным образом.

Обычно явный вид матриц U_n и V_n интересен лишь постольку, поскольку они связаны с *вероятностями поглощения*. Если система выходит, скажем, из E_5, то какова будет *вероятность* λ *того, что она в конце концов попадет в замкнутое множество, образованное состояниями* E_1 и E_2 (а не в другое замкнутое множество)? Чему равна *вероятность* λ_n *того, что это произойдет в точности на* n*-м шаге*? Ясно, что $p_{51}^{(n)}+p_{52}^{(n)}$ есть вероятность того, что рассматриваемое событие произойдет не позже чем на n-м шаге, т. е.

$$p_{51}^{(n)}+p_{52}^{(n)}=\lambda_1+\lambda_2+\ldots+\lambda_n.$$

Устремляя n к ∞, получаем λ. Предпочтительный способ вычисления λ_n состоит в следующем. На $(n-1)$-м шаге система должна попасть в состояние, отличное от E_1 и E_2, т. е. либо в E_5, либо в E_6 (поскольку из E_3 или из E_4 переход в E_1 или E_2 невозможен). Затем n-й шаг переводит систему в E_1 или в E_2. Следовательно,

$$\lambda_n=p_{55}^{(n-1)}(p_{51}+p_{52})+p_{56}^{(n-1)}(p_{61}+p_{62})=(1/4)\,p_{55}^{(n-1)}+(1/3)\,p_{56}^{(n-1)}.$$

Отметим, что λ_n вполне определяется элементами T^{n-1}, а эту матрицу легко вычислить. В настоящем случае

$$p_{55}^{(n)}=p_{56}^{(n)}=(1/4)\,(5/12)^{n-1}, \quad \text{и, следовательно,} \quad \lambda_n=(7/48)\,(5/12)^{n-2}.$$

б) *Братско-сестринское скрещивание*. Мы завершим этот параграф численным анализом цепи из примера гл. XV, 2, к). Суть последующих рассуждений состоит в том, чтобы показать, что каноническое представление

$$p_{jk}^{(n)}=\sum_{r=1}^{\nu} t_r^n c^{(r)} x_j^{(r)} y_k^{(r)} \tag{4.6}$$

остается справедливым, хотя $t=1$ является *двойным корнем* характеристического уравнения.

Система (1.9) линейных уравнений принимает вид

$$x_1=tx_1,\quad (1/4)\,x_1+(1/2)\,x_2+(1/4)\,x_3=tx_2,$$
$$(1/16)\,x_1+(1/4)\,x_2+(1/4)\,x_3+(1/4)\,x_4+(1/16)\,x_5+(1/8)\,x_6=tx_3,$$
$$(1/4)\,x_3+(1/2)\,x_4+(1/4)\,x_5=tx_4,\quad x_5=tx_5,\quad x_3=tx_6,\tag{4.7}$$

и эти уравнения выявляют форму заданной матрицы. Из первого и пятого уравнений ясно, что $x_1=x_5=0$, если не выполнено $t=1$. Стало быть, при $t\neq 1$ система на самом деле сводится к четырем уравнениям с четырьмя неизвестными, и стандартная процедура исключения переменных приводит к уравнению четвертой степени для t в качестве условия совместности этих четырех уравнений. Поскольку всего у нашей матрицы шесть собственных значений, отсюда следует, что $t=1$ является двойным корнем. Нетрудно проверить, что эти шесть собственных значений суть [1])

$$t_1=t_2=1,\quad t_3=1/2,\quad t_4=1/4,\quad t_5=1/4+\sqrt{5}/4,\quad t_6=1/4-\sqrt{5}/4.\tag{4.8}$$

Соответствующие решения $(x_1^{(r)},\ \ldots,\ x_6^{(r)})$ системы (4.7) могут быть выбраны следующим образом:

$$(1,\ 3/4,\ 1/2,\ 1/4,\ 0,\ 1/2),\quad (0,\ 1/4,\ 1/2,\ 3/4,\ 1,\ 1/2),\ (0,\ 1,\ 0,\ -1,\ 0,\ 0),$$
$$(0,\ 1,\ -1,\ 1,\ 0,\ -4),\quad (0,\ 1,\ -1+\sqrt{5},\ 1,\ 0,\ 6-2\sqrt{5}),$$
$$(0,\ 1,\ -1-\sqrt{5},\ 1,\ 0,\ 6+2\sqrt{5}).\tag{4.9}$$

Следующая задача состоит в нахождении соответствующих решений $(y_1^{(r)},\ \ldots,\ y_6^{(r)})$ системы, получаемой из (4.7) при перестановке местами строк и столбцов [в матрице коэффициентов.— *Перев.*] Для $r\geqslant 3$ это решение определено с точностью до постоянного множителя, однако для двойного корня $t_1=t_2=1$ мы должны выбирать среди бесконечного числа решений вида $(a,\ 0,\ 0,\ 0,\ b,\ 0)$. Подходящий выбор становится очевидным из вида искомого представления (4.6). В самом деле, из (4.9) ясно, что $x_1^{(r)}=0$, за исключением случая $r=1$, и поэтому из (4.6) имеем $p_{1k}^{(n)}=c^{(1)}y_k^{(1)}$ для всех k и n. Но E_1 является поглощающим состоянием, и, очевидно, $p_{1k}^{(n)}=0$ при всех $k\neq 1$. Отсюда следует, что для $r=1$ мы должны выбрать решение вида $(a,\ 0,\ 0,\ 0,\ 0,\ 0)$, и по той же причине соответствующим $r=2$ решением будет $(0,\ 0,\ 0,\ 0,\ b,\ 0)$.

Решения, соответствующие остальным характеристическим числам, находятся легко. (Те из них, которые были выбраны для наших вычислений, представляются вторыми строками приводимых ниже матриц.) Затем из (1.14) определяются нормирующие постоянные

[1]) Корень $t_3=1/2$ легко обнаружить, поскольку он соответствует простому решению $x_2=-x_4=1$ и $x_1=x_3=x_5=x_6=0$. Для остальных корней получается несложное кубическое уравнение.

$c^{(r)}$, и таким образом мы получаем все величины, входящие в представление (4.6)

Для записи окончательного результата матрицы, соответствующие $r=1$ и $r=2$, были объединены в одну. Более того, соответствующие $r=5$ и $r=6$ элементы $c^{(r)}x_j^{(r)}y_k^{(r)}$ имеют вид $a\pm b\sqrt{5}$. По техническим причинам и для ясности было необходимо перегруппировать вносимый ими вклад и представить его в виде $a[t_5^n+t_6^n]$ и $b\sqrt{5}[t_5^n-t_6^n]$:

$$P^n=\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \end{vmatrix}+\frac{2^{-n}}{4}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$+\frac{4^{-n}}{20}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -4 & 4 & -1 & -2 \\ 1 & -4 & 4 & -4 & 1 & 2 \\ -1 & 4 & -4 & 4 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & -16 & 16 & -16 & 4 & 8 \end{vmatrix}+\frac{t_5^n+t_6^n}{40}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -9 & 6 & 4 & 6 & -9 & 2 \\ -11 & 4 & 16 & 4 & -11 & -2 \\ -9 & 6 & 4 & 6 & -9 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -14 & 16 & -16 & 16 & -14 & 12 \end{vmatrix}$$

$$+\frac{t_5^n-t_6^n}{40}\sqrt{5}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -4 & 2 & 4 & 2 & -4 & 0 \\ -5 & 4 & 0 & 4 & -5 & 2 \\ -4 & 2 & 4 & 2 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -6 & 0 & 16 & 0 & -6 & -4 \end{vmatrix}.$$

Легко проверить, что эта формула справедлива при $n=0$. С другой стороны, из структуры правой части (4.6) ясно, что если (4.6) имеет место для некоторого n, то оно справедливо и для $n+1$. Таким способом истинность (4.6) может быть установлена и без обращения к общей теории § 1.

§ 5. ПРИЛОЖЕНИЕ К ВРЕМЕНАМ ВОЗВРАЩЕНИЯ

В задаче 19 гл. XIII,12 показывается, как выразить среднее μ и дисперсию σ^2 времени возвращения рекуррентного события $\mathcal{E}$ через вероятности u_n осуществления события $\mathcal{E}$ при n-м испытании. Если $\mathcal{E}$ не является периодическим, то

$$u_n\to 1/\mu \text{ и } \sum_{n=0}^{\infty}(u_n-1/\mu)=(\sigma^2-\mu+\mu^2)/(2\mu^2) \tag{5.1}$$

при условии, что σ^2 конечна.

Если отождествить $\mathcal{E}$ с возвратным состоянием E_j, то $u_n=p_{jj}^{(n)}$ (и $u_0=1$). В конечной цепи Маркова все времена возвращения имеют конечные дисперсии (см. задачу 19 гл. XV, 14), так что (5.1) применимо. Предположим, что E_j не является периодическим и применима формула (1.5). Тогда $t_1=1$ и $|t_r|<1$ при $r=2, 3, \ldots,$ так что $p_{jj}^{(n)} \to b_{jj}^{(1)} = \mu_j^{-1}$. Члену $u_n - \mu^{-1}$ суммы в (5.1) соответствует

$$p_{jj}^{(n)} - 1/\mu_j = \sum_{r=2}^{\rho} b_{jj}^{(r)} t_r^n. \tag{5.2}$$

Эта формула справедлива при $n \geqslant 1$; суммируя геометрическую прогрессию со знаменателем t_r, находим

$$\sum_{n=1}^{\infty} (p_{jj}^{(n)} - 1/\mu_j) = \sum_{r=2}^{\rho} b_{jj}^{(r)} t_r/(1-t_r). \tag{5.3}$$

Подставляя это выражение в (5.1), мы находим, что *если E_j — непериодическое возвратное состояние, то среднее время возвращения в него равно $\mu_j = 1/b_{jj}^{(1)}$, а дисперсия этого времени возвращения равна*

$$\sigma_j^2 = \mu_j - \mu_j^2 + 2\mu_j^2 \sum_{r=2}^{\rho} b_{jj}^{(r)} t_r/(1-t_r) \tag{5.4}$$

при условии, что формула (1.5) применима и $t_1 = 1$. Случай периодических состояний и наличие двойных корней потребуют лишь очевидных изменений.

ПРОСТЕЙШИЕ СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ

§ 1. ОБЩИЕ ПОНЯТИЯ. МАРКОВСКИЕ ПРОЦЕССЫ

Цепи Маркова, изучавшиеся нами в предыдущих главах, могут быть описаны очень грубо как стохастические процессы, в которых будущее зависит лишь от настоящего состояния, но не от прошлой истории или того способа, которым было достигнуто настоящее состояние. Эти процессы имеют только счетное множество значений (состояний) E_1, E_2,... и зависят от дискретного временнóго параметра, т. е. изменения могут происходить лишь в фиксированные моменты времени $t=0, 1, \ldots$. В настоящей главе мы рассмотрим такие явления, как телефонные вызовы, радиоактивный распад и расщепление хромосом, в которых изменения могут происходить в любой момент времени. С математической точки зрения мы будем иметь дело со стохастическими процессами со счетным множеством состояний, но зависящими уже от непрерывного временнóго параметра. В рамках дискретных вероятностей полное описание таких процессов невозможно, и мы на самом деле не в состоянии формально определить интересующий нас класс марковских процессов.

Действительно, чтобы описать прошлую историю процесса, мы должны указать те моменты времени, в которые происходили изменения, а для этого потребуются вероятности на континууме. Выражение «будущее развитие не зависит от прошлой истории» имеет очевидное интуитивное значение (по крайней мере по аналогии с дискретными цепями Маркова), но для формального определения потребуется понятие условной вероятности, лежащее за пределами этой книги. Однако многие задачи, связанные с такими процессами, можно изучать отдельно при помощи довольно элементарных методов, если принять на веру, что эти процессы действительно существуют. Так мы и будем теперь поступать.

Переходной вероятности $p_{jk}^{(n)}$ для цепей Маркова теперь соответствует переходная вероятность $P_{jk}(t)$, а именно условная вероятность состояния E_k в момент $t+s$ при условии, что в момент $s<t+s$ система находилась в состоянии E_j. Как показывает обозначение, предполагается, что эта вероятность зависит только от продолжительности t временнóго интервала, но не от его положения на оси времени. Такие переходные вероятности называются

*) Эта глава почти независима от гл. X—XVI. Относительно использования термина «стохастический процесс» см. примечание 5 на с. 435.

стационарными или однородными по времени. (В § 9, однако, будут рассматриваться неоднородные процессы.) Аналогом основных соотношений (3.3) гл. XV является *уравнение Колмогорова—Чэпмена*

$$P_{ik}(\tau+t)=\sum_j P_{ij}(\tau)P_{jk}(t), \tag{1.1}$$

которое основано на следующем рассуждении. Предположим, что в момент времени 0 система находится в состоянии E_i. Тогда j-й член в правой части представляет вероятность сложного события, состоящего в том, что система в момент времени τ находится в состоянии E_j, а в более поздний момент $\tau+t$ — в состоянии E_k. Но переход из состояния E_i в момент времени 0 в состояние E_k в момент $\tau+t$ с необходимостью происходит через некоторое промежуточное состояние E_j в момент времени τ, и, суммируя по всем возможным E_j, мы видим, что (1.1) должно выполняться для произвольных (фиксированных) $\tau>0$ и $t>0$.

В этой главе мы будем изучать решения основного уравнения (1.1). Будет показано, что простые постулаты, приспособленные к конкретным ситуациям, приводят к системам дифференциальных уравнений для $P_{jk}(t)$ и что из этих дифференциальных уравнений, даже не решая их, можно получить интересные результаты. И эти результаты имеют смысл, потому что наши решения действительно являются переходными вероятностями марковского процесса, который однозначно определяется этими вероятностями и начальным положением в момент времени 0. Этот интуитивно очевидный факт [1]) мы примем без доказательства.

Для фиксированных j и t переходные вероятности $P_{jk}(t)$ определяют обычное дискретное распределение вероятностей. Оно зависит от непрерывного параметра t, однако мы уже встречались с многими семействами распределений, зависящих от непрерывного параметра. Технически рассуждения последующих параграфов остаются в рамках дискретных вероятностей, но это искусственное ограничение является для многих целей слишком строгим. Этот момент может проиллюстрировать распределение Пуассона $\{e^{-\lambda t}(\lambda t)^n/n!\}$. Нулевой член $e^{-\lambda t}$ этого распределения можно интерпретировать как вероятность того, что за интервал времени фиксированной длины t не поступило ни одного телефонного вызова. Но тогда $e^{-\lambda t}$ будет также вероятностью того, что время ожидания первого вызова превышает t, и поэтому мы косвенно имеем дело с непрерывным распределением вероятностей на оси времени. Мы вернемся к этому вопросу в § 6.

[1]) Стоит отметить, однако, что могут существовать (довольно патологические) немарковские процессы с теми же переходными вероятностями. Этот вопрос подробно обсуждался в разд. 2а гл. XII в связи с процессами с независимыми приращениями (которые являются частным классом марковских процессов). См. также § 9, в частности примечание на с. 485—486.

§ 2. ПУАССОНОВСКИЙ ПРОЦЕСС

Пуассоновский процесс можно рассматривать с различных точек зрения, и здесь мы рассмотрим его в качестве прототипа всех процессов из этой главы. Последующий вывод распределения Пуассона наилучшим образом подходит для наших обобщений, однако он никоим образом не является лучшим и в других контекстах. Его следует сравнить с элементарным выводом в гл. VI, 6 и с разд. 2а гл. XII, где пуассоновский процесс фигурировал как простейший процесс с независимыми приращениями.

В качестве эмпирических предпосылок возьмем такие случайные события, как распад частиц, поступающие телефонные вызовы, расщепление хромосом под действием вредной радиации. Предполагается, что все наблюдаемые события однотипны, и мы интересуемся полным числом $\mathbf{Z}(t)$ событий, происшедших в течение произвольного интервала времени длины t. Каждое событие представляется точкой на оси времени, и поэтому мы в действительности рассматриваем некоторые случайные размещения точек на прямой. Лежащие в основе нашей математической модели физические предположения состоят в том, что силы и воздействия, управляющие процессом, остаются постоянными, так что вероятность любого отдельного события одна и та же для всех интервалов времени продолжительности t и не зависит от прошлого развития процесса. Математически это означает, что наш процесс является однородным по времени марковским процессом в смысле, описанном в предыдущем параграфе. Как уже выше говорилось, мы не стремимся к полной теории таких процессов, а удовольствуемся выводом основных вероятностей

$$P_n(t) = \mathbf{P}\{\mathbf{Z}(t) = n\}. \tag{2.1}$$

Они могут быть выведены из простых постулатов без обращения к более глубоким теоретическим соображениям.

Чтобы ввести понятия, подходящие и для других процессов из этой главы, мы выберем начало отсчета времени и будем говорить, что в *момент времени* $t>0$ *система находится в состоянии* E_n, если между 0 и t произошло ровно n скачков [функции $\mathbf{Z}(t)$.— *Перев.*]. Тогда $P_n(t)$ равняется вероятности состояния E_n в момент t, однако $P_n(t)$ может быть также описана как вероятность перехода из произвольного состояния E_j в произвольный момент времени s в состояние E_{j+n} к моменту $s+t$. Теперь наше нестрогое описание процесса мы преобразуем в свойства вероятностей $P_n(t)$.

Разобьем временно́й интервал единичной длины на N подынтервалов длины $h=N^{-1}$. Вероятность скачка внутри любого из этих подынтервалов равна $1-P_0(h)$, и поэтому математическое ожидание числа интервалов, содержащих скачки, равно $h^{-1}[1-P_0(h)]$. Интуитивно представляется, что при $h\to 0$ это число должно стремиться к математическому ожиданию числа скачков внутри произвольного интервала времени единичной длины, и поэтому естественно предположить [1]), что существует число $\lambda > 0$, такое, что

$$h^{-1}[1-P_0(h)] \to \lambda. \qquad (2.2)$$

Физическая картина процесса требует также, чтобы скачок обязательно приводил из состояния E_j в соседнее состояние E_{j+1}, и отсюда вытекает, что математическое ожидание числа подынтервалов (длины h), содержащих более чем один скачок, должно стремиться к 0. Поэтому мы должны предположить, что при $h \to 0$

$$h^{-1}[1 - P_0(h) - P_1(h)] \to 0. \qquad (2.3)$$

Чтобы окончательно сформулировать постулаты, запишем (2.2) в виде $P_0(h) = 1 - \lambda h + o(h)$, где (как обычно) $o(h)$ обозначает величину, по порядку меньшую чем h. (Точнее говоря, $o(h)$ означает такую величину, что $h^{-1}o(h) \to 0$ при $h \to 0$.) С учетом этого (2.3) эквивалентно соотношению $P_1(h) = \lambda h + o(h)$. Сформулируем теперь следующие постулаты.

Постулаты пуассоновского процесса. *Процесс начинается в момент времени 0 в состоянии* E_0. (i) *Непосредственный переход из состояния E_j возможен только в состояние* E_{j+1}. (ii) *Каково бы ни было состояние E_j процесса в момент времени t, (условная) вероятность скачка внутри последующего короткого интервала времени между t и $t+h$ равна $\lambda h + o(h)$, тогда как (условная) вероятность наличия в нем более чем одного скачка есть $o(h)$.*

Как было объяснено в предыдущем параграфе, эти условия слабее нашего исходного предположения об отсутствии влияния прошлой истории процесса на его будущую эволюцию. С другой стороны, наши постулаты носят чисто аналитический характер, и их достаточно, чтобы показать, что мы должны иметь

$$P_n(t) = [(\lambda t)^n/n!]\, e^{-\lambda t}. \qquad (2.4)$$

Для доказательства этого возьмем сперва $n \geq 1$ и рассмотрим событие, состоящее в том, что в момент времени $t+h$ система находится в состоянии E_n. Вероятность этого события равна $P_n(t+h)$, и осуществиться оно может тремя взаимоисключающими способами. Во-первых, в момент времени t система может находиться в состоянии E_n, и между t и $t+h$ не произойдет ни одного скачка. Вероятность этой возможности равна [2])

$$P_n(t)P_0(h) = P_n(t)[1-\lambda h] + o(h).$$

[1]) Предположение (2.2) вводится в основном потому, что его легко обобщить на другие процессы. В настоящем случае более естественным было бы заметить, что $P_0(t)$ должно удовлетворять функциональному уравнению $P_0(t+\tau) = P_0(t)\,P_0(\tau)$, из которого вытекает (2.2) (см. § 6).

[2]) В этом равенстве используются не входящие в постулаты пуассоновского процесса условия независимости и однородности по времени изменений процесса на непересекающихся интервалах времени. Легко убедиться в том, что для вывода (2.4) в действительности достаточно сформулированных постулатов; ср. замечание, сделанное после формулировки этих постулатов.— *Прим. перев.*

Вторая возможность состоит в том, что в момент времени t система находится в состоянии E_{n-1} и между t и $t+h$ происходит в точности один скачок. Вероятность этого равна $P_{n-1}(t)\cdot\lambda h+o(h)$. Любое другое состояние в момент t потребует [1]) более одного скачка в интервале между t и $t+h$, и вероятность подобного события есть $o(h)$. Следовательно, мы должны иметь

$$P_n(t+h)=P_n(t)(1-\lambda h)+P_{n-1}(t)\lambda h+o(h), \tag{2.5}$$

а это соотношение можно переписать в виде

$$[P_n(t+h)-P_n(t)]/h=-\lambda P_n(t)+\lambda P_{n-1}(t)+o(h)/h. \tag{2.6}$$

При $h\to 0$ последний член стремится к нулю; следовательно, предел [2]) левой части существует и равен

$$P_n'(t)=-\lambda P_n(t)+\lambda P_{n-1}(t),\quad n\geqslant 1. \tag{2.7}$$

При $n=0$ вторая и третья из упомянутых выше возможностей не возникают, и поэтому (2.5) следует заменить на

$$P_0(t+h)=P_0(t)(1-\lambda h)+o(h), \tag{2.8}$$

что приводит к

$$P_0'(t)=-\lambda P_0(t). \tag{2.9}$$

Отсюда и из $P_0(0)=1$ получаем $P_0(t)=e^{-\lambda t}$. Подставляя это значение $P_0(t)$ в (2.7) при $n=1$, мы получаем обыкновенное дифференциальное уравнение для $P_1(t)$. Поскольку $P_1(0)=0$, мы легко находим, что $P_1(t)=\lambda te^{-\lambda t}$, а это полностью согласуется с (2.4). Продолжая таким же образом, мы последовательно находим все члены (2.4).

