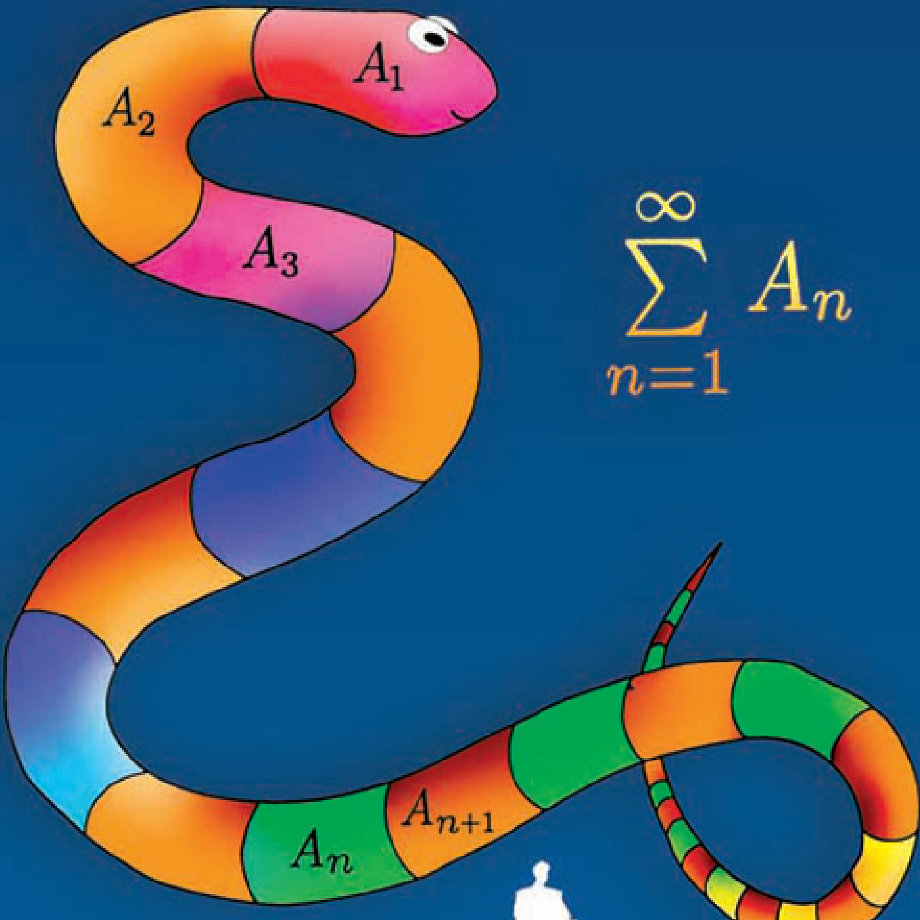
К. Л. Чжун, Ф. АитСахлиа

ЭЛЕМЕНТАРНЫЙ КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ И ФИНАНСОВАЯ МАТЕМАТИКА



$$\sum_{n=1}^{\infty} A_n$$

ЭЛЕМЕНТАРНЫЙ КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ И ФИНАНСОВАЯ МАТЕМАТИКА

Kai Lai Chung, Farid AitSahlia

ELEMENTARY PROBABILITY THEORY

WITH STOCHASTIC PROCESSES AND AN INTRODUCTION TO MATHEMATICAL FINANCE

Fourth Edition

With 57 Figures

Springer

К. Л. Чжун, Ф. АитСахлиа

ЭЛЕМЕНТАРНЫЙ КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ И ФИНАНСОВАЯ МАТЕМАТИКА

Перевод с 4-го английского
издания М. Б. Лагутина

2-е ИЗДАНИЕ (ЭЛЕКТРОННОЕ)

Москва
БИНОМ. Лаборатория знаний
2014

УДК 519.2
ББК 22.17
Ч-57

Чжун К. Л.

Ч-57 Элементарный курс теории вероятностей. Стохастические процессы и финансовая математика [Электронный ресурс] / К. Л. Чжун, Ф. АитСахлиа ; пер. с англ. — 2-е изд. (эл.). — М. : БИНОМ. Лаборатория знаний, 2014. — 455 с. : ил.

ISBN 978-5-9963-1317-4

Перевод 4-го издания популярного учебника по теории вероятностей и ее приложениям, написанного известными американскими математиками из Станфордского университета. Четвертое издание дополнено двумя новыми главами, посвященными финансовой математике.

Для студентов, преподавателей, исследователей и практиков в экономике, психологии, социологии, медицине и в других областях, где используются статистические методы и теория вероятностей.

УДК 519.2
ББК 22.17

По вопросам приобретения обращаться:
«БИНОМ. Лаборатория знаний»
Телефон: (499) 157-5272
e-mail: binom@Lbz.ru, http://www.Lbz.ru

ISBN 978-5-9963-1317-4

Предисловие к четвертому изданию

В этом издании добавлены две новые главы, 9 и 10, посвященные финансовой математике. Они написаны доктором Фаридом АитСахлиа, ancien élève, который читал курс лекций по данному предмету и работал в исследовательских отделах нескольких промышленных и финансовых организаций.

Новый текст начинается с подробного перечня специфических финансовых терминов, относящихся к разнообразным опционам и другим сделкам с ценными бумагами, подверженным влиянию колебаний курсов на финансовом рынке. Экономические процессы описаны на ясном и точном математическом языке, что позволяет исследовать их с помощью вероятностных методов, изложенных в предшествующих главах. Многочисленные практические примеры и задачи различной степени трудности иллюстрируют применимость теории к решению реальных финансовых проблем. Для понимания материала двух последних глав читателю, который интересуется в основном финансовыми вопросами, достаточно лишь поверхностно ознакомиться с содержанием первых восьми глав. Более углубленное изучение можно осуществить с помощью предметного указателя.

При переиздании я воспользовался возможностью заполнить брешь в разд. 8.1 и расширил дополнение 3, поместив в него полезное утверждение о мартингале, остановленном в случайный момент времени. Последнее понятие играет важную роль в более продвинутой теории финансового рынка и других дисциплинах. Несмотря на это, уровень нашего изложения остается *элементарным*, соответственно названию и проекту книги. Добавлены также несколько реальных эпизодов из области современных финансов, чтобы разрядить, для начинающих, многослойную атмосферу экономической науки. Мы благодарны Руфи Уильямс за прочтение черновика новых глав и внесение ценных предложений, а также Бернару Брю и Марку Барбю за информацию о законах Парето—Леви, введенных с целью описания распределения доходов. Доступный, по нашему мнению, обзор этой хорошо известной теории содержится в новом дополнении 4.

К. Л. Чжун, 3 августа 2002 г.

Предисловие к третьему изданию

Новой деталью этого издания является включение в него восьми фотографий ученых, внесших значительный вклад в современную теорию вероятностей*). Все они упоминаются в тексте, хотя, конечно, несколько ссылок не могут адекватно отразить их выдающуюся роль. Я надеюсь, что эти одухотворенные лица вызовут у читателя ощущение, что наша наука — живой организм, созданный и развиваемый реально существующими людьми. Мне посчастливилось встречаться или быть знакомым почти со всеми из них после того, как я изучил их работы, и теперь мне приятно представить этих ученых более молодому поколению. В получении фотографий мне помогали Мари-Хелен Шварц, Джоан Эллиот, Мило Кейнс и Ю. А. Розанов, которым я искренне благодарен.

Совсем недавно моя книга была издана на немецком языке. Я крайне признателен д-ру Герберту Фогту за аккуратный и точный перевод, в результате чего в текст текущего издания внесено много улучшений. Я также благодарю читателей, любезно приславших свои замечания: Марвина Гринберга, Луизу Хэй, Нору Холмквист, Х.-И. Ламана и Фреда Уолока. Еще раз выражаю свою признательность издательству «Шпрингер» за желание довести свои публикации «до совершенства».

К. Л. Чжун, 19 сентября 1978 г.

Предисловие ко второму изданию

При подготовке второго издания основное внимание было уделено исправлению ошибок, допущенных в первом издании. В этой работе участвовали Чао Хунг-по, Дж. Л. Дуб, Р. М. Экснер, В. Х. Флеминг, А. М. Глисон, Карен Кафадор, С. Х. Полит и П. ван Морбек. Г-жа Кафадор и д-р Полит аккуратно объединили все списки замечаний. Наиболее огорчительные ошибки были обнаружены в решениях задач. Все решения проверены мною лично в главах 1–5 и господином Чао — в главах 6–8. Я очень надеюсь, что ошибок там почти не осталось. Также были внесены некоторые усовершенствования и добавления, но не все рекомендации удалось реализовать к настоящему моменту. Просим читателей присылать критические замечания и комментарии для принятия их во внимание в будущем издании. Благодарю персонал издательства «Шпрингер» за предоставление возможности внести данные коррективы сразу после выхода в свет этой книги.

К. Л. Чжун

---

*) В издании на русском языке эти фотографии отсутствуют. — *Прим. ред.*

Предисловие к первому изданию

За последние полвека теория вероятностей выросла из незначительного изолированного предмета в разветвленную и бурно развивающуюся дисциплину, оказывающую влияние на многие другие направления математики. В то же время она играет важнейшую роль в математизации многих прикладных наук, таких как статистика, исследование операций, биология, экономика и психология (в названиях некоторых из них до сих пор прочно закрепилось слово «математическая»). Факт достижения теорией вероятностей совершеннолетия отразился на материале учебников по этому предмету. В прежние времена большинство таких книг демонстрировали явное «раздвоение личности»: в них излагался либо комбинаторный подход к описанию игр, в которых присутствует случайность, либо так называемая «теория ошибок» наблюдений, связанная с нормальным распределением. Данный период завершился с появлением в 1950 г. классической монографии У. Феллера [9], из которой я отбирал материал для моего первого курса по теории вероятностей. Спустя некоторое время теория вероятностей и ее приложения уверенно заняли свое место в учебных программах в качестве математической дисциплины, необходимой для многих областей знаний. Начальные сведения по этой теории теперь преподаются на разных уровнях сложности, иногда даже до изучения математического анализа. Данная книга предназначена для составления курса, ориентированного на студентов, скорее всего, второго года обучения. Не предполагается никакого предварительного знакомства с предметом, первые три главы можно читать, почти совсем не зная математического анализа. Следующие три главы уже требуют умения обращаться с бесконечными рядами и близкими к ним понятиями. При рассмотрении случайных величин, имеющих плотность, необходимо владеть основами математического анализа. Подобные темы, относящиеся к «непрерывному случаю» (как противоположности «дискретного»), легко отделяются от остальных и могут быть отложены. Материал первых шести глав, как нам представляется, должен образовывать каркас любого содержательного введения в теорию вероятностей. Рекомендуем дополнить его следующими разделами: 7.1 (распределение Пуассона, которое даже может быть включено в основной курс), 7.3, 7.4, 7.6 (нормальное распределение и закон больших чисел), 8.1 (простое случайное блуждание, которое не только вызывает интерес, но и полезно). Все это можно изложить в течение одного семестра. Для курса продолжительностью в одну четверть придется произвести некоторые сокращения. Конкретно: для такого короткого курса с гл. 1 и 3 можно ознакомиться поверхностно, не рассматривая темы, отмеченные звездочкой. Во всяком случае, глубокое изучение теоремы о нормальной аппроксимации из гл. 7 следует проводить только при условии, что времени достаточно (как в семестровом или двухчетвертном курсе). Заключительная гл. 8 представляет собой замкнутое введение в проблематику цепей Маркова и углубляет основной материал на более высоком уровне. Вместе с отмеченными звездочкой разд. 5.3, 5.4 (последовательный отбор и урновая схема Пойа), 7.2 (пуассоновский процесс) и, вероятно, с какими-то фрагментами из дополнений, данный материал обеспечивает постепенный логичный переход к изучению теории случайных процессов. Курс, включающий все указанные выше темы, рассчитан на две четверти (именно в такой форме я много раз читал его студентам математической и инженерных специализаций). Тем не менее, читатель, проработавший всего шесть глав, сможет перейти к ознакомлению с более тонкими вопросами, обсуждаемыми, например, в уже упоминавшейся монографии У. Феллера. Если же читатель имеет хорошую математическую подготовку, то он будет способен к изучению формального и строгого курса, подобного изложенному в моей более продвинутой книге [4].

Немало усилий ушло на отбор, упорядочение и представление материала с целью приспособления его к условиям преподавания. Однако автор не намеревался предложить некий поверхностный пакет, соответствующий заданному расписанию занятий или программе, который обычно требуется при ускоренной форме образования. Определенная свобода и право выбора остается у преподавателя: он лучше знает, что именно использовать для обучения студентов его группы. Каждая глава начинается с легко читаемого текста, предназначенного для того, чтобы заинтересовать читателя и проиллюстрировать материал. Благодаря этому, преподаватель имеет возможность сконцентрироваться на более формальных аспектах темы. Каждая глава включает также и более сложные разделы (например, 1.4, 2.5) для выборочного изучения. Надеюсь, они не отпугнут начинающих, а послужат приглашением к последующей проработке. Акцент делается на всестороннем и неторопливом обсуждении основных понятий и вычислительных приемов элементарной теории вероятностей с минимумом излишеств и технических сложностей. Многие примеры предназначены для предупреждения затруднений у начинающих и стимулирования мышления. Часто это осуществляется путем постановки ключевых вопросов с последующим сообщением ответов. Исторические, философские и авторские комментарии служат приправой к содержательному предмету. Хотелось бы, чтобы читатель не только узнал из этой книги что-то новое, но, возможно, также получил некоторое удовольствие от самого чтения.

Первые шесть глав включают более двухсот задач, а две последние — более восьмидесяти задач. Многие из них — простые, более сложные отмечены звездочкой. В конце книги даны ответы к задачам. Разделы, отмеченные звездочкой, содержат более специальный или трудный материал. Их можно пропустить, но рекомендуем прежде бегло их просмотреть.

Автор любого элементарного учебника, конечно же, многим обязан многочисленным предшественникам. Несколько персональных благодарностей приведено ниже. Мишель Надзела записал курс лекций, прочитанный мной в Стэнфорде в 1970 г. Жан-Карло Рота, просмотрев эти записи, дал мне начальный импульс для преобразования их в книгу. Д. Г. Кендэлл прокомментировал наброски нескольких первых глав и оказал дальнейшую моральную поддержку. Дж. Л. Дуб вызвался прочитать большую часть рукописи и внес много полезных предложений. К. Б. Эриксон использовал некоторый материал для преподавания. А. А. Балкема проверил практически окончательную версию и исправил много погрешностей. Дэн Рудольф прочитал все доказательства вместе со мной. Перфекто Мэри нарисовал замечательные иллюстрации. Гейл Лемонд напечатала текст с присущей ей быстротой и безошибочностью. Наконец, мне приятно поблагодарить издательство «Шпрингер», с которым я давно сотрудничаю, за выбор моей книги в качестве первой в новой серии учебников для студентов.

К. Л. Чжун, март 1974 г.

О введении в финансовую математику

Две новые главы, 9 и 10, представляют собой замкнутые введения, соответственно, в оптимальное инвестирование на основе средних и дисперсий и проблему определения цены опциона. Тему главы 9 иначе называют современной теорией формирования портфеля ценных бумаг. Она широко используется менеджерами крупных финансовых учреждений. Для понимания материала главы достаточно знания основ теории вероятностей и скромного владения математическим анализом. Глава 9 также содержит обсуждение устойчивых законов в рассматриваемом контексте (эту тему часто не включают в элементарные вероятностные и финансовые курсы). В свою очередь, глава 10 знакомит читателя с теорией вычисления справедливой цены опциона, проблемой, которая способствовала, как отмечено во многих недавно опубликованных книгах по данной тематике, превращению финансового анализа в активно развивающуюся математическую дисциплину. Эта глава может служить введением в мартингальную теорию опционов, выражающую на математическом языке задачу определения цены опциона на рынке. Мартингальная теория изучается в контексте биномиального случайного блуждания. Несмотря на простоту, в данной модели без арбитража проявляется сущность многих общих теоретических результатов. Ее часто используют на практике для получения начального приближения при вычислении цены сложных финансовых инструментов.

Я хотел бы поблагодарить профессора Кай Лай Чжуна за приглашение написать новый материал для четвертого издания, а также мою жену Уннур за моральную поддержку во время этой важной деятельности.

Фарид АитСахлиа,
1 ноября 2002 г.

Глава 1

Теория множеств

1.1. Множества выборочного пространства

В наше время с понятием множества знакомят уже в начальной школе. Ученика второго класса*) попросили указать «множество всех девочек в его классе». Это можно сделать, составив полный список, подобный следующему:

«Нэнси, Флоренс, Салли, Джуди, Энн, Барбара, ...».

Проблема возникает, если встречаются одинаковые имена. Чтобы различить двух Барбар, потребуется записать их фамилии или обозначить их, скажем, B_1 и B_2. Перечисляя элементы множества, нельзя использовать одно и то же название дважды.

Понятие множества находит применение во всех направлениях математики. Так, в геометрии говорят о «множестве точек, находящихся на одинаковом расстоянии от некоторой точки». Это — окружность. В алгебре изучается «множество целых чисел, не имеющих делителей, отличных от 1 и самого себя». Его называют множеством простых чисел. В математическом анализе областью определения функции обычно служит некоторое множество действительных чисел, например, интервал (a, b); то же самое можно сказать и об области значений функции, если вы помните, что это такое.

В теории вероятностей понятие множества играет основополагающую роль. В дальнейшем нам встретятся как множества абстрактного типа, так и просто наборы объектов. Для начала рассмотрим несколько примеров последних:

(a) мешок яблок;
(b) 55 пациентов, применяющих определенный вид лечения против рака;
(c) все студенты колледжа;
(d) молекулы кислорода в некотором сосуде;
(e) всевозможные исходы бросания шести игральных костей;
(f) все точки на поверхности мишени для стрельбы.

---

*) Моего сына Даниэля.

Давайте рассмотрим также и соответствующие «менее емкие» множества:

(a′) гнилые яблоки в мешке;
(b′) пациенты, которым данное лечение помогает;
(c′) студенты, специализирующиеся по математике;
(d′) молекулы, двигающиеся вверх;
(e′) исходы, при которых на всех шести костях выпали разные цифры;
(f′) точки, принадлежащие небольшой области в центре мишени, называемой «глазом быка».

Мы собираемся дать математическую модель для приведенных и многих других аналогичных примеров, которые могут прийти в голову. Другими словами, мы хотим абстрагировать и обобщить интуитивное понятие «совокупности объектов». Прежде всего станем называть объекты точками, а их совокупность — пространством точек или выборочным пространством. Пространство именуют выборочным для того, чтобы отличать от других употреблений термина «пространство», а также с целью указать на его статистическое происхождение. Таким образом, *точка выборочного пространства**) является абстракцией яблока, онкологического больного, студента, молекулы, возможного случайного исхода и обычной геометрической точки. *Выборочное пространство* состоит из некоторого количества точек и просто служит названием для полной совокупности точек. Любое множество из примеров (a)–(f) может рассматриваться как выборочное пространство. Годится также и любое из меньших множеств, определенных в примерах (a′)–(f′). Что именно взять в качестве пространства (*вероятностной «вселенной»*) — решать нам самим.

Выборочное пространство всегда будет обозначаться заглавной греческой буквой Ω. Оно может содержать произвольное количество точек, возможно, бесконечно много точек, но обязательно — не менее одной. (Вы, возможно, не раз замечали, что математические формулировки выглядят довольно педантичными!) Каждую из точек выборочного пространства станем обозначать прописной греческой буквой ω. Чтобы различать точки, используем индексы и штрихи (как в случае двух Барбар, если мы не знаем их фамилий), например: $\omega_1, \omega_2, \omega', \ldots$. Любой набор точек образует *подмножество* пространства Ω, и, коль скоро Ω фиксировано, станем просто называть его множеством. Крайними

*) В вероятностной литературе также используется термин «элементарное событие». — *Прим. перев.*

случаями подмножеств являются само Ω и *пустое множество* $\varnothing$, которое не содержит ни одной точки. Количество точек в множестве S называется его *размером* и обозначается через $|S|$. Размер множества — неотрицательное целое число или ∞. В частности, $|\varnothing| = 0$.

Множество S является корректно определенным, если для всякой точки из Ω известно, принадлежит она S или не принадлежит. Эти два случая записываются соответственно символами

$$\omega \in S \quad \text{и} \quad \omega \notin S.$$

Таким образом, множество задается с помощью некоторого правила, позволяющего устанавливать членство в нем точек выборочного пространства. Например, множества из (a$'$)–(f$'$) определены настолько корректно, насколько однозначно интерпретируются задающие их словесные описания. Конечно, можно спорить о том, что такое «гнилые яблоки» или попытаться пошутить, что при бросании шести игральных костей на тротуар какие-то из них могут провалиться в канализационный люк. Некоторые люди, обладающие псевдофилософским складом ума, любят упражняться в изобретении подобных замечаний, но мы не будем сосредоточиваться на этом.

Одним из не вызывающих сомнений способов задания множества является перечисление всех точек выборочного пространства, входящих в множество, т. е. составление списка всех его элементов (как это сделал второклассник). Однако такой способ может оказаться крайне трудоемким или вообще неосуществимым. Например, как будет показано в § 3.1, размер множества в примере (e) равен $6^6 = 46\,656$. (Попробуйте подсчитать, сколько примерно книжных страниц потребуется для того, чтобы записать все возможные исходы.) С другой стороны, каждый исход бросания шести игральных костей можно представить систематичным и корректным образом в форме упорядоченной шестерки символов:

$$(s_1, s_2, s_3, s_4, s_5, s_6),$$

где s_j, $j = 1, \ldots, 6$, принимают любое из значений от 1 до 6. Этот пример служит хорошей иллюстрацией экономии мышления (и бумаги для печати), которая достигается за счет использования математического языка.

Если каждый элемент множества A принадлежит также и множеству B, то говорят, что A *содержится* или *включено* в B, или A является *подмножеством* B. В свою очередь, B — *надмножество* по отношению к A. Мы обозначаем это двояко:

$$A \subset B, \quad B \supset A.$$

Два множества *идентичны*, если они содержат одни и те же элементы (точки выборочного пространства). В таком случае мы пишем

$$A = B.$$

Другими словами, соотношение $A = B$ равносильно выполнению двух включений: $A \subset B$ и $B \subset A$. Вероятно, последнее утверждение выглядит слишком окольным, однако часто оно дает единственно возможный способ проверки того, что два множества идентичны. Не всегда легко идентифицировать два множества, заданные разными способами. Известно ли вам, скажем, что множество четных целых чисел совпадает с множеством всех решений x уравнения $\sin(\pi x/2) = 0$? Ниже мы дадим несколько примеров установления факта совпадения множеств с помощью указанного «окольного» метода.

1.2. Операции над множествами

Будем изучать множества, производя над ними некоторые операции, подобно тому, как изучаются числа. В случае последних мы также говорим, что проводим вычисления: складываем, вычитаем, умножаем и т. д. В результате числовых операций появляются новые числа, называемые суммой, произведением и т. п. Подобным образом операции над множествами приводят к другим множествам, имеющим свои названия. Обсудим некоторые операции, а также правила, которым они подчиняются.

Дополнение. Дополнение множества A обозначается через A^c и определяется как множество точек, которые не входят в A. Обратите внимание, что мы имеем дело исключительно с точками фиксированного пространства Ω! Это утверждение записывается с помощью формулы

$$A^c = \{\omega \mid \omega \notin A\},$$

которую следует читать так: «A^c — это множество точек ω, не принадлежащих множеству A». В частности, $\Omega^c = \varnothing$ и $\varnothing^c = \Omega$. Операция взятия дополнения обладает следующим свойством: если ее применить дважды, то мы снова получим множество A:

$$(A^c)^c = A. \tag{1.2.1}$$

Объединение. Объединением $A \cup B$ множеств A и B называется множество точек, которые принадлежат хотя бы одному из этих множеств. Символически это записывается так:

$$A \cup B = \{\omega \mid \omega \in A \quad \text{или} \quad \omega \in B\},$$

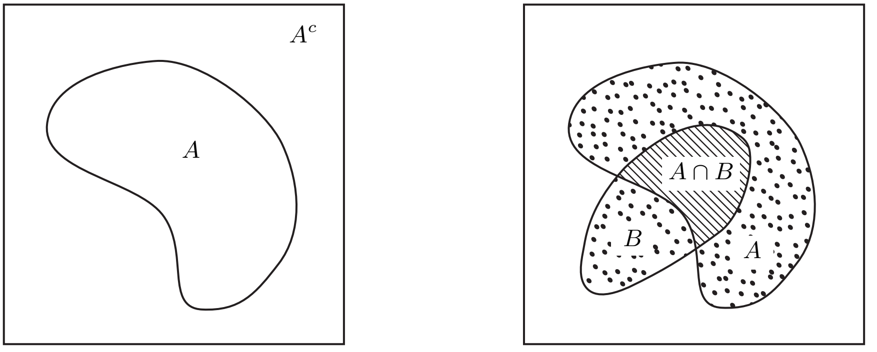


Рис. 1

где «или» обозначает «и/или» («включающее или») на языке формальной логики. Мы всегда будем использовать его именно в этом смысле.

Пересечение. Пересечением $A \cap B$ множеств A и Bу называется множество точек, принадлежащих обоим множествам. В символической форме:

$$A \cap B = \{\omega \mid \omega \in A \quad \text{и} \quad \omega \in B\}.$$

Очевидно, что справедливы следующие правила.

Коммутативный закон: $A \cup B = B \cup A, \quad A \cap B = B \cap A.$

Ассоциативный закон: $(A \cup B) \cup C = A \cup (B \cup C);$
$(A \cap B) \cap C = A \cap (B \cap C).$

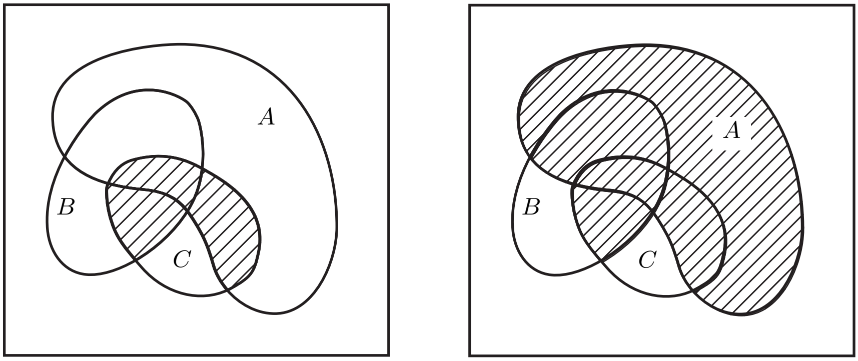
Обратите внимание, что эти соотношения служат примерами рассмотренного выше понятия идентичности множеств и требуют доказательства. Отметим аналогию*) с соответствующими законами для суммы и произведения чисел:

$$a + b = b + a, \qquad a \times b = b \times a,$$
$$(a + b) + c = a + (b + c), \qquad (a \times b) \times c = a \times (b \times c).$$

Скобки необходимы для определения порядка выполнения операций. Согласно ассоциативному закону, мы можем писать

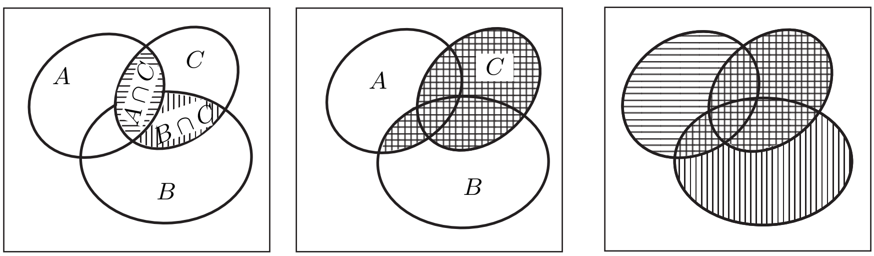
$$A \cup B \cup C, \quad A \cap B \cap C \cap D$$

*) Однако, как будет показано ниже, не каждая формула остается верной после замены сумм и произведений на объединения и пересечения. — *Прим. перев.*



$(A \cup B) \cap C$ $\qquad$ $A \cup (B \cap C)$

Рис. 2



$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ $\qquad$ $(A \cap B) \cup C$ $=$ $(A \cup C) \cap (B \cup C)$

Рис. 3

без скобок. Однако формула $A \cup B \cap C$ двусмысленна, а поэтому не является корректной. В самом деле, множество $(A \cup B) \cap C$ не совпадает с множеством $A \cup (B \cap C)$, как нетрудно убедиться с помощью рис. 2.