§ 3. ПРОЦЕСС ЧИСТОГО РАЗМНОЖЕНИЯ

Простейшее обобщение пуассоновского процесса получается при предположении, что вероятности скачков могут зависеть от текущего состояния системы. Это приводит нас к следующим требованиям.

Постулаты. (i) *Непосредственный переход из состояния E_j возможен только в состояние E_{j+1}.* (ii) *Если в момент времени t систе-*

---

[1]) Для достижения E_n. Имеется в виду любое другое состояние E_k при $k<n-1$ (при $k>n$ переход из E_k в E_n невозможен).— *Прим. перев.*

[2]) Поскольку мы рассматриваем лишь положительные h, то $P_n'(t)$ в (2.7) следует понимать как правостороннюю производную. В действительности же это обычная двусторонняя производная. В самом деле, член $o(h)$ в (2.5) не зависит от t и поэтому остается неизменным при замене t на $t-h$. Таким образом, из (2.5) вытекает непрерывность, а из (2.6) — дифференцируемость в обычном смысле. Это замечание применимо на протяжении всей главы и повторяться не будет.

ма находится в состоянии E_n, то (условная) вероятность одного скачка в последующем коротком интервале времени между t и $t+h$ равна $\lambda_n h+o(h)$, тогда как (условная) вероятность более чем одного скачка в этом интервале есть $o(h)$.

Отличительная черта этого предположения заключается в том, что время, которое система проводит в любом конкретном состоянии, не играет никакой роли; возможны внезапные изменения состояния, однако, пока система находится в одном состоянии, она не стареет.

Пусть $P_n(t)$ снова будет вероятностью того, что в момент времени t система находится в состоянии E_n. Эти функции $P_n(t)$ удовлетворяют системе дифференциальных уравнений, которую можно вывести при помощи рассуждений предыдущего параграфа с тем лишь изменением, что (2.5) заменяется на

$$P_n(t+h)=P_n(t)(1-\lambda_n h)+P_{n-1}(t)\lambda_{n-1}h+o(h). \tag{3.1}$$

Таким образом мы получим *основную систему дифференциальных уравнений*

$$\begin{aligned} P_n'(t) &= -\lambda_n P_n(t)+\lambda_{n-1}P_{n-1}(t), \quad n\geqslant 1, \\ P_0'(t) &= -\lambda_0 P_0(t). \end{aligned} \tag{3.2}$$

В пуассоновском процессе было естественно предполагать, что в момент времени 0 система выходит из начального состояния E_0. Теперь мы можем допустить более общий случай, когда система выходит из произвольного начального состояния E_i. Тогда получаем, что [1])

$$P_i(0)=1, \quad P_n(0)=0 \text{ при } n\neq i. \tag{3.3}$$

Эти *начальные условия* единственным образом определяют решение $\{P_n(t)\}$ системы (3.2). (В частности, $P_0(t)=P_1(t)=\ldots=P_{i-1}(t)=0$.) Явные формулы для $P_n(t)$ выводились независимо многими авторами, однако для нас они не представляют интереса. Легко проверить, что для произвольных заданных λ_n система $\{P_n(t)\}$ обладает всеми требуемыми свойствами, за исключением того, что при некоторых условиях $\sum P_n(t)<1$. Это явление будет обсуждаться в § 4.

Примеры. а) *Радиоактивный распад.* В результате испускания частиц или γ-лучей радиоактивный атом, скажем урана, может превратиться в атом другого вида. Каждый вид представляет собой возможное состояние, и, когда процесс протекает, мы получаем последовательность переходов $E_0\to E_1\to E_2\to\ldots\to E_m$. Согласно принятым физическим теориям, вероятность перехода $E_n\to E_{n+1}$ остается неизменной, пока атом находится в состоянии E_n, и эта гипотеза находит выражение в нашем исходном предположении. Стало быть, этот процесс описывается дифференциальными уравнениями (3.2)

[1]) Следует отметить, что здесь $P_n(t)$ — та же вероятность, что и переходная вероятность $P_{in}(t)$ в § 1.

(факт, хорошо известный физикам). Если E_m — конечное состояние, из которого невозможны никакие другие переходы, то $\lambda_m=0$ и система (3.2) обрывается при $n=m$. (При $n>m$ мы автоматически получаем $P_n(t)=0$.)

б) *Процесс Юла.* Рассмотрим совокупность, элементы которой могут (путем деления или каким-либо другим способом) порождать новых членов, но не могут умереть (исчезнуть). Предположим, что на протяжении любого короткого интервала времени длины h каждый элемент совокупности с вероятностью $\lambda h+o(h)$ производит новый элемент; постоянная λ определяет скорость разрастания совокупности. Если между ее элементами нет никакого взаимодействия и в момент времени t объем совокупности был равен n, то вероятность того, что ее увеличение произойдет в некоторый момент времени между t и $t+h$, будет равна $n\lambda h+o(h)$. Поэтому вероятность $P_n(t)$ того, что в совокупности насчитывается ровно n элементов, удовлетворяет уравнениям (3.2) с $\lambda_n=n\lambda$, т. е. уравнениям

$$\begin{aligned} &P_n'(t)=-n\lambda P_n(t)+(n-1)\lambda P_{n-1}(t), \quad n\geqslant 1,\\ &P_0'(t)=0. \end{aligned} \tag{3.4}$$

Обозначим начальный объем совокупности через i. Тогда имеют место начальные условия (3.3), и легко проверить, что при $n\geqslant i>0$

$$P_n(t)=\binom{n-1}{n-i}e^{-i\lambda t}(1-e^{-\lambda t})^{n-i} \tag{3.5}$$

и, конечно, $P_n(t)=0$ при $n<i$ и всех t. Используя обозначение (8.1) гл. VI для отрицательного биномиального распределения, мы можем переписать (3.5) в виде $P_n(t)=f(n-i;\ i,\ e^{-\lambda t})$. Отсюда следует (ср. пример гл. IX, 3, в)), что объем совокупности в момент времени t является суммой i независимых случайных величин, каждая из которых имеет распределение, получающееся из (3.5) при замене i на 1. Эти i величин представляют собой потомства исходных элементов нашей совокупности.

Процесс такого типа впервые исследовал Юл [1]) в связи с матема-

[1]) Yule G. U., A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S., Philosophical Transactions of the Royal Society, London, Ser. B, 213 (1924), 21—87. Юл не вводил дифференциальных уравнений (3.4), а получил $P_n(t)$ предельным переходом, подобным тому, который использовался в гл. VI, 5.

Гораздо более общие и более гибкие модели того же типа были предложены и использованы для анализа эпидемий и роста популяций в написанной без особых претензий, но чрезвычайно интересной статье подполковника М'Кендрика (M'Kendrick A. G., Applications of mathematics to medical problems, Proceedings Edinburgh Mathematical Society, 44 (1925), 1—34). К сожалению, эта замечательная работа прошла практически незамеченной. В частности, она была неизвестна автору настоящей книги, когда он вводил различные стохастические модели для роста популяции в статье Feller W., Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung, Acta Biotheoretica, 5 (1939), 11—40.

тической теорией эволюции. Популяция состоит из видов в пределах одного рода, и появление нового вида обусловливается мутациями. Предположение о том, что каждый вид имеет одну и ту же вероятность породить новый вид, пренебрегает разницей в размерах видов. Поскольку мы пренебрегли также возможностью вымирания вида, можно ожидать, что (3.5) даст лишь грубое приближение.

Фэрри [1]) использовал ту же модель для описания процесса, связанного с космическими лучами, однако приближение снова было довольно грубым. Дифференциальные уравнения (3.4) применимы, строго говоря, к совокупностям частиц, которые могут делиться, образуя точные копии самих себя, при условии, конечно, что между частицами нет никаких взаимодействий. ▶

§ 4*). РАСХОДЯЩИЙСЯ ПРОЦЕСС РАЗМНОЖЕНИЯ

Решение $\{P_n(t)\}$ бесконечной системы дифференциальных уравнений (3.2), удовлетворяющее начальным условиям (3.3), может быть найдено индуктивно, начиная с $P_i(t)=e^{-\lambda_i t}$. Поэтому распределение $\{P_n(t)\}$ определено единственным образом. Из известных формул для решения линейных дифференциальных уравнений следует также, что $P_n(t)\geqslant 0$. Остается открытым лишь вопрос о том, является ли $\{P_n(t)\}$ настоящим распределением вероятностей, т. е. будет ли выполняться условие

$$\sum P_n(t)=1 \tag{4.1}$$

при всех t. Мы увидим, что это не всегда так: если коэффициенты λ_n быстро возрастают, то может случиться, что

$$\sum P_n(t)<1. \tag{4.2}$$

Когда эта возможность была обнаружена, она показалась настораживающей, однако ее легко объяснить. Левую часть (4.2) можно интерпретировать как вероятность того, что на протяжении интервала времени величины t имело место лишь *конечное число* скачков. Следовательно, разность правой и левой частей (4.2) отвечает возможности бесконечного числа скачков, или своего рода взрыва. Чтобы лучше понять это явление, сравним нашу вероятностную модель роста с обычным детерминистским подходом.

Величина λ_n в (3.2) может быть названа средней скоростью роста совокупности объема n. Например, в частном случае (3.4) имеем $\lambda_n=n\lambda$, так что здесь средняя скорость роста пропорциональна фактическому размеру популяции. Если рост не подвержен случайным флуктуациям и увеличивается со скоростью, пропорциональной

---

[1]) Furry W. H., On fluctuation phenomena in the passage of high-energy electrons through lead, Physical Reviews, 52 (1937), 569.

* Этот параграф посвящен специальному вопросу и может быть опущен.

мгновенному размеру популяции $x(t)$, то последний меняется в соответствии с детерминистским дифференциальным уравнением

$$dx(t)/dt=\lambda x(t), \tag{4.3}$$

из которого следует, что

$$x(t)=ie^{\lambda t}, \tag{4.4}$$

где $i=x(0)$ есть начальный объем совокупности. Легко видеть, что математическое ожидание $\sum nP_n(t)$ распределения (3.5) совпадает с $x(t)$, так что $x(t)$ описывает не только детерминированный процесс роста, но также и средний объем совокупности в примере 3, б).

Рассмотрим теперь детерминированный процесс роста, в котором скорость роста увеличивается быстрее объема совокупности. Скорости роста, пропорциональной $x^2(t)$, соответствует дифференциальное уравнение

$$dx(t)/dt=\lambda x^2(t), \tag{4.5}$$

решение которого имеет вид

$$x(t)=i/(1-\lambda it). \tag{4.6}$$

Заметим, что $x(t)$ неограниченно возрастает при $t\to 1/(\lambda i)$. Иначе говоря, из предположения о том, что скорость роста увеличивается как квадрат объема совокупности, вытекает бесконечный рост за конечный интервал времени. Аналогично, если λ_n в (3.4) возрастают слишком быстро, то с конечной вероятностью внутри конечного интервала времени имеет место бесконечное число изменений. Точный ответ о том, каковы должны быть условия, при которых происходит такой неограниченный рост, дает следующая теорема.

Теорема. *Для того чтобы $\sum P_n(t)=1$ при всех t, необходимо и достаточно, чтобы ряд $\sum\lambda_n^{-1}$ расходился* [1]).

Доказательство. Положим

$$S_k(t)=P_0(t)+\ldots+P_k(t). \tag{4.7}$$

В силу очевидной монотонности существует предел

$$\mu(t)=\lim_{k\to\infty}[1-S_k(t)]. \tag{4.8}$$

Суммируя дифференциальные уравнения (3.2) по $n=0, \ldots, k$, мы получаем

$$S_k'(t)=-\lambda_k P_k(t). \tag{4.9}$$

---

[1]) Нетрудно видеть, что неравенство $\sum P_n(t)<1$ либо выполняется для всех $t>0$, либо не выполняется ни для одного $t>0$; см. задачу 22.

С учетом начальных условий (3.3) отсюда вытекает, что при $k \geqslant i$

$$1-S_k(t)=\lambda_k \int_0^t P_k(\tau)\,d\tau. \qquad (4.10)$$

В силу (4.8) величина в левой части лежит между μ и 1; следовательно,

$$\lambda_k^{-1}\mu(t) \leqslant \int_0^t P_k(s)\,ds \leqslant \lambda_k^{-1}. \qquad (4.11)$$

Суммируя по $k=i, \ldots, n$, мы получаем для $n \geqslant i$

$$\mu(t)\left[\lambda_i^{-1}+\ldots+\lambda_n^{-1}\right] \leqslant \int_0^t S_n(s)\,ds \leqslant \lambda_i^{-1}+\ldots+\lambda_n^{-1}. \qquad (4.12)$$

Если $\sum \lambda_n^{-1} < \infty$, то величина в самой правой части неравенства остается ограниченной при $n \to \infty$, и поэтому невозможно, чтобы подынтегральное выражение стремилось к 1 при всех s. Наоборот, если $\sum \lambda_n^{-1} = \infty$, то из первого неравенства мы заключаем, что $\mu(t)=0$ при всех t, и в силу (4.8) отсюда вытекает, что $S_n(t) \to 1$, что и утверждалось. ►

При вероятностной интерпретации этот критерий представляется весьма разумным. Система проводит некоторое время в начальном состоянии E_0, переходит из него в E_1, остается там на время, переходит в E_2 и т. д. Вероятность $P_0(t)$ того, что время пребывания в E_0 превосходит t, получается из (3.2) равной $P_0(t)=e^{-\lambda_0 t}$. Это время пребывания T_0 является случайной величиной, однако область ее значений есть положительная ось t, и поэтому эта величина формально не укладывается в рамки этой книги. Однако, поскольку переход от геометрического распределения к показательному тривиален, мы можем без особого вреда для себя немного отойти от строгого изложения. Приближение T_0 дискретной случайной величиной с геометрическим распределением показывает, что среднее время пребывания в E_0 естественно определить, как

$$\mathbf{E}(\mathrm{T}_0)=\int_0^\infty te^{-\lambda_0 t}\lambda_0\,dt=\lambda_0^{-1}. \qquad (4.13)$$

В тот момент времени, когда система попадает в E_j, состояние E_j берет на себя роль начального состояния, и поэтому такое же заключение применимо и ко времени пребывания T_j в E_j: *среднее время пребывания в E_j равно* $\mathbf{E}(\mathrm{T}_j)=\lambda_j^{-1}$. Отсюда следует, что $\lambda_0^{-1}+\lambda_1^{-1}+\ldots+\lambda_n^{-1}$ есть средняя продолжительность времени, требующегося системе для того, чтобы пройти через $E_0, E_1, \ldots, E_n$, и мы можем переформулировать критерий из § 4 следующим образом.

Для того чтобы $\sum P_n(t)=1$ для всех t, необходимо и достаточно, чтобы

$$\sum \mathbf{E}(\mathrm{T}_j)=\sum \lambda_j^{-1}=\infty; \qquad (4.14)$$

т. е. суммарная средняя продолжительность времени, проведенного в E_0, E_1, E_2, ..., должна быть бесконечной. Разумеется, $L_0(t)=1-\sum P_n(t)$ есть вероятность того, что система прошла через все состояния до момента времени t, При такой интерпретации возможность неравенства (4.2) становится понятной. Если среднее время пребывания в E_j равно 2^{-j}, то вероятность того, что система пройдет через все состояния за время $1+2^{-1}+2^{-2}+\ldots=2$, должна быть положительной. Аналогично частица, двигающаяся вдоль оси x с экспоненциально возрастающей скоростью, пройдет всю ось за конечное время.

(Мы вернемся к расходящемуся процессу размножения в примере 9, б).)

§ 5. ПРОЦЕСС РАЗМНОЖЕНИЯ И ГИБЕЛИ

Процесс чистого размножения из § 3 дает удовлетворительное описание радиоактивных превращений, однако он не может служить реалистической моделью для изменений объема совокупности, члены которой могут умирать (или исчезать). Это наводит на мысль об обобщении нашей модели, допускающем переходы из E_n не только в ближайшее сверху состояние E_{n+1}, но и в ближайшее снизу состояние E_{n-1}. (Процессы более общего вида будут определены в § 9.)

Таким образом, мы начнем со следующих постулатов.

Постулаты. *Изменения системы осуществляются только путем переходов из состояний в ближайшие к ним соседние состояния (из E_n в E_{n+1} или в E_{n-1} при $n\geqslant 1$, а из E_0 — только в E_1). Если в момент времени t система находится в состоянии E_n, то вероятность того, что между t и $t+h$ произойдет переход $E_n \to E_{n+1}$, равна $\lambda_n h+o(h)$, а вероятность перехода $E_n \to E_{n-1}$ (если $n\geqslant 1$) равна $\mu_n h+o(h)$. Вероятность более чем одного изменения на протяжении $(t, t+h)$ есть $o(h)$.*

Легко приспособить метод § 2 для вывода дифференциальных уравнений для вероятностей $P_n(t)$ того, что система находится в состоянии E_n. Чтобы вычислить $P_n(t+h)$, заметим, что состояние E_n в момент $t+h$ возможно лишь при выполнении одного из следующих условий: 1) в момент t система находится в E_n и между t и $t+h$ не происходит никаких изменений; 2) в момент t система находится в E_{n-1} и происходит переход в E_n; 3) в момент t система находится в E_{n+1} и происходит переход в E_n; 4) между t и $t+h$ происходит два или несколько переходов. По предположению вероятность последнего события есть $o(h)$. Первые три возможности взаимно исключаются, и их вероятности складываются; поэтому

$$P_n(t+h)=P_n(t)\{1-\lambda_n h-\mu_n h\}+\\ +\lambda_{n-1}hP_{n-1}(t)+\mu_{n+1}hP_{n+1}(t)+o(h). \quad (5.1)$$

Перенося член $P_n(t)$ в левую часть и деля обе части уравнения на h, получаем в левой части разностное отношение для $P_n(t)$, и в пределе при $h\to 0$ приходим к

$$P_n'(t)=-(\lambda_n+\mu_n)P_n(t)+\lambda_{n-1}P_{n-1}(t)+\mu_{n+1}P_{n+1}(t). \quad (5.2)$$

Это уравнение справедливо при всех $n \geqslant 1$. Для $n=0$ аналогичным образом получаем

$$P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t). \tag{5.3}$$

Если начальным состоянием является E_i, то ***начальными условиями*** будут

$$P_i(0)=1, \quad P_n(0)=0 \ \text{при} \ n \neq i. \tag{5.4}$$

Таким образом, мы видим, что процесс размножения и гибели зависит от бесконечной системы дифференциальных уравнений (5.2)—(5.3) с начальными условиями (5.4). Вопрос о существовании и единственности решения в этом случае отнюдь не тривиален. В процессе чистого размножения система дифференциальных уравнений (3.2) также была бесконечной, но она имела вид рекуррентных соотношений: $P_0(t)$ определялась первым уравнением, а $P_n(t)$ могла быть вычислена по $P_{n-1}(t)$. Новая система (5.2) имеет иной вид, и все $P_n(t)$ должны находиться одновременно. Здесь (как и в некоторых других случаях в этой главе) мы сформулируем свойства решений без доказательства [1]).

Для произвольных заданных коэффициентов $\lambda_n \geqslant 0$, $\mu_n \geqslant 0$ *всегда существует положительное решение* $\{P_n(t)\}$ *системы* (5.2)—(5.4), *такое, что* $\sum P_n(t) \leqslant 1$. *Если коэффициенты ограничены (или возрастают достаточно медленно), то это решение единственно и удовлетворяет условию регулярности* $\sum P_n(t) = 1$. Однако можно выбрать коэффициенты таким образом, что $\sum P_n(t) < 1$ и будет существовать бесконечно много решений. В последнем случае мы сталкиваемся с явлением, аналогичным изучавшемуся в предыдущем параграфе для процесса чистого размножения. Эта ситуация представляет значительный теоретический интерес, однако читатель может без опасения считать, что во всех практически интересных случаях условия единственности выполнены; в этом случае автоматически $\sum P_n(t) = 1$ (см. § 9).

При $\lambda_0 = 0$ переход $E_0 \to E_1$ невозможен. В терминологии цепей Маркова E_0 является *поглощающим состоянием*, выход из которого невозможен; коль скоро система окажется в E_0, она останется там навсегда. Из (5.3) следует, что в этом случае $P_0'(t) \geqslant 0$, так что $P_0(t)$ монотонно возрастает. Предел ее $P_0(\infty)$ есть вероятность *окончательного поглощения*.

[1]) Простейшее доказательство существования и критерий единственности получается как частный случай из общей теории, развитой автором (см. § 9). Недавно привлекли широкое внимание такие решения уравнений для процесса размножения и гибели, что $\sum P_n(t) < 1$. По этому вопросу см. Lederman W., Reuter G. E., Spectral theory for the differential equations of simple birth and death processes, Philos. Trans. Roy. Soc., London, Ser. A, 246 (1954), 387—391; Karlin S., McGregor J. L., The differential equations of birth-and-death processes and the Stieltjes moment problem, Trans. Amer. Math. Soc., 85 (1957), 489—546; Karlin S., McGregor J. L., The classification of birth and death processes, ibid, 86 (1957), 366—400. См. также Feller W., The birth and death processes as diffusion processes, Journal de Mathématiques Pures et Appliquées, 38(1959), 301—345.

Можно показать (либо используя явный вид решений, либо из общих эргодических теорем для марковских процессов), что в любом случае *пределы*

$$\lim_{t \to \infty} P_n(t) = p_n \tag{5.5}$$

существуют и не зависят от начальных условий (5.4); они удовлетворяют системе линейных уравнений, которая получается из (5.2)—(5.3) при замене производных в левой части нулями.

Соотношения (5.5) напоминают предельные теоремы, выведенные в гл. XV, 7 для обычных цепей Маркова, и это не просто формальное сходство. Эти соотношения становятся интуитивно почти очевидными при сравнении нашего процесса с простой цепью Маркова с переходными вероятностями

$$p_{n,\,n+1} = \lambda_n/(\lambda_n+\mu_n), \qquad p_{n,\,n-1} = \mu_n/(\lambda_n+\mu_n). \tag{5.6}$$

В этой цепи единственными возможными переходами являются $E_n \to E_{n+1}$ и $E_n \to E_{n-1}$, и они имеют те же условные вероятности, что и в нашем процессе; разница между этой цепью и нашим процессом заключается в том, что в последнем изменения могут происходить в произвольные моменты времени, так что число переходов в интервале времени длины t является случайной величиной. Однако при больших t это число заведомо будет большим, и поэтому весьма правдоподобно, что при $t \to \infty$ вероятности $P_n(t)$ будут вести себя так же, как соответствующие вероятности для простой цепи.

Если простая цепь с переходными вероятностями (5.6) невозвратна, то при всех n мы имеем $p_n = 0$; если эта цепь эргодична, то p_n определяют стационарное распределение вероятностей. В этом случае (5.5) обычно интерпретируется как «тенденция к устойчивому положению», и это двусмысленное наименование вызвало много путаницы. Следует понимать, что, за исключением того случая, когда E_0 есть поглощающее состояние, случайные флуктуации продолжаются вечно, не ослабевая, и (5.5) показывает лишь, что в конце концов влияние начальных условий исчезает. Сделанные в гл. XV, 7 замечания о статистическом равновесии применимы здесь без изменений.

Главной областью приложений процессов размножения и гибели являются задачи о времени ожидания, задачи о телефонных линиях и т. д.; см. § 6 и 7.

Примеры. а) *Линейный рост.* Предположим, что элементы совокупности могут делиться или умирать. Для любого живого элемента вероятность разделиться на два на протяжении любого короткого интервала времени длины h равна $\lambda h + o(h)$, тогда как соответствующая вероятность гибели равна $\mu h + o(h)$. Здесь λ и μ — две постоянные, характеризующие совокупность. Если между ее элементами нет никакого взаимодействия, то мы приходим к процессу размножения и гибели с $\lambda_n = n\lambda$, $\mu_n = n\mu$. Основные дифференциальные уравнения принимают вид

$$P_0'(t) = \mu P_1(t),$$
$$P_n'(t) = -(\lambda+\mu)\, nP_n(t) + \lambda(n-1)\, P_{n-1}(t) + \mu(n+1)\, P_{n+1}(t). \tag{5.7}$$

Можно найти явные решения[1]) (см. задачи 11—14), однако мы не будем обсуждать этот аспект. Пределы (5.5) здесь существуют и удовлетворяют (5.7) при $P_n'(t) = 0$. Из первого уравнения находим $p_1 = 0$, а из второго уравнения по индукции находим, что $p_n = 0$ при всех $n \geqslant 1$. Если $p_0 = 1$, то мы можем сказать, что вероятность окончательного вымирания равна 1. Если $p_0 < 1$, то из соотношений $p_1 = p_2 = \ldots = 0$ вытекает, что с вероятностью $1 - p_0$ совокупность будет безгранично расти; в конце концов она либо вымрет, либо будет бесконечно возрастать. Чтобы найти вероятность p_0 вымирания, сравним наш процесс с соответствующей цепью Маркова. В нашем случае переходные вероятности (5.6) не зависят от n, и стало быть, мы имеем обычное случайное блуждание, в котором шаги направо и налево имеют вероятности $p = \lambda/(\lambda+\mu)$ и $q = \mu/(\lambda+\mu)$ соответственно. Состояние E_0 является поглощающим. Из классической задачи о разорении (см. гл. XIV, 2) мы знаем, что вероятность вымирания равна 1 при $p \leqslant q$ и равна $(q/p)^i$, если $q < p$ и i есть начальное состояние. Отсюда мы заключаем, что *в нашем процессе вероятность* $p_0 = \lim P_0(t)$ *окончательного вымирания равна* 1 *при* $\lambda \leqslant \mu$ *и равна* $(\mu/\lambda)^i$ *при* $\lambda > \mu$. (Это легко проверить при помощи явного решения; см. задачи 11—14.)

Как и во многих подобных случаях, явный вид решения системы (5.7) довольно сложен, и поэтому хорошо было бы найти среднее и дисперсию распределения $\{P_n(t)\}$ непосредственно из дифференциальных уравнений. Для среднего имеем

$$M(t) = \sum_{n=1}^{\infty} nP_n(t). \tag{5.8}$$

Мы опустим формальное доказательство того, что $M(t)$ конечно, а следующие формальные операции обоснованы (и то, и другое легко следует из вида решения, приведенного в задаче 12). Умножая второе из уравнений (5.7) на n и суммируя по $n = 1, 2, \ldots$, мы обнаруживаем, что члены, содержащие n^2, взаимно уничтожаются,

---

[1]) Систематический подход здесь состоит в выводе уравнения в частных производных для производящей функции $\sum P_n(t)\, s^n$. Более общий процесс, в котором допускается зависимость коэффициентов λ и μ в (5.7) от времени, детально обсуждается в работе Kendall D. G., The generalized "birth and death" process, Ann. Math. Statist., 19 (1948), 1—15. См. также статью того же автора Stochastic processes and population growth, Journal of the Royal Statistical Society, B, 11 (1949), 230—265, где теория обобщается так, что она учитывает распределение возраста элементов в биологических популяциях.

и получаем

$$M'(t)=\lambda\sum(n-1)P_{n-1}(t)-\mu\sum(n+1)P_{n+1}(t)=$$
$$=(\lambda-\mu)M(t). \tag{5.9}$$

Это дифференциальное уравнение для $M(t)$. Начальный объем совокупности равен i, и поэтому $M(0)=i$. Стало быть,

$$M(t)=ie^{(\lambda-\mu)t}. \tag{5.10}$$

Мы видим, что среднее стремится к 0 при $\lambda<\mu$ и к бесконечности при $\lambda>\mu$. Дисперсию распределения $\{P_n(t)\}$ можно вычислить аналогичным образом (см. задачу 14).

б) *Очередь в случае одного канала*. В простейшем случае постоянных коэффициентов $\lambda_n=\lambda$ и $\mu_n=\mu$ процесс размножения и гибели сводится к частному случаю примера 7, б) (задачи об очередях) при $a=1$.

§ 6. ПОКАЗАТЕЛЬНЫЕ ВРЕМЕНА ОБСЛУЖИВАНИЯ [1]

Основная область приложения процессов размножения и гибели связана с расчетом числа телефонных линий [2] и различными типами очередей к телефонам, прилавкам или машинам. Этот тип задач может изучаться на различных уровнях математической абстракции. Метод процесса размножения и гибели является простейшим подходом, однако эта модель основана на математическом упрощении, известном как *предположение о показательных временах обслуживания*. Мы начнем с обсуждения этого основного предположения.

Рассмотрим для конкретности телефонный разговор и предположим, что его продолжительность обязательно равна целому числу секунд. Мы рассматриваем продолжительность разговора как случайную величину **X** и считаем известным ее распределение вероятностей $p_n=\mathbf{P}\{\mathbf{X}=n\}$. Тогда телефонная линия представляет собой физическую систему с двумя возможными состояниями «занято» (E_0) и «свободно» (E_1). Когда линия занята, вероятность изменения состояния в течение следующей секунды зависит от того, как долго уже идет разговор. Иначе говоря, прошлое влияет на будущее, и поэтому наш процесс не является марковским (см. гл. XV, 13). Это обстоятельство является источником трудностей, однако, к счастью, существует простое исключение, подробно обсуждавшееся в гл. XIII, 9.

Представим себе, что решение о том, продолжать или нет разговор, принимается каждую секунду бросанием несимметричной монеты. Иначе говоря, со скоростью одно испытание в секунду производится последовательность испытаний Бернулли, которая продол-

---

[1] В оригинале exponential holding times.— *Прим. перев.*

[2] В оригинале trunking in telephone engineering.— *Прим. перев.*

жается до первого успеха. Когда этот первый успех произойдет, разговор окончится. В этом случае общая продолжительность разговора, или «время обслуживания», имеет геометрическое распределение $p_n = q^{n-1}p$. Когда линия занята, вероятность того, что она останется занятой более одной секунды, равна q, а вероятность перехода $E_0 \to E_1$ на следующем шаге равна p. Теперь эти вероятности не зависят от того, как долго была занята линия.

Когда использование дискретного временнóго параметра оказывается нежелательным, приходится работать с непрерывными случайными величинами. Тогда роль геометрического распределения для времен ожидания берет на себя *показательное распределение*. Это единственное распределение, имеющее марковский характер, т. е. наделенное полной потерей памяти. Иначе говоря, вероятность того, что разговор, происходящий в момент времени x, продлится до $x+h$, не зависит от продолжительности предыдущего разговора тогда и только тогда, когда вероятность того, что разговор продлится более t единиц времени, равна $e^{-\lambda t}$. Мы уже встречались с этим «показательным временем обслуживания» как с нулевым членом распределения Пуассона (2.4), т. е. как с временем ожидания первого изменения.

Метод процесса размножения и гибели применим только тогда, когда изучаемые переходные вероятности не зависят от прошлого; для задач о числе занятых линий и об очередях это означает, что все времена обслуживания должны быть показательными. С практической точки зрения это предположение может на первый взгляд показаться довольно искусственным, однако опыт показывает, что оно неплохо описывает реальные явления. В частности, многочисленные измерения показали, что телефонные разговоры внутри города [1]) следуют показательному закону с поразительной степенью точности. Та же ситуация превалирует и в случае других времен обслуживания (например, продолжительности ремонта машин).

Остается охарактеризовать так называемый входной поток (поступающие вызовы, поломки станков и т. д.). Мы предположим, что вероятность поступления вызова в течение любого интервала времени длины h равна λh плюс пренебрежимо малые члены и что вероятностью более чем одного вызова в этом интервале в пределе можно пренебречь. Согласно результатам § 2, это означает, что число поступивших [до момента времени t.— *Перев.*] вызовов имеет распределение Пуассона со средним λt. Мы будем описывать эту ситуацию, говоря, что *входной поток имеет пуассоновский тип с интенсивностью* λ.

Легко проверить описанное свойство показательных времен обслуживания. Обозначим через $u(t)$ вероятность того, что разговор продлится не менее t единиц времени. Вероятность $u(t+s)$ того, что разговор, начавшийся в момент 0, окон-

[1]) Для междугородных разговоров единицей измерения обычно служат три минуты, и поэтому времена обслуживания часто являются кратными трем минутам. При таких обстоятельствах показательное распределение неприменимо.

чится после момента $t+s$, равна вероятности того, что он продлится более t единиц времени, умноженной на условную вероятность того, что разговор продлится еще s дополнительных единиц времени при условии, что его длина превышает t. Если продолжительность предыдущего разговора не оказывает никакого влияния, то последняя условная вероятность должна равняться $u(s)$, т. е. мы должны иметь

$$u(t+s)=u(t)\,u(s). \tag{6.1}$$

Для доказательства сформулированного выше характеристического свойства показательных времен обслуживания было бы достаточно показать, что *монотонные* решения этого функционального уравнения с необходимостью имеют вид $e^{-\lambda t}$. Мы докажем несколько более сильный результат, представляющий самостоятельный интерес [1]).

Теорема. *Пусть u — решение уравнения* (6.1), *определенное при $t>0$ и ограниченное на некотором интервале. Тогда либо $u(t)=0$ при всех t, либо $u(t)=e^{-\lambda t}$, где λ — некоторая постоянная.*

Доказательство. Ясно, что

$$u(a)=u^2(a/2). \tag{6.2}$$

Предположим сперва, что $u(a)=0$ для некоторого значения a. Из (6.2) мы заключаем по индукции, что $u(2^{-n}a)=0$ для всех целых n, а из (6.1) ясно, что из $u(s)=0$ вытекает равенство $u(t)=0$ при всех $t>s$. Поэтому из $u(a)=0$ следует, что u тождественно обращается в нуль. Поскольку (6.2) очевидным образом исключает отрицательные значения u, то остается рассмотреть лишь строго положительные решения уравнения (6.1).

Положим $e^{-\lambda}=u(1)$ и $v(t)=e^{\lambda t}u(t)$. Тогда

$$v(t+s)=v(t)\,v(s) \quad и \quad v(1)=1. \tag{6.3}$$

Мы должны доказать, что отсюда вытекает равенство $v(t)=1$ при всех t. Очевидно, что для произвольных положительных целых чисел m и n

$$v(m/n)=v^m(1/n)=\sqrt[n]{v^m(1)}=1, \tag{6.4}$$

и поэтому $v(s)=1$ для всех рациональных s. Кроме того, если $v(a)=c$, то $v(na)=c^n$ для любого положительного или отрицательного целого числа n. Отсюда следует, что если v принимает некоторое значение $c\neq 1$, то оно принимает также произвольно большие значения. Однако из (6.3) при $t+s=\tau$ видно, что $v(\tau-s)=v(\tau)$ для всех рациональных s. Следовательно, если значение A принимается в некоторой точке τ, то то же самое значение принимается на каждом сколь угодно малом интервале. Стало быть, ограниченность u на каком-либо заданном интервале не допускает значений v, отличных от единицы. ►

§ 7. ОЧЕРЕДИ И ЗАДАЧИ ОБСЛУЖИВАНИЯ

а) *Простейшая задача о телефонных линиях* [2]). Предположим, что имеется бесконечно много линий или каналов и что вероятность окончания разговора между моментами времени t и $t+h$ равна $\mu h+$

---

[1]) Уравнение (6.1) является лишь логарифмическим вариантом записи известного уравнения Гамеля $f(t+s)=f(t)+f(s)$. Мы доказываем, что его решения либо имеют вид at, либо неограниченны на *каждом* интервале. (Известно, что ни одно такое решение не является функцией Бэра, т. е. ни одно такое решение не может быть получено при помощи разложения в ряд или какого-либо иного предельного процесса, отправляющегося от непрерывных функций.)

[2]) Palm C., Intensitätsschwankungen in Fernsprechverkehr, Ericsson Technics (Stockholm), 44 (1943), 1—189, в частности, с. 57. Очереди и задачи обслуживания

$+o(h)$ (показательное время обслуживания). Поступающие вызовы образуют поток пуассоновского типа с параметром λ. Система находится в состоянии E_n, если занято n линий.

Предполагается, конечно, что продолжительности разговоров взаимно независимы. Если занято n линий, то вероятность того, что одна из них освободится в течение времени h, равна $n\mu h+o(h)$. Вероятность прекращения за это время двух или нескольких разговоров является, очевидно, величиной порядка h^2, и поэтому ей можно пренебречь. Вероятность поступления нового вызова равна $\lambda h+ +o(h)$. Вероятность нескольких вызовов или поступления вызова и окончания разговора тоже есть $o(h)$. Таким образом, в обозначения: § 5

$$\lambda_n=\lambda, \quad \mu_n=n\mu. \tag{7.1}$$

Основные дифференциальные уравнения (5.2) — (5.3) принимают вид

$$\begin{aligned} P_0'(t) &= -\lambda P_0(t)+\mu P_1(t), \\ P_n'(t) &= -(\lambda+n\mu)P_n(t)+\lambda P_{n-1}(t)+(n+1)\mu P_{n+1}(t) \end{aligned} \tag{7.2}$$

при $n\geqslant 1$. Явные решения могут быть получены путем вывода дифференциального уравнения в частных производных для производящей функции (см. задачу 15). Мы определим лишь величины $p_n= =\lim P_n(t)$ из (5.5). Они удовлетворяют уравнениям

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda+n\mu)\,p_n &= \lambda p_{n-1}+(n+1)\,\mu p_{n+1}. \end{aligned} \tag{7.3}$$

По индукции находим $p_n=p_0(\lambda/\mu)^n/n!$, и отсюда

$$p_n=e^{-\lambda/\mu}(\lambda/\mu)^n/n!. \tag{7.4}$$

Таким образом, *предельное распределение является распределением Пуассона с параметром λ/μ. Оно не зависит от начального состояния.*

Легко найти среднее $M(t)=\sum nP_n(t)$. Умножая n-е уравнение (7.2) на n, суммируя и учитывая, что $P_n(t)$ в сумме дают единицу, мы получаем

$$M'(t)=\lambda-\mu M(t). \tag{7.5}$$

Если начальным является состояние E_i, то $M(0)=i$ и

$$M(t)=(\lambda/\mu)(1-e^{-\mu t})+ie^{-\mu t}. \tag{7.6}$$

Читатель может проверить, что в частном случае $i=0$ величины $P_n(t)$ даются в точности распределением Пуассона со средним $M(t)$.

для телефонных станций изучались задолго до появления теории стохастических процессов и оказали стимулирующее влияние на развитие этой теории. В частности, замечательная работа Пальма оказалась полезной на протяжении многих лет. Самые ранние работы в этой области принадлежат А. К. Эрлангу (1878—1929); см. Brockmeyer E., Halström H. L., Jensen Arne, The life and works of A. K. Erlang, Transactions of the Danish Academy Technical Sciences, No. 2, Copenhagen, 1948. Ценная основополагающая работа была независимо проделана Фраем, книга которого (Fry T. C., Probability and its engineering uses, New York, Van Nostrand, 1928) много сделала для развития приложений теории вероятностей в технике.

б) *Очереди в случае конечного числа каналов* [1]). Видоизменим теперь последний пример, чтобы получить более реалистичную модель. Предположения здесь те же самые, за исключением того, что *число a линий или каналов конечно. Если все a каналов заняты, то каждый новый вызов встает в очередь и ожидает до тех пор, пока не освободится какой-либо канал.* Это означает, что все линии имеют *общую* очередь.

Слово «линия» может быть заменено на *стойку* в почтовом отделении, а «разговор» — на *обслуживание*. Практически мы рассматриваем общую задачу об очереди в предположении, что клиент должен ожидать лишь тогда, когда все a каналов заняты.

Мы говорим, что *система находится в состоянии E_n, если ровно n клиентов либо обслуживаются, либо стоят в очереди.* Такая очередь существует только при $n>a$, и тогда в ней стоят $n-a$ клиентов.

До тех пор пока хотя бы один канал свободен, ситуация здесь точно такая же, как и в предыдущем примере. Однако если система находится в состоянии E_n при $n>a$, то идет только a разговоров, и поэтому $\mu_n=a\mu$ при $n\geqslant a$. Стало быть, основная система дифференциальных уравнений состоит из уравнений (7.2) для $n<a$ и из уравнений

$$P'_n(t)=-(\lambda+a\mu)P_n(t)+\lambda P_{n-1}(t)+a\mu P_{n+1}(t) \tag{7.7}$$

для $n\geqslant a$.

В частном случае одного канала ($a=1$) эти уравнения сводятся к уравнениям для процесса размножения и гибели с коэффициентами, не зависящими от n.

Пределы $p_n=\lim P_n(t)$ удовлетворяют (7.3) при $n<a$ и

$$(\lambda+a\mu)p_n=\lambda p_{n-1}+a\mu p_{n+1} \tag{7.8}$$

при $n\geqslant a$. Последовательно находим

$$p_n=p_0\,(\lambda/\mu)^n/n!, \qquad n\leqslant a, \tag{7.9}$$

$$p_n=p_0\,(\lambda/\mu)^n/(a!a^{n-a}), \qquad n\geqslant a. \tag{7.10}$$

Ряд $\sum(p_n/p_0)$ сходится только тогда, когда

$$\lambda/\mu<a. \tag{7.11}$$

Следовательно, предельное распределение $\{p_k\}$ не может существовать при $\lambda\geqslant a\mu$. *В этом случае $p_n=0$ для всех n, и это означает, что очередь будет безгранично возрастать.* С другой стороны, если выполнено (7.11), то мы можем определить p_0 так, чтобы $\sum p_n=1$. С помощью явных выражений для $P_n(t)$ можно показать, что получаемые таким образом p_n действительно представляют собой *предельное распределение* для $P_n(t)$. Табл. 1 дает численную иллюстрацию для $a=3$, $\lambda/\mu=2$.

---

[1]) Колмогоров А. Н. О проблеме ожидания.— Математический сборник, 1931, 38 : 1—2, 101—106. Относительно аналогичных процессов см. задачи 6—8 и 20.

Таблица 1

Предельные вероятности в случае $a=3$ каналов и $\lambda/\mu=2$

n	0	1	2	3	4	5	6	7
Число занятых каналов	0	1	2	3	3	3	3	3
Число ожидающих клиентов	0	0	0	0	1	2	3	4
p_n	0,1111	0,2222	0,2222	0,1481	0,09888	0,0658	0,0439	0,0293

в) *Обслуживание станков* [1]). Для ориентировки мы начнем с простейшего случая и обобщим его в следующем примере. Задача состоит в следующем.