Операции объединения и пересечения связывает также пара *дистрибутивных законов*:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C); \tag{D$_1$}$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{D$_2$}$$

Сделаем несколько замечаний. Во-первых, аналогия с правилами арифметики справедлива для закона (D_1):

$$(a + b) \times c = (a \times c) + (b \times c),$$

но не верна в случае (D_2):

$$(a \times b) + c \neq (a + c) \times (b + c).$$

Конечно, внимательный читатель мог заметить, что аналогия нарушается уже на более ранней стадии, поскольку

$$A = A \cup A = A \cap A;$$

в то время, как единственным числом, удовлетворяющим равенству $a + a = a$, является 0, а соотношение $a \times a = a$ выполняется только для двух чисел — 0 и 1.

Во-вторых, вы вероятно уже отметили удобство использования диаграмм для доказательства или опровержения утверждений, касающихся множеств. Полезно также убедиться в истинности формул, подобных (D_1) и (D_2), с помощью хорошо подобранных примеров. Пусть, скажем,

A = «дешевые вещи»,
B = «действительно нужные вещи»,
C = «пища (продукты питания)».

Тогда формула $(A \cup B) \cap C$ представляет «(дешевые или действительно нужные) продукты». В свою очередь, $(A \cap C) \cup (B \cap C)$ означает «(дешевые продукты) или (действительно нужные продукты)». Очевидно, что это одно и то же. Рассмотренный частный пример не является доказательством формулы (D_1) — прилет одной ласточки еще не означает наступление лета. Однако, если возникает убежденность в том, что логика рассуждений и мыслительный процесс ни в коей мере не подвержены влиянию конкретной природы множеств A, B и C в том смысле, что они могут быть какими-угодно, то на самом деле получаем доказательство и для общего случая. Любопытно, что тот же самый пример, примененный к формуле (D_2), отчего-то не делает ее очевидной (по крайней мере для автора). Почему? Возможно, некоторые логические шаблоны чаще используются в повседневной жизни, чем другие.

Обсуждаемое нарушение аналогии становится еще более важным в свете очевидной двойственности двух дистрибутивных законов. Один получается из другого, если заменить символы $\cup$ на $\cap$ и наоборот. На самом деле любой из них выводится из другого на основе двойственности (задача 11).

Наконец, по причине того, что формула (D_2) представляется менее естественной с интуитивной точки зрения, воспользуемся возможностью применить для ее строгого вывода приведенный выше «окольный» метод доказательства идентичности множеств. В соответствии с данным методом требуется установить, что (i) каждая точка из левой части формулы (D_2) принадлежит ее правой части; (ii) каждая точка из правой части формулы (D_2) принадлежит ее левой части.

(i) Пусть ω принадлежит левой части соотношения (D_2). Тогда она является или элементом множества $A \cap B$, или элементом множества C. Условие $\omega \in A$ влечет $\omega \in A \cup C$; аналогично, $\omega \in B \cup C$. Следовательно, ω принадлежит правой части (D_2). С другой стороны, если $\omega \in C$, то $\omega \in A \cup C$ и $\omega \in B \cup C$, и мы приходим к прежнему заключению.

(ii) Пусть ω принадлежит правой части соотношения (D_2). Тогда ω может принадлежать, а может и не принадлежать множеству C. Рассмотрим эти два варианта по очереди. Если $\omega \in C$, тогда она обязательно входит в левую часть соотношения (D_2). Пусть теперь $\omega \notin C$. В силу того что она входит в $A \cup C$, ω должна принадлежать множеству A; аналогично, она обязана принадлежать и множеству B. Поэтому она является элементом $A \cap B$, и, таким образом, содержится в левой части соотношения (D_2). Доказательство закончено.

1.3. Разные формулы

Три операции, определенные к этому времени: дополнение, объединение и пересечение, задействованы в двух следующих формулах, называемых *законами Де Моргана*:

$$(A \cup B)^c = A^c \cap B^c; \tag{C$_1$}$$

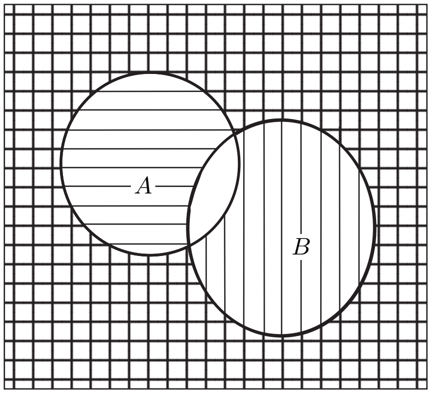
$$(A \cap B)^c = A^c \cup B^c. \tag{C$_2$}$$

Эти формулы являются двойственными в том же самом смысле, что и законы (D_1) и (D_2). Давайте проверим их с помощью нашего предыдущего примера. Если $A =$ «дешевые вещи» и $B =$ «действительно нужные вещи», то, очевидно, $(A \cup B)^c =$ «вещи, не являющиеся ни дешевыми, ни действительно нужными», т. е. дорогостоящий хлам, что совпадает с множеством $A^c \cap B^c$ недешевых и бесполезных вещей. Подобным же образом проверяется равенство (C_2).

Обратим внимание, что формулы (C_1) и (C_2) логически выводятся одна из другой. Давайте убедимся в этом. Пусть верна формула (C_1). Тогда, ввиду того что A и B — произвольные множества, мы можем подставить в (C_1) их дополнения и прийти к соотношению

$$(A^c \cup B^c)^c = (A^c)^c \cap (B^c)^c = A \cap B, \tag{1.3.1}$$

в котором мы применили соотношение (1.2.1) для того, чтобы получить второе равенство. Теперь возьмем дополнения от первого и от третьего



$(A \cap B)^c = A^c \cup B^c$

Рис. 4

множеств в формуле (1.3.1) и, снова используя равенство (1.2.1), запишем:

$$A^c \cup B^c = (A \cap B)^c.$$

Это и есть доказываемое соотношение (C_2).

Из законов Де Моргана следует, что с помощью операции дополнения объединение можно выразить через пересечение и наоборот. В самом деле,

$$A \cap B = (A^c \cup B^c)^c,$$
$$A \cup B = (A^c \cap B^c)^c,$$

и, таким образом, либо объединение, либо пересечение является избыточной операцией. С другой стороны, нельзя выразить дополнение через объединение и пересечение, хотя существует «волшебная» операция, с помощью которой выражаются все три действия (задача 14). Иногда бывает удобным использовать также некоторые другие операции над множествами, к определению которых мы сейчас и перейдем.

Разность. Множество $A \setminus B$ содержит точки, принадлежащие A и (но) не принадлежащие B. В символической форме:

$$A \setminus B = A \cap B^c = \{\omega \mid \omega \in A \quad \text{и} \quad \omega \notin B\}.$$

Эта операция не является ни коммутативной, ни ассоциативной. Давайте построим *контрпример* к ассоциативному закону, точнее, укажем некоторые множества A, B и C, для которых

$$(A \setminus B) \setminus C \neq A \setminus (B \setminus C). \tag{1.3.2}$$

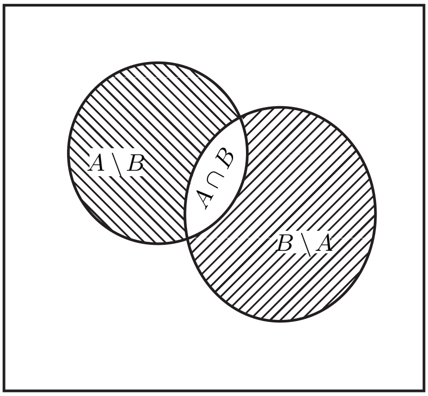


Рис. 5

Заметим, что в отличие от рассмотренного ранее доказательства идентичности, в данном случае достаточно всего лишь привести пример, показывающий отсутствие совпадения. Придумывая контрпример, обычно сначала пытаются упростить ситуацию за счет уменьшения количества «неизвестных». Поэтому возьмем $B = C$. Тогда левая часть соотношения (1.3.2) совпадет с $A \setminus B$, в то время как правая часть преобразуется в $A \setminus \varnothing = A$. Таким образом, остается убедиться, что $A \setminus B \neq A$, а это тривиально.

Если $A \supset B$, то вместо $A \setminus B$ мы будем писать $A - B$. Используя новый символ, имеем:

$$A \setminus B = A - (A \cap B)$$

и

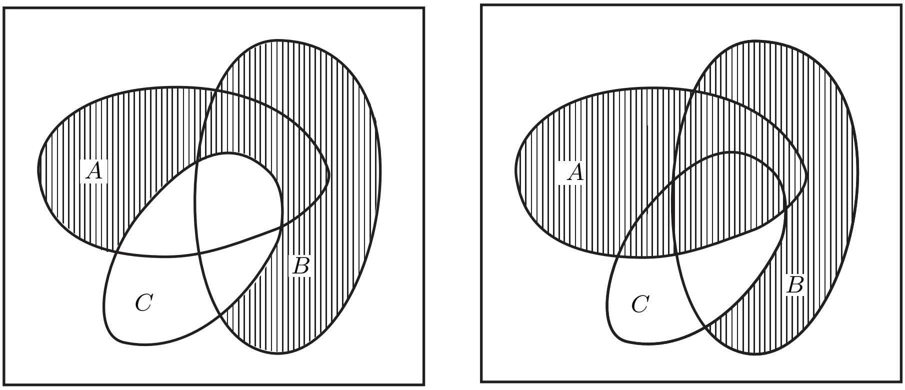
$$A^c = \Omega - A.$$

Операция «$-$» похожа на арифметическую операцию вычитания, в частности, $A - A = \varnothing$, однако сходство не распространяется слишком далеко. Например, нет аналога для формулы $(a + b) - c = a + (b - c)$.

Симметрическая разность. Множество $A \triangle B$ содержит точки, принадлежащие в точности одному из двух множеств A и B. В символической форме это можно выразить так:

$$A \triangle B = (A \cap B^c) \cup (A^c \cap B) = (A \setminus B) \cup (B \setminus A).$$

Симметрическая разность весьма полезна в некоторых более сложных вопросах теории множеств. Как показывает ее название, она симмет-



$(A \cup B) \setminus C$ $\qquad$ $A \cup (B \setminus C)$

Рис. 6

рична относительно A и B, другими словами — является коммутативной операцией. Обладает ли она ассоциативностью? Пытаясь установить это с помощью конкретных примеров или диаграмм, верно служивших нам до сих пор, вы, вероятно, так же быстро отчаетесь, как и мы. К счастью, ответ можно легко получить в результате применения индикаторного метода, изложенного в § 1.4.

Теперь, имея в распоряжении ряд операций, позволим нашему воображению представить всевозможные множества, возникающие в результате последовательного их применения и комбинирования, такие как

$$[(A \setminus C^c) \cap (B \cup C)^c]^c \cup (A^c \triangle B).$$

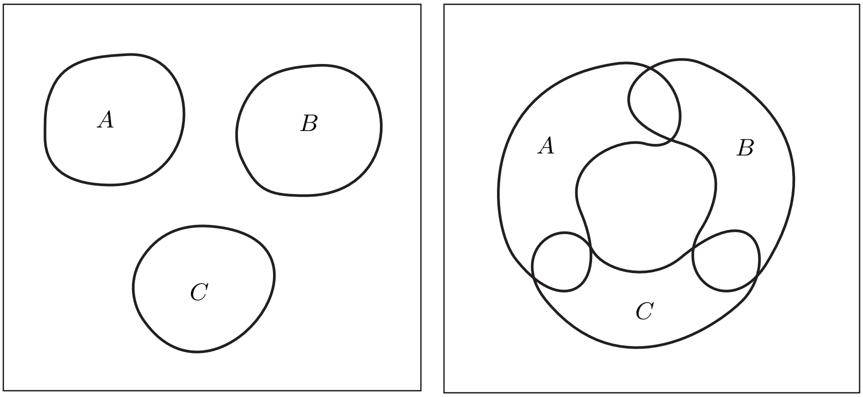
Напомним, что мы говорим о подмножествах заданного пространства Ω, и если Ω конечно, то множество его всевозможных подмножеств также конечно. Следовательно, существует огромное число взаимоотношений между теми множествами, которые мы можем сконструировать. Различные законы, рассмотренные выше, являются всего лишь наиболее важными из этих взаимоотношений. Некоторые другие встретятся в задачах, приводимых ниже.

Определим еще одно весьма полезное отношение между множествами. Два множества A и B называются *непересекающимися*, если они не имеют общих точек:

$$A \cap B = \varnothing.$$

Это отношение равносильно любому из двух следующих включений:

$$A \subset B^c; \quad B \subset A^c.$$



A, B, C не пересекаются $\qquad$ $A \cap B \cap C = \varnothing$

Рис. 7

Несколько множеств называются непересекающимися, если все они попарно не пересекаются. Таким образом, утверждение «A, B, C являются непересекающимися» означает больше, чем $A \cap B \cap C = \varnothing$ (см. рис. 7). Оно означает, что

$$A \cap B = \varnothing, \quad A \cap C = \varnothing, \quad B \cap C = \varnothing.$$

Начиная с этого места, мы будем опускать знак пересечения и писать просто

$$AB \quad \text{вместо} \quad A \cap B,$$

точно так же, как мы пишем ab вместо $a \times b$. Когда A и B не пересекаются, будем иногда использовать запись

$$A + B \quad \text{вместо} \quad A \cup B.$$

Однако будьте осторожны: знак «+» может использоваться не только для суммирования чисел и объединения непересекающихся множеств A и B. Иногда $A + B$ может обозначать векторную сумму двух множеств.

Для любого множества A справедливо очевидное *разложение*

$$\Omega = A + A^c. \tag{1.3.3}$$

Это тождество можно понимать так: множество A задает *разбиение* всех точек ω пространства Ω на два класса в зависимости от того, входит ω в A или в A^c. Так, студента колледжа можно классифицировать на основе того, специализируется он по математике или нет. Однако его также можно классифицировать в зависимости от того, является ли он первокурсником или нет, достиг ли он возраста, дающего право голосовать, или нет, есть ли у него машина или нет, ..., является ли студент девушкой или нет. Каждая двухуровневая классификация разбивает выборочное пространство на два непересекающихся множества, а если наложить несколько разбиений одно на другое, то получим, в частности, следующие тождества:

$$\Omega = (A + A^c)(B + B^c) = AB + AB^c + A^cB + A^cB^c, \tag{1.3.4}$$

$$\begin{aligned}\Omega &= (A + A^c)(B + B^c)(C + C^c) = \\ &= ABC + ABC^c + AB^cC + AB^cC^c + A^cBC + \\ &\quad + A^cBC^c + A^cB^cC + A^cB^cC^c.\end{aligned} \tag{1.3.5}$$

Мы станем называть части подобного разложения *атомами*. Таким образом возникают 2, 4, 8 атомов в случае использования соотвественно 1, 2, 3 множеств (см. рис. 8). В общем случае n множеств порождают 2^n атомов. Эти атомы обладают замечательным свойством, которое может быть проиллюстрировано на примере трех множеств: вне зависимости от того, какие операции вы применяете к множествам A, B, C, и сколько раз вы делаете это, множество, появляющееся в результате, всегда представляется как объединение некоторых из атомов, участву-

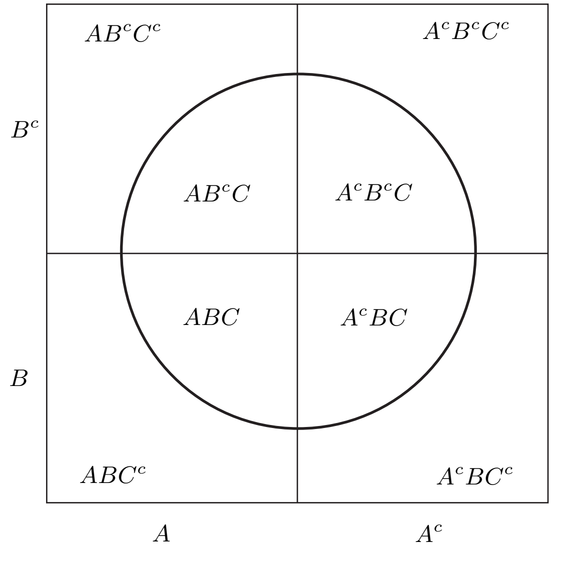


Рис. 8

ющих в правой части (1.3.5). Вот некоторые примеры:

$$A \cup B = ABC + ABC^c + AB^cC + AB^cC^c + A^cBC^c + A^cBC,$$
$$(A \setminus B) \setminus C^c = AB^cC,$$
$$(A \triangle B)C^c = AB^cC^c + A^cBC^c.$$

Как вы думаете, почему?

До сих пор мы рассматривали объединения и пересечения только конечного числа множеств. Не возникает трудностей в обобщении этих операций на случай их бесконечного числа. Допустим, что задана конечная или бесконечная последовательность множеств A_n, $n = 1, 2, \ldots$. Дадим следующие общие определения объединения и пересечения:

$$\bigcup_n A_n = \{\omega \mid \omega \in A_n \quad \text{хотя бы для одного } n\};$$
$$\bigcap_n A_n = \{\omega \mid \omega \in A_n \quad \text{для всех } n\}.$$

Если последовательность бесконечна, их можно считать результатами предельного перехода, примененного к конечным объединениям и пересечениям:

$$\bigcup_{n=1}^{\infty} A_n = \lim_{m\to\infty} \bigcup_{n=1}^{m} A_n; \quad \bigcap_{n=1}^{\infty} A_n = \lim_{m\to\infty} \bigcap_{n=1}^{m} A_n.$$

Обратите внимание, что, когда m увеличивается, множество $\cup_{n=1}^m A_n$ не уменьшается, в то время как $\cap_{n=1}^m A_n$ не увеличивается, и мы можем говорить, что первое множество *раздувается* до $\cup_{n=1}^\infty A_n$, а последнее — *сжимается* до $\cap_{n=1}^\infty A_n$.

Дистрибутивные законы и законы Де Моргана очевидным образом обобщаются на конечное или бесконечное число множеств:

$$\left(\bigcup_n A_n\right) \cap B = \bigcup_n (A_n \cap B), \tag{1.3.6}$$

$$\left(\bigcap_n A_n\right)^c = \bigcup_n A_n^c. \tag{1.3.7}$$

Интересные новые множества возникают в случае последовательного применения операций бесконечного объединения и пересечения. Наиболее известными из таких множеств являются:

$$\bigcap_{m=1}^{\infty} \left(\bigcup_{n=m}^{\infty} A_n\right); \quad \bigcup_{m=1}^{\infty} \left(\bigcap_{n=m}^{\infty} A_n\right).$$

Они изучаются в более сложных курсах (см., например, [4, § 4.2]) и приведены здесь лишь для того, чтобы вызвать ваше любопытство.

1.4. Индикатор*)

Обсуждавшаяся в § 1.3 идея классификации точек ω посредством дихотомии (быть или не быть во множестве A) лежит в основе рассматриваемого ниже индикаторного метода. Дальнейшее обобщение этого подхода в гл. 4 приведет нас к основному понятию теории вероятностей — понятию случайной величины.

Представьте, что Ω — это мишень, а A — некоторая отмеченная область на мишени, как в примерах (f) и (f′) выше (см. также рис. 9). Выбор «наудачу» точки ω из Ω осуществляется путем бросания дротика в мишень. Если дротик попадает в A, звонит колокольчик (или зажигается лампочка), в противном случае ничего не происходит. Это можно выразить с помощью следующей математической формулы:

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

Таким образом, I_A — это функция, заданная на всем выборочном пространстве Ω и принимающая значения 0 и 1 в случае тишины или звона колокольчика соответственно. Возможно, при изучении математического анализа вы уже встречались с тем, что важно различать функцию (иногда называемую отображением) и значение, принимаемое функцией. В нашем случае функция I_A однозначно определяется множеством A, и поэтому она называется индикаторной функцией или, кратко, *индикатором.* Другое множество B имеет свой собственный индикатор I_B. Две функции I_A и I_B совпадают (что это значит?) тогда и только тогда, когда множества A и B идентичны.

Чтобы продемонстрировать, как применяются индикаторы, найдем их для некоторых из рассмотренных выше множеств. Нам понадобятся новые символы $\vee$ и $\wedge$, которые, возможно, не знакомы читателю. Для двух действительных чисел a и b они определяются так:

$$\begin{aligned} a \vee b &= \text{максимум из } a \text{ и } b; \\ a \wedge b &= \text{минимум из } a \text{ и } b. \end{aligned} \tag{1.4.1}$$

*) Этот параграф может быть пропущен после знакомства с первыми тремя абзацами.

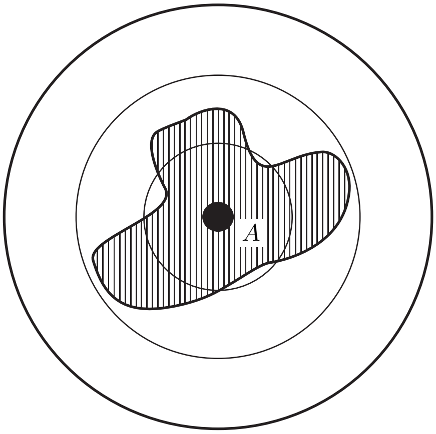


Рис. 9

В случае $a = b$ каждое из чисел является и максимумом, и минимумом. Важнейшие свойства индикаторов представляются формулами

$$I_{A\cap B}(\omega) = I_A(\omega) \wedge I_B(\omega) = I_A(\omega) \cdot I_B(\omega); \tag{1.4.2}$$

$$I_{A\cup B}(\omega) = I_A(\omega) \vee I_B(\omega). \tag{1.4.3}$$

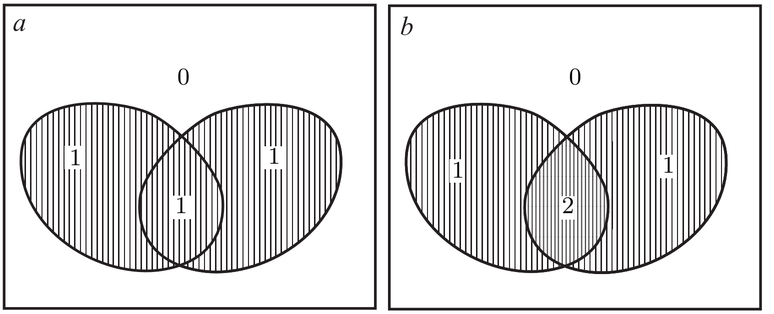
Нетрудно убедиться в справедливости приведенных равенств, поскольку каждая из функций принимает всего два значения: 0 и 1. Так как равенства верны для всех ω, их можно понимать как равенства *функций*:

$$I_{A\cap B} = I_A \wedge I_B = I_A \cdot I_B, \tag{1.4.4}$$

$$I_{A\cup B} = I_A \vee I_B. \tag{1.4.5}$$

Здесь, например, функция $I_A \wedge I_B$ представляет собой отображение, сопоставляющее каждой ω значение $I_A(\omega) \wedge I_B(\omega)$, точно так же, как в математическом анализе функцией $f + g$ называется отображение, приписывающее каждому x число $f(x) + g(x)$.

Обратив внимание на произведение $I_A(\omega) \cdot I_B(\omega)$ в правой части (1.4.2), читатель, вероятно, удивится тому, что в формуле (1.4.3) не появляется сумма $I_A(\omega) + I_B(\omega)$. Но в этом случае стало бы возможным значение 2, которое недопустимо для стоящего слева индикатора $I_{A\cup B}(\omega)$. Тем не менее, нельзя ли все-таки как-нибудь интерпретировать сумму $I_A(\omega) + I_B(\omega)$? Давайте снова вернемся к примеру с мишенью, но на этот раз нарисуем на ней две перекрывающиеся области A и B. Вместо звона колокольчика теперь назначается выигрыш в размере 1 пенни, если вы попадете в A, а также — если попадете в B. Что происхо-



$I_{A \cup B}$ $\qquad$ $I_A + I_B$

Рис. 10

дит в случае попадания в AB? Это зависит от правил игры. Может, вы по-прежнему будете получать 1 пенни, а может — 2 пенни. Оба правила законны. Формула (1.4.3) соответствует первому из правил. Если же вы предпочитаете применять второе правило, то уже будете иметь дело не просто с множеством $A \cup B$, как на рис. 10, *а*, но с тем, что изображено на рис. 10, *б*.

Приведем также и электрическую интерпретацию. Представим, что один электрический заряд равномерно распределен в области A, а другой — в области B. Тогда общий заряд будет иметь распределение, изображенное на рис. 10, *б*. Здесь заряд представляет собой переменную величину, задаваемую функцией $I_A + I_B$. Эта сумма индикаторов является весьма частным случаем суммы случайных величин, изучаемых в следующих главах.

Давайте теперь вернемся к формуле (1.4.5) и заметим, что если множества A и B не пересекаются, то правая часть на самом деле сводится к сумме индикаторов, поскольку при этом условии только один из двух индикаторов может иметь значение 1. Таким образом, в данном случае максимум совпадает с суммой, а именно:

$$0 \vee 0 = 0 + 0, \quad 0 \vee 1 = 0 + 1, \quad 1 \vee 0 = 1 + 0.$$

Итак, мы имеем:

$$I_{A+B} = I_A + I_B \quad \text{при условии, что} \quad A \cap B = \varnothing. \tag{1.4.6}$$

В частности, для любого множества A верно тождество

$$I_\Omega = I_A + I_{A^c}.$$

Здесь $I_\Omega \equiv 1$ (на Ω). Поэтому можно переписать последнее тождество в виде

$$I_{A^c} = 1 - I_A. \tag{1.4.7}$$

Все готово для того, чтобы вывести некоторые интересные формулы. Так как $(A \cup B)^c = A^c B^c$, с учетом соотношений (1.4.7), (1.4.4) и еще раз (1.4.7), находим, что

$$I_{A\cup B} = 1 - I_{A^c B^c} = 1 - I_{A^c} I_{B^c} = 1 - (1 - I_A)(1 - I_B).$$

Раскрывая скобки в правой части последнего равенства (мы имеем дело с функциями, принимающими числовые значения!) и перенося $I_{A\cap B}$ налево, приходим к формуле

$$I_{A\cup B} + I_{A\cap B} = I_A + I_B. \tag{1.4.8}$$

Наконец, изучим индикатор $I_{A\triangle B}$. Для этого потребуются некоторые сведения из арифметики (называемой также теорией чисел). Известно, что целые числа делятся на четные и нечетные в зависимости от того, равен 0 или 1 остаток от деления их на 2. Таким образом, каждое целое число может быть заменено на (сведено к) 0 или 1 при условии, что для нас важна только его четность, а не значение. Когда целые числа складываются или вычитаются на основе такой замены, говорят, что производятся операции *по модулю* 2. Например,

$$5 + 7 + 8 - 1 + 3 = 1 + 1 + 0 - 1 + 1 = 2 = 0 \quad \text{по модулю } 2.$$

Хорошо известным применением такого метода счета служит гадание на ромашке: девушка отрывает лепестки цветка один за другим и приговаривает: «любит», «не любит», ... Надеемся, вы без особого труда сможете установить справедливость следующего равенства при всех ω:

$$\begin{aligned} I_{A\triangle B}(\omega) &= I_A(\omega) + I_B(\omega) - 2I_{AB}(\omega) = \\ &= I_A(\omega) + I_B(\omega) \quad \text{по модулю } 2. \end{aligned} \tag{1.4.9}$$

Теперь мы готовы к тому, чтобы разрешить проблему из § 1.3, заключающуюся в проверке тождества

$$(A \triangle B) \triangle C = A \triangle (B \triangle C). \tag{1.4.10}$$

Доказательство. Применяя соотношение (1.4.9) дважды, получим:

$$I_{(A\triangle B)\triangle C} = I_{A\triangle B} + I_C = (I_A + I_B) + I_C \quad \text{по модулю } 2. \tag{1.4.11}$$

Если вы понимаете смысл сложения по модулю 2, то для вас должно быть очевидно, что это — ассоциативная операция. Следовательно, правая часть равенства (1.4.11) представляется как

$$I_A + (I_B + I_C) = I_A + I_{B\triangle C} = I_{A\triangle(B\triangle C)} \quad \text{по модулю } 2.$$

Тем самым, мы убедились, что оба множества в формуле (1.4.10) имеют одинаковые индикаторы, а поэтому являются идентичными. Доказательство закончено.