Рассмотрим станки-автоматы, которые обычно не требуют ухода за исключением того, что они могут сломаться и потребовать обслуживания. Время, требующееся для обслуживания станка, вновь считается случайной величиной с показательным распределением. Иначе говоря, станок характеризуется двумя постоянными λ и μ со следующими свойствами. Если в момент времени t станок находится в рабочем состоянии, то вероятность того, что он потребует обслуживания до момента $t+h$, равняется λh плюс члены, пренебрежимые в пределе при $h\to 0$. Наоборот, когда станок находится на обслуживании, вероятность того, что время обслуживания окончится до момента $t+h$ и станок вернется в рабочее состояние, равняется $\mu h+o(h)$. Для производительного станка λ должно быть сравнительно малым, а μ — сравнительно большим. Отношение λ/μ называется *коэффициентом обслуживания*.

Мы предполагаем, что *m независимо работающих станков с одними и теми же параметрами λ и μ обслуживаются одним рабочим*. Сломавшийся станок обслуживается немедленно, если рабочий не занят обслуживанием другого станка, а в последнем случае образуется очередь. Мы говорим, что *система находится в состоянии E_n, если не работает n станков*. При $1\leqslant n\leqslant m$ это означает, что один станок находится на обслуживании, а $n-1$ — в очереди; в состоянии E_0 все станки работают, и рабочий не занят.

Переход $E_n\to E_{n+1}$ вызывается поломкой одного из $m-n$ работающих станков, тогда как переход $E_n\to E_{n-1}$ происходит тогда, когда обслуживаемый станок возвращается в рабочее состояние. Следовательно, мы имеем процесс размножения и гибели с коэффи-

[1]) Примеры в) и г), включающие численные иллюстрации, взяты из статьи Пальма «Распределение числа рабочих при обслуживании автоматов» (на шведском языке) из Industritidningen Norden, 75 (1947), 75—80, 90—94, 119—123. Пальм приводит таблицы и графики для наиболее экономичного числа рабочих.

циентами

$$\lambda_n = (m-n)\lambda, \quad \mu_0 = 0, \quad \mu_1 = \mu_2 = \ldots \mu_m = \mu. \tag{7.12}$$

При $1 \leqslant n \leqslant m-1$ основные дифференциальные уравнения (5.2) принимают вид

$$P_n'(t) = -\{(m-n)\lambda + \mu\} P_n(t) + \\ + (m-n+1)\lambda P_{n-1}(t) + \mu P_{n+1}(t), \tag{7.13}$$

в то время как для крайних состояний $n=0$ и $n=m$

$$\begin{aligned} P_0'(t) &= -m\lambda P_0(t) + \mu P_1(t), \\ P_m'(t) &= -\mu P_m(t) + \lambda P_{m-1}(t). \end{aligned} \tag{7.13a}$$

Мы пришли к конечной системе дифференциальных уравнений, которая может быть решена обычными методами. Пределы $p_n = \lim P_n(t)$ определяются из уравнений

$$\begin{aligned} m\lambda p_0 &= \mu p_1, \\ \{(m-n)\lambda + \mu\} p_n &= (m-n+1)\lambda p_{n-1} + \mu p_{n+1}, \\ \mu p_m &= \lambda p_{m-1}, \end{aligned} \tag{7.14}$$

которые можно привести к рекуррентному виду:

$$(m-n)\lambda p_n = \mu p_{n+1}. \tag{7.15}$$

Подставляя последовательно $n=m-1,\ m-2,\ \ldots,\ 1,\ 0$, получаем

$$p_{m-k} = (1/k!)(\mu/\lambda)^k p_m.$$

Оставшаяся неизвестной постоянная p_m может быть найдена из того условия, что p_j в сумме дают единицу. Результат известен как *формула Эрланга*:

$$p_m = \{1 + (1/1!)(\mu/\lambda)^1 + \ldots + (1/m!)(\mu/\lambda)^m\}^{-1}. \tag{7.16}$$

Типичные численные значения приведены в табл. 2.

Таблица 2

Формула Эрланга
Вероятности p_n для случая
$\lambda/\mu = 0{,}1,\ m = 6$

n	Число станков в очереди	p_n	n	Число станков в очереди	p_n
0	0	0,4845	4	3	0,0175
1	0	0,2907	5	4	0,0035
2	1	0,1454	6	5	0,0003
3	2	0,0582			

Вероятность p_0 можно интерпретировать как вероятность того, что рабочий будет свободен (в примере из табл. 2 он должен быть свободен примерно половину всего времени). *Среднее число станков в очереди равно*

$$w=\sum_{k=1}^{m}(k-1)\,p_k=\sum_{k=1}^{m}kp_k-(1-p_0). \tag{7.17}$$

Эту величину можно вычислить суммированием соотношений (7.15) при $n=0, 1, \ldots, m$. Используя тот факт, что p_n в сумме составляют единицу, мы получаем

$$m\lambda-\lambda w-\lambda(1-p_0)=\mu(1-p_0),$$

или

$$w=m-[(\lambda+\mu)/\lambda](1-p_0). \tag{7.18}$$

В примере из табл. 2 мы имеем $w=6\cdot 0{,}0549$. Таким образом, 0,0549 есть средний вклад одного станка в очередь.

г) *Продолжение: несколько рабочих.* Мы не будем менять основные предположения предыдущей задачи, за исключением того, что *теперь m станков обслуживаются r рабочими* ($r<m$). Таким образом, при $n\leqslant r$ состояние E_n означает, что $r-n$ рабочих не занято, n станков обслуживаются и в очереди на ремонт нет ни одного станка. При $n>r$ состояние E_n означает, что r станков обслуживаются и $n-r$ станков стоят в очереди. Мы можем воспользоваться здесь рассуждениями из предыдущего примера с тем исключением, что (7.12) должно быть, очевидно, заменено на

$$\begin{aligned}
&\lambda_0=m\lambda, && \mu_0=0, &&\\
&\lambda_n=(m-n)\,\lambda, && \mu_n=n\mu, && 1\leqslant n\leqslant r,\\
&\lambda_n=(m-n)\lambda, && \mu_n=r\mu, && r\leqslant n\leqslant m.
\end{aligned} \tag{7.19}$$

Мы не будем выписывать основную систему дифференциальных уравнений, а ограничимся лишь уравнениями для предельных вероятностей p_n. При $1\leqslant n<r$

$$\{(m-n)\,\lambda+n\mu\}\,p_n=(m-n+1)\,\lambda p_{n-1}+(n+1)\,\mu p_{n+1}, \tag{7.20а}$$

а при $r\leqslant n\leqslant m$

$$\{(m-n)\,\lambda+r\mu\}\,p_n=(m-n+1)\,\lambda p_{n-1}+r\mu p_{n+1}. \tag{7.20б}$$

При $n=0$, очевидно, $m\lambda p_0=\mu p_1$. Это равенство определяет отношение p_1/p_0, и из (7.20а) мы получаем по индукции, что при $n<r$

$$(n+1)\,\mu p_{n+1}=(m-n)\,\lambda p_n; \tag{7.21}$$

наконец, при $n\geqslant r$ мы видим из (7.20б), что

$$r\mu p_{n+1}=(m-n)\,\lambda p_n. \tag{7.22}$$

Эти уравнения позволяют последовательно вычислить все отношения p_n/p_0. Наконец, p_0 определяется из условия $\sum p_k=1$. Значения, приведенные в табл. 3, получены именно таким способом.

Таблица 3

Вероятности p_n для случая $\lambda/\mu = 0{,}1$, $m = 20$, $r = 3$

n	Число обслуживаемых станков	Число станков в очереди	Число незанятых рабочих	p_n
0	0	0	3	0,13625
1	1	0	2	0,27250
2	2	0	1	0,25888
3	3	0	0	0,15533
4	3	1	0	0,08802
5	3	2	0	0,04694
6	3	3	0	0,02347
7	3	4	0	0,01095
8	3	5	0	0,00475
9	3	6	0	0,00190
10	3	7	0	0,00070
11	3	8	0	0,00023
12	3	9	0	0,00007

При сравнении табл. 2 и 3 выясняются удивительные факты. Эти таблицы относятся к одним и тем же станкам ($\lambda/\mu = 0{,}1$), однако во втором случае мы имеем $m = 20$ станков и $r = 3$ рабочих. Число станков на одного рабочего возросло с 6 до $6^2/_3$, и все же станки обслуживаются более эффективно. Определим *коэффициент простоя для станков*

$$\frac{w}{m} = \frac{\text{Среднее число станков в очереди}}{\text{Общее число станков}} \tag{7.23}$$

и *коэффициент простоя для рабочих*

$$\frac{\rho}{r} = \frac{\text{Среднее число незанятых рабочих}}{\text{Общее число рабочих}}. \tag{7.24}$$

Для практических целей мы можем отождествить вероятности $P_n(t)$ с их пределами p_n. Тогда в случае табл. 3 мы имеем $w = p_4 + 2p_5 + 3p_6 + \ldots + 17p_{20}$ и $\rho = 3p_0 + 2p_1 + p_2$. Табл. 4 убе-

Таблица 4

Сравнение эффективности двух систем, рассмотренных в примерах «в» и «г»

	«в»	«г»
Число станков	6	20
Число рабочих	1	3
Число станков на одного рабочего	6	$6^2/_3$
Коэффициент простоя для рабочих	0,4845	0,4042
Коэффициент простоя для станков	0,0549	0,01694

дительно доказывает, что для наших конкретных станков (с $\lambda/\mu=0{,}1$) *иметь трех рабочих на двадцать станков намного экономичнее, чем одного рабочего на шесть станков.*

д) *Задача о снабжении энергией*[1]. Одна электрическая цепь снабжает энергией a сварщиков, которые используют эту энергию лишь время от времени. Если в момент времени t сварщик использует энергию, то вероятность того, что он перестанет пользоваться ею до момента $t+h$, равна $\mu h+o(h)$; если в момент времени t электроэнергия ему не требуется, то вероятность того, что она потребуется ему до момента $t+h$, равна $\lambda h+o(h)$. Сварщики работают независимо друг от друга.

Мы говорим, что система находится в состоянии E_n, если электроэнергию используют n сварщиков. Таким образом, мы имеем лишь конечное число состояний $E_0, \ldots, E_a$.

Если система находится в состоянии E_n, то $a-n$ сварщиков не потребляют электроэнергию, и вероятность возникновения на протяжении интервала времени длины h новой потребности в энергии равна $(a-n)\lambda h+o(h)$; с другой стороны, вероятность того, что один из n сварщиков перестанет пользоваться энергией, равна $n\mu h+$ $+o(h)$. Следовательно, мы имеем процесс размножения и гибели с

$$\lambda_n=(a-n)\lambda, \quad \mu_n=n\mu, \quad 0\leqslant n\leqslant a. \tag{7.25}$$

Основные дифференциальные уравнения принимают вид

$$\begin{aligned} P_0'(t) &= -a\lambda P_0(t)+\mu P_1(t), \\ P_n'(t) &= -\{n\mu+(a-n)\lambda\}P_n(t)+(n+1)\mu P_{n+1}(t)+ \\ &\qquad +(a-n+1)\lambda P_{n-1}(t), \\ P_a' &= -a\mu P_a(t)+\lambda P_{a-1}(t). \end{aligned} \tag{7.26}$$

(Здесь $1\leqslant n\leqslant a-1$.) Легко проверить, что *предельные вероятности даются биномиальным распределением*

$$p_n=\binom{a}{n}\left(\frac{\lambda}{\lambda+\mu}\right)^n\left(\frac{\mu}{\lambda+\mu}\right)^{a-n}, \tag{7.27}$$

результат, который можно было бы здесь интуитивно ожидать (явные выражения для $P_n(t)$ приведены в задаче 17).

§ 8. ОБРАТНЫЕ (ОБРАЩЕННЫЕ В ПРОШЛОЕ) УРАВНЕНИЯ

В предыдущих параграфах мы изучали вероятности $P_n(t)$ того, что система в момент времени t находится в состоянии E_n. Это обозначение удобно, однако оно может и ввести в заблуждение, так

---

[1]) Этот пример был подсказан задачей, изучавшейся (хотя и неудачно) в работе Adler H. A., Miller K. W., A new approach to probability problems in electrical engineering, Transactions of the American Institute of Electrical Engineers, 65 (1946), 630—632.

как в нем не указывается начальное состояние E_i системы в момент времени нуль. Для дальнейшего развития нашей теории предпочтительнее вернуться к обозначениям § 1 и воспользоваться *переходными вероятностями*. В соответствии с этим мы обозначим через $P_{in}(t)$ (условную) вероятность состояния E_n в момент времени $t+s$ при условии, что в момент s система находилась в состоянии E_i. Мы будем по-прежнему обозначать через $P_n(t)$ (безусловную) вероятность состояния E_n в момент времени t. Если задано начальное состояние E_i, то безусловная вероятность $P_n(t)$ совпадает с $P_{in}(t)$, однако, когда начальное состояние выбирается в соответствии с распределением вероятностей $\{a_i\}$, мы имеем

$$P_n(t) = \sum_i a_i P_{in}(t). \tag{8.1}$$

Для рассматривавшихся до сих пор частных процессов мы показали, что при фиксированном i переходные вероятности $P_{in}(t)$ *удовлетворяют основным дифференциальным уравнениям* (3.2) *и* (5.2). Индекс i появляется только в начальных условиях, а именно

$$P_{in}(0) = \begin{cases} 1 & \text{при } n=i, \\ 0 & \text{в противном случае}. \end{cases} \tag{8.2}$$

Чтобы подготовиться к теории более общих процессов, мы покажем сейчас, что эти же переходные вероятности удовлетворяют также и второй системе дифференциальных уравнений. Для объяснения основных идей начнем с процесса чистого размножения из § 3. При выводе дифференциальных уравнений (3.2) мы продолжали интервал времени $(0, t)$ до $(0, t+h)$ и рассматривали возможные изменения на протяжении короткого промежутка времени $(t, t+h)$. Но мы могли таким же образом продолжать интервал $(0, t)$ и в направлении прошлого и рассматривать изменения на протяжении $(-h, 0)$. Таким образом мы получаем новую систему дифференциальных уравнений, в которой остается фиксированным n (вместо i). Действительно, переход из E_i в момент времени $-h$ в состояние E_n в момент времени t может произойти тремя взаимоисключающими способами: 1) между $-h$ и 0 нет никаких скачков, и система переходит в E_n из состояния E_i в момент времени 0; 2) между $-h$ и 0 происходит ровно один скачок, и система переходит из E_{i+1} в момент времени 0 в состояние E_n в момент t; 3) между $-h$ и 0 происходит более одного скачка. Вероятность первой возможности равна $1-\lambda_i h+o(h)$, второй $\lambda_i h+o(h)$, тогда как третья возможность имеет вероятность $o(h)$. Как и в § 2 и 3, мы заключаем, что

$$P_{in}(t+h) = P_{in}(t)(1-\lambda_i h) + P_{i+1,n}(t)\lambda_i h + o(h). \tag{8.3}$$

Следовательно, при $i \geqslant 0$ новая основная система принимает вид

$$P_{in}'(t) = -\lambda_i P_{in}(t) + \lambda_i P_{i+1,n}(t). \tag{8.4}$$

Эти уравнения называются *обратными уравнениями*, а уравнения (3.2) называются для различия *прямыми уравнениями*. Начальные условия имеют вид (8.2). (Интуитивно следовало бы ожидать, что

$$P_{in}(t)=0 \quad при \quad n<i, \tag{8.5}$$

однако существуют патологические исключения; см. пример 9, б).)

В случае процесса размножения и гибели основные *прямые уравнения* (при фиксированном i) имеют вид (5.2) — (5.3). То же рассуждение, которое привело к (8.4), приводит теперь к соответствующим *обратным уравнениям*

$$P'_{in}(t)=-(\lambda_i+\mu_i)P_{in}(t)+\lambda_i P_{i+1,n}(t)+\mu_i P_{i-1,n}(t). \tag{8.6}$$

Должно быть ясно, что прямые и обратные уравнения не являются независимыми друг от друга: решение обратных уравнений с начальными условиями (8.2) автоматически удовлетворяет прямым уравнениям, за исключением тех редких случаев, когда решение не единственно.

Пример. *Пуассоновский процесс.* В § 2 мы интерпретировали выражение Пуассона (2.4) как вероятность того, что на протяжении произвольного интервала времени длины t поступило ровно n вызовов. Будем говорить, что в момент времени t система находится в состоянии E_n, если за интервал времени от 0 до t поступило в точности n вызовов. Переход из E_i в момент времени t_1 в состояние E_n в момент t_2 означает, что между t_1 и t_2 поступило $n-i$ вызовов. Это возможно лишь при $n\geqslant i$, и поэтому для переходных вероятностей пуассоновского процесса мы имеем

$$\begin{aligned} &P_{in}(t)=e^{-\lambda t}(\lambda t)^{n-i}/(n-i)!, && \text{если } n\geqslant i,\\ &P_{in}(t)=0, && \text{если } n<i. \end{aligned} \tag{8.7}$$

Они удовлетворяют прямым уравнениям

$$P'_{in}(t)=-\lambda P_{in}(t)+\lambda P_{i,n-1}(t), \tag{8.8}$$

а также обратным уравнениям

$$P'_{in}(t)=-\lambda P_{in}(t)+\lambda P_{i+1,n}(t). \tag{8.9}$$

▶

§ 9. ПРОЦЕССЫ ОБЩЕГО ВИДА

До сих пор наша теория ограничивалась процессами, в которых непосредственные переходы из состояния E_n возможны только в соседние состояния E_{n+1} и E_{n-1}. Кроме того, эти процессы были однородными по времени, т. е. переходные вероятности $P_{in}(t)$ были одними и теми же для всех интервалов времени длины t. Теперь мы рассмотрим процессы более общего вида, в которых оба этих предположения опущены.

Как и в теории обычных цепей Маркова, мы допустим непосредственные переходы из любого состояния E_i в любое состояние E_n. Переходные вероятности могут меняться с течением времени. Это вынуждает указывать для любого интервала времени оба его конца, а не только его длину. В соответствии с этим мы будем обозначать через $P_{in}(\tau,\ t)$ *условную вероятность того, что система в момент времени t находится в состоянии E_n при условии, что в предшествующий ему момент τ состоянием системы было E_i.* Символ $P_{in}(\tau,\ t)$ утрачивает смысл, если не выполняется условие $\tau < t$. Если процесс однороден по времени, то $P_{in}(\tau,\ t)$ зависит лишь от разности $t-\tau$, и вместо $P_{in}(\tau,\ \tau+t)$ (которое не зависит тогда от τ) мы можем писать $P_{in}(t)$.

В § 1 мы видели, что переходные вероятности однородного по времени марковского процесса удовлетворяют *уравнению Колмогорова — Чэпмена*

$$P_{in}(s+t)=\sum_{\nu} P_{i\nu}(s)\, P_{\nu n}(t). \tag{9.1a}$$

Для неоднородных процессов аналогичное тождество имеет вид

$$P_{in}(\tau,\ t)=\sum_{\nu} P_{i\nu}(\tau,\ s)\, P_{\nu n}(s,\ t) \tag{9.1б}$$

и справедливо при $\tau < s < t$. Это соотношение выражает тот факт, что переход из состояния E_i в момент времени τ в состояние E_n в момент t происходит через некоторое состояние E_ν в промежуточный момент времени s, а для марковских процессов вероятность $P_{\nu n}(s,\ t)$ перехода из E_ν в E_n не зависит от предшествующего состояния E_i. Стало быть, переходные вероятности марковского процесса со счетным множеством состояний являются решениями уравнения Колмогорова—Чэпмена (9.1б), удовлетворяющими дополнительным условиям

$$P_{ik}(\tau,\ t)\geqslant 0, \quad \sum_{k} P_{ik}(\tau,\ t)=1. \tag{9.2}$$

Мы примем без доказательства тот факт, что и, обратно, каждое такое решение представляет собой переходные вероятности некоторого марковского процесса [1]). Отсюда следует, что основная задача

---

[1]) Понятие марковского процесса включает требование, чтобы при заданном состоянии E_ν в момент времени s развитие процесса до момента s не оказывало бы никакого влияния на его будущую эволюцию. Как было отмечено в § 1, уравнение Колмогорова—Чэпмена отражает это требование лишь частично, поскольку в него входят лишь один момент $\tau < s$ и один момент $t > s$. На остававшийся долгое время нерешенным вопрос о существовании немарковских процессов, переходные вероятности которых удовлетворяют (9.1), получен теперь утвердительный ответ; простейший из известных таких процессов однороден по времени и имеет лишь три возможных состояния E_j; см. Feller W., Ann. Math. Statist., 30 (1959), 1252—1253. [См. также пример

теории марковских процессов состоит в нахождении всех решений уравнения Колмогорова — Чэпмена, удовлетворяющих условиям (9.2).

Главная цель настоящего параграфа состоит в том, чтобы показать, что постулаты процесса размножения и гибели допускают естественное обобщение, разрешающее произвольные непосредственные переходы $E_i \to E_j$. Из этих постулатов мы выведем две системы обыкновенных дифференциальных уравнений, которые будут названы соответственно прямыми и обратными уравнениями. При обычных обстоятельствах каждая из двух этих систем определяет переходные вероятности единственным образом. Прямые уравнения с вероятностной точки зрения являются более естественными, однако, как ни странно, при их выводе требуются более сильные и менее интуитивно ясные предположения.