Мы не собираемся в дальнейшем использовать этот результат. Нам просто хотелось показать с его помощью, что индикаторный метод может оказаться полезнее диаграмм.

Задачи

1. Почему набор чисел $\{1, 2, 1, 2, 3\}$ не является множеством?

2. Будут ли идентичными два множества, если их размеры совпадают?

3. Могут ли множество и его строгое подмножество иметь одинаковые размеры? (*Строгим* называется подмножество, которое не является надмножеством!)

4. Если дополнения множеств идентичны, то и сами множества идентичны. Докажите это утверждение двумя способами: (i) используя определения, (ii) с помощью формулы (1.2.1).

5. Пусть A, B, C имеют тот же смысл, что и в конце § 1.2. Выясните, что представляют собой множества

$$A \cup (B \cap C); \quad (A \setminus B) \setminus C; \quad A \setminus (B \setminus C).$$

6. Покажите, что в общем случае

$$(A \cup B) \cap C \neq A \cup (B \cap C);$$

Приведите частные примеры, когда равенство выполняется.

7. Выразите через атомы, участвующие в правой части разложения (1.3.5), следующие множества:

$$A \cup B \cup C; \quad (A \cup B)(B \cup C); \quad A \setminus B; \quad A \bigtriangleup B;$$

множество, состоящее из элементов ω, принадлежащих в точности одному (в точности двум, по крайней мере двум) из множеств A, B, C.

8. Покажите, что включение $A \subset B$ равносильно каждому из двух условий $AB = A$ и $A \cup B = B$, т. е. отношение включения можно определить через идентичность и операцию пересечения (или объединения).

9. Убедитесь, что множества A и B не пересекаются тогда и только тогда, когда выполняется любое из условий: $A \setminus B = A$; $A \cup B = A \bigtriangleup B$. (С учетом задачи 8 это можно осуществить, используя исключительно символику и не возвращаясь на уровень определений.)

10. Проверьте, что дистрибутивный закон справедлив также для разности:

$$(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C).$$

Является ли верной двойственная формула

$$(A \cap B) \setminus C = (A \setminus C) \cap (B \setminus C)?$$

11. Выведите закон (D_2) из равенства (D_1), применяя соотношения (C_1) и (C_2).

12.* Покажите, что

$$(A \cup B) \setminus (C \cup D) \subset (A \setminus C) \cup (B \setminus D).$$

13.* Определим новую операцию «/» так:

$$A/B = A^c \cup B.$$

Проверьте, что

(i) $(A \,/\, B) \cap (B \,/\, C) \subset A \,/\, C$;

(ii) $(A \,/\, B) \cap (A \,/\, C) = A \,/\, BC$;

(iii) $(A \,/\, B) \cap (B \,/\, A) = (A \triangle B)^c$.

С точки зрения интуитивной логики «A/B» равносильно утверждению «A влечет B». Используйте это для интерпретации приведенных соотношений.

14.* Если вы — любитель ловких трюков, то эта задача — для вас. Оказывается, существует некая операция над множествами A и B, на основе которой могут быть получены все определенные выше операции. [*Указание.* Достаточно выразить через нее дополнение и объединение. Поищите ее среди комбинаций, включающих эти две операции. Она не единственна.]

15. Покажите, что включение $A \subset B$ эквивалентно неравенству $I_A \leqslant I_B$, а $A \cap B = \varnothing$ эквивалентно равенству $I_A I_B = 0$.

16. Придумайте какие-нибудь содержательные примеры для иллюстрации формулы (1.4.8).

17. Дайте непосредственное доказательство формулы (1.4.8), проверяя ее для каждой точки ω. Можно использовать атомы из разложения (1.3.4), если вы предпочитаете методический подход.

18. Убедитесь, что для произвольных действительных чисел a и b верно тождество

$$a + b = (a \vee b) + (a \wedge b).$$

Докажите равенство (1.4.8) еще раз с его помощью.

19. Выразите индикаторы $I_{A \setminus B}$ и I_{A-B} в терминах функций I_A и I_B.

20. Представьте индикатор $I_{A \cup B \cup C}$ в виде *многочлена* от индикаторов I_A, I_B, I_C. [*Указание.* Рассмотрите функцию $1 - I_{A \cup B \cup C}$.]

21.* Покажите, что справедливо тождество

$$I_{ABC} = I_A + I_B + I_C - I_{A \cup B} - I_{A \cup C} - I_{B \cup C} + I_{A \cup B \cup C}.$$

Вы можете проверить это непосредственно, однако удобнее вывести его из задачи 20, используя двойственность.

Глава 2

Вероятность

2.1. Подсчет вероятностей

В первой главе мы узнали кое-что о множествах, а здесь мы собираемся измерять их. Так как самым простым способом измерения является подсчет количества элементов множества, начнем именно с такого примера.

Пример 1. Представим, что количество гнилых яблок в мешке из примера (a$'$) в § 1.1 равно 28. Тем самым, множеству A из (a$'$) приписывается мера, называемая размером множества и обозначаемая через $|A|$. Заметим, что данный размер, вообще говоря, никак не связан с общим числом яблок в мешке, т. е. с размером всего выборочного пространства Ω, определенного в примере (a). При покупке мешка яблок нас, скорее всего, будет интересовать не количество гнилых яблок в нем, а их относительная *доля* или *пропорция*. Предположим, что в мешке лежат 550 яблок. Условно обозначая пропорцию буквой «P», запишем это в следующем виде:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{28}{550}\,. \tag{2.1.1}$$

Теперь рассмотрим множество незрелых яблок в мешке. Пусть их количество равно 47. Аналогично имеем:

$$P(B) = \frac{|B|}{|\Omega|} = \frac{47}{550}\,.$$

Естественно предположить, что яблоко не может быть одновременно и гнилым, и незрелым (на самом деле — это вопрос строгого определения данных прилагательных). Тогда соответствующие множества не пересекаются, в том смысле, что у них нет общих элементов. Следовательно, количество «гнилых или незрелых яблок» совпадает с суммой количеств «гнилых яблок» и «незрелых яблок»: $28 + 47 = 75$. В символической форме это записывается так:

$$|A + B| = |A| + |B|. \tag{2.1.2}$$

Поделив на $|\Omega|$, получаем

$$P(A + B) = P(A) + P(B). \tag{2.1.3}$$

С другой стороны, если некоторые яблоки все же могут быть сразу и гнилыми, и незрелыми (когда, например, червяк забрался в зеленое яблоко), то равенство (2.1.2) необходимо заменить на неравенство

$$|A \cup B| \leqslant |A| + |B|,$$

из которого вытекает соответствующее неравенство для вероятностей

$$P(A \cup B) \leqslant P(A) + P(B). \tag{2.1.4}$$

Насколько $|A| + |B|$ превосходит $|A \cup B|$? В точности на число «гнилых и незрелых яблок», т. е. на $|A \cap B|$. Отсюда

$$|A \cup B| + |A \cap B| = |A| + |B|,$$

что влечет красивое тождество

$$P(A \cup B) + P(A \cap B) = P(A) + P(B). \tag{2.1.5}$$

Пример 2. Более сложным способом измерения множества является вычисление площади некоторой области на плоскости, как в примерах (f) и (f′) из § 1.1, или же объема тела. Известно, что необходимость измерения площадей земельных участков привела к возникновению в далеком прошлом геометрии и тригонометрии. Хотя кочевники до сих пор используют для счета пальцы рук и ног, уже древние китайцы и египтяне, а также другие народы, делили свои пахотные земли на участки, измеряли их площади в определенных единицах и вели учет земель на глиняных табличках или папирусах. Названия мер площадей были свои у каждой из цивилизаций (кто из читателей помнит значение коэффициента пересчета, скажем, акра в гектар?). Однако нередко представляют интерес не сами площади, а их отношение, как, скажем, в случае стрельбы по мишени. Используя обозначение $|\ |$ для площади, запишем пропорцию площадей подмножества A и всего Ω:

$$P(A) = \frac{|A|}{|\Omega|}. \tag{2.1.6}$$

Другими словами, если выбрать единицу измерения так, чтобы вся площадь Ω равнялась 1, то в этом масштабе площадь множества A совпадет с пропорцией $P(A)$. Формула (2.1.6) выглядит так же, как и формула (2.1.1), благодаря умышленному использованию одинаковых обозначений с целью подчеркнуть сходство двух ситуаций. Более того, соотношения (2.1.3) — (2.1.5) остаются верными и для новой интерпретации множеств A и B.

Пример 3. При бросании игральной кости может осуществится один из шести исходов. Сравнивая выпадение определенного номера (грани) с выбором наудачу яблока в примере 1, мы приходим к тому, чтобы взять в качестве Ω множество $\{1,2,3,4,5,6\}$ и задать

$$P(\{k\}) = \frac{1}{6}, \quad k = 1,2,3,4,5,6. \tag{2.1.7}$$

Здесь мы полагаем все шесть исходов «равновероятными» в том смысле, что всем им приписана одна и та же мера, так же, как мы по умолчанию считали в примере с яблоками. Это предположение обычно выражают говоря, что кость «идеальна». В действительности, конечно, таких костей не бывает. Уже само нанесение номеров на грани может привести к нарушению полной симметрии; и даже в случае, если кость — идеальный куб, исход все равно зависит от того, как производилось бросание. Следовательно, мы должны потребовать также, чтобы бросание осуществлялось абсолютно симметричным образом, и т. п. Подобные условия могут быть довольно точно соблюдены, что позволяет обосновать предположение о равном правдоподобии исходов на почве симметрии.

Здравый смысл подсказывает необходимость обсуждения эмпирического подхода к пониманию «вероятности», заданной в формуле (2.1.7). Она представляет собой меру *правдоподобия* того, что может произойти, а это интуитивно связано с наблюдаемой в эксперименте частотой осуществления. Конкретно, пусть кость брошена несколько раз. Какова частота выпадения определенной грани? С более общих позиций, пусть A — некоторое событие, определяемое исходом, скажем, «выпадение номера, не превосходящего 5» (или «выпадение нечетного номера»). Обозначим через $N_n(A)$ количество наступлений события A в серии из n бросаний. Тогда *относительная частота* A в данных испытаниях задается отношением

$$Q_n(A) = \frac{N_n(A)}{n}. \tag{2.1.8}$$

Естественно рассматривать $Q_n(A)$ как меру множества A. Действительно, возьмем другое событие B, такое, что A и B являются *несовместными* или *взаимно исключающими* в том смысле, что оба не могут осуществиться при одном и том же испытании. Очевидно, что

$$N_n(A+B) = N_n(A) + N_n(B),$$

и следовательно,

$$\begin{aligned} Q_n(A+B) &= \frac{N_n(A+B)}{n} = \frac{N_n(A)+N_n(B)}{n} = \\ &= \frac{N_n(A)}{n} + \frac{N_n(B)}{n} = Q_n(A) + Q_n(B). \end{aligned} \tag{2.1.9}$$

Аналогично, для любых двух событий A и B, заданных на множестве исходов, причем не обязательно несовместных, соотношения (2.1.4) и (2.1.5) остаются верными после замены P на Q_n. Конечно, Q_n зависит от n и будет колебаться, возможно даже значительно, при увеличении n. Но будут ли наблюдаться уменьшение колебаний и сходимость последовательности Q_n к некоторому числу при стремлении n к бесконечности? На этот вопрос нельзя получить ответ эмпирическим путем, так как в соответствии с самим смыслом предела мы не можем ограничиться конечным числом испытаний. Поэтому предположение о существовании предела является математической идеализацией, которая выражается так:

$$Q(A) = \lim_{n\to\infty} Q_n(A). \tag{2.1.10}$$

Назовем Q *предельной частотой* наблюдения события A. Если вы умеете обращаться с пределами, то сразу увидите, что равенство (2.1.9) остается верным при предельном переходе. Точнее: устремив $n \to \infty$ всюду в этой формуле и используя определение (2.1.10), получаем соотношение (2.1.3) с заменой P на Q. То же самое можно проделать с формулами (2.1.4) и (2.1.5).

Однако предельная частота Q все еще остается зависящей от конкретной последовательности испытаний, используемой для нахождения ее значения. Ввиду этого нет никаких гарантий, что на другой последовательности испытаний, даже произведенных в тех же самых условиях, реализуется прежнее значение. Интуитивно представляется важным, чтобы мера правдоподобия событий, подобных A, представляла собой нечто большее, чем запись результатов одного эксперимента. Плодотворная теория, основанная на частотах, должна включать в себя условие постоянства определенной выше величины Q на всех *похожих* последовательностях испытаний. Несмотря на неопределенность, заключенную в слове «похожих», можно достаточно далеко продвинуться, опираясь на данное условие. Такие попытки производились и даже имели некоторый успех. В значительной степени они полагались на здравый смысл. Однако мы не станем их здесь излагать. Напротив, будем использовать подход, при котором всем исходам приписываются одинаковые вероятности, как в формуле (2.1.7). Тогда для произвольного подмножества A пространства Ω, имеющего размер $|A|$, получаем:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6}. \tag{2.1.11}$$

Например, событие $A =$ «выпал нечетный номер» отождествляется с множеством $\{1, 3, 5\}$. При этом $P(A) = 3/6 = 1/2$.

В теории вероятностей доказано основополагающее утверждение, в силу которого при выполнении определенных условий (повторные *независимые* бросания *одной и той же* кости) для «практически всех» мыслимых последовательностей испытаний предел частоты в формуле (2.1.10) действительно существует и равен вероятности $P(A)$, определенной согласно равенству (2.1.11). Эта замечательная теорема, известная как *закон больших чисел*, служит краеугольным камнем всех эмпирических наук. В некотором смысле она дает интуитивное обоснование для обсуждаемого выше понимания вероятности как предела частот. Точная формулировка и доказательство теоремы будут приведены в гл. 7. Мы упомянули ее здесь, чтобы уменьшить возможную неудовлетворенность читателя, вызванную краткостью обсуждения проблем частотного подхода. В следующих параграфах мы намерены сконцентрироваться на дальнейшем изучении множеств и вероятностей.

2.2. Определение и примеры

Прежде всего отметим, что вероятность является числом, назначаемым или приписываемым множеству, с целью измерения последнего в некотором смысле. Мы собираемся рассматривать одновременно несколько множеств (именно для этого мы изучали гл. 1), каждому из которых сопоставлена вероятность. Таким образом, вероятность представляет собой «функцию множеств». Осваивая математику, вы, конечно же, не раз встречались с понятием функции; на самом деле, оно уже использовалось нами в гл. 1. Тем не менее, давайте еще раз определим его в привычных обозначениях: функцией f, заданной на некотором подмножестве или на всем множестве действительных чисел, называется правило соответствия, согласно которому числу x сопоставляется число $f(x)$. Это иногда записывается как $f(\cdot)$ или, более подробно, в виде

$$f : x \to f(x). \tag{2.2.1}$$

Другими словами, говоря, что вероятность есть функция множеств, мы подразумеваем аналогичное соответствие, за исключением того, что число x заменяется на множество S:

$$P : S \to P(S). \tag{2.2.2}$$

Значение $P(S)$ по-прежнему является числом, более того, числом между 0 и 1. Приведенное определение функции нельзя считать совершенно строгим ввиду того, что не указано множество тех x, при которых справедливо соответствие (2.2.1). Это множество может представлять собой

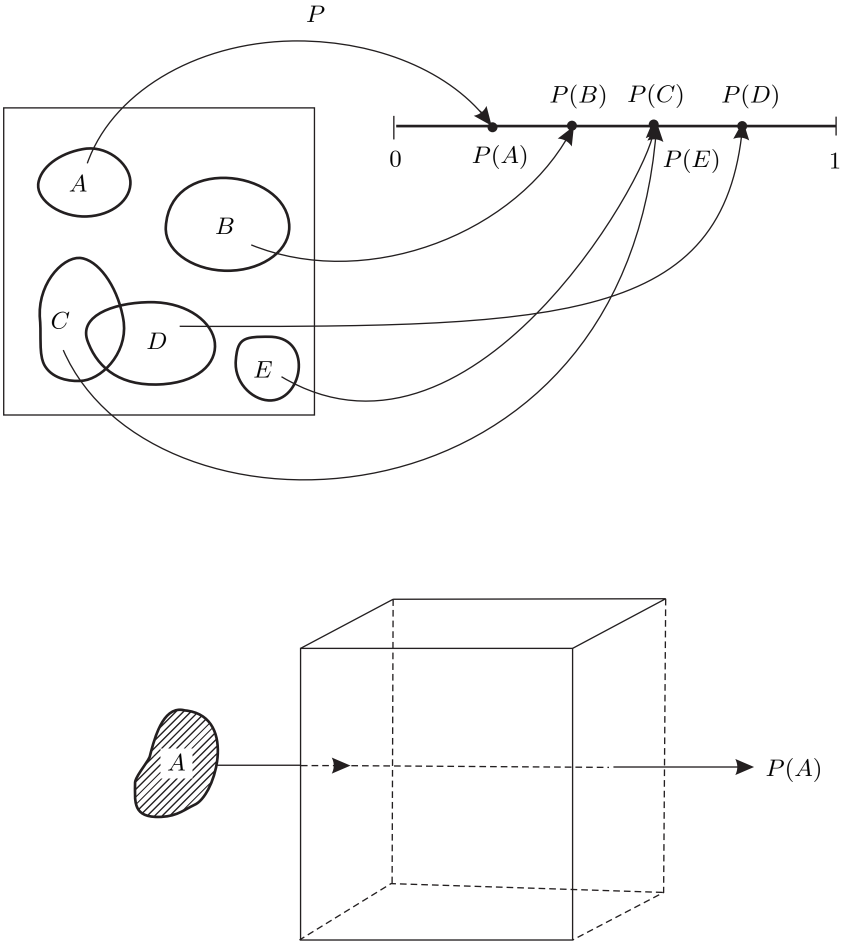


Рис. 11

интервал (a, b), луч $(0, \infty)$ или какое-то более сложное множество, называемое областью определения функции f. А что служит областью определения вероятностной функции $P(S)$? Это должно быть *множество множеств* или, во избежание повторов, *семейство* (*класс*) множеств. В гл. 1 мы изучали подмножества заданного вероятностного пространства Ω. Было бы замечательно, если бы удалось использовать семейство *всех* подмножеств пространства Ω, но в случае полного отсутствия ограничений на Ω возникает ряд неожиданных затруднений. Подробнее: если Ω очень велико, более точно — содержит несчетное количество точек, то оно имеет слишком большое число подмножеств, и становится невозможным приписать каждому из них вероятность так, чтобы вы-

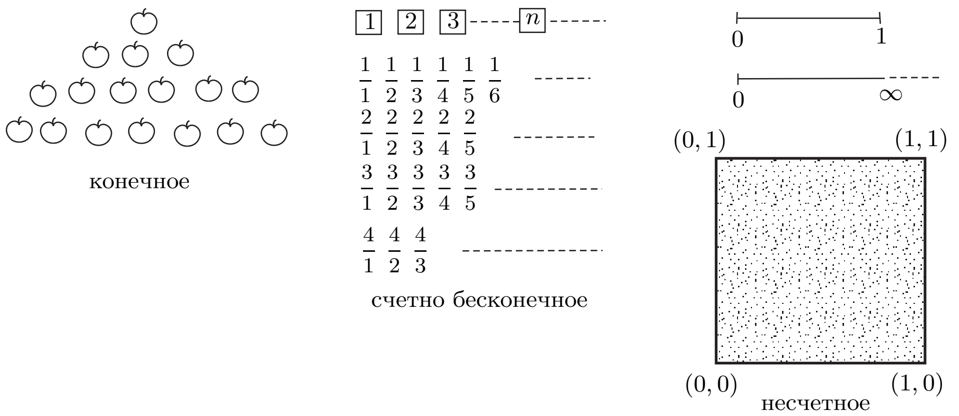


Рис. 12

полнялось такое важное свойство вероятности, как приведенная ниже аксиома (ii*). Однако, если Ω — конечное или счетное бесконечное множество, то подобных проблем не возникает, и мы можем корректно приписать вероятность каждому из подмножеств. Это будет установлено в начале § 2.4. Предполагается, что читатель представляет, что такое конечное множество (хотя трудно дать логически безупречное определение: высказывание «множество содержит конечное число элементов» является просто тавтологией). Давайте теперь обсудим термин «счетно бесконечное множество». Это довольно важное понятие, несмотря на то, что обычно оно присутствует как бы на заднем плане.

Множество называется счетно бесконечным, если можно установить взаимно однозначное соответствие между ним и множеством натуральных чисел. Построение соответствия можно представлять себе как приклеивание ярлыков с номерами элементам множества, скажем, $\{s_1, s_2, \ldots, s_n, \ldots\}$. Конечно, существуют разные способы нумерации: например, достаточно просто поменять местами ярлыки какой-то пары элементов (или места элементов, если соответствие устанавливается путем выстраивания их в ряд). Множество положительных рациональных чисел является счетно бесконечным, так как эти числа возможно некоторым образом записать в виде последовательности $\{r_1, r_2, \ldots\}$. Но не следует думать, что при этом сохраняется их упорядоченность по возрастанию, как в случае самих натуральных чисел: $1 < 2 < \ldots < n < \ldots$. Начиная с этого момента, мы станем называть множество *счетным*, если оно конечно или счетно бесконечно. В противном случае множество называется *несчетным* (см. рис. 12). Несчетные множества нам встретятся позже. Мы приведем также некоторые свойства счетных множеств, когда они нам понадобятся. Пока же будем предполагать, что выборочное пространство Ω является счетным. Это допущение позволит нам дать определение вероятностной меры в его простейшей форме без отвлекающих усложнений. На самом деле, можно даже считать Ω конечным, как в примерах (a)–(e) из § 1.1, не утрачивая существа последующего обсуждения.

Определение. *Вероятностной мерой на выборочном пространстве Ω называется заданная на подмножествах Ω функция, удовлетворяющая следующим трем аксиомам:*

(i) для произвольного множества $A \subset \Omega$ значением функции является неотрицательное число: $P(A) \geqslant 0$.

(ii) значение функции на объединении $A + B$ любых двух непересекающихся множеств A и B равно сумме значений на этих множествах:

$$P(A + B) = P(A) + P(B) \quad \text{при условии, что } AB = \varnothing.$$

(iii) значение функции на всем Ω (как на подмножестве) есть 1:

$$P(\Omega) = 1.$$

Заметьте, что мы постоянно подчеркиваем различие между функцией $P(\cdot)$ и ее значениями, такими как $P(A)$, $P(B)$, $P(A + B)$, $P(\Omega)$. Эти значения — суть вероятности. В свою очередь, для самой функции используется термин «вероятностная мера».

Пример 1 из § 2.1 показывает, что определенная там пропорция P является вероятностной мерой на выборочном пространстве — мешке с 550 яблоками. Каждому подмножеству, состоящему из яблок, ставится в соответствие вероятность. При этом соответствие удовлетворяет всем трем аксиомам. В примере 2 в качестве Ω возьмем всю землю, принадлежащую фараону. К сожалению, Ω не является счетным множеством. Несмотря на это, возможно определить понятие площади для весьма широкого класса подмножеств, называемых «измеримыми». Если ограничиться только такими множествами, то «функция площади» становится вероятностной мерой (что было показано в примере 2 без учета данного ограничения). Обратим также внимание, что аксиома (iii) представляет собой просто соглашение о нормировке меры. Обсудим теперь, за счет чего некоторый участок земли может оказаться неизмеримым. Это достаточно сложный с математической точки зрения вопрос, в который мы не намерены углубляться в данной книге. Однако нетрудно

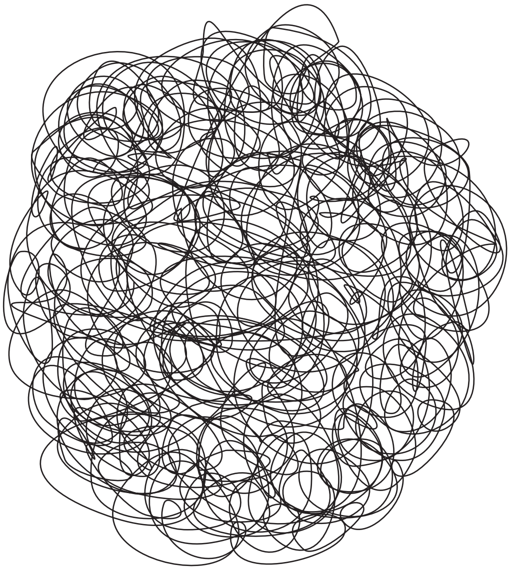


Рис. 13

понять, в чем на самом деле заключается причина неизмеримости: участок земли может быть очень зазубренным, извилистым и непроходимым (см. рис. 13).

В примере 3 было установлено, что эмпирическая относительная частота является вероятностной мерой. Но мы не станем опираться на частотный подход для формализации понятия вероятности. Вместо этого будем использовать первое определение, данное в начале примера 3, которое и исторически появилось раньше. Итак, все готово для того, чтобы дать общую формулировку.

Пример 4. Классическим определением вероятности является следующее: вероятность события есть отношение количества исходов*), при которых данное событие происходит, к общему числу исходов при условии, что все исходы считаются одинаково правдоподобными.

Переведем его на привычный язык. Пусть выборочным пространством служит конечное множество возможных исходов:

$$\{\omega_1, \omega_2, \ldots, \omega_m\},$$

где ω_i — исходы (случаи). Событие A — это подмножество

$$\{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_n}\}$$

*) Называемых также элементарными событиями или случаями. — *Прим. перев.*

«благоприятных случаев», т. е. таких ω_{i_j}, при которых данное событие происходит. Тогда вероятностью события A называется отношение

$$P(A) = \frac{|A|}{|\Omega|} = \frac{n}{m}\,. \tag{2.2.3}$$

Из обсуждения, приведенного в примере 1, вытекает, что данное определение задает вероятностную меру P на пространстве Ω. Поэтому требование одинакового правдоподобия исходов является излишним с аксиоматической точки зрения. Что же оно в действительности означает? Выглядит оно почти как тавтология, и, вообще, как проверить, одинаково правдоподобны все случаи или нет?

Для объяснения рассмотрим известный пример. Подбросим две монеты. Даламбер (математик, философ и энциклопедист, 1717–1783) констатировал, что есть только три возможных исхода, а именно:

(i) две решки, (ii) два орла, (iii) орел и решка.

Отсюда он пришел к заключению, что вероятность исхода «орел и решка» равна 1/3. Если бы Даламбер понимал, что *вероятность* должна быть как-то связана с экспериментальной *частотой* наблюдения события, то, возможно, он изменил бы свое мнение после нескольких подбрасываний. (История умалчивает, делал ли он это, но, говорят, на протяжении столетий люди полагали, что у мужчин больше зубов, чем у женщин, только потому, что так сказал Аристотель, и, по-видимому, никто не озаботился проверить это, заглянув в несколько ртов.) Рассматриваемые три исхода не являются одинаково правдоподобными. Исход (iii) необходимо разделить на два новых:

(iii a) на первой из монет выпала решка, на второй — орел,

(iii b) на первой из монет выпал орел, на второй — решка.

Именно исходы (i), (ii), (iii a) и (iii b) одинаково правдоподобны в силу симметрии и практического опыта. Это становится очевидным, когда монеты бросаются не одновременно, а одна за другой. Тем не менее, есть важное обстоятельство, которое необходимо прояснить. Монеты могут быть физически неразличимы в том смысле, что в эксперименте реально удается наблюдать только три исхода, указанные Даламбером. При бросании монет данные исходы не являются одинаково правдоподобными с точки зрения здравого смысла и экспериментального опыта. Однако в аналогичной модели для некоторых элементарных частиц*),

*) Например, для фотонов. — *Прим. перев.*

подчиняющихся так называемой статистике Бóзе — Эйнштейна (см. задачу 24 в гл. 3), предположение о равновероятности подобных исходов позволяет объяснить наблюдаемые физические явления. Таким образом, вопрос, что именно считать одинаково правдоподобным, находится вне пределов аксиоматики. Другими словами, используя формулу (2.2.3) в качестве определения вероятности, мы фактически рассматриваем все исходы ω как равновозможные в том смысле, что подсчитываем только их количество и не приписываем им разные веса.

Пример 5. Какова вероятность того, что при бросании шести игральных костей номера всех выпавших граней будут различными?

Это — пример (e) и (e′). Он приведен снова, чтобы вы привыкали к задачам подобного рода. Уже указывалось, что полное число возможных исходов равно $6^6 = 46\,656$. Исходы предполагались равновероятными, хотя до сих пор об этом речи не шло. Увы, нельзя решить сформулированную задачу без подобных предположений. Необходимо иметь дополнительную информацию о костях прежде, чем приступить к решению (это приводит к определенным проблемам, когда аналогичные задачи решаются на практике). Пусть все кости абсолютно симметричны. Механизм, подбрасывающий их, также совершенен. В частности, не происходят столкновений костей между собой во время движения. Тогда нашу гипотезу о равновозможности исходов можно считать оправданной. Подобные условия предполагаются выполняющимися, если о костях ничего не известно. Решение в таком случае дается формулой (2.2.3) при $n = 6^6$ и $m = 6!$ (соответствующие вычисления см. в примере 2 из § 3.1):

$$\frac{6!}{6^6} = \frac{720}{46\,656} \approx 0.015432.$$

Отметим, что если кости неотличимы одна от другой, то для наблюдателя существует в точности один вариант, при котором номера выпавших граней на всех шести костях окажутся разными. Соответственно, в этом случае и общее количество вариантов будет намного меньше, чем 6^6 (см. пример 3 в § 3.2). Однако при вычислении вероятностей для результатов бросания реальных костей необходимо считать, что кости различимы, как будто они раскрашены в разные цвета. Такой подход весьма важен для понимания комбинаторных методов из гл. 3.

В некоторых ситуациях требуется выяснить, какие именно исходы являются одинаково правдоподобными. Проиллюстрируем эту проблему на примере известной исторической задачи о справедливом дележе ставки.

Пример 6. Игроки A и B состязаются в игре, состоящей из серии партий. Вероятность выигрыша каждого в отдельной партии равна $1/2$, безотносительно (независимо) от результатов других партий. Например, они могут играть в теннис, в котором оба одинаково искусны, или просто в «орла или решку», подбрасывая симметричную монету. Игрок, выигравший очередную партию, зарабатывает одно очко, а проигравший — ничего. Допустим, что состязание было прервано, когда игроку A не хватало до выигрыша ставки двух очков, а игроку B — трех очков. Как справедливо разделить ставку?

Понятно, что победитель обязательно был бы выявлен, продлись игра еще всего 4 партии. В этих партиях либо A набрал бы $\geqslant 2$ очков, либо B заработал бы $\geqslant 3$, но не оба сразу. Давайте перечислим все исходы, возможные в этих четырех партиях, используя буквы A и B для обозначения победителя в каждой из партий:

$$\begin{array}{ccccc} AAAA & AAAB & AABB & ABBB & BBBB \\ & AABA & ABAB & BABB & \\ & ABAA & ABBA & BBAB & \\ & BAAA & BAAB & BBBA & \\ & & BABA & & \\ & & BBAA & & \end{array}$$

Они равно правдоподобны в силу симметрии. Всего существует*) $\binom{4}{4} + \binom{4}{3} + \binom{4}{2} = 11$ исходов, при которых A выигрывает ставку, и $\binom{4}{3} + \binom{4}{4} = 5$ исходов, когда побеждает B. Следовательно, ставку следует разделить в отношении 11:5. Например, если она равна \$64 000, то A должен получить \$44 000, а B — \$20 000. (Мы допустили определенную вольность, использовав доллар в качестве валюты; когда была поставлена рассматриваемая задача, США еще не существовали.)

Приведенное решение принадлежит Б. Паскалю. Оно содержалось в его письме к П. Ферма, датированном 24 августа 1654 г. (Блез Паскаль (1623–1662), Пьер Ферма (1601–1665); оба являются выдающимися математиками всех времен.) Другие ученые, современники Паскаля и Ферма, высказали следующие возражения (с тех пор постоянно повторяемые): приведенное выше перечисление случаев некорректно, так как серия партий должна быть завершена сразу, как только выявится

*) Используемое обозначение $\binom{m}{n}$ определяется формулой (3.2.3).

победитель, тем самым, в некоторых случаях не придется проводить все 4 партии. Таким образом, в действительности исходы таковы:

$$\begin{array}{ll} AA & ABBB \\ ABA & BABB \\ ABBA & BBAB \\ BAA & BBB \\ BABA & \\ BBAA & \end{array}$$

Однако данные исходы не являются одинаково правдоподобными. В современным терминах: если эти 10 исходов рассматривать как элементарные события, образующие выборочное пространство, то

$$\begin{gathered} P(AA) = \frac{1}{4}, \quad P(ABA) = P(BAA) = P(BBB) = \frac{1}{8}, \\ P(ABBA) = P(BABA) = P(BBAA) = P(ABBB) = \\ = P(BABB) = P(BBAB) = \frac{1}{16} \end{gathered}$$

в силу того, что события A и B независимы и происходят с вероятностью $1/2$ каждое (см. § 2.4). Если сложить указанные вероятности, получим, конечно же, что

$$\begin{aligned} P(A \text{ выигрывает ставку}) &= \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{11}{16}, \\ P(B \text{ выигрывает ставку}) &= \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{8} = \frac{5}{16}. \end{aligned}$$

Паскаль не объяснял свой метод решения подобным образом, а лишь отметил, что «совершенно все равно, заканчивается ли игра, как только один из игроков наберет требуемое число очков, или же она продолжается все 4 партии». Более позднее его письмо свидетельствует о том, что он испытывал неуверенность, сталкиваясь с той же проблемой, когда решал аналогичную задачу для трех игроков. То обстоятельство, что изложенные рассуждения вызывали затруднение даже у известных математиков прошлого, вероятно, придаст уверенности читателю в собственных способностях.

2.3. Следствия аксиом

В этом параграфе мы будем иметь дело с аксиоматикой, хотя и на элементарном уровне. Другими словами, мы намерены вывести ряд свойств вероятностной меры, опираясь исключительно на аксиомы, ее определяющие. В какой-то степени аксиомы математической теории подобны конституции государства. До тех пор, пока в нее не внесены поправки или изменения, каждый *закон* обязан быть согласован с ней, следовать из нее. Математика отличается тем, что в ней нет расхождений во взглядах на то, как конституция должна быть устроена.

Следствиям из аксиом присвоим номера от (iv) до (viii). Прежде всего, давайте установим, что вероятность на самом деле является числом, лежащим между 0 и 1.

(iv) Для произвольного множества A имеем

$$P(A) \leqslant 1.$$

В этом нетрудно убедиться, тем не менее, в ходе доказательства будут применены все три аксиомы. Рассмотрим дополнение A^c. Множества A и A^c не пересекаются, а их объединением служит все Ω:

$$A + A^c = \Omega. \tag{2.3.1}$$

До сих пор использовалась только теория множеств, и не было никаких вероятностей. Теперь применим аксиому (ii) к левой, а аксиому (iii) — к правой части равенства (2.3.1):

$$P(A) + P(A^c) = P(\Omega) = 1. \tag{2.3.2}$$

Наконец, в силу аксиомы (i) для A^c получим:

$$P(A) = 1 - P(A^c) \leqslant 1.$$

Действительно, данное неравенство — не что иное, как аксиома (i). Возможно, детальность доказательства вызвала у вас недовольство, так как вам представляется очевидным, что *включение* A в Ω влечет неравенство $P(A) \leqslant P(\Omega) = 1$. Это утверждение, конечно же, верно, однако мы обязаны вывести его из аксиом, и этот вывод составляет часть проведенных рассуждений. Оно вытекает также из следующего более общего свойства.

(v) Для любых множеств $A \subset B$

$$P(A) \leqslant P(B) \quad \text{и} \quad P(B - A) = P(B) - P(A).$$

Доказательство повторяет предыдущее с заменой Ω на B. В самом деле,

$$\begin{gathered} B = A + (B - A), \\ P(B) = P(A) + P(B - A) \geqslant P(A). \end{gathered}$$

Следующее утверждение настолько близко к аксиоме (ii), что мы могли бы использовать его в качестве аксиомы.

(vi) Для любых непересекающихся множеств $A_1, \ldots, A_n$ выполняется равенство

$$P(A_1 + \ldots + A_n) = P(A_1) + \ldots + P(A_n). \tag{2.3.3}$$

Данное свойство вероятностной меры называется *конечной аддитивностью.* Его легко получить, если вспомнить, что означает термин «непересекающиеся», и применить аксиому (ii) несколько раз или же доказать формально, опираясь на индукцию. Справедливо и важное обобщение формулы (2.3.3) на счетное число множеств, но его нельзя вывести из конечной аддитивности с помощью индукции!

Как ранее устанавливалось в частных случаях, существует формула, обобщающая аксиому (ii), а следовательно и равенство (2.3.3), на множества, которые могут пересекаться. Вероятно, вам она покажется банальной, но у нее есть собственное название — *неравенство Буля.* Буль (1815–1864) известен как первооткрыватель «законов мышления» и автор *теории логики и вероятностей.*

(vii) Для конечного числа множеств $A_1, \ldots, A_n$ верно неравенство

$$P(A_1 \cup \cdots \cup A_n) \leqslant P(A_1) + \ldots + P(A_n). \tag{2.3.4}$$

Давайте сначала установим его при $n = 2$. Представим объединение двух произвольных множеств A и B в виде суммы непересекающихся множеств следующим образом:

$$A \cup B = A + A^c B. \tag{2.3.5}$$

В силу аксиомы (ii) имеем:

$$P(A \cup B) = P(A) + P(A^c B). \tag{2.3.6}$$

Так как $A^c B \subset B$, мы можем применить свойство (v) и получить неравенство (2.3.4).

Доказательство в общем случае нетрудно провести с помощью математической индукции. Рекомендуем читателю записать его в качестве полезного упражнения на применение индуктивного метода. Вы увидите, что потребуется использовать ассоциативный закон как для операции объединения множеств, так и для сложения чисел.

Изучим теперь вопрос о том, что представляет собой разность между частями неравенства (2.3.4). Вопрос не является вполне корректным, так как ответ зависит от того, в какой форме мы желаем выразить эту разность. Тем не менее при $n = 2$ существует простое решение.

(viii) Для любых двух множеств A и B верно равенство

$$P(A \cup B) + P(A \cap B) = P(A) + P(B). \tag{2.3.7}$$

Его можно вывести из соотношения (2.3.6) с учетом тождества $A^c B = B - AB$ и свойства (v):

$$P(A \cup B) = P(A) + P(B - AB) = P(A) + P(B) - P(AB),$$

что равносильно равенству (2.3.7). Другое доказательство приведено в задаче 12.

Мы отложим рассмотрение случая нескольких множеств до § 6.2. На практике неравенство часто оказывается более полезным, чем соответствующее достаточно громоздкое равенство.

Нельзя не отметить очевидное сходство соотношения (2.3.7) с формулой (1.4.8) из § 1.4, которую мы повторим здесь для удобства сравнения:

$$I_{A \cup B} + I_{A \cap B} = I_A + I_B. \tag{2.3.8}$$

Это сходство опирается на следующую глубокую связь. Вероятность $P(S)$ каждого множества S может быть получена из индикаторной функции I_S с помощью процедуры (операции), называемой взятием математического ожидания или интегрированием. Если применить ее к каждому из членов соотношения (2.3.8), то придем к формуле (2.3.7). Описанная процедура играет важную роль в теории вероятностей и будет детально изучена в гл. 6. Частный случай рассматривается в задаче 19.

В завершение обсуждения аксиоматики усилим аксиому (ii) или ее непосредственное следствие (vi), т. е. условие конечной аддитивности меры P, потребовав выполнения новой аксиомы.

(ii*) Аксиома счетной аддитивности: для любого счетно бесконечно набора непересекающихся множеств A_k, $k = 1, 2, \ldots$, верно равенство

$$P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k). \tag{2.3.9}$$

Эта аксиома содержит аксиому (vi) в качестве частного случая, так как для получения формулы (2.3.3) достаточно положить $A_k = \varnothing$ при $k > n$ в равенстве (2.3.9). Пустое множество не пересекается с любым множеством и имеет вероятность нуль (почему?). Если Ω — конечное множество, то новая аксиома сводится к старой. Но важно показать, почему формула (2.3.9) *не может* быть выведена из соотношения (2.3.3) при устремлении $n \to \infty$. Давайте попытаемся убедиться в этом, переписав формулу (2.3.3) так:

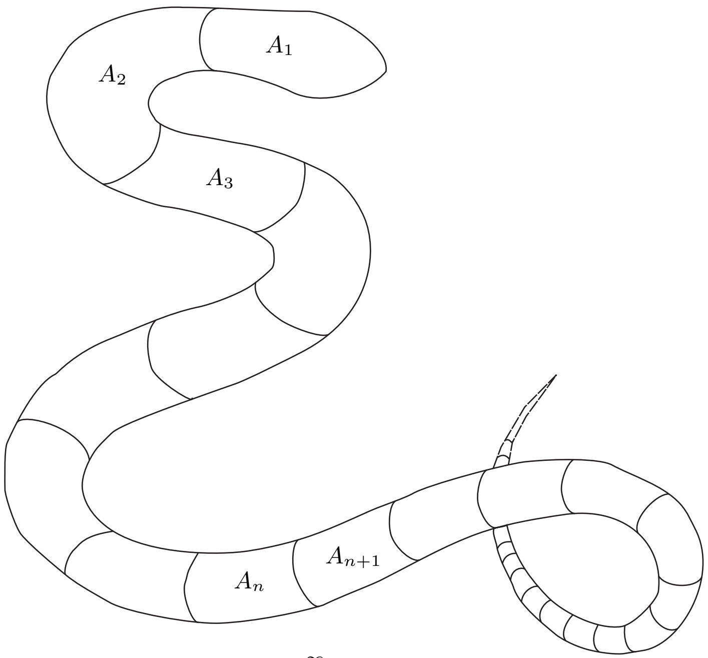
$$P\left(\sum_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k). \tag{2.3.10}$$

Поскольку при всех n левая часть не может превосходить 1, стоящий справа ряд должен сходиться. Отсюда получаем:

$$\lim_{n\to\infty} P\left(\sum_{k=1}^{n} A_k\right) = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{\infty} P(A_k). \tag{2.3.11}$$

Сравнивая установленный результат с желаемым (2.3.9), видим, что вопрос сводится к проверке равенства

$$\lim_{n\to\infty} P\left(\sum_{k=1}^{n} A_k\right) = P\left(\sum_{k=1}^{\infty} A_k\right),$$



$\sum\limits_{n=1}^{\infty} A_n$

Рис. 14

которое представляется также в следующем, более выразительном виде:

$$\lim_{n\to\infty} P\left(\sum_{k=1}^{n} A_k\right) = P\left(\lim_{n\to\infty}\sum_{k=1}^{n} A_k\right). \tag{2.3.12}$$

См. заключительную часть § 1.3 и рис. 14.

Таким образом, проблема заключается в обосновании возможности перестановки операций «lim» и «P» или, другими словами, возможности перехода к пределу под знаком вероятности. Читатель, достаточно хорошо владеющий математическим анализом, знает, что, как правило, доказать допустимость подобной перестановки не так легко. Подчас перестановка остается необоснованной или даже вообще оказывается неверной. Роль новой аксиомы заключается в констатации в данной ситуации законности предельного перехода, который служит важнейшим инструментом для получения фундаментальных теорем теории вероятностей.

2.4. Независимые события

Далее мы будем предполагать, что вероятностная мера удовлетворяет аксиомам (i), (ii*) и (iii). Подмножества пространства Ω, которым приписана такая мера, станем называть *событиями*.

Покажем, каким образом можно *задать* вероятностную меру на любом счетном пространстве $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$. Каждой точке ω_n выборочного пространства сопоставим некоторый «вес» p_n. При этом от весов требуется лишь подчинение следующим условиям:

$$\forall n:\quad p_n \geqslant 0,\ \sum_n p_n = 1. \tag{2.4.1}$$

Другими словами, веса должны быть неотрицательными и в сумме составлять 1. Теперь определим вероятность произвольного подмножества A из Ω как *сумму весов всех точек, принадлежащих* A. В символической форме: если

$$\forall n:\quad P(\{\omega_n\}) = p_n, \tag{2.4.2}$$

то для любого $A \subset \Omega$

$$P(A) = \sum_{\omega_n\in A} p_n = \sum_{\omega_n\in A} P(\{\omega_n\}).$$

Последнюю сумму можно записать и в таком виде:

$$P(A) = \sum_{\omega\in A} P(\{\omega\}). \tag{2.4.3}$$

Тем самым, вероятность P представляет собой функцию, заданную на подмножествах пространства Ω. Остается только убедиться, что она удовлетворяет аксиомам (i), (ii*) и (iii). Для этого не нужно ничего, кроме небольшого размышления на свежую голову, которое мы советуем читателю выполнить самостоятельно. Поскольку веса совершенно произвольны, если не считать простых условий (2.4.1), понятно, что на счетном выборочном пространстве разных вероятностных мер «пруд пруди». В действительности, все они могут быть заданы с помощью рассмотренного метода. Так, если у нас есть некоторая вероятностная мера P, неважно как заданная, мы с учетом соотношений (2.4.2) можем взять $P(\{\omega_n\})$ в качестве p_n, при этом вероятность $P(A)$ определяется в соответствии с формулой (2.4.3) в силу аксиомы (ii*). Более того, из аксиом легко вывести, что веса p_n должны удовлетворять условиям (2.4.1). Иначе говоря, произвольная вероятностная мера P допускает описание в терминах нашей конструкции.

В очень частном случае, когда Ω конечно и состоит в точности из m точек, можно приписать им всем одинаковые веса, т. е. положить

$$p_n = \frac{1}{m}, \quad n = 1, 2, \ldots, m.$$

Тогда мы вернемся к ситуации равного правдоподобия исходов, рассмотренной в примере 4 из § 2.2. Но в общем случае веса p_n не обязаны быть одинаковыми, а если Ω является счетно бесконечным, то они не могут быть все одинаковыми (почему?). Проведенное обсуждение дает некоторое представление о степени произвольности, заключенной в устройстве общей вероятностной меры.

Важным примером вероятностного пространства служит модель, связанная с *последовательностью независимых испытаний*: она используется для описания бросаний монеты или игральной кости, многократных извлечений карты из колоды (с возвращением). Более того, можно бросать одновременно несколько монет или костей. Давайте начнем с примера.

Пример 7. Сначала подбросим монету, затем бросим игральную кость, наконец, извлечем карту из колоды для покера. С каждым из испытаний связано свое событие:

$A =$ «монета выпала решкой»;

$B =$ «сверху оказалась грань кости с номером 5 или 6»;

$C =$ «извлечена пика».

Будем считать, что монетка симметрична, кость имеет совершенную форму, колода тщательно перетасована. Кроме того, предположим, что все три испытания проводятся «независимо» друг от друга. На уровне интуиции это означает, что результат любого из них никак не влияет на результаты остальных. Например, данное условие можно считать выполняющимся, если испытания осуществляются разными людьми в разных местах, либо одним человеком, но в разные месяцы! Тогда всевозможные совместные исходы допустимо рассматривать как одинаково вероятные. Всего имеется 2, 6 и 52 случаев соответственно для каждого из испытаний. Общее число возможностей равно произведению $2\cdot 6\cdot 52$ (как вы скоро поймете, его лучше не вычислять). Этот подсчет базируется на правилах комбинаторики, детально обсуждаемых в § 3.1, с которыми можно ознакомиться и сейчас, если хотите. (Вообще, многие части этой книги допустимо читать в произвольном порядке.) Тот же самый подход позволяет вычислить количества благоприятных исходов для событий A, B, C, AB, AC, BC, ABC, приведенные ниже (здесь символ $|\ldots|$ обозначает размер события):

$$\begin{aligned}
|A| &= 1\cdot 6\cdot 52, & |B| &= 2\cdot 2\cdot 52, & |C| &= 2\cdot 6\cdot 13,\\
|AB| &= 1\cdot 2\cdot 52, & |AC| &= 1\cdot 6\cdot 13, & |BC| &= 2\cdot 2\cdot 13,\\
|ABC| &= 1\cdot 2\cdot 13.
\end{aligned}$$

Поделив эти числа на $|\Omega| = 2\cdot 6\cdot 52$, после сокращений общих множителей получаем:

$$\begin{aligned}
P(A) &= \frac{1}{2}, & P(B) &= \frac{1}{3}, & P(C) &= \frac{1}{4},\\
P(AB) &= \frac{1}{6}, & P(AC) &= \frac{1}{8}, & P(BC) &= \frac{1}{12},\\
P(ABC) &= \frac{1}{24}.
\end{aligned}$$

Нетрудно проверить, что верны следующие равенства:

$$\begin{aligned}
P(AB) &= P(A)P(B), & P(AC) &= P(A)P(C),\\
P(BC) &= P(B)P(C), & P(ABC) &= P(A)P(B)P(C).
\end{aligned} \tag{2.4.4}$$

Надеемся, у читателя уже сложилось понимание, что данный набор равенств справедлив для любых таких событий A, B, C, что A определяется *исключительно* испытанием, связанным с бросанием монеты, B — с выпадением кости, C — с извлечением карты. В этом случае говорят, что подобные испытания и события являются *стохастически независимыми*. Наречие «стохастически» обычно опускают ради краткости.

Внимательный читатель здесь может заметить, что мы рассуждаем о независимых испытаниях, еще не определив формально само понятие испытания! Логическая конструкция таких объектов достаточно проста, но, вероятно, несколько абстрактна для упоминания по ходу дела. Она носит название «прямое произведение пространств», см. задачу 29. Проще определить свойство независимости событий, к чему мы сейчас и перейдем.

События A и B называются *независимыми*, если $P(AB) = P(A)P(B)$. Три события A, B и C считаются независимыми, когда для них выполняются равенства (2.4.4). Понятие независимости связано с конкретной вероятностной мерой (в отличие, скажем, от понятия несовместности событий, не имеющего отношения к их вероятностям). В общем случае, n событий $A_1, A_2, \ldots, A_n$ называются независимыми, если вероятность пересечения (совместного осуществления) любого набора из них равна произведению вероятностей отдельных событий. Если вы находите данную формулировку слишком длинной и сложной, то, возможно, вы предпочтете следующий символизм. Для любого подмножества $(i_1, i_2, \ldots, i_k)$ множества $(1, 2, \ldots, n)$ должно выполняться равенство

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}). \tag{2.4.5}$$

Здесь предполагается, что индексы $i_1, i_2, \ldots, i_k$ различны и $1 \leqslant k \leqslant n$.

Дальнейшее изучение понятия независимости отложим до § 5.5, ввиду того, что его лучше объяснять в терминах случайных величин. Тем не менее, мы собираемся сейчас кратко рассмотреть классическую схему повторных испытаний, которая служила объектом исследования таких известных ученых, как Я. Бернулли, А. де Муавр, П. С. Лаплас, Э. Борель и др.

Пример 8 (*схема бросаний монеты)).** Монета бросается n раз. Совместный результат может быть записан в виде последовательности букв H и T, где H = «орел», T = «решка».**) Часто бывает удобным использовать числа вместо букв, положив $H = 1$ и $T = 0$ или $H = 1$ и $T = -1$. Мы выберем первый способ. Таким образом, результат является набором из 0 и 1 длины n (подобным набору 110010110 при $n = 9$). Поскольку в каждом из испытаний возможны 2 варианта, всего существует 2^n совместных исходов. Это — еще один пример применения основного комбинаторного правила из § 3.1. Если все исходы предполагаются одинаково правдоподобными, то каждому из них следует припи-

*) Называемая также схемой испытаний Бернулли. — *Прим. перев.*

**) Обозначения происходят от английских слов «heads» и «tails». — *Прим. перев.*

сать вероятность $1/2^n$. В данном случае, рассуждая, как в примере 7, можно убедиться, что испытания независимы и монетка симметрична. Вы увидите, что эта проверка — достаточно скучное упражнение, но мы советуем выполнить его, если, конечно, ваша голова еще не склонилась к бумаге. Однако сейчас мы собираемся посмотреть на дело иначе, *предположив априори*, что подбрасывания монеты образуют последовательность независимых испытаний. При этом не требуется, чтобы монета была симметричной. Предполагается только постоянство вероятностей выпадения орла и решки во всех испытаниях. С практической точки зрения, это условие может выполняться лишь приближенно, так как ничто не может быть неизменным долгое время. Нам понадобятся строгие обозначения для записи сложных утверждений, обычные слова часто оказываются нечеткими и двусмысленными. Пусть X_i обозначает исход i-го испытания, ε_i принимает значения 0 или 1 при каждом i, но, конечно, меняется в зависимости от индекса. Тогда вышеупомянутое предположение может быть записано так:

$$P(X_i = 1) = p; \quad P(X_i = 0) = 1 - p; \quad i = 1, 2, \ldots, n, \tag{2.4.6}$$

где p есть вероятность выпадения решки в каждом из испытаний. Для отдельной, т. е. фиксированной, последовательности $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ из 0 и 1 вероятность соответствующей последовательности исходов равна

$$\begin{aligned} &P(X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots, X_n = \varepsilon_n) = \\ &\quad = P(X_1 = \varepsilon_1) P(X_2 = \varepsilon_2) \ldots P(X_n = \varepsilon_n) \end{aligned} \tag{2.4.7}$$

в силу независимости. Отметим, что каждый из множителей в правой части равен либо p, когда $\varepsilon_i = 1$, либо $1 - p$, когда $\varepsilon_i = 0$. Пусть всего имеется j единиц и $n - j$ нулей. Тогда значением вероятности (2.4.7) будет

$$p^j (1 - p)^{n-j}. \tag{2.4.8}$$

Заметьте, что для всякой последовательности испытаний количество «решек» задается суммой $\sum\limits_{i=1}^{n} X_i$. Важно понимать, что число (2.4.8) не есть вероятность наблюдения j решек в n бросаниях, но вероятность получения определенной последовательности из решек и орлов, в которой присутствуют ровно j решек. Чтобы найти первую вероятность, необходимо подсчитать полное количество таких последовательностей, поскольку все они имеют одинаковую вероятность, задаваемую формулой (2.4.8). Это количество равно биномиальному коэффициенту $\binom{n}{j}$ (подробно обсуждаемому в § 3.2). Каждая из $\binom{n}{j}$ последовательностей соответствует одной возможности наблюдения j решек в n испытаниях, причем все возможности несовместны. Следовательно, из аддитивности вероятности P вытекает, что

$$P\left(\sum_{i=1}^{n} X_i\right) = P(\text{в точности } j \text{ «решек» в } n \text{ испытаниях}) =$$
$$= \binom{n}{j} P(\text{любая фиксированная последовательность}$$
$$\text{из } n \text{ испытаний, содержащая ровно } j \text{ «решек»}) =$$
$$= \binom{n}{j} p^j (1-p)^{n-j}.$$

Этот замечательный результат известен как *формула Бернулли*. Мы не раз будем ссылаться на него.

2.5. Арифметическая плотность*)

В этом параграфе приводится поучительный пример из области арифметики.