В однородном по времени процессе размножения и гибели из § 5 исходные постулаты относились к поведению переходных вероятностей $P_{jk}(h)$ при малых h; там просто требовалось существование в нуле производных P'_{jk}. В случае неоднородных процессов мы наложим те же условия на $P_{jk}(t, t+x)$, рассматриваемые как функции от x. Эти производные будут иметь аналогичную вероятностную интерпретацию, однако теперь они будут функциями от t.

Предположение 1. *Каждому состоянию E_n соответствует непрерывная функция $c_n(t) \geqslant 0$, такая, что при $h \to 0$*

$$[1 - P_{nn}(t, t+h)]/h \to c_n(t). \tag{9.3}$$

Предположение 2. *Каждой паре состояний E_j, E_k при $j \neq k$ соответствуют переходные вероятности $p_{jk}(t)$ (зависящие от времени), такие, что при $h \to 0$*

$$P_{jk}(t, t+h)/h \to c_j(t)\, p_{jk}(t), \quad j \neq k. \tag{9.4}$$

Величины $p_{jk}(t)$ непрерывны по t, и для любых фиксированных t и j

$$\sum_k p_{jk}(t) = 1, \quad p_{jj}(t) = 0. \tag{9.5}$$

Вероятностная интерпретация (9.3) очевидна: если в момент времени t система находится в состоянии E_n, то вероятность того, что между t и $t+h$ произойдет изменение, равна $c_n(t)h + o(h)$. Коэффициент $p_{jk}(t)$ можно интерпретировать как условную вероятность того, что *если* между t и $t+h$ система меняет свое состояние E_j, то

гл. XV, 13, е). — *Перев.*] Подобные процессы, однако, имеют довольно патологический характер, и их существование не противоречит тому утверждению, что *каждое* решение уравнения Колмогорова — Чэпмена, удовлетворяющее (9.2), соответствует (единственным образом) некоторому марковскому процессу.

это изменение переводит систему из E_j в E_k. В процессе размножения и гибели $c_n(t)=\lambda_n+\mu_n$,

$$p_{j,\ j+1}(t)=\lambda_j/(\lambda_j+\mu_j), \qquad p_{j,\ j-1}(t)=\mu_j/(\lambda_j+\mu_j) \tag{9.6}$$

и $p_{jk}(t)=0$ для всех других комбинаций j и k. Для каждого фиксированного t величины $p_{jk}(t)$ можно интерпретировать как переходные вероятности цепи Маркова.

Двух приведенных предположений достаточно для вывода системы обратных уравнений для $P_{jk}(\tau, t)$, однако для вывода прямых уравнений потребуется дополнительное предположение.

Предположение 3. *При фиксированном k переход к пределу в* (9.4) *равномерен по j.*

Необходимость этого предположения представляет значительный теоретический интерес и будет обсуждаться нами несколько позже.

Перейдем теперь к выводу дифференциальных уравнений для $P_{ik}(\tau, t)$ как функций от t и k (прямые уравнения). Из (9.1) имеем

$$P_{ik}(\tau,\ t+h)=\sum_j P_{ij}(\tau,\ t)\,P_{jk}(t,\ t+h). \tag{9.7}$$

Подставив вместо $P_{kk}(t,\ t+h)$ в правой части его выражение из (9.3), мы получим

$$\frac{P_{ik}(\tau,\ t+h)-P_{ik}(\tau,\ t)}{h}=$$
$$=-c_k(t)\,P_{ik}(\tau,\ t)+h^{-1}\sum_{j\neq k}P_{ij}(\tau,\ t)\,P_{jk}(t,\ t+h)+\ldots, \tag{9.8}$$

где опущенные члены стремятся к 0 вместе с h, а суммирование проводится по всем j, за исключением $j=k$. Применим теперь (9.4) к членам этой суммы. Поскольку (по предположению 3) переход к пределу равномерен по j, правая часть имеет предел. Следовательно, левая часть также имеет предел, а это означает, что $P_{ik}(\tau,\ t)$ имеет частную производную по t и

$$\partial P_{ik}(\tau,\ t)/\partial t=-c_k(t)\,P_{ik}(\tau,\ t)+\sum_j P_{ij}(\tau,\ t)\,c_j(t)\,p_{jk}(t). \tag{9.9}$$

Это основная система прямых дифференциальных уравнений. Здесь i и τ фиксированы, так что мы имеем (несмотря на формальное появление частной производной) систему *обыкновенных дифференциальных уравнений* [1]) для функций $P_{ik}(\tau,\ t)$. Параметры i и τ входят

---

[1]) В стандартной форме имеющей вид

$$x_k'(t)=-c_k(t)\,x_k(t)+\sum_j x_j(t)\,c_j(t)\,p_{jk}(t).$$

только в *начальное условие*

$$P_{ik}(\tau, \tau)=\begin{cases}1 & \text{при } k=i,\\ 0 & \text{в противном случае.}\end{cases} \tag{9.10}$$

Обратимся теперь к обратным уравнениям. В них постоянными остаются k и t, так что переходные вероятности $P_{ik}(\tau, t)$ рассматриваются как функции от начальных данных E_i и τ. При формулировке наших исходных предположений эти начальные переменные были фиксированы, однако для вывода обратных уравнений предпочтительнее сформулировать те же условия по отношению к интервалу времени от $t—h$ до t. Иначе говоря, естественнее начать со следующей *альтернативной формы условий* (9.3) *и* (9.4):

$$[1-P_{nn}(t-h, t)]/h \to c_n(t), \tag{9.3a}$$

$$P_{jk}(t-h, t)/h \to c_j(t)\, p_{jk}(t), \quad j\neq k. \tag{9.4a}$$

Нетрудно доказать эквивалентность этих двух наборов условий (или представить их в единой форме), но мы удовлетворимся тем, что начнем с этой альтернативной формы. Замечательная черта последующего вывода состоит в том, что нам не понадобится никакого аналога предположения 3.

Согласно уравнению Колмогорова—Чэпмена (9.1б),

$$P_{ik}(\tau-h, t)=\sum_{\nu} P_{i\nu}(\tau-h, \tau)\, P_{\nu k}(\tau, t), \tag{9.11}$$

и, используя (9.3a) при $n=i$, мы получаем

$$[P_{ik}(\tau-h, t)-P_{ik}(\tau, t)]/h = \\ = -c_i(\tau)\, P_{ik}(\tau, t)+h^{-1}\sum_{\nu\neq i} P_{i\nu}(\tau-h, \tau)\, P_{\nu k}(\tau, t)+\ldots . \tag{9.12}$$

Здесь $h^{-1}P_{i\nu}(\tau-h, \tau)\to c_i(\tau)\, p_{i\nu}(\tau)$, и переход к пределу в сумме в правой части (9.12) всегда равномерен. Действительно, если $N>i$, то

$$0\leqslant h^{-1}\sum_{\nu=N+1}^{\infty} P_{i\nu}(\tau-h, \tau)\, P_{\nu k}(\tau, t)\leqslant h^{-1}\sum_{\nu=N+1}^{\infty} P_{i\nu}(\tau-h, \tau)= \\ =h^{-1}\left\{1-\sum_{\nu=0}^{N} P_{i\nu}(\tau-h, \tau)\right\}\to c_i(\tau)\left\{1-\sum_{\nu=0}^{N} p_{i\nu}(\tau)\right\}. \tag{9.13}$$

В силу условия (9.5) правая часть здесь может быть сделана сколь угодно малой за счет выбора достаточно большого N. Отсюда следует, что в (9.12) возможен почленный переход к пределу, и мы получаем

$$\partial P_{ik}(\tau, t)/\partial\tau=c_i(\tau)\, P_{ik}(\tau, t)-c_i(\tau)\sum_{\nu} p_{i\nu}(\tau)\, P_{\nu k}(\tau, t). \tag{9.14}$$

Это и есть основные *обратные дифференциальные уравнения*. Здесь k и t фигурируют как фиксированные параметры, и поэтому (9.14) представляет собой систему обыкновенных дифференциальных уравнений. Параметры k и t входят лишь в *начальные условия*

$$P_{ik}(t, t) = \begin{cases} 1 & \text{при } i = k, \\ 0 & \text{в противном случае.} \end{cases} \tag{9.15}$$

Пример. а) *Обобщенный пуассоновский процесс.* Рассмотрим случай, когда *все* $c_i(t)$ *равны одной и той же постоянной*, $c_i(t) = \lambda$, *а* p_{jk} *не зависят от* t. В этом случае p_{jk} являются переходными вероятностями обыкновенной цепи Маркова, и (как и в гл. XV) ее вероятности перехода за несколько шагов мы обозначим через $p_{jk}^{(n)}$.

Из $c_i(t) = \lambda$ следует, что вероятность осуществления перехода за время между t и $t+h$ не зависит от состояния системы и равна $\lambda h + o(h)$. Отсюда вытекает, что число переходов между τ и t имеет распределение Пуассона с параметром $\lambda(t-\tau)$. При условии, что произошло ровно n переходов, (условная) вероятность перехода из j в k равна $p_{jk}^{(n)}$. Следовательно,

$$P_{ik}(\tau, t) = e^{-\lambda(t-\tau)} \sum_{n=0}^{\infty} [\lambda^n (t-\tau)^n/n!]\, p_{ik}^{(n)} \tag{9.16}$$

(где, как обычно, $p_{jj}^{(0)} = 1$ и $p_{jk}^{(0)} = 0$ при $j \neq k$). Легко проверить, что (9.16) действительно является решением обеих систем (9.9) и (9.14) дифференциальных уравнений и удовлетворяет нужным граничным условиям.

В частности, если

$$p_{jk} = 0 \text{ при } k < j, \quad p_{jk} = f_{k-j} \text{ при } k \geqslant j, \tag{9.17}$$

то (9.16) сводится к *обобщенному распределению Пуассона* (см. гл. XII, 2). ►

Обе наши системы дифференциальных уравнений впервые были выведены А. Н. Колмогоровым в работе, в которой были развиты основания теории марковских процессов [1]). Предполагая, что последовательность коэффициентов $c_n(t)$ остается ограниченной при каждом t, В. Феллер показал, что существует единственное общее для обеих систем решение $\{P_{jk}(\tau, t)\}$ и что это решение удовлетворяет уравнению Колмогорова — Чэпмена (9.16), а также условиям (9.2). Более того, в этом случае ни у одной из двух наших систем нет других решений, и поэтому эти системы по существу эквивалентны. Однако конкретные задачи вскоре привели уравнениям с неограниченными последовательностями $\{c_n\}$, и, как показано в § 4, в таких случаях мы иногда сталкиваемся с неожиданными решения-

[1]) Kolmogoroff A., Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, Mathematische Annalen, 104 (1931), 415—458.

ми, для которых в соотношении

$$\sum_k P_{jk}(\tau, t) \leqslant 1 \tag{9.18}$$

имеет место строгое неравенство. Было показано [1]) (без каких-либо ограничений на коэффициенты $c_n(t)$), что всегда существует *минимальное решение* $\{P_{jk}(\tau, t)\}$, удовлетворяющее обеим системам дифференциальных уравнений, а также уравнению Колмогорова — Чэпмена (9.16) и соотношению (9.18). Это решение называется минимальным потому, что

$$\overline{P}_{jk}(\tau, t) \geqslant P_{jk}(\tau, t), \tag{9.19}$$

если $\overline{P}_{jk}(\tau, t)$ удовлетворяет либо обратным, либо прямым дифференциальным уравнениям (вместе с банальными начальными условиями (9.10)). Когда минимальное решение удовлетворяет (9.18) со знаком равенства при всех t, это означает, что ни у обратных, ни у прямых уравнений не может быть никаких имеющих вероятностный смысл решений, кроме $P_{jk}(\tau, t)$. Иначе говоря, когда минимальное решение не имеет дефекта, наш процесс однозначно определяется любой из двух этих систем уравнений. Как уже говорилось выше, это имеет место тогда, когда для каждого фиксированного t коэффициенты $c_n(t)$ остаются ограниченными.

Ситуация становится совершенно другой, когда минимальное решение дефектно, т. е. когда в (9.18) для некоторого (а потому и для любого) t имеет место знак неравенства. В этом случае существует бесконечное число настоящих переходных вероятностей, удовлетворяющих обратным уравнениям и уравнению Колмогорова — Чэпмена, и, следовательно, существует бесконечно много марковских процессов, удовлетворяющих предположениям 1 и 2, лежащим в основе обратных уравнений. Некоторые из них могут удовлетворять также и прямым уравнениям, однако в остальных случаях решение прямых уравнений единственно [2]).

---

[1]) Feller W., On the integro-differential equations of purely discontinuous Markoff processes, Trans. Amer. Math. Soc., 48 (1940), 488—515. В этой статье рассматриваются более общие пространства состояний, однако счетные пространства состояний упоминаются в ней как наиболее интересный частный случай. Это было не замечено последующими авторами, которые дали более сложные и менее полные доказательства. Минимальное решение в однородном по времени случае получается в гл. XIV, 7 тома 2 при помощи преобразований Лапласа. Более полное изложение см. в работе Feller W., On boundaries and lateral conditions for the Kolmogorov differential equations, Ann. Math., 65 (1957), 527—570.

[2]) Напомним, что только предположения 1 и 2 имеют вероятностный смысл, тогда как предположение 3 носит чисто аналитический характер и было введено лишь для удобства. Оно не является естественным в том смысле, что даже не все решения прямых уравнений удовлетворяют наложенному условию равномерности. Таким образом, обратные уравнения выражают имеющие вероятностный смысл условия и приводят к интересной теории, однако о прямых уравнениях этого сказать нельзя. Это объясняет, почему вся теория марковских процессов должна

Пример. б) *Процесс размножения*. Дифференциальные уравнения (3.2) для однородного по времени процесса размножения имели вид

$$x_0'(t) = -\lambda_0 x_0(t), \quad x_k'(t) = -\lambda_k x_k(t) + \lambda_{k-1} x_{k-1}(t). \quad (9.20)$$

Это прямые уравнения. Поскольку они образуют рекуррентную систему, их решение однозначно определяется начальными значениями при $t=0$. Стало быть, для переходных вероятностей мы последовательно получаем $P_{ik}(t)=0$ при всех $k<i$,

$$P_{ii}(t) = e^{-\lambda_i t}, \quad P_{i,\,i+1}(t) = \frac{\lambda_i}{\lambda_i - \lambda_{i+1}}\left(e^{-\lambda_{i+1}t} - e^{-\lambda_i t}\right), \quad (9.21)$$

и, наконец, при $k>i$

$$P_{ik}(t) = \lambda_{k-1}\int_0^t e^{-\lambda_k s} P_{i,\,k-1}(t-s)\,ds. \quad (9.22)$$

Для того чтобы увидеть, что эти переходные вероятности удовлетворяют уравнению Колмогорова — Чэпмена (9.1а), достаточно заметить, что при фиксированных i и s обе части этого уравнения представляют собой решения дифференциальных уравнений (9.20), имеющие одинаковые начальные значения.

Обратные уравнения были выведены в (8.4) и имеют вид

$$y_i'(t) = -\lambda_i y_i(t) + \lambda_i y_{i+1}(t). \quad (9.23)$$

Мы должны показать, что этому уравнению удовлетворяют $P_{ik}(t)$ при фиксированном k. При $k<i$ это тривиально, ибо в этом случае все три члена в (9.23) обращаются в нуль. Используя (9.21), видим, что это утверждение справедливо также и при $k-i=0$ и $k-i=1$. Теперь мы можем продолжать доказательство по индукции, используя тот факт, что при $k>i+1$

$$P_{ik}'(t) = \lambda_{k-1}\int_0^t e^{-\lambda_k s} P_{i,\,k-1}'(t-s)\,ds. \quad (9.24)$$

Допустим, что $P_{ik}(t)$ удовлетворяет (9.23) при $k-i\leqslant n$. При $k=i+1+n$ мы можем выразить подынтегральное выражение в (9.24), используя правую часть (9.23), и получим в результате, что (9.23) справедливо и при $k-i=n+1$.

Таким образом, мы доказали, что система переходных вероятностей $P_{ik}(t)$ однозначно определяется прямыми уравнениями и что эти вероятности удовлетворяют обратным уравнениям, а также уравнению Колмогорова — Чэпмена.

основываться на обратных уравнениях (или — в абстрактной формулировке — на полугруппах преобразований функций, а не на вероятностных мерах).

Обратные уравнения (9.23) могут иметь и другие решения. Упоминавшееся выше свойство минимальности (9.19) наших переходных вероятностей можно переформулировать следующим образом. Для произвольных неотрицательных решений уравнений (9.23) имеем:

$$\textit{если } y_i(0) = P_{ik}(0), \quad \textit{то } y_i(t) \geqslant P_{ik}(t) \tag{9.25}$$

при всех $t > 0$. Здесь k произвольно, но фиксировано. Это утверждение тривиально для $k < i$, поскольку в этом случае правые части в (9.25) обращаются в нуль. Если y_{i+1} задано, то решение y_i уравнения (9.23) может быть представлено в явном виде как интеграл, аналогичный (9.22), и теперь справедливость (9.25) доказывается рекуррентным образом при $i = k, k-1, \ldots$.

Предположим теперь, что $\sum \lambda_k^{-1} < \infty$. В § 4 было показано, что в этом случае величины

$$L_i(t) = 1 - \sum_{k=0}^{\infty} P_{ik}(t) \tag{9.26}$$

не обращаются тождественно в нуль. Очевидно, $L_i(t)$ можно интерпретировать как вероятность того, что при выходе из E_i «бесконечность» достигается до момента времени t. Ясно также, что L_i являются решениями дифференциальных уравнений (9.23) с начальными условиями $L_i(0) = 0$. Рассмотрим теперь произвольные неотрицательные функции A_k и определим

$$\overline{P}_{ik}(t) = P_{ik}(t) + \int_0^t L_i(t-s)\, A_k(s)\, ds. \tag{9.27}$$

Легко проверить, что при фиксированном k *функции* $\overline{P}_{ik}(t)$ *удовлетворяют обратным дифференциальным уравнениям и* $\overline{P}_{ik}(0) = P_{ik}(0)$. Встает вопрос о том, можно ли определить $A_k(t)$ таким образом, чтобы $\overline{P}_{ik}(t)$ были переходными вероятностями, удовлетворяющими уравнению Колмогорова — Чэпмена. Ответ на этот вопрос утвердительный. Мы воздержимся от доказательства этого утверждения, но приведем его вероятностную интерпретацию.

Вероятности $P_{ik}(t)$ определяют так называемый *процесс с поглощающим экраном*: *когда система достигает бесконечности, процесс останавливается*. Дуб [1]) впервые исследовал *процесс с возвращением*, в котором по достижении бесконечности система мгновенно возвращается в E_0 (или в некоторое другое заранее заданное состояние), и процесс начинается заново. В таком процессе система может перейти из E_0 в E_5 либо за пять, либо за бесконечное число шагов, совершая один или несколько проходов от E_0 до «бесконечности».

---

[1]) Doob J. L., Markoff chains — denumerable case, Trans. Amer. Math. Soc., 58 (1945), 455—473.

Переходные вероятности этого процесса имеют вид (9.27). Они *удовлетворяют обратным уравнениям* (8.4) *или* (9.23), *но не удовлетворяют прямым уравнениям* (9.24) *или* (8.5). ▶

Этим объясняется, почему при выводе прямых уравнений мы были вынуждены ввести странное на первый взгляд предположение 3, которое не является необходимым для обратных уравнений: интуитивно понятные и имеющие простой вероятностный смысл предположения 1—2 совместимы с процессами с возвращением, для которых прямые уравнения (9.24) не выполняются. Иначе говоря, если мы будем исходить из предположений 1—2, то *обратные уравнения Колмогорова будут выполнены, однако к прямым уравнениям нужно будет прибавить еще один член* [1]).

Конечно, пример с процессом чистого размножения слишком банален, чтобы быть действительно интересным, но описанная здесь ситуация типична и для наиболее общего случая уравнений Колмогорова. Однако встречаются и два новых явления. Во-первых, у процесса размножения существует только один путь ухода на «бесконечность», или, в абстрактной терминологии, одна *граничная* точка. В противоположность этому у общих процессов могут быть границы сложной топологической структуры. Во-вторых, в процессе размножения движение направлено к границе, так как возможны лишь переходы $E_n \to E_{n+1}$. Но можно построить процессы и другого типа; например, направление может быть изменено на противоположное, и мы получим процесс, в котором будут возможны только переходы $E_{n+1} \to E_n$. Такой процесс может *начинаться* на границе, вместо того чтобы оканчиваться на ней. В процессе размножения и гибели, как и в одномерной диффузии, переходы возможны в обоих направлениях. Оказывается, что в этом случае существуют процессы, аналогичные процессам с упругими и отражающими экранами из теории диффузии, однако их описание вывело бы нас за рамки этой книги.

§ 10. ЗАДАЧИ

1. Пусть в процессе чистого размножения, определенного уравнением (3.2), коэффициенты $\lambda_n > 0$ при всех n. Доказать, что для каждого фиксированного $n \geqslant 1$ функция $P_n(t)$ сперва возрастает, а затем убывает до 0. Если t_n — точка ее максимума, то $t_1 < t_2 < t_3 < \ldots$. *Указание*. Воспользоваться индукцией; продифференцировать (3.2).

2. *Продолжение*. Показать, что если $\sum \lambda_n^{-1} = \infty$, то $t_n \longrightarrow \infty$. *Указание*. Если $t_n \longrightarrow \tau$, то при фиксированном $t > \tau$ последовательность $\lambda_n P_n(t)$ возрастает. Использовать (4.10).