Пример 9. Пусть Ω состоит из 120 *натуральных чисел* $\{1, 2, \ldots, 120\}$. В качестве вероятностной меры используем пропорцию, определенную в примере 1 из § 2.1. Рассмотрим множества

$$A = \{\,\omega \mid \omega \text{ делится нацело на } 3\},$$
$$B = \{\,\omega \mid \omega \text{ делится нацело на } 4\}.$$

Таким образом, каждое третье число в Ω принадлежит A и каждое четвертое принадлежит B. Следовательно, получаем пропорции:

$$P(A) = \frac{1}{3}, \quad P(B) = \frac{1}{4}.$$

Что представляет собой множество AB? Это множество чисел, делящихся нацело и на 3, и на 4. Если вы не совсем забыли школьную арифметику, то вы знаете, что это множество образовано числами, кратными $3 \cdot 4 = 12$. Поэтому $P(AB) = 1/12$. Тогда мы можем применить формулу (viii) и найти $P(A \cup B)$:

$$P(A \cup B) = P(A) + P(B) - P(AB) = \frac{1}{3} + \frac{1}{4} - \frac{1}{12} = \frac{1}{2}. \tag{2.5.1}$$

Что это означает? $A \cup B$ содержит элементы Ω, которые делятся нацело на 3 или на 4 (или на оба числа сразу). Мы можем пересчитать их

*) Этот параграф может быть пропущен.

одно за одним, но если вы достаточно сообразительны, то поймете, что нет нужды выполнять эту процедуру целиком. Все, что требуется, — это досчитать до 12 (что составляет 10 % всего множества Ω) и отметить подходящие варианты, например, так:

$$\begin{array}{cccccccccccc}
1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & 10, & 11, & 12. \\
 & & \checkmark & \checkmark & & \checkmark & & \checkmark & \checkmark & & & \checkmark \\
 & & & & & & & & & & & \checkmark
\end{array}$$

Всего отмечено 6 вариантов (один из них — дважды). Поэтому доля $A \cup B$ среди 12 чисел равна $6/12 = 1/2$, что соответствует формуле (2.5.1).

Внимательный читатель может заметить, что в данном случае также справедливо равенство

$$P(AB) = \frac{1}{12} = \frac{1}{3} \cdot \frac{1}{4} = P(A) \cdot P(B).$$

Оно верно благодаря тому, что числа 3 и 4 являются *взаимно простыми*, т. е. не имеют общих делителей, отличных от 1. Давайте рассмотрим другое множество:

$$C = \{\omega \mid \omega \text{ делится нацело на } 6\}.$$

Тогда $P(C) = 1/6$, но чему равна вероятность $P(BC)$? Множество BC включает те целые числа, которые делятся и на 4 и на 6, т. е. делятся на их *наименьшее общее кратное* (помните, что это такое?). В данном случае оно равно не произведению $4 \cdot 6 = 24$, а 12. Поэтому $P(BC) = 1/12$. Кроме того, учитывая, что 12 есть наименьшее общее кратное, можно и теперь ограничиться первыми двенадцатью числами при подсчете пропорции для множества $B \cup C$. Подсчет приводит к ответу $4/12 = 1/3$, который также вытекает из формулы (2.3.7):

$$P(B \cup C) = P(B) + P(C) - P(BC) = \frac{1}{4} + \frac{1}{6} - \frac{1}{12} = \frac{1}{3}\,. \tag{2.5.2}$$

Этот пример служит иллюстрацией для проблемы, возникшей при обсуждении примера 3 из § 2.1. Вместо того, чтобы говорить о пропорции, образуемой числами, кратными, скажем, 3, мы можем изучать соответствующую частоту. Здесь не требуется бросать игральную кость. Бог дал нам натуральные числа (как выразился известный математик Кронекер), и среди них делящиеся на 3 встречаются вполне регулярно — периодически с частотой 1/3. Действительно, пусть $N_n(A)$ обозначает количество чисел, принадлежащих множеству A, среди первых n натуральных. Нетрудно показать, что

$$\lim_{n\to\infty} \frac{N_n(A)}{n} = \frac{1}{3}\,.$$

Обозначим эту предельную частоту через $P(A)$. Интуитивно она понимается как шанс выбора наудачу натурального числа, кратного трем (можно представить, что все числа хранятся в мешке, откуда мы можем их извлекать подобно неразличимым шарам из урны). Конечно, аналогичные пределы существуют для множеств B, C, AB, BC и т. п. Их значения найдены выше. Однако теперь, имея дело с бесконечным выборочным пространством всех натуральных чисел (обозначим его через Ω^*), мы вправе применить тот же самый подход к любому множеству вида

$$A_m = \{\omega \mid \omega \text{ делится нацело на } m\}, \tag{2.5.3}$$

где m — произвольное натуральное число. Почему же тогда мы не использовали эту более естественную и понятную модель?

Ответ, возможно, вас удивит. Согласно определению вероятностной меры из § 2.2, требуется, чтобы каждому подмножеству пространства Ω^* была приписана некоторая вероятность при условии, что Ω^* счетно (что верно в нашем случае). Давайте возьмем, например, множество, состоящее из единственного числа $\{1971\}$, или, если хотите, множество

$$Z = \{\text{все числа от 1 до 1971}\}.$$

Его вероятность определяется как значение предела $\lim\limits_{n\to\infty} N_n(Z)/n$ в соответствии с подходом, использованным для определения вероятности множества A. Но $N_n(Z) = 1971$ для всех $n \geqslant 1971$. Следовательно, предел равен 0. Понятно, что при данном подходе любое конечное множество имеет вероятность 0. Если бы вероятностная мера P была счетно-аддитивной, как требует аксиома (ii*) из § 2.3, то $P(\Omega^*)$ равнялась бы 0, а не 1. Это противоречие показывает, что P не может рассматриваться как вероятностная мера на Ω^*. Тем не менее, она вполне годится для множеств, подобных A_m.

Существует выход из этой парадоксальной ситуации. Мы должны отказаться от требования, чтобы мера была определена на всех подмножествах (натуральных чисел). Пусть задано некоторое конечное число множеств вида A_m. Давайте рассмотрим *производные множества*, образованные из заданных с помощью операций дополнения, объединения и пересечения. Будем говорить, что такой класс множеств *порожден* исходными множествами. Оказывается, тогда удается определить P с помощью описанного выше подхода для всех множеств из данного класса. Множествам, не входящим в него, вероятность вообще не приписывается. Например, множество Z не содержится в классе, порожденном множествами A, B, C. Поэтому его вероятность не равна 0, а просто не определена. В контексте примера 2 из § 2.1, можно также сказать, что множество Z *неизмеримо*. Это разрешает проблему, но мы не станем сейчас углубляться в эту тему, а только приведем еще один пример.

Пример 10. Какова вероятность того, что натуральное число делится нацело на 3, не делится на 5, но при этом делится на 4 или на 6?

Во введенных ранее обозначениях нас интересует множество

$$AD^c(B \cup C),$$

где $D = A_5$. С учетом дистрибутивного закона представим его в виде $AD^cB \cup AD^cC$. Мы также имеем:

$$(AD^cB)(AD^cC) = AD^cBC = ABC - ABCD.$$

Отсюда, ввиду формулы (v),

$$P(AD^cBC) = P(ABC) - P(ABCD) = \frac{1}{12} - \frac{1}{60} = \frac{1}{15}\,.$$

Аналогично находим:

$$P(AD^cB) = P(AB) - P(ABD) = \frac{1}{12} - \frac{1}{60} = \frac{4}{60} = \frac{1}{15}\,,$$

$$P(AD^cC) = P(AC) - P(ACD) = \frac{1}{6} - \frac{1}{30} = \frac{4}{30} = \frac{2}{15}\,.$$

Наконец, в силу соотношения (viii), получаем:

$$P(AD^cB \cup AD^cC) = P(AD^cB) + P(AD^cC) - P(AD^cBC) =$$
$$= \frac{1}{15} + \frac{2}{15} - \frac{1}{15} = \frac{2}{15}\,.$$

Советуем вам проверить полученный результат для пространства Ω из примера 9.

Решение можно несколько упростить, опираясь на следующее арифметическое утверждение: интересующее нас множество состоит из чисел, делящихся нацело на 2 и 3, но не на 5. Тогда наш метод приведет к ответу более коротким путем.

Задачи

1. Вернемся к примеру 1 из § 2.1. Допустим, что каждое хорошее яблоко стоит 1 цент, а гнилое — ничего. Обозначим множество гнилых яблок через R. Переложим некоторую часть яблок из мешка в корзину и обозначим множество переложенных яблок через S. Определим

$$Q(S) = |S \setminus R| \,/\, |\Omega - R|.$$

Тогда Q представляет собой относительную стоимость корзины сравнительно со всем мешком. Покажите, что она является вероятностной мерой.

2. Представим, что земля квадратного королевства разделена на три полосы A, B, C равной площади. Предположим, что цены некоторой единицы площади этих полос относятся как 1:3:2. Тогда относительная цена произвольного (измеримого) участка S сравнительно с ценой всей земли королевства задается формулой

$$V(S) = \frac{P(SA) + 3P(SB) + 2P(SC)}{2},$$

где P определена в примере 2 из § 2.1. Убедитесь, что V — это вероятностная мера.

3*. Обобщим задачу 2. Пусть $a_1, \ldots, a_n$ — произвольные положительные числа, $A_1 + \ldots + A_n = \Omega$ — произвольное разбиение, P — вероятностная мера на Ω и

$$Q(S) = [a_1P(SA_1) + \ldots + a_nP(SA_n)]/[a_1P(A_1) + \ldots + a_nP(A_n)],$$

для любого подмножества S пространства Ω. Покажите, что Q — вероятностная мера.

4. Обозначим через A и B две чашечки кофе, которые вы заказали у бармена во время завтрака. Допустим, что первая чашечка стоила 15 центов, а вторая — 10 центов. Используя букву P для цены, запишите формулу, подобную аксиоме (ii), но со знаком неравенства (P «субаддитивна»).

5. Предположим, что на распродаже покупатель может приобрести 2 рубашки по цене 4$ каждая, в то время как цена одной рубашки равна 5$. Пусть покупатель купил 4 рубашки $S_1, \ldots, S_4$. Запишите формулу, подобную аксиоме (ii), и сравните ее с неравенством из задачи 4. Забудьте о налоге с продаж! (Здесь P «супераддитивна».)

6. Покажите, что если P и Q — две вероятностные меры, заданные на одном и том же (счетном) выборочном пространстве, то $aP + bQ$ также является вероятностной мерой для любых неотрицательных чисел a и b, связанных равенством $a + b = 1$. Приведите конкретный пример такой *смеси*.

7*. Пусть P — вероятностная мера. Убедитесь, что функция $P/2$ удовлетворяет аксиомам (i) и (ii), но не аксиоме (iii). В свою очередь, для функции P^2 справедливы соотношения (i) и (iii), но не обязательно выполняется аксиома (ii). Приведите контрпример к формуле (ii), используя пример 1.

8*. Покажите, что для произвольных множеств A, B, C верны неравенства:
(a) $P(A \cap B \cap C) \leqslant P(A) \wedge P(B) \wedge P(C)$;
(b) $P(A \cup B \cup C) \geqslant P(A) \vee P(B) \vee P(C)$.

9*. Докажите, что для любых двух множеств A и B

$$P(AB) \geqslant P(A) + P(B) - 1.$$

Придумайте конкретный пример для этого неравенства. [*Указание.* Используйте формулу (2.3.4) и законы Де Моргана.]

10. Очевидно, $A \cap A = A$. А в каком случае выполняется соотношение $P(A) \cdot P(A) = P(A)$? Возможно ли, что $P(A) = 0$, но $A \neq \varnothing$?

11. Приведите пример, когда $P(AB) < P(A)P(B)$.

12. Докажите равенство (2.3.7), установив предварительно, что

$$(A \cup B) - A = B - (A \cap B).$$

13. Даны две группы людей. Некоторые люди принадлежат обеим группам. Пусть к группе A относятся 123 человека, к в группе B — 78. Всего в двух группах 184 человека. Сколько человек входят сразу в обе группы?

14. Группы A, B, C включают, соответственно, 57, 49, 43 человека. При этом в группах A и B вместе 13 человек, в A и C — 7, в B и C — 4, а один человек принадлежит всем трем группам сразу. Найдите общее количество людей во всех трех группах.

15*. Решите задачу 14 для произвольных численностей групп (которые, конечно, должны удовлетворять некоторым очевидным неравенствам). В результате, поделив на общее число людей, вы получите формулу, обобщающую равенство (2.3.7) на случай трех множеств (в которой некоторые пересечения могут оказаться пустыми!).

16. Выразите $P(A \bigtriangleup B)$ через $P(A)$, $P(B)$ и $P(AB)$ или через $P(A)$, $P(B)$ и $P(A \cup B)$.

17*. Используя обозначение (2.5.3) и приведенное в контексте определение вероятности, покажите, что для любых m и n

$$P(A_m A_n) \geqslant P(A_m)P(A_n).$$

Когда это неравенство обращается в равенство?

18*. Вспомните, как вычисляются площади с помощью двойного интеграла в математическом анализе. Для областей простого вида (прямоугольника, трапеции, круга) верно представление

$$\text{Площадь области } S = \iint\limits_S 1\,dx\,dy.$$

Покажите, что его можно записать с использованием индикатора I_S так:

$$A(S) = \iint\limits_\Omega I_S(x, y)\,dx\,dy,$$

где Ω — вся плоскость и $I_S(x, y)$ есть значение функции I_S в точке (x, y) (обозначаемой через ω в § 1.4). Удостоверьтесь также, что для двух областей S_1 и S_2 справедливо равенство

$$A(S_1) + A(S_2) = \iint (I_{S_1} + I_{S_2}),$$

в котором опущены некоторые необязательные символы.

19*. Теперь несложно продемонстрировать связь между соотношениями (2.3.7) и (2.3.8), упоминавшуюся в контексте, на примере областей на плоскости.

20. Придумайте несколько последовательностей $\{p_n\}$, удовлетворяющих условиям (2.4.1). Приведите по крайней мере два примера, где все $p_n > 0$.

21*. Выведите из аксиомы (ii*) следующие утверждения.

(a) Если множества A_n расширяются, т. е. $A_n \subset A_{n+1}$ для $n \geqslant 1$, $A_\infty = \bigcup_n A_n$, то $P(A_\infty) = \lim\limits_{n\to\infty} P(A_n)$.

(b) Если множества A_n сжимаются, т. е. $A_n \supset A_{n+1}$ для $n \geqslant 1$, $A_\infty = \bigcap_n A_n$, то $P(A_\infty) = \lim\limits_{n\to\infty} P(A_n)$.

[*Указание.* В случае (a) рассмотрите $A_1 + (A_2 - A_1) + (A_3 - A_2) + \ldots$; в случае (b) перейдите к дополнениям.]

22. Чему равна вероятность (понимаемая в смысле примера 10) того, что выбранное наудачу натуральное число не делится нацело на любое из чисел 3, 4, 6, но делится на 2 или на 5?

23*. Покажите, что если $(m_1, \ldots, m_n)$ — взаимно простые натуральные числа, то события $(A_{m_1}, \ldots, A_{m_n})$, определенные в § 2.5, независимы.

24. Что вы можете сказать о событии A, которое не зависит само от себя? Если A и B не пересекаются и независимы, что можно сказать о них?

25. Докажите, что если (A, B) — пара независимых событий, то независимыми являются также события в парах (A, B^c), (A^c, B), (A^c, B^c). Обобщите этот результат на случай трех независимых событий.

26. Проверьте, что если A, B, C — независимые события, то A и $B \cup C$ независимы, $A \setminus B$ и C независимы.

27. Установите тождество

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - \\ - P(AB) - P(AC) - P(BC) + P(ABC)$$

в случае, когда A, B, C независимы, рассматривая $P(A^cB^cC^c)$. (Тождество справедливо и без предположения о независимости, см. § 6.2.)

28. Бросаются 5 монет. Предположим, что исходы независимы, но вероятности выпадения решки могут отличаться у разных монет. Запишите вероятность осуществления конкретного набора исходов $TTHTH$. Найдите вероятность выпадения в точности трех решек.

29*. Каким образом, не опираясь на условие независимости, можно построить математическую модель произвольных последовательных испытаний? Другими словами, опишите выборочное пространство, подходящее для записи таких испытаний. Что служит математическим определением события, зависящего только от одного испытания, двух испытаний и т. д.? Для ответа вам не понадобится вероятностная мера. А теперь придумайте, как на построенном пространстве задать такую меру, чтобы сделать испытания независимыми. Решение приведено в книге [9, § V.4], но вы поймете его лучше, если прежде попытаетесь придумать его самостоятельно.

Глава 3

Комбинаторика

3.1. Основное правило

Часто для вычисления вероятностей необходимо уметь подсчитывать число возможных случаев, как в примерах 4 и 5 из § 2.2. Такой подсчет образует основу классической теории, произрастающей из поиска оптимальных стратегий в азартных играх. Однако комбинаторная техника применяется также при случайном отборе, ранжировании, разбиении на части, размещении, программировании, построении моделей и во многих других ситуациях. В этой главе мы рассмотрим ряд простых комбинаторных задач и методов их решения.

Иногда обсуждение «перестановок и сочетаний» начинается с вопроса, подобного следующему: «Сколькими способами может одеться человек, комбинируя три рубашки и два галстука?» Здесь участвуют всего два числа, 2 и 3. Ясно, что ответ представляет собой некоторую комбинацию этих чисел. Должны ли мы сложить их $(2+3)$, умножить (2×3) или возвести в степень (2^3 или 3^2)? Вопрос звучит, как риторический, однако на практике встречаются самые разные неправильные ответы. Ввиду этого мы собираемся обсудить данный вопрос подробнее, чем вам, вероятно, представляется необходимым.

Прежде всего, в таком простом случае, как у нас, легко перебрать все возможные варианты, подсчитав их в уме. Обычно используется табличная форма представления вариантов:

Г \ Р	1	2	3
1	11	21	31
2	12	22	32

(3.1.1)

Так как математика есть способ экономии мышления, можно представить их в более компактном виде:

$$(s_1,t_1)(s_1,t_2)(s_2,t_1)(s_2,t_2)(s_3,t_1)(s_3,t_2),$$

и, наконец, легко видеть, что достаточно просто записать

$$(1,1)(1,2)(2,1)(2,2)(3,1)(3,2), \qquad (3.1.2)$$

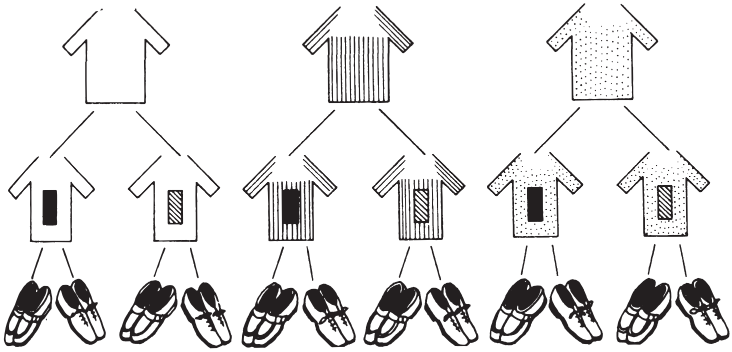


Рис. 15

предполагая, что первая координата указывает вариант выбора рубашки, а вторая — галстука. Тем самым, мы пришли к математическому методу описания совокупности исходов, представленному в виде (3.1.2). Эта совокупность является множеством всех упорядоченных пар (a, b), где $a = 1, 2, 3$; $b = 1, 2$. Теперь понятно, что правильным ответом служит $3 \times 2 = 6$.

В более общей ситуации рассматриваются упорядоченные числовые наборы $(a_1, \ldots, a_k)$ длины k, где для каждого j от 1 до k символ a_j указывает значение (выбранный вариант) j-й координаты, которое является числом между 1 и некоторым m_j. Так, в нашем примере $k = 2$, $m_1 = 3$, $m_2 = 2$ и совокупность всех наборов (a_1, a_2) приведена в формуле (3.1.2).

Символьный способ описания очень удобен. Скажем, если у человека также имеются две пары ботинок, мы просто дополним каждый набор из двух координат, добавив в качестве третьей координаты 1 или 2. Тем самым, каждая первоначальная «двойка» преобразуется в две «тройки», откуда общее число «троек» равно $3 \times 2 \times 2 = 12$. Это и есть количество вариантов выбора человеком рубашки, галстука и пары обуви (см. рис. 15). Как видите, подсчет формализован до автоматизма. По сути, именно на таком языке происходит управление компьютером, в отличие от неконкретности, присущей человеческому общению («по крайней мере», «еще нет»).

Описанная выше идея расщепления на координаты хорошо согласуется с визуальным представлением исходов. Она показывает, почему 3 «рубашки» в результате *умножения* преобразуются в 6 наборов вида «рубашка — галстук» и 12 наборов вида «рубашка — галстук — пара обуви». Итак, будьте внимательны — мы формулируем общее утверждение.

Основное правило. Пусть необходимо несколько раз произвести выбор. Существует m_1 вариантов при первом выборе, m_2 — при втором, m_3 — при третьем и т. д. Если каждый раз выбор производится без всяких ограничений, тогда общее число возможностей для всей последовательности выборов равно

$$m_1 \times m_2 \times m_3 \times \ldots .$$

Формальное доказательство повторяет приведенные выше рассуждения в несколько более «усеченном и высушенном» виде. Его мы оставляем читателю в качестве упражнения. Однако стоит отметить, что выражение «без всяких ограничений» означает в нашем примере, что не принимается во внимание возможное совпадение расцветок рубашек или галстуков.

Пример 1. Меню в ресторане выглядит так:

Выберите одно из блюд:

суп, сок, фруктовый коктейль.

Выберите одно из блюд:

рубленая говядина;
жареная свинина;
цыплята жареные;
спагетти с фрикадельками.

Выберите одно из блюд:

картофельное пюре, брокколи[*)], лимская фасоль.

Выберите одно из блюд:

мороженое, яблочный пирог.

Выберите один из напитков:

кофе, чай, молоко.

Предположим, что вы выбираете ровно одно блюдо из каждого списка без пропусков и замен. Сколько существует разных вариантов? Или, если вы предпочитаете язык, применяемый теперь при принятии более важных решений в подобных ситуациях, сколько *сценариев* комплексного обеда вы можете составить из пяти указанных в меню списков блюд?

Легко подсчитать, что общее количество блюд в меню равно

$$3 + 4 + 3 + 2 + 3 = 15.$$

[*)] Капуста спаржевая. — *Прим. перев.*

Но вы не будете есть их все. В свою очередь, число различных возможных обедов равняется

$$3 \times 4 \times 3 \times 2 \times 3 = 216$$

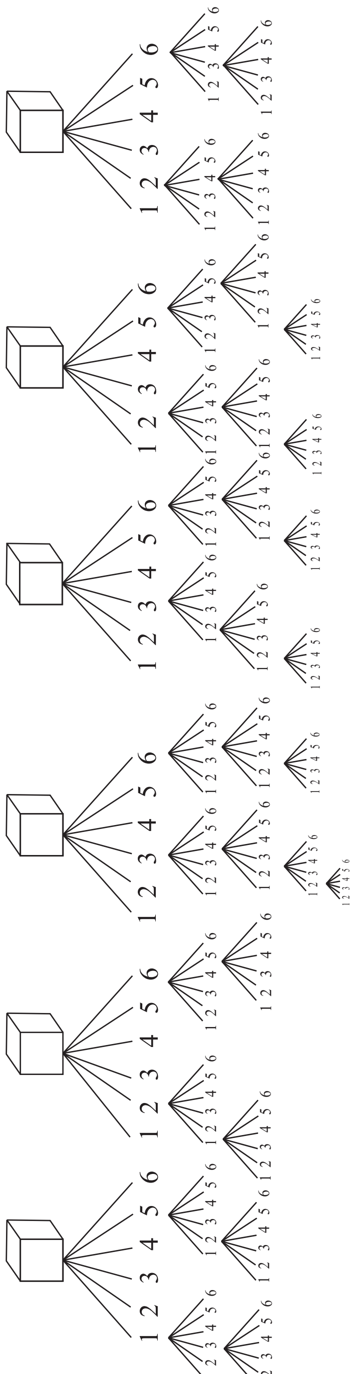
в соответствии с основным правилом. Конечно, за один раз вы съедаете лишь один обед. Однако вы вполне можете перебрать все 216 вариантов, если обладаете широким вкусом и достаточно часто посещаете этот ресторан. Более реалистичным и статистически значимым является следующее соображение: все 216 вариантов обеда могут быть в действительности сервированы для разных клиентов за определенный период времени, возможно даже в течение одного дня. Эта возможность представляет собой эмпирическую основу комбинаторных вычислений и их использования для вычисления вероятностей.

Пример 2. Теперь мы готовы решить задачи, встретившиеся в примерах (e) и (e$'$) из § 1.1: сколько всего существует разных исходов при бросании шести игральных костей, а также каково количество исходов, у которых все номера выпавших граней различны.

С помощью каждой из костей осуществляется случайный выбор одного из шести номеров. В случае первой из задач нет ограничений на номера, наблюдаемые на каждой из 6 костей. Поэтому основное правило прямо приводит к ответу $6^6 = 46656$. Во второй задаче номера нельзя свободно комбинировать ввиду того, что они обязаны все быть различными. Напрямую основное правило не применимо, однако мы можем воспользоваться рассуждениями, на которые оно опирается. Подчас важное значение в математике имеет не слепое следование правилам, а истинное понимание их смысла. (Вероятно, по этой причине для многих студентов «перестановки и сочетания» менее понятны, чем алгебра и математический анализ.) Рассмотрите диаграмму на рис. 16, демонстрирующую процесс ветвления.

На первой из костей может выпасть любой из номеров, но вторая должна иметь другой номер. Следовательно, после того, как сделан первый выбор, для второго остается только 5 возможностей. От номера на первой кости зависит, какие именно 5 вариантов допустимы, но *не их количество.* Поэтому существует всего 6×5 возможностей для первого и второго выборов вместе. После того, как они сделаны, остается 4 варианта для третьего выбора и т. д. Тогда для *последовательности* из шести выборов имеем

$$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6! = 720 \text{ возможностей.}$$



Брошены 6 костей

Рис. 16

Между прочим, убедитесь с помощью диаграммы, что первая кость не имеет никаких преимуществ перед другими. И вообще, какая из них «первая»?

Конечно, можно сформулировать более общее правило с целью охватить также и рассмотренную ситуацию, но надо ли это делать, коль скоро понятен сам принцип подсчета?

3.2. Модели случайного выбора

Давайте познакомимся с несколькими стандартными методами вычислений, используемыми при решении большинства комбинаторных задач. Их удобно изучать либо на моделях случайного выбора, либо на моделях случайного размещения. Начнем с первых.

Представьте, что урна содержит m различимых шаров с номерами от 1 до m. Из нее извлекаются n шаров при соблюдении некоторых условий на способ извлечения. Для каждой модели вычисляются количества всех возможных исходов.

I. Упорядоченный выбор с возвращением

Шары извлекаются наудачу один за другим, причем каждый вынутый шар возвращается назад в урну прежде, чем будет извлечен следующий. При этом записываются номера шаров в порядке их появления. Таким образом, мы имеем дело с упорядоченными наборами $(a_1, \ldots, a_n)$, в которых каждое a_j может принимать любое значение от 1 до m. Основное правило сразу приводит к ответу m^n для полного числа исходов. Такая модель соответствует свободному от ограничений бросанию шести костей. Аналогия будет еще более близкой, если представить, что одна и та же кость последовательно бросается 6 раз. При этом каждое бросание соответствует одному извлечению.

II. Упорядоченный выбор без возвращения

Процесс извлечения такой же, как и в модели I, с тем отличием, что извлеченные шары не возвращаются обратно в урну. Мы снова имеем дело с упорядоченными наборами $(a_1, \ldots, a_n)$, но уже с ограничением, что в них все a_j различны. Конечно, должно выполняться неравенство $n \leqslant m$. Основное правило напрямую не применимо. Тем не менее, используя «метод ветвления» из примера 2, приходим к ответу

$$m \cdot (m-1) \cdot (m-2) \ldots (m-n+1) = (m)_n. \tag{3.2.1}$$

Обратите внимание, что в левой части равенства (3.2.1) присутствуют ровно n сомножителей, причем последним является множитель $m-(n-1)$, а не $m-n$ (почему?). Мы ввели обозначение $(m)_n$ для сокращенной записи «длинного» произведения из левой части формулы (3.2.1).

Модель II имеет важный частный случай — модель перестановок.

IIa. Перестановка из m различимых шаров

Рассмотрим модель II при $m=n$. Тогда все m шаров извлекаются один за другим без возвращений. Результатом выбора является набор из m занумерованных шаров, расставленных в некотором порядке. Полное количество возможностей совпадает с числом всех расположений (упорядочений, ранжировок, перестановок) элементов множества $\{1,2,\ldots,m\}$. Это число называется *факториалом* от m и обозначается

$$m! = (m)_m = m(m-1)\ldots 2\cdot 1.$$

Ниже приведены значения $m!$ для $m=1,2,\ldots,10$.

m	$m!$
1	1
2	2
3	6
4	24
5	120
6	720
7	5 040
8	40 320
9	362 880
10	3 628 800

III. Неупорядоченный выбор без возвращения

В данной модели вынутые шары не возвращаются назад в урну, а также не фиксируется порядок их номеров в процессе извлечения. Другими словами, можно представлять себе, что все n шаров вынимаются сразу за одно извлечение. Следовательно, мы имеем дело с выбором произвольного подмножества размера n из множества (*популяции*) размера m. Для подсчета числа исходов модифицируем модель II за счет дополнительного ранжирования по возрастанию номеров вынутых шаров. Тогда согласно модели IIa произвольный упорядоченный по возрастанию набор из n номеров может появиться в результате $n!$ разных последовательных извлечений шаров с этими номерами. Таким образом, упорядоченная выборка размера n порождает $n!$ неупорядоченных, по каждой из которых можно однозначно восстановить исходную. Например, при $m = 5$ и $n = 3$ подмножество $\{2,3,5\}$ получается в результате извлечения следующих $3! = 6$ последовательных наборов:

$$(2,3,5)(2,5,3)(3,2,5)(3,5,2)(5,2,3)(5,3,2).$$

В общем случае из обсуждения модели II нам известно, что количество последовательных наборов размера n равно $(m)_n$. Давайте обозначим через x искомое число исходов (подмножеств размера n). Проведенные выше рассуждения показывают, что

$$n!\,x = (m)_n.$$

Отсюда сразу получаем искомый ответ, который будем обозначать

$$\binom{m}{n} = \frac{(m)_n}{n!}\,. \tag{3.2.2}$$

Если умножить числитель и знаменатель на $(m-n)!$, то с учетом соотношения (3.2.1) видим, что

$$\begin{aligned}\binom{m}{n} &= \frac{(m)_n(m-n)!}{n!\,(m-n)!} = \\ &= \frac{m(m-1)\dots(m-n+1)(m-n)\dots 2\cdot 1}{m!\,(m-n)!} = \frac{m!}{n!\,(m-n)!}\,.\end{aligned} \tag{3.2.3}$$

Когда $n = m$, существует в точности одно подмножество размера n — само множество. Чтобы формула оставалась верной и в этом случае, значение (3.2.3) должно равняться 1. Поэтому необходимо положить $0!=1$. С учетом данного соглашения формула (3.2.3) справедлива при $0 \leqslant n \leqslant m$. Число $\binom{m}{n}$ называется *биномиальным коэффициентом* и играет важную роль в теории вероятностей. Заметьте, что верно тождество

$$\binom{m}{n} = \binom{m}{m-n}, \tag{3.2.4}$$

которое немедленно вытекает из определения (3.2.3). Оно очевидно и без этой ссылки, если интерпретировать обе его части как вычислительные формулы (почему?).

Рассуждение, изложенное для модели III, приводит к следующему обобщению IIIa.

IIIa. Перестановка из m шаров, неразличимых внутри групп

Допустим, что у нас есть m_1 шаров цвета номер 1, m_2 шаров цвета номер 2, $\dots$, m_r шаров цвета номер r. Цвета различаются, а шары одного цвета — нет. Конечно, $m_1 + m_2 + \ldots + m_r = m$. Сколько существует отличающихся перестановок таких шаров?

Например, если $m_1 = m_2 = 2$, $m = 4$, цвета — черный и белый, то всего существует 6 следующих перестановок:

●●○○ ●○●○ ●○○● ○●●○ ○●○● ○○●●

Для получения ответа в общем случае сравним данную модель с моделью IIa, в которой все шары различимы. Дополнительно занумеруем шары цвета 1 числами от 1 до m_1, шары цвета 2 — числами от 1 до m_2 и т. д. Тогда все шары станут различимыми, и, согласно модели IIa, полное число перестановок после нумерации будет равно $m!$. При этом m_1 шаров цвета 1 можно с учетом их номеров разместить $m_1!$ способами, m_2 шаров цвета 2 — $m_2!$ способами и т. д. Каждое размещение шаров одного цвета может быть скомбинировано с произвольным размещением шаров другого цвета. Следовательно, для каждой первоначальной перестановки без различения шаров одного цвета в силу основного правила существует

$$m_1!\, m_2! \ldots m_r!$$

новых способов размещения с учетом нумерации. Рассуждения, аналогичные проведенным для модели III, показывают, что искомое число ненумерованных перестановок равно частному

$$\frac{m!}{m_1!\, m_2! \ldots m_r!}\,.$$

Оно называется *мультиномиальным**) коэффициентом. Когда $r = 2$, коэффициент сводится к биномиальному $\binom{m}{m_1} = \binom{m}{m_2}$.

IV. Неупорядоченный выбор с возвращением

Из урны извлекаются один за другим n шаров, каждый вынутый шар возвращается назад прежде, чем будет извлечен следующий. При этом (возможно, повторяющиеся) номера всех вынутых шаров регистрируются в виде неупорядоченного набора (группы), т. е. без обращения внимания на порядок их появления. Это — не совсем привычная ситуация, поэтому мы начнем с числового примера. Возьмем $m = n = 3$. Все возможные исходы для данной модели указаны в первом из столбцов

*) Или полиномиальным. — *Прим. перев.*

следующей таблицы:

$$
\begin{array}{r|l|l}
111 & \checkmark\checkmark\checkmark|\quad | & \checkmark\checkmark\checkmark|| \\
112 & \checkmark\checkmark|\quad \checkmark| & \checkmark\checkmark|\checkmark| \\
113 & \checkmark\checkmark|\quad |\checkmark & \checkmark\checkmark||\checkmark \\
122 & \checkmark|\quad \checkmark\checkmark| & \checkmark|\checkmark\checkmark| \\
123 & \checkmark|\quad \checkmark|\checkmark & \checkmark|\checkmark|\checkmark \\
133 & \checkmark|\quad |\checkmark\checkmark & \checkmark||\checkmark\checkmark \\
222 & |\checkmark\checkmark\checkmark| & |\checkmark\checkmark\checkmark| \\
223 & |\quad \checkmark\checkmark|\checkmark & |\checkmark\checkmark|\checkmark \\
233 & |\quad \checkmark|\checkmark\checkmark & |\checkmark|\checkmark\checkmark \\
333 & |\quad |\checkmark\checkmark\checkmark & ||\checkmark\checkmark\checkmark
\end{array}
\tag{3.2.5}
$$

Понятен ли вам организационный принцип, лежащий в ее основе?

В общем случае рассмотрим «итоговый список», в котором в верхней строке перечислены номера шаров:

1	2	3	4		m
✓✓	✓		✓✓✓		

После очередного извлечения поставим «галочку» под номеров вынутого шара. Таким образом, по окончании процесса извлечения количество «галочек» в списке станет равно n (которое, вообще говоря, может быть больше, чем m). Может случиться так, что все «галочки» будут стоять только под одним из номеров, или под некоторыми номерами окажутся пробелы. Удалив необязательную первую строку с номерами, представим в данном виде исходы из первого столбца таблицы (3.2.5). В результате получим второй столбец. Убедитесь, что при этом не происходит потери информации. Наконец, выравнивая символы «галочек» и разделительных «черточек», образуем третий столбец. Теперь забудем о первых двух столбцах и сконцентрируемся на изучении третьего. Понятно ли вам, как восстановить на основе этой несложной криптограммы исходные итоговые списки? Видите ли вы, что здесь перечислены все возможные варианты размещения трех «галочек» и двух «черточек»? Таким образом, общее число вариантов (согласно модели IIIа при $m = 5$, $m_1 = 3$, $m_2 = 2$ или модели III при $m = 5$, $n = 3$) равно $5!/(3!2!) = 10$ в полном соответствии с таблицей. Оно, конечно, совпадает с числом всех возможных итоговых списков.

Для произвольных m и n подсчет числа исходов в модели IV сводится тем же самым методом к нахождению числа размещений n «галочек» и $m - 1$ «черточек» (для разделения m промежутков требуются $m - 1$ границ). Рекомендуем вам дополнительно обдумать этот вопрос и удостовериться во взаимной однозначности данного соответствия в общем случае. Из модели IIIa известно, что ответом является

$$\binom{m+n-1}{n} = \binom{m+n-1}{m-1}. \tag{3.2.6}$$

Это и есть искомое число исходов в модели IV.

Пример 3. Метод подсчета Даламбера, упоминавшийся в примере 4 из § 2.2, эквивалентен применению модели IV при $m = n = 2$. Бросание монеты соответствует выпадению орла или решки. Результат бросания двух монет учитывается целиком, без фиксации очередности, что приводит к трем возможным исходам:

✓✓|	✓|✓	|✓✓
HH	HT = TH	TT

Аналогично, при бросании шести неотличимых игральных костей общее число различных вариантов подсчитывается с помощью формулы (3.2.6) при $m = n = 6$:

$$\binom{6+6-1}{6} = \binom{11}{6} = 462.$$

Это количество составляет менее 1 % от числа 46 656 из модели I.

Проиллюстрируем на простом числовом примере отличие разных методов подсчета для рассмотренных четырех процедур выбора при $m = 4$, $n = 2$:

Модель (I)

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$

Модель (II)

	$(1,2)$	$(1,3)$	$(1,4)$
$(2,1)$		$(2,3)$	$(2,4)$
$(3,1)$	$(3,2)$		$(3,4)$
$(4,1)$	$(4,2)$	$(4,3)$	

Модель (IV)

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$
	$(2,2)$	$(2,3)$	$(2,4)$
		$(3,3)$	$(3,4)$
			$(4,4)$

Модель (III)

$(1,2)$	$(1,3)$	$(1,4)$
	$(2,3)$	$(2,4)$
		$(3,4)$

3.3. Модели размещения. Биномиальные коэффициенты.

Источником как воодушевления, так и разочарования в комбинаторике является то обстоятельство, что некоторые задачи встречаются в разном обличии, причем подчас непросто установить их идентичность. Примерами могут служить модель IV или модель IIIb, обсуждаемая ниже. Люди, обладающие разными типами мышления, часто предпочитают одно из представлений другому. Однако полезно ознакомиться с альтернативами (по тем же соображениям, по которым полезно изучать иностранные языки). Выше мы рассмотрели некоторые основные методы подсчета вариантов в моделях случайного выбора. Другой подход, которому отдают предпочтение физики и инженеры, — размещение шаров по ящикам. Так как теперь шары играют совсем другую роль, чем прежде, мы станем вместо них использовать *жетоны*, чтобы в дальнейшем не возникло путаницы.

Итак, у нас есть m ящиков, занумерованных числами от 1 до m, и n жетонов с номерами от 1 до n. Жетоны размещаются по ящикам с условием, что ни один ящик не может содержать более одного жетона (или без этого условия). Мы записываем результат размещения (заполнения), отмечая количества жетонов в каждом из ящиков, при этом учитывая (или нет) номера жетонов. Следующие четыре случая соответствуют рассмотренным выше четырем моделям выбора:

I$'$. Каждый ящик может содержать произвольное количество жетонов. Номера жетонов учитываются (жетоны различимы).

II$'$. Ни один ящик не может содержать более одного жетона. Номера жетонов учитываются (жетоны различимы).

III$'$. Ни один ящик не может содержать более одного жетона. Номера жетонов не учитываются (жетоны не различимы).

IV$'$. Каждый ящик может содержать произвольное количество жетонов. Номера жетонов не учитываются (жетоны не различимы).

Мы не видим большой пользы в изложении *доказательства* идентичности каждого из случаев соответствующей модели, так как оно представляет собой рассуждение, которое лучше провести самостоятельно, чтобы убедиться в его справедливости. (Некоторые учителя заходят настолько далеко, что утверждают, что вообще нельзя научить комбинаторному мышлению.) Тем не менее, укажем некоторые ключевые соответствия, возникающие при сопоставлении двух представлений.

Выбор	**Размещение**
Шар	Ящик
Количество извлечений	Количество жетонов
j-м вынут шар с номером k	j-й жетон попал в ящик с номером k

В некотором смысле новое представление является более гибким, благодаря легкости наложения дополнительных условий на размещение жетонов. Например, можно потребовать, чтобы ни один из ящиков не оставался пустым при $n \geqslant m$, или задать «веса» некоторых ящиков. Конечно, подобные ограничения возможно перевести на язык случайного выбора, но там они будут выглядеть менее естественно. Укажем один из важных примеров такого сорта, который есть не что иное, как модель IIIa в новом обличии.

IIIb. Разбиение на занумерованные группы

Разобьем множество из m объектов на r *подмножеств* или, проще говоря, групп так, что m_1 объектов окажутся в группе с номером 1, m_2 объектов — в группе с номером 2, ..., m_r объектов — в группе с номером r, где $m_1+m_2+\ldots+m_r=m$ и все $m_j \geqslant 1$. Очевидно, что данное разбиение соответствует размещению m жетонов по r ящикам, где m_j — количество жетонов, попавших в ящик с номером j. Важно отметить, что эта процедура не совпадает с разбиением объектов на r групп размеров m_1, m_2, ...,m_r, так как в нашем случае группы занумерованы (т. е. различимы). Простой пример поможет прояснить различие.

Пример 4. Сколькими способами можно разбить четырех человек на две пары?

Этот вопрос, конечно, не является однозначно понимаемым, но, скорее всего, в качестве ответа первыми на ум приходят следующие три варианта:

$$(12)(34) \quad (13)(24) \quad (14)(23). \tag{3.3.1}$$

Такая интерпретация годится, если две пары участвуют в соревновании по шахматам или настольному теннису, причем для обеих пар доступны два одинаковых стола. Напротив, давайте представим, что четыре человека собираются играть в теннис парами, и пара, считающаяся «первой», обладает правом выбора стороны корта или подачи. Тогда имеет значение, какая из пар считается первой: (12) или (34). Поэтому каждый вариант в записи (3.3.1) следует преобразовать в два новых, меняя порядок пар. В результате получаем 6 возможностей:

$$(12)(34) \quad (34)(12) \quad (13)(24) \quad (24)(13) \quad (14)(23) \quad (23)(14).$$

Именно эта ситуация является частным случаем разбиения вида IIIb, к подсчету числа возможных исходов в котором мы сейчас переходим.

Представьте, что жетоны (люди) распределяются по ящикам (группам). В соответствии с моделью III существуют $\binom{m}{m_1}$ способов выбора m_1 жетонов, помещаемых в ящик 1; после этого, имеется $\binom{m-m_1}{m_2}$ способов выбора m_2 жетонов, помещаемых в ящик 2, из оставшихся $m - m_1$ и т. д. Основное правило здесь не применимо, но мы можем использовать его модификацию, рассмотренную при изучении модели II. Результат таков:

$$\binom{m}{m_1}\binom{m-m_1}{m_2}\binom{m-m_1-m_2}{m_3}\ldots\binom{m-m_1-m_2-\ldots-m_{r-1}}{m_r} =$$
$$= \frac{m!}{m_1!\,(m-m_1)!}\,\frac{(m-m_1)!}{m_2!\,(m-m_1-m_2)!}\,\frac{(m-m_1-m_2)!}{m_3!\,(m-m_1-m_2-m_3)!}\times\ldots$$
$$\ldots\times\frac{(m-m_1-\ldots-m_{r-1})!}{m_r!\,0!} = \frac{m!}{m_1!\,m_2!\ldots m_r!}\,. \qquad (3.3.2)$$

Обратите внимание, что данный метод не приводит к повторному учету исходов даже в случае, когда некоторые группы имеют одинаковые размеры, как в нашем примере игры в теннис. Это происходит благодаря фиксации номеров ящиков (групп). С другой стороны, важно подчеркнуть отличие термина «занумерованные группы» от термина «упорядоченные группы», который не совсем корректно используется в данной ситуации некоторыми авторами. Дело в том, что здесь группы не расставляются по порядку. Для прояснения этого (скорее лингвистического) недоразумения рассмотрим еще один простой пример.

Пример 5. Шесть альпинистов решили разделиться на три группы для совершения финального восхождения на вершину. Группы будут иметь размеры 1, 2, 3. Важен также порядок развертывания*). Каково полное число возможных вариантов разбиения на группы с учетом порядка развертывания?

Число способов разбиения на группы G_1, G_2, G_3, где индекс обозначает размер группы, задается формулой (3.3.2):

$$\frac{6!}{1!\,2!\,3!} = 60.$$

Сформировав группы, необходимо решить, какая из них будет лидировать, какая пойдет в середине, какая окажется последней. Такая задача встречалась при обсуждении модели IIa: количество соответствующих

*) То есть, какая из групп пойдет первой, какая — второй и т. д. — *Прим. перев.*

вариантов равно $3! = 6$. При этом каждое разбиение на группы можно комбинировать с любым упорядочением групп. Применяя основное правило, приходим к искомому ответу: $60 \cdot 6 = 360$.

Что произойдет, если некоторые группы будут иметь одинаковые размеры? Подумайте еще раз об игроках в теннис.

Возвращаясь к формуле (3.3.2), вспомним, что стоящий справа мультиномиальный коэффициент был ранее получен как решение перестановочной проблемы IIIa. Теперь же он появился в результате основанного на многократном использовании модели III подсчета, относящегося к перебору комбинаций. Тщетно пытаться наклеивать ярлыки на проблемы с целью отделить перестановочные от комбинационных. Большинство комбинаторных вычислений включают в себя смесь разных методов, рассмотренных выше. Мы собираемся сейчас продемонстрировать это на ряде технических задач.

В оставшейся части параграфа устанавливаются некоторые полезные формулы, связанные с биномиальными коэффициентами. Прежде всего, введем соглашение

$$\binom{m}{n} = 0 \quad \text{если } m < n \text{ или если } n < 0. \tag{3.3.3}$$

Затем покажем, что справедливо тождество

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n}, \quad 0 \leqslant n \leqslant m. \tag{3.3.4}$$

Оно легко проверяется непосредственно с помощью выражения (3.2.3) для $\binom{m}{n}$. Тем не менее, приведем чисто комбинаторное доказательство без использования вычислений. Напомним, что $\binom{m}{n}$ есть количество разных способов выбора n объектов из m объектов «за одно извлечение». Теперь давайте считать какой-то из объектов «особенным». Этот объект может попасть в группу извлеченных, а может и не попасть. Если он попал, то число способов выбора еще $n-1$ объектов из остальных $m-1$ объектов равно $\binom{m-1}{n-1}$. Если не попал, то число способов выбора всех n объектов из остальных $m-1$ объектов равно $\binom{m-1}{n}$. Сумма вариантов двух альтернатив дает общее количество способов выбора, что и утверждается тождеством (3.3.4). Изящно, не правда ли?

Из формулы (3.3.4) получаем следующий метод вычисления коэффициентов $\binom{m}{n}$, $0 \leqslant n \leqslant m$, одного за другим:

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 \\
. & & . & & . & & . & & . & & . & & . & & .
\end{array}
\tag{3.3.5}
$$

Например, каждое число в последней строке получается сложением его двух соседей из предпоследней строки (пропуски считаются нулями):

$$
\begin{array}{cccc}
1 = 0 + 1, & 7 = 1 + 6, & 21 = 6 + 15, & 35 = 15 + 20, \\
35 = 20 + 15, & 21 = 15 + 6, & 7 = 6 + 1, & 1 = 1 + 0.
\end{array}
$$

Другими словами,

$$\binom{7}{n} = \binom{6}{n-1} + \binom{6}{n} \quad \text{для } 0 \leqslant n \leqslant 7.$$

Последовательность чисел в форме (3.3.5) называется *треугольником Паскаля*, хотя, безусловно, он не был первым, кто ее использовал.

*) Заметьте, что подобно тому, как мы расщепили коэффициент $\binom{m}{n}$, можно расщепить и последний член $\binom{m-1}{n}$ в равенстве (3.3.4), применив доказанную формулу к числу $m-1$. Продолжая эту процедуру, получаем последовательно:

$$
\begin{aligned}
\binom{m}{n} &= \binom{m-1}{n-1} + \binom{m-1}{n} = \binom{m-1}{n-1} + \binom{m-2}{n-1} + \binom{m-2}{n} = \\
&= \binom{m-1}{n-1} + \binom{m-2}{n-1} + \binom{m-3}{n-1} + \binom{m-3}{n} = \ldots .
\end{aligned}
$$

Окончательный результат имеет вид

$$
\begin{aligned}
\binom{m}{n} &= \binom{m-1}{n-1} + \binom{m-2}{n-1} + \ldots + \binom{n}{n-1} + \binom{n}{n} = \\
&= \sum_{k=n-1}^{m-1} \binom{k}{n-1} = \sum_{k \leqslant m-1} \binom{k}{n-1},
\end{aligned}
\tag{3.3.6}
$$

*) Заключительная часть параграфа может быть опущена.

поскольку последнее слагаемое есть $\binom{n}{n} = \binom{n-1}{n-1}$, а слагаемые для $k < n-1$ равны нулю в силу соглашения (3.3.3).

Пример. $35 = \binom{7}{4} = \binom{6}{3} + \binom{5}{3} + \binom{4}{3} + \binom{3}{3} = 20 + 10 + 4 + 1$. Посмотрите, как располагаются эти числа на треугольнике Паскаля.

В качестве применения формулы (3.3.6) дадим еще одно решение проблемы подсчета числа исходов в модели IV из § 3.2. Согласно альтернативной формулировке IV$'$, оно совпадает с числом размещений n неразличимых (не имеющих номеров) жетонов по m ящикам без ограничений на количество жетонов в одном ящике и равно $\binom{m+n-1}{m-1}$ в силу соотношения (3.2.6). Однако приведенное ранее доказательство этого факта представляется довольно хитроумным. Предположим, что мы не настолько проницательны, чтобы установить результат таким методом, но смогли угадать правильный ответ путем экспериментирования с небольшими значениями m и n. Используем теперь нашу догадку для вывода формулы в общем случае. (Вероятно, хитроумное доказательство в действительности было открыто после того, как ответ был угадан.)

Применим математическую индукцию по параметру m. Для $m = 1$, очевидно, существует только один вариант размещения всех жетонов, неважно к каком количестве, в единственном ящике, что согласуется с доказываемой формулой, поскольку $\binom{1+n-1}{1-1} = \binom{n}{0} = 1$. Далее допустим, что формула верна для любого количества жетонов, когда число ящиков равно $m-1$. Добавим еще один ящик, в который мы можем поместить произвольное количество жетонов. Если положить в него j жетонов, то остальные $n-j$ необходимо разместить в других $m-1$ ящиках. В соответствии с индукционным предположением имеется $\binom{m-2+n-j}{m-2}$ способов сделать это. Суммируя по всем возможным значениям j, получаем

$$\sum_{j=0}^{n} \binom{m-2+n-j}{m-2} = \sum_{k=m-2}^{m+n-2} \binom{k}{m-2},$$

где мы заменили индекс суммирования на $k = m-2+n-j$. В силу формулы (3.3.6) сумма справа равняется $\binom{m+n-1}{m-1}$, что и требовалось доказать.

Теперь давайте установим справедливость тождества

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} = \sum_{k=0}^{n} \binom{n}{k} = 2^n; \tag{3.3.7}$$

другими словами, сумма чисел в n-й строке треугольника Паскаля равна 2^n (считается что верхняя строка в (3.3.5) имеет номер 0). Если вы знакомы с биномом Ньютона, то сразу выведите тождество (3.3.7) из равенства

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{3.3.8}$$

подставив $a = b = 1$. Приведем комбинаторное доказательство. Слагаемые в левой части соотношения (3.3.7) представляют собой количества способов выбора $0, 1, 2, \ldots, n$ объектов из n объектов. Следовательно, вся сумма равна полному числу способов выбора *произвольного* подмножества (включая пустое и само множество) из множества размера n. При таком выборе каждый объект может быть как включен в подмножество, так и не включен, причем включения или исключения объектов могут свободно комбинироваться друг с другом. Поэтому основное правило дает в качестве полного числа вариантов величину

$$\underbrace{2 \times 2 \times \ldots \times 2}_{n \text{ раз}} = 2^n.$$

Она совпадает с правой частью формулы (3.3.7). Таким образом, мы нашли число всех подмножеств множества размера n.

Пример. При $n = 2$ подмножествами множества $\{a, b\}$ являются

$$\varnothing, \{a\}, \{b\}, \{ab\}.$$

При $n = 3$ все подмножества множества $\{a, b, c\}$ указаны в списке

$$\varnothing, \{a\}, \{b\}, \{c\}, \{ab\}, \{ac\}, \{bc\}, \{abc\}.$$

В заключение параграфа докажем, что для натуральных $k \leqslant m$ верно равенство

$$\binom{m}{n} = \sum_{j=0}^{k} \binom{k}{j} \binom{m-k}{n-j}. \tag{3.3.9}$$

Обратите внимание, что верхние (нижние) индексы в правой части в сумме дают верхний (нижний) индекс в левой части:

$$k + (m-k) = m; \quad j + (n-j) = n.$$

Отметим также, что нет необходимости точно указывать диапазон изменения индекса j. Можно считать, что j пробегает все целые числа, поскольку лишние члены автоматически равны нулю в силу соглашения (3.3.3). Чтобы убедиться в справедливости тождества (3.3.9), представим, что m объектов разделены на две группы, одна из которых содержит k объектов, а другая — $m - k$ объектов. Для выбора n объектов из всего множества объектов можно взять j объектов из первой группы,

$n-j$ объектов из второй группы и объединить их. Согласно основному правилу, для каждого фиксированного значения j количество комбинаций равно $\binom{k}{j}\binom{m-k}{n-j}$. Позволяя j принимать все возможные значения и суммируя результаты, получаем полное число вариантов выбора, которое равно $\binom{m}{n}$. Не стоит беспокоиться о «недопустимых» значениях j, когда $j > n$ или $n-j > m-k$, поскольку соответствующие члены равны нулю ввиду нашего соглашения.

Пример.

$$\binom{7}{3} = \binom{3}{0}\binom{4}{3} + \binom{3}{1}\binom{4}{2} + \binom{3}{2}\binom{4}{1} + \binom{3}{3}\binom{4}{0},$$
$$\binom{7}{5} = \binom{3}{1}\binom{4}{4} + \binom{3}{2}\binom{4}{3} + \binom{3}{3}\binom{4}{2}.$$

В частности, если в соотношении (3.3.9) $k = 1$, возвращаемся к тождеству (3.3.4), причем наше доказательство сводится к приведенному выше доказательству формулы (3.3.4).

Алгебраический вывод формулы (3.3.9), вместе с обобщением ее на случай, когда верхние индексы не являются натуральными числами, будет дан в гл. 6.

3.4. Как решать комбинаторные задачи

Этот параграф можно также озаглавить «Как считать варианты». У многих студентов такие задачи вызывают затруднения. Причина в том, что они привыкли к обычным для многих элементарных математических курсов шаблонным задачам типа «решите уравнение $x^2 - 5x + 10 = 0$, «продифференцируйте (может быть, дважды) функцию xe^{-x}» и т. п. Каждый человек, запомнив несколько правил, способен решать подобные задачи без всяких дополнительных размышлений. Конечно, задачи такого вида встречаются также и при подсчете «перестановок и комбинаций». Вы найдете их среди упражнений в конце главы. Например, к ним относится известная формула решения проблемы «круглого стола»: «Сколькими различными способами можно рассадить 8 человек за круглый стол?». Если вы ее знаете, то можете дать ответ даже не задумываясь, что означает слово «различными». Однако небольшое изменение условий, вероятно, вызовет большие затруднения. Хотя это и звучит банально, но правда состоит в том, что не существует альтернативы глубокому пониманию. Тем не менее, сложно усваивать принципы без конкретных применений. Надеемся, что многочисленные примеры, приводимые ниже, послужат хорошей проверкой понимания. Другие задачи вы найдете в конце главы. Прежде, чем перейти к детальному обсуждению примеров, дадим некоторые рекомендации, которые помогут вам самостоятельно решать комбинаторные задачи. Они неизбежно являются довольно общими и легко забываемыми, но иногда будут полезными.