3. *Процесс Юла*. Вывести среднее и дисперсию распределения, определенного уравнениями (3.4). (Использовать только эти дифференциальные уравнения, а не явный вид решения (3.5).)

4. *Процесс чистой гибели*. Вывести дифференциальные уравнения процесса

---

[1]) Дальнейшие подробности см. в гл. XIV, 8 тома 2.

типа Юла с переходами только из E_n в E_{n-1}. Найти распределение $P_n(t)$, его среднее и дисперсию, предполагая, что начальным состоянием является E_i.

5. *Стоянки автомашин.* На автостоянку с числом мест N прибывают автомобили, образующие поток пуассоновского типа с интенсивностью λ, до тех пор, пока имеются свободные места. Времена стоянки автомобилей имеют пуассоновское распределение (как и времена обслуживания в § 7). Вывести соответствующие дифференциальные уравнения для вероятностей $P_n(t)$ того, что ровно n мест оказываются занятыми.

6. *Различные дисциплины обслуживания.* Рассмотрим очередь к одному каналу, подчиняющуюся правилам, приведенным в примере 7, б). На этот раз мы будем рассматривать процесс исключительно с точки зрения мистера Смита, чей вызов поступает в момент времени 0. Его время ожидания зависит от дисциплины обслуживания, а именно от того, в каком порядке обслуживаются ожидающие вызовы. Наибольший интерес представляют следующие дисциплины:

а) *последним прибыл — последним обслуживается*, т. е. вызовы обслуживаются в порядке поступления;

б) *случайный порядок*, т. е. члены очереди имеют равные вероятности быть обслуженными в первую очередь;

в) *последним прибыл — первым обслуживается*, т. е. вызовы обслуживаются в порядке, обратном порядку поступления[1]).

Удобно перенумеровать состояния, начиная с -1. На протяжении времени действительного обслуживания мистера Смита система находится в состоянии E_0, а по истечении этого времени она переходит в E_{-1}, где и остается навсегда. При $n \geqslant 1$ система находится в состоянии E_n, если вызов мистера Смита все еще находится в очереди вместе с $n-1$ другими вызовами, которые будут (или могут быть) обслужены до него. (Обслуживаемый в этот момент вызов в очередь не включается.) Обозначим $P_n(t)$ вероятность E_n в момент времени t. Доказать, что во всех трех случаях

$$P'_{-1}(t) = \mu P_0(t).$$

Кроме того,

а) при дисциплине *последним прибыл — последним обслуживается*

$$P'_n(t) = -\mu P_n(t) + \mu P_{n-1}(t), \quad n \geqslant 0;$$

б) при дисциплине *со случайным порядком* для $n \geqslant 2$

$$P'_n(t) = -(\lambda+\mu) P_n(t) + [n\mu/(n+1)] P_{n+1}(t) + \lambda P_{n-1}(t),$$
$$P'_1(t) = -(\lambda+\mu) P_1(t) + (1/2)\mu P_2(t),$$
$$P'_0(t) = -\mu P_0(t) + \mu P_1(t) + (1/2)\mu P_2(t) + (1/3)\mu P_3(t) + \ldots;$$

в) при дисциплине *последним прибыл — первым обслуживается* для $n \geqslant 2$

$$P'_n(t) = -(\lambda+\mu) P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t),$$
$$P'_1(t) = -(\lambda+\mu) P_1(t) + \mu P_2(t),$$
$$P'_0(t) = -\mu P_0(t) + \mu P_1(t).$$

(См. также задачу 20.)

7. *Продолжение.* Предположим, что действует дисциплина обслуживания последним прибыл — последним обслуживается (случай а)), и что $P_r(0) = 1$. Показать, что

$$P_k(t) = [(\mu t)^{r-k}/(r-k)!]\, e^{-\mu t}, \quad 0 \leqslant k \leqslant r.$$

[1]) Эта дисциплина имеет смысл в обрабатывающих информацию машинах, когда последняя информация (или наблюдение) имеет наибольший вес. Она была предложена в работе Vaulot E., Delais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, Comptes Rendues, Académie des Sciences, Paris, 238 (1954), 1188—1189.

8. *Продолжение.* Обобщить задачу 6 на случай a каналов.

9. *Процесс Пойа* [1]. Это нестационарный процесс чистого размножения, у которого λ_n зависят от времени:

$$\lambda_n(t) = (1+an)/(1+at). \tag{10.1}$$

Показать, что решение, удовлетворяющее начальному условию $P_0(0)=1$, имеет вид

$$P_0(t) = (1+at)^{-1/a},$$

$$P_n(t) = \frac{(1+a)(1+2a)\ldots\{1+(n-1)a\}}{n!}t^n(1+at)^{-n-1/a}. \tag{10.2}$$

Показать (при помощи дифференциальных уравнений), что среднее и дисперсия здесь равны t и $t(1+at)$ соответственно.

10. *Продолжение.* Процесс Пойа можно получить предельным переходом из урновой схемы Пойа (пример гл. V, 2, в)). Если состояние системы определяется числом красных шаров, то вероятность перехода $E_k \to E_{k+1}$ при $(n+1)$-м извлечении равна

$$p_{k,n} = \frac{r+kc}{r+b+nc} = \frac{p+k\gamma}{1+n\gamma}, \tag{10.3}$$

где $p=r/(r+b)$, $\gamma=c/(r+b)$.

Пусть, как и при переходе от испытаний Бернулли к распределению Пуассона, интервалы между извлечениями шаров составляют h единиц времени, и пусть $h\to 0$, $n\to\infty$ так, что $np\to t$, $n\gamma\to at$. Показать, что в пределе (10.3) приводит к (10.1). Показать также, что распределение Пойа (2.3) гл. V переходит в (10.2).

11. *Линейный рост.* Если в процессе, определенном уравнениями (5.7), $\lambda=\mu$ и $P_1(0)=1$, то

$$P_0(t) = \lambda t/(1+\lambda t), \qquad P_n(t) = (\lambda t)^{n-1}/(1+\lambda t)^{n+1}. \tag{10.4}$$

Вероятность окончательного вымирания равна 1.

12. *Продолжение.* Взяв в качестве пробного решения системы (5.7) функции вида $P_n(t) = A(t)B^n(t)$, доказать, что решение с $P_1(0)=1$ имеет вид

$$P_0(t) = \mu B(t), \qquad P_n(t) = \{1-\lambda B(t)\}\{1-\mu B(t)\}\{\lambda B(t)\}^{n-1} \tag{10.5}$$

при

$$B(t) = (1-e^{(\lambda-\mu)t})/(\mu-\lambda e^{(\lambda-\mu)t}) \tag{10.6}$$

13. *Продолжение.* Доказать, что производящая функция $P(s,t)=\sum P_n(t)s^n$ удовлетворяет уравнению в частных производных

$$\partial P/\partial t = \{\mu-(\lambda+\mu)s+\lambda s^2\}\,\partial P/\partial s. \tag{10.7}$$

14. *Продолжение.* Пусть $M_2(t)=\sum n^2P_n(t)$ и $M(t)=\sum nP_n(t)$ (как и в § 5). Показать, что

$$M_2'(t) = 2(\lambda-\mu)M_2(t)+(\lambda+\mu)M(t). \tag{10.8}$$

Вывести отсюда, что при $\lambda>\mu$ *дисперсия* распределения $\{P_n(t)\}$ дается формулой

$$e^{2(\lambda-\mu)t}\{1-e^{(\mu-\lambda)t}\}(\lambda+\mu)/(\lambda-\mu). \tag{10.9}$$

15. Доказать, что производящая функция $P(s,t)=\sum P_n(t)s^n$ для про-

[1] Lundberg O., On random processes and their applications to sickness and accident statistics, Uppsala, 1940.

цесса (7.2) удовлетворяет уравнению в частных производных

$$\partial P/\partial t = (1-s)\{-\lambda P + \mu\partial P/\partial s\}, \tag{10.10}$$

решением которого является

$$P(s, t) = e^{-\lambda(1-s)(1-e^{-\mu t})/\mu}\{1-(1-s)e^{-\mu t}\}^{i}.$$

При $t=0$ это распределение Пуассона с параметром $\lambda(1-e^{-\mu t})/\mu$. *При $t \to \infty$ распределение $\{P_n(t)\}$ стремится к распределению Пуассона с параметром λ/μ.*

16. Доказать, что производящая функция $P(s) = \sum p_n s^n$ для стационарного состояния процесса, определенного уравнениями (7.26), удовлетворяет уравнению в частных производных

$$(\mu + \lambda s)\,\partial P/\partial s = a\lambda P, \tag{10.11}$$

которое имеет решение $P = \{(\mu+\lambda s)/(\lambda+\mu)\}^a$.

17. Рассмотрим для дифференциальных уравнений (7.26) пробное решение вида

$$P_n(t) = \binom{a}{n} A^n (1-A)^{a-n}.$$

Доказать, что эти функции являются решением тогда и только тогда, когда

$$A = [\lambda/(\lambda+\mu)]\,(1-e^{-(\lambda+\mu)t}).$$

18. Пусть в простейшей задаче о телефонных линиях (пример 7, а)) $Q_n(t)$ — вероятность, того, что, выходя из E_n, система достигнет состояния E_0 до момента времени t. Доказать справедливость дифференциальных уравнений

$$\begin{aligned} Q_n'(t) &= -(\lambda+n\mu)\,Q_n(t) + \lambda Q_{n+1}(t) + n\mu Q_{n-1}(t), \quad n \geqslant 2, \\ Q_1'(t) &= -(\lambda+\mu)\,Q_1(t) + \lambda Q_2(t) + \mu \end{aligned} \tag{10.12}$$

с начальными условиями $Q_n(0) = 0$.

19. *Продолжение.* Рассмотреть ту же задачу для процесса, определяемого произвольной системой прямых уравнений. Показать, что тогда $Q_n(t)$ удовлетворяют соответствующим *обратным уравнениям* (при фиксированном k), в которых $P_{0k}(t)$ заменено на 1.

20. Показать, что дифференциальные уравнения из задачи 6 по существу представляют собой прямые уравнения для переходных вероятностей. Вывести соответствующие обратные уравнения.

21. Допустим, что решение хотя бы одной из двух систем (прямых и обратных) уравнений единственно. Доказать, что переходные вероятности, удовлетворяющие этой системе, удовлетворяют уравнению Колмогорова — Чэпмена (1.1). *Указание.* Показать, что выражения, стоящие в обеих частях этого уравнения, удовлетворяют одной и той же системе дифференциальных уравнений с одинаковыми начальными условиями.

22. Пусть $P_{ik}(t)$ удовлетворяют уравнению Колмогорова — Чэпмена (1.1). Предполагая, что $P_{ik}(t) > 0$ и $S_i(t) = \sum_k P_{ik}(t) \leqslant 1$, доказать, что либо $S_i(t) = 1$ при всех t, либо $S_i(t) < 1$ при всех t.

23. *Эргодические свойства.* Рассмотрим стационарный процесс с конечным числом состояний, т. е. предположим, что система дифференциальных уравнений (9.9) конечна и что коэффициенты c_j и p_{jk} суть постоянные величины. Доказать, что решениями здесь являются линейные комбинации показательных членов $e^{\lambda(t-\tau)}$, где действительная часть λ отрицательна, если $\lambda \neq 0$. Вывести отсюда, что асимптотическое поведение переходных вероятностей такое же, как и в случае *конечных* цепей Маркова, за тем исключением, что периодический случай теперь невозможен.

Глава I

1. а) 3/5; б) 3/5; в) 3/10.
2. События S_1, S_2, $S_1 \cup S_2$ и S_1S_2 содержат 12, 12, 18 и 6 точек соответственно.
4. Пространство элементарных событий содержит две точки $ГГ$ и $РР$ с вероятностями 1/4, две точки $ГРР$ и $РГГ$ с вероятностями 1/8 и вообще по две точки с вероятностями 2^{-n} при $n \geq 2$. Сумма этих вероятностей равна единице, так что нет необходимости рассматривать возможность бесконечной серии бросаний. Искомые вероятности равны 15/16 и 2/3 соответственно.
9. $\mathbf{P}\{AB\} = 1/6$, $\mathbf{P}\{A \cup B\} = 23/36$, $\mathbf{P}\{AB'\} = 1/3$.
12. $x = 0$ в случае событий а), б) и ж); $x = 1$ в случае событий д) и е); $x = 2$ в случае события г); $x = 4$ в случае события в).
15. а) A; б) AB; в) $B \cup (AC)$.
16. Правильными являются соотношения в), г), д), е), з), и), л) и м). Соотношение а) бессмысленно, за исключением случая $C \subset B$. Вообще говоря, оно неверно даже и в этом случае, но будет правильным при дополнительных условиях $C \subset B$, $AC = 0$. Соотношение б) правильно в случае $C \supset AB$. Соотношение ж) должно выглядеть так: $(A \cup B) - A = A'B$. Наконец, правильным вариантом к) является л).
17. а) $AB'C'$; б) ABC'; в) ABC; г) $A \cup B \cup C$; д) $AB \cup AC \cup BC$; е) $AB'C' \cup A'BC' \cup A'B'C$; ж) $ABC' \cup AB'C \cup A'BC = (AB \cup AC \cup BC) - ABC$; з) $A'B'C'$; и) $(ABC)'$.
18. $A \cup B \cup C = A \cup (B - AB) \cup \{C - C(A \cup B)\} = A \cup BA' \cup CA'B'$.

Глава II

1. а) 26^3; б) $26^2 + 26^3 = 18\,252$; в) $26^2 + 26^3 + 26^4$. В городе с 20 000 жителями или некоторые люди имеют одинаковые инициалы, или по меньшей мере 1748 людей имеют более трех инициалов.
2. $2(2^{10} - 1) = 2046$.
3. $\binom{n}{2} + n = \dfrac{n(n+1)}{2}$. 4. а) $\dfrac{1}{n}$; б) $\dfrac{1}{n(n-1)}$.
5. $q_A = (5/6)^6$, $q_B = (5/6)^{12} + 12\,(5/6)^{11} \cdot (1/6)$.
6. а) $p_1 = 0{,}01$, $p_2 = 0{,}27$, $p_3 = 0{,}72$;
б) $p_1 = 0{,}001$, $p_2 = 0{,}063$, $p_3 = 0{,}432$, $p_4 = 0{,}504$.
7. $p_r = (10)_r\, 10^{-r}$. Например, $p_3 = 0{,}72$, $p_{10} = 0{,}00036288$. Формула Стирлинга дает $p_{10} = 0{,}0003598\ldots$.
8. а) $(9/10)^k$; б) $(9/10)^k$; в) $(8/10)^k$; г) $2\,(9/10)^k - (8/10)^k$; д) AB и $A \cup B$.
9. $\binom{n}{2} n!\, n^{-n}$. 10. $9\binom{12}{8}^{-1} = \dfrac{1}{55}$.
11. Вероятность ровно r испытаний равна $(n-1)_{r-1}/(n)_r = n^{-1}$.
12. а) $[1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)]^{-1} = 2^n n!/(2n)!$;
б) $n!\,[1 \cdot 3 \cdot \ldots \cdot (2n-1)]^{-1} = 2^n \binom{2n}{n}^{-1}$.

13. В предположении случайности вероятность того, что все двенадцать штрафов придутся на вторники или четверти, равна $(2/7)^{12} = 0{,}0000003\ldots$. Имеется лишь $\binom{7}{2} = 21$ пара дней, так что вероятность остается крайне малой даже для любых двух дней. Следовательно, разумно предположить, что полиция действует по системе.

14. В предположении случайности вероятность события равна $(6/7)^{12}$, или приблизительно 1/6. Никакого уверенного заключения сделать нельзя.

15. $(90)_{10}/(100)_{10} = 0{,}330476\ldots$.

16. $25!(5!)^{-5}5^{-25} = 0{,}00209\ldots$.

17. $2(n-2)_r(n-r-1)!/n! = 2(n-r-1)/[n(n-1)]$.

18. а) 1/216; б) 83/3888.

19. Вероятности равны $1-(5/6)^4 = 0{,}517747\ldots$ и $1-(35/36)^{24} = 0{,}491404\ldots$.

20. а) $(n-N)_r/(n)_r$; б) $(1-N/n)^r$. При $r=N=3$ вероятности равны а) $0{,}911812\ldots$; б) $0{,}912673\ldots$. При $r=N=10$ они равны а) $0{,}330476$; б) $0{,}348678\ldots$.

21. а) $(1-N/n)^{r-1}$; б) $(n)_{Nr}/((n)_N)^r$.

22. $(1-2/n)^{2^r-2}$; для медианы приблизительно $2^{r+1} = 0{,}7n$.

23. В предположении случайности вероятности того, что три или четыре тарелки разбиты а) одной девочкой, б) самой младшей девочкой, равны соответственно $13/64 \approx 0{,}2$ и $13/256 \approx 0{,}05$.

24. а) $12!/12^{12} = 0{,}000054$; б) $\binom{12}{2}(2^6-2)\,12^{-6} = 0{,}00137$.

25. $\dfrac{30!}{2^6 6^6}\binom{12}{6}12^{-30} \approx 0{,}00035\ldots$.

26. а) $\binom{n}{2r}2^{2r}\binom{2n}{2r}^{-1}$; б) $n\binom{n-1}{2r-2}2^{2r-2}\binom{2n}{2r}^{-1}$;

в) $\binom{n}{2}\binom{n-2}{2r-4}2^{2r-4}\binom{2n}{2r}^{-1}$.

27. $\binom{N-3}{r-1}\binom{N-1}{r-1}^{-1}$.

28. $p = \binom{2N}{N}^2\binom{4N}{2N}^{-1} \approx \sqrt{2/(N\pi)}$.

29. $$p = \binom{4}{k}\binom{48}{13-k}\binom{39}{13}\binom{26}{13}\binom{52}{13}^{-1}\binom{39}{13}^{-1}\binom{26}{13}^{-1} = \binom{4}{k}\binom{48}{13-k}\binom{52}{13}^{-1}.$$

30. См. задачу 29. Вероятность равна
$$\binom{13}{m}\binom{39}{13-m}\binom{13-m}{n}\binom{26+m}{13-n}\binom{52}{13}^{-1}\binom{39}{13}^{-1}.$$

31. $\binom{4}{k}\binom{48}{26-k}\binom{52}{26}^{-1}$.

32. $$\binom{13}{a}\binom{39}{13-a}\binom{13-a}{b}\binom{26+a}{13-b}\binom{13-a-b}{c}\binom{13+a+b}{13-c}\times \binom{52}{13}^{-1}\binom{39}{13}^{-1}\binom{26}{13}^{-1}.$$

33. а) $24p(5,4,3,1)$; б) $4p(4,4,4,1)$; в) $12p(4,4,3,2)$.

34. $\binom{13}{a}\binom{13}{b}\binom{13}{c}\binom{13}{d}\binom{52}{13}^{-1}$. (См. задачу 33, где найдена вероятность того, что у одного игрока имеется a карт одной масти, b — другой и т. д.)

35. $p_0(r) = (52-r)_4/(52)_4$; $p_1(r) = 4r\,(52-r)_3/(52)_4$;
$p_2(r) = 6r\,(r-1)\,(52-r)_2/(52)_4$;
$p_3(r) = 4r\,(r-1)\,(r-2)\,(52-r)/(52)_4$; $p_4(r) = (r)_4/(52)_4$.

36. Вероятности того, что времена ожидания первого, ... четвертого туза *превосходят* r, равны $w_1(r) = p_0(r)$; $w_2(r) = p_0(r) + p_1(r)$; $w_3(r) = p_0(r) + p_1(r) + p_2(r)$; $w_4(r) = 1 - p_4(r)$. Далее, $f_i(r) = w_i(r-1) - w_i(r)$. Медианы равны 9, 20, 33, 44.

37. а) $\binom{4}{k}\binom{4-k}{k}\binom{48}{r-k}\binom{48-r+k}{r-k}\binom{52}{r}^{-1}\binom{52-r}{r}^{-1}$ при $k \leqslant 2$;

б) $\left\{\binom{4}{k}\binom{48}{r-k}\binom{52}{r}^{-1}\right\}^2$ при $k \leqslant 4$.

39. $\binom{r_1+n-1}{r_1}\binom{r_2+n-1}{r_2}$. **40.** $\binom{r_1+5}{5}(r_2+1)$.

41. $(r_1+r_2+r_3)!/(r_1!\,r_2!\,r_3!)$. **42.** $(49)_4/(52)_4$.

43.
$$\mathbf{P}\{(7)\} = 10\cdot 10^{-7} = 0{,}000001,$$
$$\mathbf{P}\{(6,1)\} = \frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{1!\,6!}\cdot 10^{-7} = 0{,}000063,$$
$$\mathbf{P}\{(5,2)\} = \frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{2!\,5!}\cdot 10^{-7} = 0{,}000189,$$
$$\mathbf{P}\{(5,1,1)\} = \frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{1!\,1!\,5!}\cdot 10^{-7} = 0{,}001512,$$
$$\mathbf{P}\{(4,3)\} = \frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{3!\,4!}\cdot 10^{-7} = 0{,}000315,$$
$$\mathbf{P}\{(4,2,1)\} = \frac{10!}{7!\,1!\,1!}\cdot\frac{7!}{1!\,2!\,4!}\cdot 10^{-7} = 0{,}007560,$$
$$\mathbf{P}\{(4,1,1,1)\} = \frac{10!}{6!\,3!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,4!}\cdot 10^{-7} = 0{,}017640,$$
$$\mathbf{P}\{(3,3,1)\} = \frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{1!\,3!\,3!}\cdot 10^{-7} = 0{,}005040,$$
$$\mathbf{P}\{(3,2,2)\} = \frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{2!\,2!\,3!}\cdot 10^{-7} = 0{,}007560,$$
$$\mathbf{P}\{(3,2,1,1)\} = \frac{10!}{6!\,2!\,1!\,1!}\cdot\frac{7!}{1!\,1!\,2!\,3!}\cdot 10^{-7} = 0{,}105840,$$
$$\mathbf{P}\{(3,1,1,1,1)\} = \frac{10!}{5!\,4!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,1!\,3!}\cdot 10^{-7} = 0{,}105840,$$
$$\mathbf{P}\{(2,2,2,1)\} = \frac{10!}{6!\,3!\,1!}\cdot\frac{7!}{1!\,2!\,2!\,2!}\cdot 10^{-7} = 0{,}052920,$$
$$\mathbf{P}\{(2,2,1,1,1)\} = \frac{10!}{5!\,3!\,2!}\cdot\frac{7!}{1!\,1!\,1!\,2!\,2!}\cdot 10^{-7} = 0{,}317520,$$
$$\mathbf{P}\{(2,1,1,1,1,1)\} = \frac{10!}{4!\,5!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,1!\,1!\,2!}\cdot 10^{-7} = 0{,}317520,$$
$$\mathbf{P}\{(1,1,1,1,1,1,1)\} = \frac{10!}{3!\,7!}\cdot 7!\cdot 10^{-7} = 0{,}060480.$$

44. Пусть S, D, T, Q означает, что в какой-то день родились один, два, три и четыре человека соответственно. Тогда
$$\mathbf{P}\{22S\} = \frac{365!}{343!}\cdot 365^{-22} = 0{,}52430,$$
$$\mathbf{P}\{20S+1D\} = \frac{365!}{1!\,344!}\cdot\frac{22!}{20!\,2!}\cdot 365^{-22} = 0{,}35208,$$
$$\mathbf{P}\{18S+2D\} = \frac{365!}{2!\,345!}\cdot\frac{22!}{18!\,2!\,2!}\cdot 365^{-22} = 0{,}09695,$$

$$\mathbf{P}\{16S+3D\} = \frac{365!}{3!\,346!}\cdot\frac{22!}{16!\,2!\,2!\,2!}\cdot 365^{-22} = 0{,}01429,$$

$$\mathbf{P}\{19S+1T\} = \frac{365!}{345!}\cdot\frac{22!}{19!\,3!}\cdot 365^{-22} = 0{,}00680,$$

$$\mathbf{P}\{17S+1D+1T\} = \frac{365!}{346!}\cdot\frac{22!}{17!\,2!\,3!}\cdot 365^{-22} = 0{,}00336,$$

$$\mathbf{P}\{14S+4D\} = \frac{365!}{347!}\cdot\frac{22!}{14!\,2!\,2!\,2!\,2!}\cdot 365^{-22} = 0{,}00124,$$

$$\mathbf{P}\{15S+2D+1T\} = \frac{365!}{347!}\cdot\frac{22!}{15!\,2!\,2!\,3!}\cdot 365^{-22} = 0{,}00066,$$

$$\mathbf{P}\{18S+1Q\} = \frac{365!}{346!}\cdot\frac{22!}{18!\,4!}\cdot 365^{-22} = 0{,}00009.$$

45. Пусть $q=\binom{52}{5}=2\,598\,960$. Вероятности равны а) $4/q$; б) $13\cdot 12\cdot 4\cdot q^{-1}=1/4165$; в) $13\cdot 12\cdot 4\cdot 6\cdot q^{-1}=6/4165$; г) $9\cdot 4^5\cdot q^{-1}=768/216580$; д) $13\cdot\binom{12}{2}\cdot 4\cdot 4^2\cdot q^{-1}=88/4165$; е) $\binom{13}{2}\cdot 11\cdot 6\cdot 6\cdot 4\cdot q^{-1}=198/4165$; ж) $13\cdot\binom{12}{3}\cdot 6\cdot 4^3\cdot q^{-1}=1760/4165$.

Глава IV

1. 99/323. **2.** 0,21... . **3.** 1/4. **4.** $7/2^6$. **5.** 1/81 и $31/6^6$.

6. Если A_k — событие, состоящее в том, что (k, k) не появится, то из (1.5) следует, что

$$1-p_r = 6\left(\frac{35}{36}\right)^r-\binom{6}{2}\left(\frac{34}{36}\right)^r+\binom{6}{3}\left(\frac{33}{36}\right)^r-\binom{6}{4}\left(\frac{32}{36}\right)^r+6\left(\frac{31}{36}\right)^r-\left(\frac{30}{36}\right)^r.$$

7. Положим $p^{-1}=\binom{52}{13}$. Тогда $S_1=13\binom{48}{9}p$; $S_2=\binom{13}{2}\binom{44}{5}p$; $S_3=40\binom{13}{3}p$. Приближенные численные значения: $P_{[0]}=0{,}09658$; $P_{[1]}=0{,}0341$; $P_{[2]}=0{,}0001$.

8. $u_r=\sum_{k=0}^{N}(-1)^k\binom{N}{k}\left(1-\frac{k}{n}\right)^r$.

9. $u_r=\sum_{k=0}^{N}(-1)^k\binom{N}{k}\frac{(n-k)_r}{(n)_r}$. Для доказательства того, что эти два результата согласуются, использовать формулу (12.18) гл. II.

10. Общий член имеет вид $a_{1k_1}a_{2k_2}\dots a_{Nk_N}$, где $(k_1, \dots, k_N)$ — перестановка $(1, 2, \dots, N)$. Для диагонального элемента $k_\nu=\nu$.

12. $u_r=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{(ns-ks)_r}{(ns)_r}$.

14. Заметить, что по определению $u_r=0$ при $r<n$ и $u_n=n!\,s^n/(ns)_n$.

15. $u_r-u_{r-1}=\sum_{k=1}^{n}(-1)^{k-1}\binom{n-1}{k-1}\frac{(ns-ks)_{r-1}}{(ns-1)_{r-1}}$.

Предел равен $\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\left(1-\frac{k+1}{n}\right)^{r-1}$.

16. $\binom{N}{2}^{-r}\binom{N}{m}\sum_{k=2}^{m}(-1)^{m-k}\binom{m}{k}\binom{k}{2}^r$.

17. Воспользоваться тем, что $\binom{52}{5}S_k=\binom{4}{k}\binom{52-13k}{5}$. Приближенные численные значения: $P_{[0]}=0{,}264$, $P_{[1]}=0{,}588$, $P_{[2]}=0{,}146$, $P_{[3]}=0{,}002$.

18. Воспользоваться тем, что $\binom{52}{13}S_k=\binom{4}{k}\binom{52-2k}{13-2k}$. Приближенные численные значения: $P_{[0]}=0{,}780217$, $P_{[1]}=0{,}204606$, $P_{[2]}=0{,}014845$, $P_{[3]}=0{,}000330$, $P_{[4]}=0{,}000002$.

19. $m!\,N!\,u_m=\sum_{k=0}^{N-m}(-1)^k(N-m-k)!/k!$.

20. См. формулу в ответе к задаче 21 при $r=2$.

21. $(rN)!\,x=$
$$=\binom{N}{2}r^2(rN-2)!-\binom{N}{3}r^3(rN-3)!+-\ldots+(-1)^N r^N(rN-N)!.$$

24. $P_{[m]}=\binom{n}{m}\binom{n+r-1}{r}^{-1}\sum_{k=0}^{n-m}(-1)^k\binom{n-m}{k}\binom{n-m+r-1-k}{r}$.

25. Воспользоваться формулами (12.16) и (12.4) гл. II.

26. Положить $U_N=A_1\cup\ldots\cup A_N$ и заметить, что $U_{N+1}=U_N\cup A_{N+1}$ и $U_NA_{N+1}=(A_1A_{N+1})\cup\ldots\cup(A_NA_{N+1})$.

Глава V

1. $1-(5)_3/(6)_3=1/2$. 2. $p=1-10\cdot 5^9/(6^{10}-5^{10})=0{,}61\ldots$.

3. а) $\binom{35}{13}\binom{39}{13}^{-1}=0{,}182\ldots$. Вероятность ровно одного туза равна $4\binom{35}{12}\binom{39}{13}^{-1}=0{,}411\ldots$; б) приближенно $1-0{,}182-0{,}411=0{,}407$.

4. а) $2\binom{23}{10}\binom{26}{13}^{-1}=\frac{11}{50}$; б) $2\binom{23}{12}\binom{26}{13}^{-1}=\frac{13}{50}$.

6. 125/345; 140/345; 80/345. 7. 20/21. 9. $(5/6)^2$. 10. $1-(5/6)^2$.

12. $p/(2-p)$. 13. б) 3/5; в) $2^n(1+2^n)^{-1}$.

14. г) Положим $a_n=x_n-4/7$, $b_n=y_n-1/7$, $c_n=z_n-2/7$; тогда
$$|a_n|+|b_n|+|c_n|=(1/2)\{|a_{n+1}|+|b_{n+1}|+|c_{n+1}|\}.$$
Следовательно, $|a_n|+|b_n|+|c_n|$ возрастает со скоростью геометрической прогрессии.

15. $p=(1-p_1)(1-p_2)\ldots(1-p_n)$.

16. Использовать неравенство $1-x<e^{-x}$ при $0<x<1$ или разложение в ряд Тейлора функции $\log(1-x)$; см. формулу (12.26) гл. II.

18. $(b+c)/(b+c+r)$.

19. Предположим, что утверждение справедливо для n-го извлечения при любых b, r и c. Рассматривая две возможности при *первом* извлечении, находим, что вероятность черного шара в $(n+1)$-м испытании равна
$$\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}+\frac{r}{b+r}\cdot\frac{b}{b+r+c}=\frac{b}{b+r}.$$

20. В предыдущей задаче устанавливается, что утверждение верно при $m=1$ и любом n. Для применения индукции рассмотреть две возможности в первом испытании.
23. Воспользоваться формулой (12.9) гл. II.
24. Биномиальный коэффициент в правой части является пределом первого множителя в числителе (8.2). Заметить, что

$$\binom{-1/\gamma}{n} \sim \binom{-1/\gamma}{n_2} (1+\rho)^{n_1}.$$

26. В силу (5.2) $2v=2p(1-p)\leqslant 1/2$.
28. а) u^2; б) $u^2+uv+v^2/4$; в) $u^2+(25uv+9v^2+vw+2uw)/16$.
33. $p_{11}=p_{32}=2p_{21}=p$, $p_{12}=p_{33}=2p_{23}=q$, $p_{13}=p_{31}=0$, $p_{22}=1/2$.

Глава VI

1. 5/16. 2. Вероятность равна 0,02804... . 3. $(0{,}9)^x\leqslant 0{,}1$, $x\geqslant 22$.
4. $q^x\leqslant 1/2$ и $(1-4p)^x\leqslant 1/2$, где $p=\binom{48}{9}\binom{52}{13}^{-1}$. Следовательно, $x\geqslant 263$ и $x\geqslant 66$ соответственно.
5. $1-(0{,}8)^{10}-2(0{,}8)^9=0{,}6242\ldots$.
6. $\{1-(0{,}8)^{10}-2(0{,}8)^9\}/\{1-(0{,}8)^{10}\}=0{,}6993\ldots$.
7. $\binom{26}{2}\binom{26}{11}\binom{52}{13}^{-1}=0{,}003954\ldots$ и $\binom{13}{2}\frac{1}{2^{13}}=0{,}00952\ldots$.
8. $\binom{12}{2}\{6^{-6}-2\cdot 12^{-6}\}$.
9. Истинные значения: 0,6651..., 0,40187.., и 0,2009...; пуассоновские приближения: $1-e^{-1}=0{,}6321\ldots$, 0,3679.., и 0,1839... .
10. $e^{-2}\sum\limits_{4}^{\infty}2^k/k!=0{,}143\ldots$. 11. $e^{-1}\sum\limits_{3}^{\infty}1/k!=0{,}080\ldots$.
12. $e^{-x/100}\leqslant 0{,}05$, или $x\geqslant 300$.
13. $e^{-1}=0{,}3679\ldots$, $1-2\cdot e^{-1}=0{,}264\ldots$.
14. $e^{-x}\leqslant 0{,}01$, $x\geqslant 5$. 15. $1/p=649\,740$.
16. $1-p^n$, где $p=p(0;\lambda)+\ldots+p(k;\lambda)$.
18. q^3 при $k=0$; pq^3 при $k=1, 2, 3$ и pq^3-pq^6 при $k=4$.
19. $\sum\limits_{k=0}^{n}\binom{n}{k}^2 2^{-2n}=\binom{2n}{n}2^{-2n}\approx 1/\sqrt{\pi n}$ для больших n.
20. $\sum\limits_{k=a}^{a+b-1}\binom{a+b-1}{k}p^kq^{a-b-1-k}$. Этот результат можно записать также в виде $p^a\sum\limits_{k=0}^{b-1}\binom{a+k-1}{k}q^k$, где k-е слагаемое равно вероятности того, что a-й успех имеет место непосредственно после $k\leqslant b-1$ неудач.
21. $x_r=\binom{2N-1-r}{N-r}\cdot 2^{-2N+r+1}$.
22. а) $x=\sum\limits_{r=1}^{N}x_r 2^{-r-1}=|2^{-2N}\sum\limits_{r=1}^{N}\binom{2N-1-r}{N-1}$; б) Воспользоваться формулой (12.6) гл. II.
23. $k_i\approx np_i$, $k_{12}\approx np_{12}$; следовательно, $n\approx k_1k_2/k_{12}$.

24. $\binom{n}{n_1}\cdot\binom{n-s_1}{n_2}\dots\binom{n-s_{r-1}}{n_r}\cdot q^{s_r}p^{(rn-s_1-\dots-s_r)}$, где $s_l=n_1+\dots+n_l$.
25. $p=p_1q_2(p_1q_2+p_2q_1)^{-1}$.
31. Из разложения логарифма в ряд Тейлора следует, что

$$b(0;\ n,\ p)=q^n=(1-\lambda/n)^n<e^{-\lambda}=p(0;\ \lambda).$$

Члены каждого распределения дают в сумме единицу, поэтому невозможно, чтобы *все* члены одного распределения были больше соответствующих членов другого.

32. Только конечное число членов распределения Пуассона превосходят ε, а остальные члены оказываются больше соответствующих членов биномиального распределения.

Глава VII

1. Действовать так же, как в § 1. 2. Воспользоваться формулой (1.7).
3. $\mathfrak{N}(-32/30)=0{,}143\dots$. 4. 0,99. 5. 511. 6. 66 400.
7. Конечно. Из неравенств гл. VI следует, что вероятность превышения восьмикратного стандартного отклонения крайне мала.
8. $(2\pi n)^{-1}\{p_1p_2(1-p_1-p_2)\}^{-1/2}$.

Глава VIII

1. $\beta=21$.
2. $x=pu+qv+rw$, где u, v, w — решения системы

$$u=p^{\alpha-1}+(qv+rw)(1-p^{\alpha-1})/(1-p),$$
$$v=(pu+rw)(1-q^{\beta-1})/(1-q),$$
$$w=pu+qv+rw=x.$$

3. $u=p^{\alpha-1}+(qv+rw)(1-p^{\alpha-1})/(1-p)$,
$v=(pu+rw)(1-q^{\beta-1})/(1-q)$,
$w=(pu+qv)(1-r^{\gamma-1})/(1-r)$.
4. Заметить, что $\mathbf{P}\{A_n\}<(2p)^n$, но

$$\mathbf{P}\{A_n\}>1-(1-p^n)^{2^n/(2n)}>1-e^{(-2p)^n/(2n)}.$$

Если $p=1/2$, то последняя величина имеет порядок $1/(2n)$, если $p>1/2$, то $\mathbf{P}\{A_n\}$ даже не стремится к нулю.

Глава IX

1. Возможны комбинации (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (2, 0), (2, 1), (3, 0). Их вероятности равны 0,047539, 0,108883, 0,017850, 0,156364, 0,214197, 0,321295, 0,026775, 0,107098.
2. а) Совместное распределение задается (6×6)-матрицей. Главная диагональ состоит из элементов q, $2q$, ..., $6q$, где $q=1/36$. По одну сторону от главной диагонали все элементы равны нулю, по другую все они равны q. б) $\mathbf{E}(\mathbf{X})=7/2$, $\mathrm{Var}(\mathbf{X})=35/12$, $\mathbf{E}(\mathbf{Y})=161/36$, $\mathrm{Var}(\mathbf{Y})=2555/1296$, $\mathrm{Cov}(\mathbf{X},\ \mathbf{Y})=105/72$.
3. Строки совместного распределения равны (перед всеми строками подразумевается множитель 32^{-1})

для $\mathbf{X}$, $\mathbf{Y}$: (1, 0, 0, 0, 0, 0), (0, 5, 4, 3, 2, 1), (0, 0, 6, 6, 3, 0), (0, 0, 0, 1, 0, 0);
для $\mathbf{X}$, $\mathbf{Z}$: (1, 0, 0, 0, 0, 0), (0, 5, 6, 1, 0, 0), (0, 0, 4, 6, 1, 0), (0, 0, 0, 3, 2, 0), (0, 0, 0, 0, 2, 0), (0, 0, 0, 0, 0, 1);

для Y, Z: (1, 0, 0, 0), (0, 5, 6, 1), (0, 4, 7, 0), (0, 3, 2, 0), (0, 2, 0, 0), (0, 1, 0, 0);
распределение X + Y: (1, 0, 5, 4, 9, 8, 5), значения X + Y меняются от 0 до 6; распределение XY: (1, 5, 4, 3, 8, 1, 6, 0, 3, 1), значения XY меняются от 0 до 9.
$\mathbf{E}(\mathbf{X}) = 5/2$, $\mathbf{E}(\mathbf{Y}) = 3/2$, $\mathbf{E}(\mathbf{Z}) = 31/16$, $\operatorname{Var}(\mathbf{X}) = 5/4$, $\operatorname{Var}(\mathbf{Y}) = 3/8$, $\operatorname{Var}(\mathbf{Z}) = 303/256$.

4. а) $p/(1+q)$; б) $1/(1+q+q^2)$; в) $1/(1+q)^2$.
8. Распределение $\mathbf{V}_n$ задается формулой (3.5), для $\mathbf{U}_n$ оно получается по симметрии.
9. а) $\mathbf{P}\{\mathbf{X} \leqslant r,\ \mathbf{Y} \geqslant s\} = N^{-n}(r-s+1)^n$ для $r \geqslant s$;

$$\mathbf{P}\{\mathbf{X}=r,\ \mathbf{Y}=s\} = \begin{cases} N^{-n}\{(r-s+1)^n - 2(r-s)^n + (r-s-1)^n\}, & \text{если } r > s, \\ N^{-n}, & \text{если } r = s, \end{cases}$$

б) $x = \dfrac{r^{n-2}-(r-1)^{n-2}}{r^n-(r-1)^n}$, если $j < r$ и $k < r$;

$x = \dfrac{r^{n-2}}{r^n-(r-1)^n}$, если $j \leqslant r$ и $k = r$ или $j = r$ и $k \leqslant r$;

$x = 0$, если $j > r$ или $k > r$.

в) $\sigma^2 \approx nN^2/[(n+1)^2(n+2)]$.

10. Вероятность того, что будет n двойных бросаний, равна $2pq(p^2+q^2)^{n-1}$. Математическое ожидание равно $1/(2pq)$.
12. $\mathbf{P}\{\mathbf{N}=n,\ \mathbf{K}=k\} = \binom{n}{k} p^{n-k}(qq')^k \cdot qp'$ для $k \leqslant n$;

$\mathbf{P}\{\mathbf{N}=n\} = (1-qp')^n qp'$;

$\mathbf{P}\{\mathbf{K}=k\} = (qq')^k qp' \sum \binom{-k-1}{\nu}(-p)^\nu = p'q'^k$.

13. $$\mathbf{E}\left(\frac{\mathbf{K}}{\mathbf{N}+1}\right) = \sum_{k,\,n} kp_{k,\,n}/(n+1) = q^2p'q' \sum_{n=1}^{\infty}\left(1-\frac{1}{n+1}\right)(p+qq')^{n-1} =$$
$$= qq'/(1-qp') - q^2p'q'/(1-qp')^2 \log[1/(qp')];$$
$$\mathbf{E}(\mathbf{K}) = q'/p';\quad \mathbf{E}(\mathbf{N}) = (1-qp')/(qp');\quad \operatorname{Cov}(\mathbf{K},\ \mathbf{N}) = q'/(qp'^2),$$
$$\rho(\mathbf{K},\ \mathbf{N}) = \sqrt{q'/(1-qp')}.$$

14. а) $p_k = p^kq + q^kp$; $\mathbf{E}(\mathbf{X}) = pq^{-1}+qp^{-1}$; $\operatorname{Var}(\mathbf{X}) = pq^{-2}+qp^{-2}-2$.
б) $q_k = p^2q^{k-1}+q^2p^{k-1}$; $\mathbf{P}\{\mathbf{X}=m,\ \mathbf{Y}=n\} = p^{m+1}q^n + q^{m+1}p^n$ с $m, n \geqslant 1$; $\mathbf{E}(\mathbf{Y}) = 2$; $\sigma^2 = 2(pq^{-1}+qp^{-1}-1)$.
17. $\binom{n}{k} 364^{n-k}365^{1-n}$.
18. а) $365\{1-364^n \cdot 365^{-n} - n364^{n-1} \cdot 365^{-n}\}$; б) $n \geqslant 28$.
19. а) $\mu = n$, $\sigma^2 = (n-1)n$; б) $\mu = (n+1)/2$, $\sigma^2 = (n^2-1)/12$.
20. $\mathbf{E}(\mathbf{X}) = np_1$; $\operatorname{Var}(\mathbf{X}) = np_1(1-p_1)$; $\operatorname{Cov}(\mathbf{X},\ \mathbf{Y}) = -np_1p_2$.
21. $-n/36$. Это частный случай задачи 20.
25. $\mathbf{E}(\mathbf{Y}_r) = \sum_{k=1}^{r} \dfrac{N}{r-k+1}$; $\operatorname{Var}(\mathbf{Y}_r) = \sum_{k=1}^{r} \dfrac{N(N-r+k-1)}{(r-k+1)^2}$.
26. а) $1-q^k$; б) $\mathbf{E}(\mathbf{X}) = N\{1-q^k+k^{-1}\}$; в) $d\mathbf{E}(\mathbf{X})/dk = 0$.
27. $\sum(1-p_j)^n$. Положить $\mathbf{X}_j = 1$ или 0 в соответствии с тем, представлен или нет j-й класс.
28. $\mathbf{E}(\mathbf{X}) = \dfrac{r_1(r_2+1)}{r_1+r_2}$; $\operatorname{Var}(\mathbf{X}) = \dfrac{r_1r_2(r_1-1)(r_2+1)}{(r_1+r_2-1)(r_1+r_2)^2}$.