(a) Если вы *не представляете себе ясно* проблему, рассмотрите некоторые частные (но не слишком частные) случаи с небольшими числами, чтобы прояснить ситуацию. Это позволит понять, что именно требуется подсчитать, и особенно поможет вам обнаружить возможные дублирования и пропуски вариантов.

(b) Разделите задачу на части, которые представляются вам проще и яснее, чем сама задача, на которых легко сосредоточиться. Это иногда удается сделать, зафиксировав одну из «переменных». Если повезет, части могут оказаться однотипными подзадачами.

(c) Не пытайтесь разбивать подсчет на шаги, если трудности быстро возрастают от шага к шагу. По опыту авторов из всех негативных рекомендаций на данный совет студенты меньше всего обращают внимание, хотя он больше других его заслуживает. Вычисления шаг за шагом могут оказаться простыми для первой пары шагов, но часто не понятно, как пройти весь путь до конца.

(d) Не сдавайтесь, если в формулировке задачи присутствует двусмысленность. Это — семантическая проблема, а не математическая. Попробуйте найти решения для всех интерпретаций, если необходимо. Возможно, такой подход не является наилучшей стратегией, но он весьма уместен, если вы хотите понять суть дела. Во всяком случае, не пытайтесь облегчить задачу, ограничиваясь наиболее удобной интерпретацией, или воспользоваться оплошностью преподавателя, позволившего вам превратить содержательную задачу в тривиальную (см. задачу 13).

Задача 1 (*контроль качества*). Пусть из 550 яблок в мешке 2 % гнилых. Чему равна вероятность того, что случайная выборка из 25 яблок содержит 2 гнилых?

Эта задача лежит в основе проверки качества продукции методом выборочного контроля. Если, в предположении, что доля брака составляет некоторый заданный процент, вероятность окажется слишком малой, то данное предположение следует считать сомнительным и допустить, что истинная доля брака больше. Задачу 1 нетрудно решить для произвольных чисел, поэтому сформулируем ее в общем виде. Пусть среди m изделий k являются бракованными. Чему равна вероятность того, что случайная выборка размера n содержит j бракованных изделий? Слово «случайная» означает, что все выборки размера n, в соответствии с моделью III из § 3.2, считаются одинаково правдоподобными. Поэтому их полное число есть $\binom{m}{n}$. Сколько из них содержат ровно j бракованных изделий? Чтобы сформировать такую выборку, надо выбрать любые j изделий из k бракованных и скомбинировать их с любыми $n-j$ изделиями из $m-k$ качественных. Согласно модели III, первый выбор может быть осуществлен $\binom{k}{j}$ способами, второй — $\binom{m-k}{n-j}$ способами. В силу основного правила количество выборок размера n, содержащих j бракованных изделий, равно произведению этих биномиальных коэффициентов, и, следовательно, искомая вероятность есть отношение

$$\binom{k}{j}\binom{m-k}{n-j}\Big/\binom{m}{n}. \tag{3.4.1}$$

В случае с яблоками имеем $m = 550$, $k = 11$, $n = 25$, $j = 2$. Отсюда вероятность равна

$$\binom{11}{2}\binom{539}{23}\Big/\binom{550}{25}.$$

Вычислить это число непросто. Позже мы покажем, как его можно достаточно точно аппроксимировать. Есть также соответствующие таблицы.

Если сложить вероятности вида (3.4.1) по всем j, $0 \leqslant j \leqslant n$, то в результате должна получиться 1, так как перебираются все возможности. Тем самым, мы вывели формулу

$$\sum_{j=0}^{k}\binom{k}{j}\binom{m-k}{n-j} = \binom{m}{n}$$

с помощью вероятностного подхода. Она совпадает с тождеством (3.3.9). Небольшое размышление позволит вам убедиться, что два доказательства на самом деле эквивалентны.

Задача 2. Чему равна вероятность того, что при выкладывании на стол всех карт из тщательно претасованной колоды для покера появятся четыре туза подряд?

Всего в колоде 52 карты, среди которых 4 туза. Тщательная перетасовка означает, что все перестановки карт одинаково правдоподобны. Согласно модели IIа, для выкладывания на стол всех карт колоды существует $(52)!$ исходов. Во скольких из них четыре туза окажутся вместе?

Используя совет (b), разобьем задачу на части в зависимости от того, где располагается группа тузов. Так как тузы идут подряд, достаточно указать место первого из них. Он может быть в колоде первым, вторым, и т. д. до 49-го места. Следовательно, для группы из 4 тузов имеется 49 позиций. Когда позиция фиксирована, тузы можно переставлять между собой 4! способами. То же самое относится и к остальным 48 картам (модель IIIa при $r = 2$, $m_1 = 4$, $m_2 = 48$). Применяя основное правило, получаем ответ:

$$\frac{49 \cdot 4!\,(48)!}{(52)!} = \frac{24}{52 \cdot 51 \cdot 50}\,.$$

На примере данной задачи полезно посмотреть, как работает совет (a). Возьмите 4 карты, среди которых 2 туза. Полное число перестановок с двумя тузами подряд равно всего лишь $3 \cdot 2!2! = 12$, поэтому вы легко можете их все записать и убедиться, что ответ верен.

Задача 3. 15 студентов разбиты на 3 равные по численности учебные группы. Предположим, что среди студентов — 3 вундеркинда. Какова вероятность того, что в каждой из групп будет один вундеркинд? А вероятность того, что все вундеркинды окажутся в одной группе?

Надеемся, читатель понимает, что в данной ситуации необходимо использовать модель IIIb разбиения на группы с параметрами $m = 15$, $m_1 = m_2 = m_3 = 5$. При этом общее число исходов равно

$$\frac{15!}{5!\,5!\,5!}\,.$$

Для подсчета количества разбиений, при которых в каждой группе окажется ровно один вундеркинд, прежде всего надо распределить по группам самих вундеркиндов. Это можно осуществить 3! способами ввиду IIa. Остальные студенты распределяются поровну по группам на основе модели IIIb с $m = 12$, $m_1 = m_2 = m_3 = 4$. Применяя основное правило, вычисляем искомую вероятность:

$$3!\,\frac{12!}{4!\,4!\,4!} \bigg/ \frac{15!}{5!\,5!\,5!} = \frac{6 \cdot 5^3}{15 \cdot 14 \cdot 13}\,.$$

Далее, если все вундеркинды попали в одну группу, то существует 3 возможности для выбора этой группы. Остальные рассуждения аналогичны. Таким образом, нам надо только заменить числитель в предыдущей формуле:

$$3 \cdot \frac{12!}{5!\,5!\,2!} \bigg/ \frac{15!}{5!\,5!\,5!} = \frac{5 \cdot 4 \cdot 3^2}{15 \cdot 14 \cdot 13}\,.$$

Кроме того, легко найти и вероятность оставшейся возможности, а именно, того, что вундеркинды распределятся по группам в количестве $2, 1, 0$.

Задача 4. Бросаются 6 игральных костей. Чему равна вероятность получить три пары?

Вы сразу спросите: «Какие именно три пары?» Для этого надо выбрать 3 номера из 6 чисел от 1 до 6. Количество вариантов такого выбора подсчитано в модели III: $\binom{6}{3} = 20$. Теперь мы можем сосредоточиться на выбранной тройке, скажем, $\{2, 3, 5\}$, и вычислить вероятность получения «пары двоек, пары троек и пары пятерок». Задача стала немного проще, тем самым, совет (b) оказался полезным. Одним из методов подсчета числа способов выпадения на 6 костях набора $\{2, 2, 3, 3, 5, 5\}$ основывается на модели распределения шести занумерованных жетонов (кости считаются различимыми) по трем ящикам с номерами $\boxed{2}$, $\boxed{3}$, $\boxed{5}$, при котором в каждый из ящиков попадают 2 жетона. Согласно IIIb это число равно

$$\frac{6!}{2!\,2!\,2!} = 90.$$

Другой метод заключается в воображаемой замене игральной кости на очередь из шести актеров, ожидающих за кулисами момента выхода на сцену для исполнения рутинных номеров. Для каждого из номеров 2, 3, 5 нужны два актера, но кто именно выйдет на сцену в этот раз — решает Господин Случай. Тогда задача становится перестановочной проблемой, описываемой моделью IIIa и имеющей, конечно, то же самое число исходов. Наконец, умножив это число на количество способов выбора трех номеров, найдем искомое количество вариантов:

$$\binom{6}{3}\frac{6!}{2!\,2!\,2!} = 20 \cdot 90.$$

Быть может, вы рассматриваете данное умножение как еще одно применение вездесущего основного правила, но на самом деле оно возникло благодаря сложению возможностей двадцати взаимно исключающих категорий, каждая из которых состоит из 90 исходов. Искомая вероятность равна

$$\frac{20 \cdot 90}{6^6} = \frac{25}{648}\,.$$

На примере этой задачи удобно показать, как негативный совет (c) позволяет сэкономить ваше время. Разберем один из иногда используемых студентами подходов к решению. Если мы хотим разобраться с тремя парами, две первые кости могут выпасть как угодно; номер на третьей кости должен совпадать с одним из первых двух номеров, если они различны, либо отличаться от них, если они одинаковы. Вероятность того, что первые два номера различны, равна 5/6. В этом случае вероятность выпадения нужного номера на третьей кости есть 4/6. В свою очередь, вероятность совпадения номеров первых двух костей равна 1/6. В этом случае для третьей кости вероятность имеет значение 5/6. Вам еще не надоело? Как насчет следующего шага? А еще одного?

Тем не менее, анализ последовательного типа, базирующийся на условных вероятностях, будет обсуждаться в гл. 5. Иногда он работает так же хорошо, как в следующей задаче.

Задача 5 (*дни рождения*). Чему равна вероятность того, что по крайней мере двое среди n человек имеют одинаковый день рождения?

Мы предполагаем, что дни рождения встречаются *независимо* один от другого. Другими словами, их распределение таково, будто они выбираются подобно шарам с номерами от 1 до 365 (мы игнорируем високосные годы) из урны в модели I. Таким образом, все исходы являются одинаково правдоподобными, а их полное количество равно $(365)^n$. Теперь надо подсчитать число исходов, при которых некоторые из шаров имеют одинаковые номера. Эта задача представляется довольно сложной, однако легко найти вероятность дополнительного события, что все n номеров различны. Данный вопрос предполагает применение модели II и имеет ответ $(365)_n$. Отсюда для искомой вероятности получаем представление:

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

Довольно удивительным является тот факт, что эта вероятность превосходит 1/2 уже при $n \geqslant 23$ (см. таблицу ниже)*).

n	p_n
5	0.03
10	0.12
15	0.25
20	0.41
21	0.44
22	0.48
23	0.51
24	0.54

n	p_n
25	0.57
30	0.71
35	0.81
40	0.89
45	0.94
50	0.97
55	0.99

*) Вычисления от $n = 2$ до $n = 55$ выполнены на карманном калькуляторе с точностью до 5 знаков после запятой за несколько минут.

Можно также решить данную задачу с помощью следующего простого, но при этом корректного рассуждения. Для нахождения вероятности того, что все n человек имеют разные дни рождения, упорядочим их произвольным образом и подсчитаем отдельно для каждого его возможности (аналогично тому, как мы вычисляли вероятность выпадения разных номеров на всех 6 костях в примере 2 из § 3.1). Первый может иметь день рождения в любой из дней года, для него вероятность равна 1; второй — в любой день, кроме одного, для него вероятность равна 364/365; третий — в любой день, кроме двух, следовательно, вероятность равна 363/365 и т. д. Поэтому итоговая вероятность есть произведение

$$\frac{365}{365}\,\frac{364}{365}\,\frac{363}{365}\cdots \quad (n \text{ сомножителей}),$$

что является просто другой записью для числа $(365)_n/(365)^n$. Использованная здесь интуитивная идея последовательных условных вероятностей эквивалента диаграмме из § 3.1, демонстрирующей процесс ветвления. Диаграмма начинается с 365 вариантов, каждый из которых расщепляется на 364, затем — на 363 и т. д. Если делить на 365 на каждом шаге, то придем к указанному произведению дробей.

Задача 6 (*совпадения*). Четыре карточки с номерами от 1 до 4 кладутся на стол цифрами вниз. Некий человек утверждает, что он ясновидящий и может назвать эти номера. Какова вероятность того, что будет угадана хотя бы одна цифра, если он обманщик и полагается на удачу?

Существует решение этой известной задачи в общем случае в виде явной формулы. Мы выведем ее позже в § 6.2. Но для таких малых значений, как 4, можно применить грубую силу, а заодно, по ходу дела узнать нечто новое. Предположим, что обманщик просто выбирает случайно одну из 4! одинаково вероятных перестановок. Используя совет (b), подсчитаем сначала число случаев, при которых наблюдается *в точности* одно совпадение. Тогда остальные три номера называются неправильно, поэтому необходимо найти количество «полных несовпадений» для трех карточек. Это можно сделать путем перечисления всех $3! = 6$ возможностей, приведенных в таблице:

На самом деле	(abc)	(abc)	(abc)	(abc)	(abc)	(abc)
Были названы	(abc)	(acb)	(bac)	(bca)	(cab)	(cba)

Всего есть два случая «полных несовпадений»: 4-й и 5-й в таблице. Мы укажем все исходы, в которых наблюдается ровно одно верное угадывание и три ошибочных, комбинируя 4 варианта для правильно названного номера с двумя найденными вариантами несовпадения. В силу модифицированного основного правила имеем $4 \cdot 2 = 8$ возможностей получить в точности 1 совпадение на 4 карточках. Далее, исследуем случай двух совпадений и двух ошибок. Для последнего есть только один вариант, поэтому число угадываний в точности 2 номеров на 4 карточках совпадает с числом вариантов выбора этой пары номеров: $\binom{4}{2} = 6$.

Наконец, очевидно, что если три номера названы верно, то и четвертый номер обязан совпадать. Следовательно, остается единственная возможность правильного угадывания всех четырех цифр. Объединим все результаты в таблицу:

Число совпадений	Число исходов	Вероятность
4	1	1/24
3	0	0
2	6	1/4
1	8	1/3
0	9	3/8

Последняя строка, соответствующая случаю отсутствия совпадений, получена вычитанием суммы исходов других случаев из полного числа исходов:

$$24 - (1 + 6 + 8) = 9.$$

Вероятность по крайней мере одного совпадения равна $15/24 = 5/8$, по крайней мере двух совпадений — $7/24$.

Рекомендуем читателю выполнить подсчет напрямую, выписав все 24 возможных исхода для 4 карточек, аналогично тому, как мы проделали для 3 карточек. Это полезно сделать не только для того, чтобы убедиться в правильности ответа, но также для проверки промежуточных результатов из рассуждений выше. Наш переход от 3 к 4 карточкам служит иллюстрацией эмпирического индуктивного метода и позволяет также перейти от 4 к 5 и т. д. Фактически, именно этот подход лежит в основе компьютерных расчетов. Компьютеры, в действительности, не слишком сообразительны, всегда действуют шаг за шагом, но прекрасно организованы и считают поразительно быстро. В данном случае небольшое использование алгебры предпочтительнее: позже будет установлена следующая явная формула для вычисления количества исходов, в которых встречается хотя бы одно совпадение на n карточках:

$$n!\left(1 - \frac{1}{2!} + \frac{1}{3!} - + \ldots + (-1)^{n-1}\,\frac{1}{n!}\right).$$

Задача 7. Сколько существует способов размещения n шаров по n занумерованным ящикам, при которых ровно один ящик остается пустым?

Эта задача поучительна тем, что позволяет проиллюстрировать несколько ранее отмеченных деталей. Прежде всего, в формулировке не уточнено, являются ли шары различимыми или нет. Следуя совету (d), изучим обе возможности.

Гипотеза 1. *Шары неразличимы.*

Тогда ясно, что достаточно выбрать два ящика: пустой и тот, где будут лежать 2 шара. Эта проблема выбора, относящаяся к модели II. Ответом служит $(n)_2 = n(n-1)$.

Приведенное простое решение, вероятно, выглядит лучше, чем пространные рассуждения, но мы узнаем больше, если попробуем пройти также и длинным путем.

Гипотеза 2. *Шары различимы.*

После точно такого же выбора двух ящиков, как в случае гипотезы 1 (назовем его шагом 1), надо еще указать, какой из шаров оказался в каком из ящиков. Это — задача разбиения из модели IIIb с параметрами $m = n$, $m_1 = 2$, $m_2 = \ldots = m_{n-1} = 1$, где пустой ящик не учитывается. Следовательно, ответом является

$$\frac{n!}{2!\,1!\ldots 1!} = \frac{n!}{2!}\,. \tag{3.4.2}$$

Вам не обязательно знать эту формулу ввиду того, что можно прийти к ответу иначе. Действительно, задача заключается в размещении n занумерованных шаров по $n-1$ занумерованным ящикам так, чтобы 2 шара попали в некоторый ящик (уже выбранный на шаге 1), а остальные распределились по одному в остальных ящиках. Всего есть $\binom{n}{2}$ способов выбора двух шаров, помещаемых в фиксированный ящик, после чего оставшиеся $n-2$ шара можно распределить по $n-2$ ящикам $(n-2)!$ способами. Произведение этих двух чисел равно величине (3.4.2). Итак, полное количество исходов, удовлетворяющих гипотезе 2, задается выражением

$$n(n-1)\cdot\frac{n!}{2}\,. \tag{3.4.3}$$

Наши рассуждения состояли из двух шагов. Кто-то, возможно, предпочтет следующее трехшаговое решение. Сначала выберем пустой ящик. Затем выберем $n-1$ шаров из n и положим их по одному в каждый из $n-1$ других ящиков. Наконец, поместим последний шар в любой из этих $n-1$ ящиков. Количества возможных вариантов для каждого из шагов равны соответственно n, $(n)_{n-1}$ (выбор в модели II) и $n-1$.

Если их перемножить, то в результате получим ответ

$$n \cdot n!\,(n-1), \tag{3.4.4}$$

который в два раза больше, чем (3.4.3). Какой из ответов верен?

Это как раз та ситуация, когда может пригодиться совет (a). Возьмем $n = 3$ и предположим, что пустой ящик уже выбран. Тогда задача состоит в размещении шаров с номерами 1, 2, 3 в ящиках A и B. Для наглядности пусть A будет квадратным, а B — круглым. Выберем 2 шара из трех и поместим их в наши ящики. Вот 6 возможных размещений:

$$\boxed{1}\,2 \qquad \boxed{2}\,1 \qquad \boxed{1}\,3 \qquad \boxed{3}\,1 \qquad \boxed{2}\,3 \qquad \boxed{3}\,2$$

Затем бросим оставшийся шар в один из двух ящиков. Тем самым, каждый из только что перечисленных случаев распадается на два новых в зависимости от того, где оказался третий шар:

$$\begin{array}{cccccc}
\boxed{13}\,2 & \boxed{23}\,1 & \boxed{12}\,3 & \boxed{32}\,1 & \boxed{21}\,3 & \boxed{31}\,2 \\
\boxed{1}\,23 & \boxed{2}\,13 & \boxed{1}\,32 & \boxed{3}\,12 & \boxed{2}\,31 & \boxed{3}\,21
\end{array}$$

Теперь вы видите, в чем дело: каждое итоговое размещение встречается дважды, так как в ящик с двумя шарами может сначала попасть шар с меньшим номером, а может — с большим номером. То же самое происходит и в общем случае. Поэтому для исключения двойного учета необходимо разделить число (3.4.4) на 2. В результате получим тот же ответ, что и в записи (3.4.3). Гармония восстановлена.

Задачи

(Для вычисления вероятностей в некоторых задачах важно понимать, какие из исходов являются одинаково правдоподобными. Это обычно ясно из контекста. Если возникают сомнения, следуйте совету (d).)

1. Юноша решает, что ему надеть на день рождения: пиджак или жилет? У него есть три жилета и два пиджака. Сколько существует вариантов выбора, если он собирается воспользоваться только одной из этих вещей? А если он захочет надеть и жилет, и пиджак?

2. В продаже имеются рубашки трех видов. (a) Сколько существует возможностей купить каждому из двух мужчин по одной рубашке? (b) Сколько существует вариантов продажи двух рубашек?

3. В условиях задачи 2 для 3 рубашек и 2 мужчин задайте соответствующие вопросы, ответами на которые являются числа 2^3 и $\binom{3+2-1}{3}$.

4. Представим, что в меню из § 3.1 указаны три вида мороженного и два вида пирога. Сколько существует разных вариантов комплексного обеда?

А если принять в расчет, что посетитель может не заказывать овощи или десерт, или и то и другое вместе?

5. Найдите количество разных инициалов, составленных из двух или трех букв алфавита. Насколько велик должен быть алфавит, чтобы 1 миллион человек можно было бы идентифицировать с помощью трехбуквенных инициалов?

6. Подсчитайте количество целых чисел между 1 миллионом и 10 миллионами, в десятичном представлении которых не встречается одинаковых рядом стоящих цифр.

7. Тест состоит из 12 вопросов, предполагающих ответы «да» или «нет». Студентка решила выбрать случайно 6 из них для проверки. Сколько у нее возможностей выбора?

8. Сколькими способами 4 мальчика и 4 девочки могут разбиться на пары? Каково число вариантов построения их в ряд в порядке чередования пола?

9. Сколько существует возможностей образования из 20 человек комитета, состоящего из трех представителей? Сколькими способами могут быть назначены председатель, секретарь и казначей?

10. Допустим вы располагаете 2 долларами, 2 монетами достоинством в 25 центов и 3 монетами по 5 центов. Сколько разных цен способны вы заплатить без сдачи? Замените монеты в четверть доллара на десятицентовики и снова ответьте на вопрос.

11. Из машины, в которой используются шурупы трех разных размеров, пропали два шурупа. На замену высланы 3 шурупа разных размеров. Чему равна вероятность того, что они такие, какие нужны?

12. Дверь заперта на два замка, ключи от которых лежат у вас в кармане вместе с 4 другими. В спешке вы потеряли где-то один из ключей. Какова вероятность того, что, несмотря на это, вы сможете открыть дверь? Чему равна вероятность того, что подойдут первые два ключа, которые вы попробуете?

13. Игральная кость бросается трижды. Какова вероятность того, что с каждым разом вы получаете все больший номер? (Эта простая задача была включена одним из авторов в экзаменационный тест, но по недосмотру в ней использовалась формулировка «…что выпадающие номера увеличиваются постоянно». Подумайте о возможной неправильной интерпретации этих слов.)

14*. Три игральные кости бросаются дважды. Найдите вероятность того, что все их выпавшие грани будут одинаковыми, если кости (a) различимы, (b) неразличимы. [*Указание.* Разбейте исходы в зависимости от результата первого бросания: выпала пара, тройка или все номера различны; затем сравните их, соответственно, с результатом второго бросания.]

15. Вы собрались на вечеринку, где никого не знаете. Ожидается, что в ней примут участие шесть женщин и четверо мужчин, причем известно, что среди них есть четыре семейные пары. Сколько существует вариантов предполагаемого разбиения на эти пары? Дайте ответ в случае трех пар.

16. Четыре ботинка выбраны случайно из пяти разных пар. Какова вероятность того, что среди них есть хотя бы одна пара?

17. Водитель из Калифорнии решил, что он будет менять ряд движения каждую минуту. Представим, что он едет по четырехполосному шоссе и перестраивается случайно. Чему равна вероятность возвращения на первоначальный ряд в течение 4 минут (в предположении отсутствия аварии)? [*Указание.* Ответ зависит от того, с какого из рядов начинается процесс: внутреннего или внешнего.]

18. Вычислите вероятность того, что определенный шар ни разу не появится при n извлечениях в условиях моделей I или II из § 3.2. Предполагается, что $n < m$.

19. Вам сообщили, что из четырех карт, лежащих на столе рубашкой вверх, две принадлежат красной масти и две — черной. Если вы называете цвет наудачу, то с какой вероятностью дадите 0, 2, 4 верных ответа?

20. Автобус в аэропорту должен высадить 15 пассажиров на 4 остановках. Какова вероятность того, что все пассажиры выйдут на одной остановке? Какова вероятность того, что хоть кто-нибудь (по крайней мере один человек) будет выходить на каждой из остановок?

21. Десять книг раскладываются в две стопки. Сколько существует способов раскладки, если и книги, и стопки могут быть как различимыми, так и неразличимыми. Рассмотрите все четыре случая с условием, что стопки не являются пустыми.

22. Десять разных книг родители намеревались распределить между детьми: Даниэлем, Филиппом, Полом и Джоном, в количествах 3, 3, 2, 2 соответственно. Сколькими способами это можно сделать? Так как Пол и Джон закричали, что «это нечестно», то было решено, что надо бросить жребий для определения тех двоих, кто получит по 3 книги. Сколько теперь существует вариантов распределения? В конце концов, Марильда и Корина также захотели участвовать в розыгрыше. Поэтому решили, что все шесть детей будут тянуть жребий для определения двоих, кто получит по 3 книги, двоих, кому достанутся 2 книги, и двоих, кому ничего не достанется. Сколько тогда имеется вариантов? (Затруднительно сформулировать данную задачу в общем случае. В подобной ситуации лучше быть многословным, чем лаконичным. Попробуйте раскладывать жетоны по ящикам.)

23. В чековой лотерее разыгрываются 366 дней года (включая 29 февраля). Какова вероятность того, что первые «вытащенные» (без возвращения) 180 дней распределятся поровну между 12 месяцами? Чему равна вероятность того, что среди первых 30 дней не окажется ни одного дня из августа или сентября? [*Указание.* Выберите сначала по 15 дней из каждого месяца.]

24. Туристическое агентство выяснило, что туристы распределяются по 20 отелям некоторого курорта подобно неразличимым жетонам, помещаемым в занумерованные ящики. В предположении, что эта теория верна, вычислите вероятность того, что прибывшие первыми 30 туристов разместятся так, что ни один из отелей не окажется пустым. (Такая модель в физике известна под названием «статистика Бóзе—Эйнштейна».) Если считать туристов различимыми объектами, то получаем классическую «статистику Максвелла—Больцмана» (см. [9, § II.5]).

25. Сто форелей были выловлены в небольшом озере, помечены и выпущены обратно. Позже были пойманы еще 100 форелей, среди которых оказались 7 помеченных рыб. Найдите вероятность того, что в озере живут n форелей? (Какое значение n наиболее вероятно? Данный вопрос типичен для *математической статистики*.)

26. Установите взаимно однозначное соответствие между исходами в моделях IIIa и IIIb для параметров $m = 4$, $m_1 = 2$, $m_2 = m_3 = 1$.

27*. (Исключительно для игроков в покер.) Предположим, что все игроки одинаково удачливы (т. е. применима модель III). Вычислите вероятность иметь на руках (a) флэш, (b) стрейт, (c) стрейт флэш, (d) четыре одного типа, (b) фул хаус[*)].

28*. Докажите, что

$$\binom{2n}{n} = \sum_{k=0}^{\infty} \binom{n}{k}^2 .$$

[*Указание.* Примените формулу (3.3.9).]