30. $\mathbf{E}(\mathbf{S}_n)=\frac{nb}{b+r}$; $\quad \operatorname{Var}(\mathbf{S}_n)=\frac{nbr\,(b+r+nc)}{(b+r)^2\,(b+r+c)}$.

33. $$\mathbf{E}\left(\frac{r}{\mathbf{X}}\right)=r\sum_{k=r}^{\infty}k^{-1}\binom{k-1}{r-1}p^r q^{k-r}=$$
$$=\sum_{k=1}^{r-1}(-1)^{k-1}\frac{r}{r-k}\left(\frac{p}{q}\right)^k+\left(\frac{-p}{q}\right)^r r\log p.$$

Чтобы вывести последнюю формулу из первой, положить $f(q)=$ $=r\sum k^{-1}\binom{k-1}{r-1}q^k$. Используя формулу (12.4) гл. II, найти $f'(q)=$ $=rq^{r-1}(1-q)^{-r}$. Утверждение теперь доказывается повторным интегрированием по частям.

Глава XI

1. $sP(s)$ и $P(s^2)$.
2. а) $(1-s)^{-1}P(s)$; б) $(1-s)^{-1}sP(s)$; в) $\{1-sP(s)\}/(1-s)$;
   г) $p_0s^{-1}+\{1-s^{-1}P(s)\}/(1-s)$; д) $(1/2)\{P(\sqrt{s})+P(-\sqrt{s})\}$.
3. $U(s)=pqs^2/[(1-ps)(1-qs)]$. Математическое ожидание равно $1/(pq)$; дисперсия равна $(1-3pq)/(p^2q^2)$.
6. Производящая функция удовлетворяет квадратному уравнению $A(s)=$ $=A^2(s)+s$. Поэтому $A(s)=(1/2)-(1/2)\sqrt{1-4s}$ и $a_n=n^{-1}\binom{2n-2}{n-1}$.
10. а) $\Phi^r(s)F^k(s)\,|\,p-q\,|$;
    б) $\Phi^r(s)\,[1+F(s)+\ldots+F^k(s)]\,|\,p-q\,|$.
11. а) $(q/p)^r\,\Phi^{2r}(s)$; б) $(q/p)^r\,\Phi^{2r}(s)\,U(s)$.
12. Используя производящую функцию случайной величины $\mathbf{X}_\nu$, имеющей геометрическое распределение, сразу получаем
$$P_r(s)=s^r\cdot\frac{N-1}{N-s}\,\frac{N-2}{N-2s}\cdots\frac{N-r+1}{N-(r-1)s}.$$
13. $P_r(s)\{N-(r-1)s\}=P_{r-1}(s)(N-r-1)s$.
14. $$P_r(s)=\frac{s}{N-(N-1)s}\,\frac{2s}{N-(N-2)s}\cdots\frac{rs}{N-(N-r)s}.$$
15. $\mathbf{S}_r$ является суммой r независимых случайных величин с общим геометрическим распределением. Поэтому
$$P_r(s)=\left(\frac{q}{1-ps}\right)^r,\quad p_{r,k}=q^rp^k\binom{r+k-1}{k}.$$
16. а) $$\mathbf{P}\{\mathbf{R}=r\}=\sum_{k=0}^{\nu-1}\mathbf{P}\{\mathbf{S}_{r-1}=k\}\,\mathbf{P}\{\mathbf{X}_r\geqslant\nu-k\}=$$
$$=\sum_{k=0}^{\nu-1}q^{r-1}p^k\binom{r+k-2}{k}p^{\nu-k}=p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1},$$
$$\mathbf{E}(\mathbf{R})=1+q^\nu/p,\quad \operatorname{Var}(\mathbf{R})=\nu q/p^2;$$
    б) $$(p_1p_2)^N\sum_{\nu=1}^{\infty}\binom{N+\nu-2}{\nu-1}^2(q_1q_2)^{\nu-1},$$

17. Учесть, что

$$1+s+\ldots+s^{ab-1}=(1+s+\ldots+s^{a-1})\,(1+s^a+s^{2a}+\ldots+s^{(b-1)\,a}).$$

21. $u_n=q^n+\sum\limits_{k=3}^{n}\binom{k-1}{2}p^3q^{k-3}u_{n-k}$, где $u_0=1$, $u_1=q$, $u_2=q^2$, $u_3=p^3+q^3$. Используя тот факт, что это рекуррентное соотношение является соотношением типа свертки, получаем

$$U(s)=\frac{1}{1-qs}+\frac{(ps)^3}{(1-qs)^3}\,U(s).$$

22. $u_n=pw_{n-1}+qu_{n-1}$, $v_n=pu_{n-1}+qv_{n-1}$, $w_n=pv_{n-1}+qw_{n-1}$. Поэтому $U(s)-1=psW(s)+qsU(s)$, $V(s)=psU(s)+qsV(s)$, $W(s)=psV(s)+qsW(s)$.

Глава XIII

1. Достаточно показать, что если $F(s)=1$ и $s\neq 1$, то $|s|\geqslant 1$, причем равенство $|s|=1$ возможно только в периодическом случае.

2. $u_{2n}\left\{\binom{2n}{n}2^{-2n}\right\}^r\sim\frac{1}{\sqrt{(\pi n)^r}}$. Поэтому $\mathscr{E}$ возвратно только при $r=2$. При $r=3$ численное интегрирование методом касательных дает [1])

$$u-1=\sum_{n=1}^{\infty}u_{2n}\approx(1/\sqrt{\pi^3})\int\limits_{1/2}^{\infty}\left[1/\sqrt{x^3}\right]dx=\sqrt{(2/\pi)^3}\approx 1/2.$$

3. $u_{6n}\sim\sqrt{6/(2\pi n)^5}$. Поэтому $u-1\approx\sqrt{6/(2\pi)^5}\int\limits_{1/2}^{\infty}x^{-5/2}\,dx$. Значит [2]), $u\approx$ $\approx 1{,}047$ и $f\approx 0{,}045$.

4. $u_{(\lambda+1)\,n}=\binom{(\lambda+1)\,n}{n}p^{\lambda n}q^n$. Верхняя грань отношений последовательных членов меньше 1, если $p\neq\lambda/(\lambda+1)$. (Это утверждение следует также из закона больших чисел.)

6. Из оценки $\sum f_i+\mathbf{P}\{\mathbf{X}_1>0\}\leqslant 1$ вывести, что $f<1$, если $\mathbf{P}\{\mathbf{X}_1>0\}\neq 0$. Если $\mathbf{P}\{\mathbf{X}_1>0\}=0$, то все $\mathbf{X}_i\leqslant 0$ и $\mathscr{E}$ либо происходит при первом испытании, либо не происходит никогда.

7. $\mathbf{Z}_n=\min\{m\colon m\geqslant(n-\mathbf{N}_n)/r,\ m$ целое$\}$. Далее, $\mathbf{E}(\mathbf{Z}_n)\sim np/(q+pr)$, $\operatorname{Var}\mathbf{Z}_n\sim npq\,(q+pr)^{-3}$.

8. $G(s)=\dfrac{(1-qs)\,q^rs^r}{1-s+pq^rs^{r+1}}$, $F(s)=qs+psG(s)$, $\mu=q^{-r}$.

9. $G(s)=\dfrac{(1-qs)\,B(qs)}{1-s+psB(qs)}$, $F(s)$ та же, что в задаче 8.

11. $\mathbf{N}_n^*\approx(\mathbf{N}_n-714{,}3)/22{,}75$; $\mathfrak{N}(2/3)-\mathfrak{N}(-2/3)\approx 1/2$.

12. $r_0=r_1=r_2=1$, $r_n=r_{n-1}-(1/4)\,r_{n-2}+(1/8)\,r_{n-3}$,

$$R(s)=(8+2s^2)\,(8-8s+2s^2-s^3)^{-1},\qquad r_n\sim 1{,}444248\,(1{,}139680)^{n-1}.$$

14. Если a_n — вероятность появления A-серии длины r при n-м испытании, то $A(s)$ определяется формулой (7.5) с заменой p на α и q на $1-\alpha$.

---

[1]) Подынтегральное выражение плохо приближает первые члены суммы, дающие в нее основной вклад. Более громоздкие вычисления показывают, что $u-1\sim 0{,}393$, и поэтому $f=1-u^{-1}\approx 0{,}282$.— *Прим. перев.*

[2]) См. примечание к предыдущему ответу. Более точные значения: $u\approx 1{,}0225$ и $f\approx 0{,}0220$.— *Прим. перев.*

Пусть $B(s)$ и $C(s)$ — соответствующие функции для B- и C-серий. Искомыми производящими функциями являются $[(1-s)U(s)]^{-1}$, где в случае а) $U(s)=A(s)$, в случае б) $U(s)=A(s)+B(s)-1$, в случае в) $U(s)=A(s)+B(s)+C(s)-2$.

15. Скомбинировать методы, использованные в примере 8, б) и в решении задачи 14.
16. Математическое ожидание для возраста k равно Npq^k.
18. $w_k(n)=v_{n-k}r_k$, если $n>k$, и $w_k(n)=\beta_{k-n}r_k/r_{k-n}$, если $n\leqslant k$.
19. Учесть, что $1-F(s)=(1-s)Q(s)$ и $\mu-Q(s)=(1-s)R(s)$, причем $Q(1)=\mu$, $2R(1)=\sigma^2-\mu+\mu^2$. Степенной ряд для $Q^{-1}(s)=\sum(u_n-u_{n-1})s^n$ при $s=1$ сходится.

Глава XIV

1. $((q/p)^b-1)/((q/p)^{a+b}-1)$ при $p\neq q$ и $b/(a+b)$ при $p=q$.
3. При $q<p$ число попаданий является дефектной случайной величиной.
4. Математическое ожидание числа попаданий равно $p(1-q_1)/(qq_0-1)= =(p/q)^a$.
5. Вероятность разорения по-прежнему задается формулой (2.4) при $p=\alpha(1-\gamma)^{-1}$, $q=\beta(1-\gamma)^{-1}$. Ожидаемая продолжительность игры равна $D_z(1-\gamma)^{-1}$, где D_z определяется по (3.4) или (3.5).
6. Граничные условия (2.2) заменяются условиями $q_0-\delta q_1=1-\delta$, $q_a=0$. Решению (2.4) соответствует решение
$$q_z=\frac{\{(q/p)^a-(q/p)^z\}(1-\delta)}{(q/p)^a(1-\delta)+\delta q/p-1}.$$
Граничные условия (3.2) превращаются в $D_0=\delta D_1$, $D_a=0$.
7. Соотношению (2.1) здесь соответствует уравнение $q_z=pq_{z+2}+qq_{z-1}$, и $q_z=\lambda^z$ будет частным решением последнего, если $\lambda=p\lambda^3+q$, т. е. если $\lambda=1$ или $\lambda^2+\lambda=qp^{-1}$. Вероятность разорения равна
$$q_z=\begin{cases}1, & \text{если } q\geqslant 2p,\\ (\sqrt{1/4+q/p}-1/2)^z, & \text{если } q\leqslant 2p.\end{cases}$$
10. $w_{z,n+1}(x)=pw_{z+1,n}(x)+qw_{z-1,n}(x)$ с граничными условиями 1) $w_{0,n}(x)= =w_{a,n}(x)=0$, 2) $w_{z,0}(x)=0$ при $z\neq x$ и $w_{x,0}(x)=1$.
11. Заменить 1) на $w_{0,n}(x)=w_{1,n}(x)$ и $w_{a,n}(x)=w_{a-1,n}(x)$.
12. Граничное условие: $u_{a,n}=u_{a-1,n}$. Производящая функция:
$$\frac{\lambda_1^z(s)\lambda_2^{a-1/2}(s)+\lambda_2^z(s)\lambda_1^{a-1/2}(s)}{\lambda_1^{a-1/2}(s)+\lambda_2^{a-1/2}(s)}=\frac{\lambda_1^{a-z-1/2}(s)+\lambda_2^{a-z-1/2}(s)}{\lambda_1^{a-1/2}(s)+\lambda_2^{a-1/2}(s)}.$$
18. $\mathbf{P}\{\mathbf{M}_n<z\}=\sum_{x=1}^{\infty}(v_{x-z,n}-v_{x+z,n})$, $\mathbf{P}\{\mathbf{M}_n=z\}=\mathbf{P}\{\mathbf{M}_n<z+1\}-\mathbf{P}\{\mathbf{M}_n<z\}$.
19. Первое достижение x должно произойти в момент $k\leqslant n$, и за последующие $n-k$ шагов частица возвращается в x.
31. Соотношение (8.2) заменяется на
$$U_z(s)=s\sum_{x=1}^{a-1}U_x(s)p_{x-z}+sr_z.$$
Характеристическое уравнение имеет вид $s\sum p_k\sigma^k=1$.

Глава XV

1. P имеет строки $(p, q, 0, 0)$, $(0, 0, p, q)$, $(p, q, 0, 0)$ и $(0, 0, p, q)$. При $n>1$ строки таковы: (p^2, pq, pq, q^2).

2. а) Цепь неприводима и эргодична; $p_{jk}^{(n)} \to 1/3$ при всех j, k. (Заметим, что P дважды стохастическая.)
б) Цепь имеет период 3, класс G_0 содержит E_1 и E_2; состояние E_4 образует класс G_1, а E_3 — класс G_2. Имеем $u_1=u_2=1/2$, $u_3=u_4=1$.
в) Состояния E_1 и E_3 образуют замкнутое множество S_1, а E_4 и E_5 — другое замкнутое множество S_2, тогда как E_2 невозвратно. Матрицы, соответствующие замкнутым множествам, суть (2×2)-матрицы с элементами $1/2$. Следовательно, $p_{jk}^{(n)} \to 1/2$, если E_j и E_k принадлежат одному и тому же S_r; $p_{j2}^{(n)} \to 0$; наконец, $p_{2k}^{(n)} \to 1/2$, если $k=1,\ 3$, и $p_{2k}^{(n)} \to 0$, если $k=2,\ 4,\ 5$.
г) Цепь имеет период 3. Полагая $a=(0,\ 0,\ 0,\ 1/3,\ 1/3,\ 1/3)$, $b=(1,\ 0,\ 0,\ 0,\ 0,\ 0)$, $c=(0,\ 1/2,\ 1/2,\ 0,\ 0,\ 0)$, мы находим, что строки матриц $P^2=P^5=\ldots$ суть $a,\ b,\ b,\ c,\ c,\ c$, строки $P^3=P^6=\ldots$ суть $b,\ c,\ c,\ a,\ a,\ a$, а строки $P=P^4=\ldots$ суть $c,\ a,\ a,\ b,\ b,\ b$.

3. $p_{jj}^{(n)}=(j/6)^n$, $p_{jk}^{(n)}=(k/6)^n-((k-1)/6)^n$ при $k>j$ и $p_{jk}^{(n)}=0$ при $k<j$.

4. $x_k=(3/4,\ 1/2,\ 1/4,\ 1/2)$, $y_k=(1/4,\ 1/2,\ 3/4,\ 1/2)$.

6. При $n\geqslant j$

$$f_{j0}^{(n)}=\binom{n-1}{j-1}p^{n-j}q^j=\binom{j}{n-j}(-p)^j q^j.$$

Производящая функция $(qs)^j(1-ps)^{-j}$. Математическое ожидание равно j/q.

7. $$f_{00}^{(n)}=\sum_{k=1}^{n=1} v_k\binom{n-2}{k-1}p^{n-1-k}q^k \quad\text{при } n>1.$$

8. Состояния с четными номерами образуют неприводимое замкнутое множество. Вероятность возвращения в E_0 не более чем за n шагов равна

$$1-v_0+v_0(1-v_2)+v_0v_2(1-v_4)+\ldots+v_0v_2\ldots v_{2n-2}(1-v_{2n})=\\ =1-v_0v_2v_4\ldots v_{2n}.$$

Поэтому состояния с четными номерами возвратны тогда и только тогда, когда последнее произведение стремится к 0. Вероятность того, что выходящая из E_{2r+1} система останется навсегда среди состояний с нечетными номерами (невозвратных состояний), равна $v_{2r+1}v_{2r+3}\ldots$.

9. $u_r=[1-p/q]\,(p/q)^{r-1}\,[1-(p/q)^\rho]^{-1}$.

10. Возможные состояния $E_0,\ \ldots,\ E_w$. При $j>0$

$$p_{j,\,j-1}=j(\rho-w+j)\rho^{-2},\quad p_{j,\,j+1}=(\rho-j)(w-j)\rho^{-2},$$
$$p_{jj}=j(w-j)\rho^{-2}+(\rho-j)(\rho-w+j)\rho^{-2},$$
$$u_k=\binom{w}{k}\binom{b}{\rho-k}\binom{2\rho}{\rho}^{-1}.$$

13. $$P=\begin{bmatrix} q & p & 0 & 0 & \ldots & 0 & 0\\ 0 & 0 & 1 & 0 & \ldots & 0 & 0\\ 0 & 0 & 0 & 1 & \ldots & 0 & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & 0 & \ldots & 0 & 1\\ q & p & 0 & 0 & \ldots & 0 & 0\end{bmatrix}.$$

14. Заметим, что матрица дважды стохастическая; использовать пример 7, з).

15. Положить $p_{k,\,k+1}=1$ при $k=1,\ \ldots,\ N-1$ и $p_{Nk}=p_k$.

16. $\sum u_jp_{jk}=u_k$; тогда $U(s)=u_0(1-s)\,P(s)\,\{P(s)-s\}^{-1}$. Для эргодичности необходимо и достаточно, чтобы $\mu=P'(1)<1$. По правилу Лопиталя $U(1)=u_0(1-\mu)$, откуда $u_0=(1-\mu)^{-1}$.

25. Если $n\geqslant m-2$, то величины $\mathbf{X}^{(m)}$ и $\mathbf{X}^{(n)}$ независимы и, следовательно,

все три строки матрицы $p_{jk}^{(m, n)}$ совпадают с распределением $\mathbf{X}^{(n)}$, а именно с (1/4, 1/2, 1/4). При $n = m+1$ строки матрицы суть (1/2, 1/2, 0), (1/4, 1/2, 1/4), (0, 1/2, 1/2).

Глава XVII

3. $\mathbf{E}(\mathbf{X}) = ie^{\lambda t}$, $\quad \mathrm{Var}(\mathbf{X}) = ie^{\lambda t}(e^{\lambda t} - 1)$.
4. $P_n' = -\lambda n P_n + \lambda(n+1) P_{n+1}$,
$$P_n(t) = \binom{i}{n} e^{-i\lambda t}(e^{\lambda t} - 1)^{i-n} \quad (n \leqslant i),$$
$$\mathbf{E}(\mathbf{X}) = ie^{-\lambda t}, \quad \mathrm{Var}(\mathbf{X}) = ie^{-\lambda t}(1 - e^{-\lambda t}).$$
6. $P_n'(t) = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t)$ при $n \leqslant N-1$ и $P_N'(t) = -N\mu P_N(t) + \lambda P_{N-1}(t)$.
10. Обычный метод решения линейных дифференциальных уравнений приводит к системе линейных уравнений.

ИМЕННОЙ УКАЗАТЕЛЬ

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

Вильям Феллер

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ
И ЕЕ ПРИЛОЖЕНИЯ

В 2-х томах

Том 1

Ст. научный редактор Г. М. Ильичева
Мл. научный редактор Т. А. Денисова
Художник Е. И. Волков
Художественный редактор В. И. Шаповалов
Технический редактор Н. И. Борисова
Корректор Е. Г. Литвак

ИБ № 3127

Сдано в набор 26.09.83. Подписано к печати 18.04.84. Формат 60×90 1/16. Бумага типографская № 2. Объем 16,50 бум. л. Усл. кр.-отт. 33,00. Гарнитура литературная. Печать высокая. Усл. печ. л. 33,00. Уч.-изд. л. 33,28. Изд. № 1/1766. Тираж 40 000 экз. Заказ № 221. Цена 2 р. 60 к.

ИЗДАТЕЛЬСТВО «МИР»
129820, Москва, И-110, ГСП, 1-й Рижский пер., 2

Набрано и сматрицировано в ордена Октябрьской Революции и ордена Трудового Красного Знамени Первой Образцовой типографии имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28

Отпечатано в Ленинградской типографии № 2 головном предприятии ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.