29*. Количество способов, которыми натуральное число n записывается в виде суммы (не превосходящих его) натуральных слагаемых в комбинаторике называется *числом разбиений* n. Например,

$6 = 6$	шестерка
$= 5 + 1$	пятерка
$= 4 + 2$	четверка и пара
$= 4 + 1 + 1$	четверка
$= 3 + 3$	две тройки
$= 3 + 2 + 1$	тройка и пара
$= 3 + 1 + 1 + 1$	тройка
$= 2 + 2 + 2$	три пары
$= 2 + 2 + 1 + 1$	две пары
$= 2 + 1 + 1 + 1 + 1$	одна пара
$= 1 + 1 + 1 + 1 + 1 + 1$	все различны

[*)] Флэш (flush) — любые 5 карт одной масти, стрейт (straight) — любые 5 последовательных карт, стрейт флэш (straight flush) — 5 любых последовательных карт одной масти, четыре одного типа (four of a kind) — 4 однотипные карты (например, 4 короля), фул хаус (full house) — 3 карты одного типа и пара. — *Прим. перев.*

Таким образом, число разбиений 6 равно 11 (сравните его с числами 46656 и 462 из примеров 2 и 3). Его также можно считать количеством отличающихся «шаблонов совпадений» при бросании шести игральных костей. В приведенном выше списке этим шаблонам были присвоены простые (однако, не сразу понятные) имена. Вычислите соответствующие вероятности. (Представляется удивительным, что вероятность шаблона «две пары» больше, чем у шаблона «одна пара», и превосходит 1/3. Сомнения одного из авторов в правильности вычислений развеялись только после того, как он бросил 6 костей 100 раз. В Китае сохранился старинный обычай играть в кости во время новогодних каникул. Насколько автор помнит, «две пары» имеют более высокий ранг (цену), чем «одна пара». Это несправедливо с точки зрения вероятностей выпадения. Через некоторое время после экспериментирования автор обнаружил, что Феллер*) вычислял аналогичные вероятности для семи костей. То, что он взял в качестве примера число 7, тем самым проигнорировав существование старинной игры, возможно, послужило причиной того, что автор раньше не обратил внимание на эти его вычисления.)

30.* (*Задача Банаха о спичках.*) Польский математик Банах носил с собой два коробка спичек в разных карманах. В каждом из коробков было n спичек. Когда Банаху требовалась спичка, он доставал случайно один из коробков. В некоторый момент вынутый коробок оказывался пустым. Какое распределение имеет число спичек, оставшихся в другом коробке? [*Указание.* Рассмотрите два случая в зависимости от того, левый или правый коробок оказался пустым. Будьте внимательны при учете возможности, когда оба коробка пусты.]

*) Уильям Феллер (William Feller) (1906–1970) — известный специалист в области теории вероятностей.

Глава 4

Случайные величины

4.1. Что такое случайная величина?

Мы уже видели, что точки выборочного пространства могут представлять конкретные объекты, такие как яблоки, молекулы, люди. Раз так, то они обладают разнообразными свойствами, некоторые из которых можно измерить. У яблока есть вес и объем; можно определить химический состав его сока; даже вкус яблока может быть отнесен экспертами к одной из категорий. Молекула имеет массу и скорость, через которые выражаются с помощью физических формул ее импульс и кинетическая энергия. В свою очередь, человек обладает такими физиологическими характеристиками, как возраст, рост, вес. Кроме того, у него много других качеств, которые поддаются количественному измерению: IQ[*)], количество лет обучения, число братьев и сестер, ежегодный доход и размер уплачиваемых налогов и т. п. Мы рассмотрим еще несколько аналогичных примеров, а затем дадим математическое определение случайной величины в общих терминах.

Пример 1. Пусть Ω обозначает человеческую популяцию, состоящую из n человек. Это можно записать так:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}. \tag{4.1.1}$$

Предположим, нас интересует распределение людей по возрасту. Обозначим через $A(\omega)$ возраст ω. Таким образом, каждой точке ω приписано число $A(\omega)$, выраженное в некоторых единицах измерения, скажем, в годах. Другими словами, отображение

$$\omega \to A(\omega)$$

является *функцией* с областью определения Ω. В качестве значений функции выступают целые числа. Однако возраст можно измерять и более точно, записывая результат в виде дробей, десятичного представления или давая словесное описание, например, «18 лет 5 месяцев и 1 день». Не будет никакого вреда, если мы станем считать областью значений функции A все множество целых или действительных положительных чисел, несмотря на то, что нам потребуется только очень малая

[*)] Коэффициент интеллекта. — *Прим. перев.*

его часть. Соответственно, будем говорить, что A является целочисленной или действительнозначной функцией. Подобно возрасту, представим рост, вес и доход в виде следующих функций:

$$\omega \to H(\omega),$$
$$\omega \to W(\omega),$$
$$\omega \to I(\omega).$$

Заметим, что функция I может принимать отрицательные значения! Для некоторых медицинских исследований представляет интерес такая характеристика, как линейная комбинация роста и веса

$$\omega \to \lambda H(\omega) + \mu W(\omega),$$

где λ и μ — действительные числа. Она также является функцией от ω. Аналогично, если ω представляет «главу семьи», иначе — кормильца, то чиновники статистического бюро, проводящие перепись населения, могут захотеть вычислить функцию

$$\omega \to \frac{I(\omega)}{N(\omega)},$$

где $N(\omega)$ обозначает количество членов семьи, другими словами, количество едоков. Данное отношение называется «доходом, приходящимся на каждого члена семьи».

Давайте введем удобный символизм для обозначения разнообразных подмножеств выборочного пространства, порождаемых случайными величинами. Например, множество тех ω из Ω, чей возраст заключен между 20 и 40 годами, будем обозначать так:

$$\{\omega \mid 20 \leqslant A(\omega) \leqslant 40\}$$

или более кратко, когда нет опасности неправильного понимания,

$$\{20 \leqslant A \leqslant 40\}.$$

Множество людей, имеющих рост от 65 до 75 дюймов и вес от 120 до 180 фунтов, может быть записано в виде следующих вариантов:

$$\begin{aligned}\{\omega \mid 65 \leqslant H(\omega) \leqslant 75\} \cap \{\omega \mid 120 \leqslant W(\omega) \leqslant 180\} &= \\ = \{\omega \mid 65 \leqslant H(\omega) \leqslant 75; 120 \leqslant W(\omega) \leqslant 180\} &= \\ = \{65 \leqslant H \leqslant 75; 120 \leqslant W \leqslant 180\}.&\end{aligned}$$

Пример 2. Пусть Ω состоит из молекул газа в некотором сосуде. Мы по-прежнему можем использовать для представления Ω форму (4.1.1),

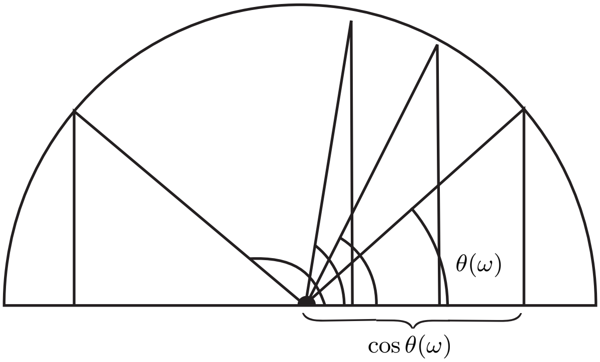


Рис. 17

несмотря на то, что теперь n — очень большое число порядка 10^{25}. Введем обозначения: m для массы, v для скорости, M для импульса, E для кинетической энергии. Тогда имеем, соответственно, функции

$$\begin{aligned}
&\omega \to m(\omega),\\
&\omega \to v(\omega),\\
&\omega \to M(\omega) = m(\omega)v(\omega),\\
&\omega \to E(\omega) = \frac{1}{2}\, m(\omega)v(\omega)^2.
\end{aligned}$$

В экспериментах с газом в действительности измеряются m и v, а интересовать исследователя могут M и E, которые вычисляются с помощью указанных формул. Аналогично, если обозначить через θ угол между вектором скорости молекулы и осью абсцисс, то $\omega \to \theta(\omega)$ является функцией от ω, а

$$\omega \to \cos\theta(\omega)$$

представляет собой суперпозицию функции «cos» с функцией «θ». Множество молекул, движущихся направо, определяется записью

$$\{\omega \mid \cos\theta(\omega) > 0\}.$$

Пример 3. Пусть Ω обозначает пространство исходов двух бросаний игральной кости. Тогда оно состоит из следующих $6^2 = 36$ точек:

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

Таким образом, каждая точка ω задается упорядоченной парой чисел или двумерным *вектором*:

$$\omega_k = (x_k, y_k), \quad k = 1, 2, \ldots, 36,$$

где x_k и y_k принимают значения от 1 до 6. Первая *координата* x отражает результат первого бросания, вторая координата y — второго бросания. Обе координаты определяются точкой ω. Другими словами, они являются *функциями* от ω:

$$\omega \to x(\omega), \quad \omega \to y(\omega). \tag{4.1.2}$$

В свою очередь, каждая точка ω однозначно задается парой своих координат, поэтому можно считать, что ω и есть эта самая пара:

$$\omega \equiv (x(\omega), y(\omega)).$$

Важно усвоить отмеченное только что изменение точки зрения. Например, пусть последовательные результаты n бросаний кости обозначаются через

$$x_1(\omega), x_2(\omega), \ldots, x_n(\omega);$$

тогда не только каждая из величин $x_k(\omega)$, $k = 1, \ldots, n$, называемая *k-й координатой*, является функцией от ω, но и сам набор из всех n координат, в свою очередь, однозначно определяет ω. Другими словами, ω есть не что иное, как n-мерный вектор

$$\omega \equiv (x_1(\omega), x_2(\omega), \ldots, x_n(\omega)). \tag{4.1.3}$$

В общем случае каждая $x_k(\omega)$ представляет собой определенную числовую характеристику исхода ω. Поскольку ω может обладать самыми разными свойствами, в большинстве случае в расчет принимаются только некоторые из них. При этом вполне годится форма представления (4.1.1). Например, на типичном конкурсе красоты, учитываются только три измерения фигуры, скажем, $(36, 29, 38)$ в дюймах. В таких конкурсах (без «пения и танцев») каждая претендентка характеризуется всего лишь тройкой чисел:

$$\text{претендентка} = (x, y, z).$$

Другим примером такого же типа может служить тестирование. Предположим, что студент должен выполнить 4 тестовых задания, каждое из которых оценивается в процентах. Обозначим самого студента через ω, а его результаты в четырех тестах — через $x_1(\omega)$, $x_2(\omega)$, $x_3(\omega)$,

$x_4(\omega)$. Для проверяющего (или для компьютера, если задания обрабатываются с его помощью) каждый студент характеризуется исключительно 4-мерным вектором $(x_1(\omega), x_2(\omega), x_3(\omega), x_4(\omega))$. Студенты, имеющие одинаковые результаты, не различаются. Допустим, что критерием успешного выполнения теста является условие, что сумма четырех оценок превышает 200 процентов. Тогда студенты, прошедшие тестирование, образуют множество

$$\{\omega \mid x_1(\omega) + x_2(\omega) + x_3(\omega) + x_4(\omega) > 200\}.$$

Другой вариант критерия получается, когда отдельным заданиям назначаются разные веса, скажем, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. При этом критерий зависит от линейной комбинации $\lambda_1 x_1(\omega) + \ldots + \lambda_4 x_4(\omega)$. Еще один возможный вариант критерия определяется множеством

$$\{\omega \mid \min(x_1(\omega), x_2(\omega), x_3(\omega), x_4(\omega)) > 35\}.$$

Что это означает на обычном языке?

4.2. Как образуются случайные величины?

Теперь мы готовы дать общее определение для числовых характеристик точек выборочного пространства Ω в предположении, что оно является *счетным множеством.* Это требование, как вы увидите, приводит к существенному упрощению. Несчетные выборочные пространства будут рассмотрены позже.

Определение случайной величины. Определенная на пространстве Ω и принимающая числовые значения функция X от ω

$$\omega \in \Omega: \quad \omega \to X(\omega) \tag{4.2.1}$$

называется случайной величиной (на Ω).

Термин «случайная величина» является общепринятым, поэтому мы будем использовать его в этой книге. Однако термины «стохастическая переменная» или «размер шансов» также считаются вполне пригодными. Прилагательное «случайный» напоминает, что мы имеем дело с выборочным пространством и пытаемся описывать явления, обычно называемые случайными или зависящими от случая событиями. Важно отметить, что вся случайность функции $X(\omega)$ заключается в выборе наудачу аргумента ω, как при бросании игральной кости или случайном отборе индивидуума из популяции. Коль скоро аргумент ω выбран, значение функции $X(\omega)$ немедленно становится полностью определенным, всякая случайность исчезает. Например, после того, как яблоко ω достали из мешка, его можно взвесить, поэтому вес $W(\omega)$ можно считать известным. Заодно отметим, что термин «величина» здесь понимается в широком смысле как «зависимая величина», т. е. функция от ω (см. § 4.1). Можно сказать, что точка ω выборочного пространства играет роль «независимой величины», подобно роли x в $\sin x$. Однако лучше не использовать этот язык ввиду того, что понятие независимости имеет в теории вероятностей совсем другое, причем очень важное, значение (см. § 5.5).

Наконец укажем, что традицией (не всегда соблюдаемой) является использование заглавных букв (X, Y, N или S) для обозначения случайных величин. Тем не менее, нет веских причин для отказа от использования строчных букв x и y, как мы делали в примерах из § 4.1.

Обратите внимание, что случайные величины могут быть определены на выборочном пространстве без всякого упоминания о вероятности. Позже мы увидим, что, задавая на пространстве вероятностную меру, мы тем самым порождаем вероятностные распределения случайных величин.

Имея в запасе несколько случайных величин, можно получать другие величины, производя разные операции над имеющимися. Некоторые примеры операций уже были приведены в § 4.1. Более общее утверждение имеет следующий вид.

Утверждение 1. *Если X и Y — случайные величины, то таковыми являются и*

$$X + Y, \quad X - Y, \quad XY, \quad X/Y \ (Y \neq 0), \tag{4.2.2}$$

а также $aX + bY$, где a и b — действительные числа.

Это утверждение немедленно вытекает из общего определения, так как, например,

$$\omega \to X(\omega) + Y(\omega)$$

есть функция на Ω в силу того, что она является суммой функций X и Y. Ситуация здесь совершенно такая же, как в математическом анализе: если f и g — функции, то функциями являются и

$$f + g, \quad f - g, \quad fg, \quad \frac{f}{g} \ (g \neq 0), \quad af + bg.$$

Единственное отличие заключается в том, что в математическом анализе их аргументом служит действительное число x, а в записи (4.2.2) они суть функции от ω — точки выборочного пространства. Так же, как и в анализе, где константа представляет собой весьма частный случай функции, теперь константа — весьма частный вариант случайной величины. Например, вполне возможно, что все ученики некоторого класса имеют один и тот же возраст. Тогда случайная величина $A(\omega)$, определенная в примере 1 из § 4.1, равна константе, скажем, 9 (годам) для четвертого класса.

В анализе *функция от функции* снова является функцией, например, $\ln(\sin x)$ или $f(\varphi(x)) = (f \circ g)(x)$. Функция от случайной величины также является случайной величиной, подобно $\cos\theta$ в примере 2 из § 4.1. Можно обобщить это утверждение на случай нескольких случайных величин.

Утверждение 2. *Если φ есть функция двух переменых, а X и Y — случайные величины, то отображение*

$$\omega \to \varphi(X(\omega), Y(\omega)) \tag{4.2.3}$$

также является случайной величиной, которая более кратко записывается в виде $\varphi(X, Y)$.

Хорошим примером служит функция $\varphi(x, y) = \sqrt{x^2 + y^2}$. Пусть $X(\omega)$ и $Y(\omega)$ обозначают, соответственно, длины горизонтальной и вертикальной составляющих вектора скорости молекулы газа. Тогда

$$\varphi(X, Y) = \sqrt{X^2 + Y^2}$$

выражает ее *абсолютную скорость*.

По ходу дела заметим, что утверждение 2 содержит в себе утверждение 1 в качестве частного случая. Например, если взять $\varphi(x, y) = x + y$, то $\varphi(X, Y) = X + Y$. Оно также включает в качестве частного случая произвольные функции от одной случайной величины вида $f(X)$. Как вы думаете, почему? Наконец, утверждение 2, очевидно, обобщается на функции от произвольного числа переменных. Особенно важным случаем является сумма n случайных величин:

$$S_n(\omega) = X_1(\omega) + \ldots + X_n(\omega) = \sum_{k=1}^{n} X_k(\omega). \tag{4.2.4}$$

Например, если $X_1, \ldots, X_n$ обозначают последовательные результаты бросания кости, то S_n — общая сумма результатов всех n бросаний. Мы не раз будем иметь дело с такими *частными суммами* S_n.

Проиллюстрируем использование случайных величин в некоторых повседневных ситуациях. Нередко интуитивное представление о некой случайной характеристике предшествует формированию представления о выборочном пространстве. Действительно, часто можно говорить о случайных величинах X, Y и т. д., не задумываясь над тем, как устроено пространство Ω. Довольно формальная (и трудно воспринимаемая?)

математическая структура служит необходимой логической опорой, но ее не стоит извлекать на свет в каждом из случаев применения вероятностного языка.

Пример 4. Себестоимость издания некоторой книги составляет \$3 за экземпляр при тираже не более 1 000, затем \$2 за каждую дополнительную копию с номером от 1001 до 5 000 и \$1 за каждую последующую копию. На практике, конечно, допечатка осуществляется не поэкземплярно, а определенными крупными партиями. Будем считать, что вся допечатка обязательно распродается, и издательство не несет убытков от неудовлетворенного спроса. Предположим, что первоначально издано 1 000 экземпляров по цене \$5 за книгу. Случайным в данной ситуации является количество книг, которые удастся продать. Обозначим его через X. Очевидно, что как только значение X станет известным, можно подсчитать размер доходов или потерь от продажи, назовем его Y. Таким образом, Y является функцией от X, которая случайна исключительно благодаря случайности X. Приведем формулу, связывающую Y и X (см. рис. 18):

$$Y = \begin{cases} 5X - 3\,000, & \text{если } X \leqslant 1\,000, \\ 2\,000 + 3(X - 1\,000), & \text{если } 1\,000 < X \leqslant 5\,000, \\ 14\,000 + 4(X - 5\,000), & \text{если } X > 5\,000. \end{cases}$$

Какова вероятность того, что данная книга окажется убыточной? Она равна вероятности события, представленного множеством

$$\{5X - 3\,000 < 0\} = \{X < 600\}.$$

Какова вероятность того, что доход превзойдет \$10 000? Она равна вероятности события

$$\begin{aligned} \{2\,000 + 3(X - 1\,000) \geqslant 10\,000\} \cup \{X > 5\,000\} &= \\ &= \left\{X \geqslant \frac{8\,000}{3} + 1\,000\right\} \cup \{X > 5\,000\} = \\ &= \{X \geqslant 3667\}. \end{aligned}$$

Но какие значения имеют эти вероятности? Они зависят от поведения случайной величины X. Можно только попытаться предугадать эту величину, ведь она случайна. Ее точное значение станет известным лишь тогда, когда торговля будет закончена, подобно тому, что мы должны бросить кость, чтобы узнать результат. Вероятности разных значений называются распределением случайной величины X. Подробнее это понятие будет обсуждаться в § 4.3.

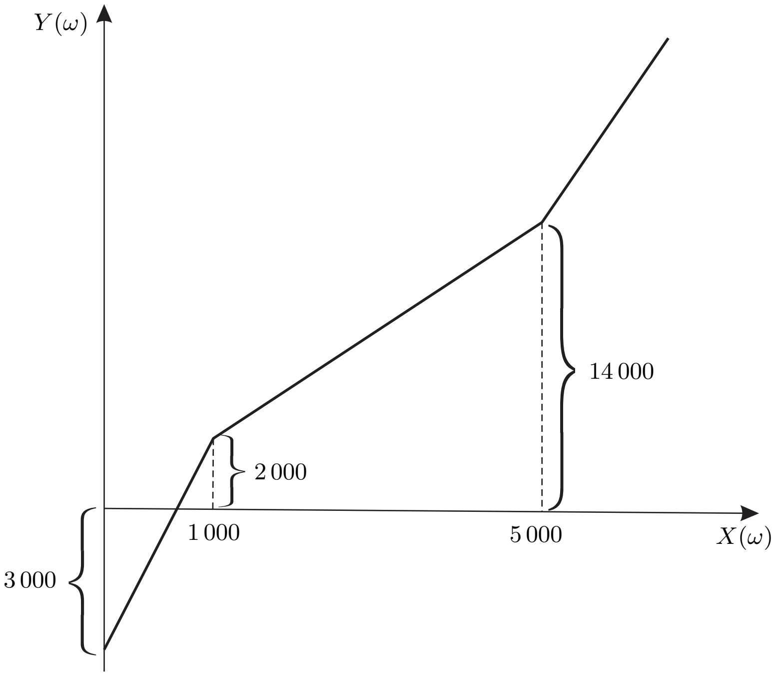


Рис. 18

Что здесь представляет собой выборочное пространство? Так как основным интересующим нас объектом является величина X, мы вполне можем рассматривать ее саму в качестве ω вместо того, чтобы подгонять обозначения к привычному виду. Тогда каждая точка ω есть какое-то натуральное число, при этом $\omega \to Y(\omega)$ — случайная величина на множестве Ω натуральных чисел. Выбрать наудачу ω означает в нашем случае сделать предположение (или выдвинуть гипотезу) относительно количества продаж, на базе которого можно вычислить доход по формуле, приведенной выше. Нет ничего плохого в данной модели выборочного пространства, хотя она кажется слегка надуманной.

Более поучительная точка зрения заключается в ассоциировании каждой точки ω с записью в книге учета продаж. Подчас издателя интересует не только информация об общем числе продаж. Важным фактором, до сих пор не принятым к рассмотрению, является время реализации тиража. Понятно, что есть разница между продажей 5 000 книг за 1 год или за 10 лет. Если книга — университетский учебник, как эта, то может оказаться важной информация о том, как она реализуется в разных учебных заведениях или разных регионах. Напротив, для беллетристики или пьесы весьма значимым (даже из коммерческих соображений) является мнение критики. По этой причине, скорее всего, будут фиксироваться записи, содержащие сведения о рекламных успехах, а не о продажах. Все подобные характеристики можно упаковать в капсулу, называемую точкой выборочного пространства ω. Вы можете представлять ее себе как *исчерпывающую запись* всех битов информации, имеющей отношение к книге, где $X(\omega)$ и $Y(\omega)$ — всего лишь две грани. Тогда, что же такое Ω? Это — совокупность всех таких мыслимых записей. Данное понимание выборочного пространства выглядит странно и громоздко (можно ли утверждать, что Ω счетно?), но оно дает соответствующую картину, когда речь идет, скажем, о пути частицы при броуновском движении или о траектории случайного процесса (см. гл. 8). С другой стороны, оно также демонстрирует целесообразность работы с некоторыми отдельными случайными величинами вместо того, чтобы беспокоиться обо всей вселенной.

Пример 5. Страховая компания время от времени получает заявки, содержащие требования о возмещении ущерба. Как моменты поступления требований, так и суммы страховых выплат заранее не известны и определяются случаем. При этом, конечно, общее количество заявок, скажем, за год, также случайно. Понятно, что оно будет определено, как только мы узнаем ответы на вопросы «когда» и «сколько» обо всех требованиях. Занумеруем требования в порядке поступления. Пусть S_n — дата поступления n-й заявки. Например, $S_3 = 33$ означает, что третье требование пришло 2-го февраля. Итак, мы имеем:

$$1 \leqslant S_1 \leqslant S_2 \leqslant \ldots \leqslant S_n \leqslant \ldots,$$

и равенства встречаются, когда в течение одного дня поступают несколько требований. Обозначим через C_n размер выплаты (в долларах), выставленной в n-й заявке. Сколько требований поступит за год? Ответ N задается следующей формулой:

$$N = \max\{n \mid S_n \leqslant 365\}.$$

Очевидно, величина N случайна. Однако она полностью определяется последовательностью из S_n. Теоретически нам надо знать всю последовательность, поскольку значение N может оказаться сколь угодно большим. Зная N и последовательность из C_n, мы сумеем найти общую сумму всех выплат за год

$$C_1 + \ldots + C_N. \tag{4.2.5}$$

Эта запись похожа на формулу (4.2.4), но в сумме (4.2.5) случайными являются не только сами слагаемые, но и их количество. Конечно, эта сумма также случайна. Она зависит как от S_n, так и от C_n.

В данном случае легко представить, что требования поступают в офис одно за другим и регистрируются в виде записей в книге учета, как в дневнике. В какие-то из дней совсем не будет поступлений, в другие — сразу несколько требований на разные суммы выплат. Такая книга обычно хранится годами и может выглядеть по-разному в различные временные периоды. Другая страховая компания ведет свою книгу учета, которая в чем-то похожа, но и в чем-то отличается от первой. Каждый мыслимый список требований в таких книгах может рассматриваться в качестве ω, и достаточно большое число таких списков могут служить выборочным пространством. К примеру, списки, содержащие миллион заявок за один день или выплаты в размере 95 центов, следует исключить из рассмотрения. Это позволит сохранить наше представление о выборочном пространстве в рамках реальности.

Если принять такую точку зрения, то в голову сразу приходят и другие случайные величины. Например, обозначим через Y_k суммарный размер требуемых выплат в заявках, полученных в течение k-го дня. Возможно, что $Y_k = 0$. Тогда полная сумма поступивших счетов за первые n дней года равна

$$Z_n = \sum_{k=1}^{n} Y_k = Y_1 + Y_2 + \ldots + Y_n. \tag{4.2.6}$$

Требуемые выплаты за период $[s, t]$ представляются в виде:

$$Z_t - Z_{s-1} = \sum_{k=s}^{t} Y_k = Y_s + Y_{s+1} + \ldots + Y_t. \tag{4.2.7}$$

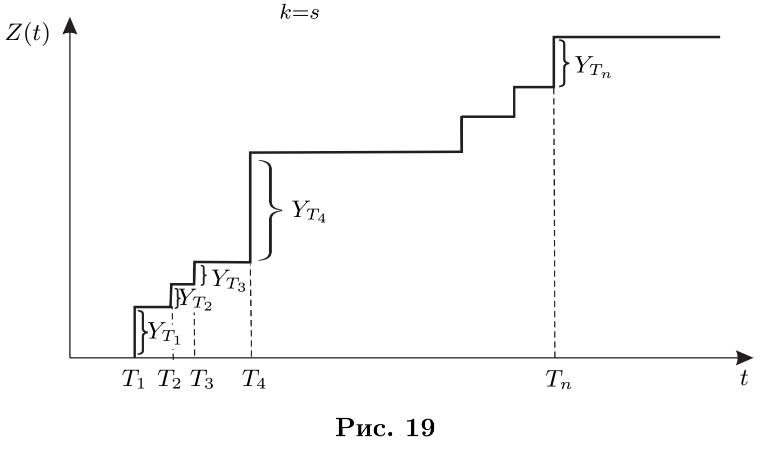


Рис. 19

Можно нарисовать график, подобный изображенному на рис. 19, для *накопленного* за время t размера требуемых выплат Z_t.

График имеет скачки при тех t, когда имелись заявки в течение t-го дня, причем величина скачка равна сумме требуемых выплат по всем этим заявкам. Другими словами, последовательные скачки соответствуют тем Y_k, которые больше нуля. Понятно, что можно считать с такого графика информацию о полной сумме поступивших счетов за любой интересующий период, а также установить длины интервалов, свободных от заявок. Однако нельзя узнать размеры требуемых выплат, содержащихся в отдельных заявках, поступивших в течение одного дня. Если вся нужная вам информация может быть извлечена из такого графика, то каждый мыслимый график можно рассматривать как точку выборочного пространства. Это приводит к более узкому пространству по сравнению с описанным выше, однако для наших целей его достаточно. С математической точки зрения идентификация точки выборочного пространства с графиком (также называемым *выборочной кривой, путем* или *траекторией*) очень удобна ввиду того, что кривая — более точный (и привычный!) объект, чем книга бухгалтерского учета или какая-нибудь запись о продажах или рекламных успехах.

4.3. Распределение и математическое ожидание

В главе 2 обсуждались вероятности, приписываемые множествам точек выборочного пространства. Обычно сами множества определяются с помощью случайных величин. Типичным примером служит запись

$$\{a \leqslant X \leqslant b\} = \{\omega \mid a \leqslant X(\omega) \leqslant b\}, \tag{4.3.1}$$

где X — случайная величина, a и b — константы. Частные случаи такого способа задания множеств встречались в примерах из § 4.1. Так как каждому подмножеству счетного выборочного пространства Ω приписана некоторая вероятность, множество (4.3.1) тоже имеет вероятность, обозначаемую

$$P(a \leqslant X \leqslant b). \tag{4.3.2}$$

Немного обобщим: пусть A является некоторым подмножеством множества действительных чисел (или точек прямой $\mathbb{R}^1 = (-\infty, \infty)$). Тогда можно записать:

$$P(X \in A) = P(\{\omega \mid X(\omega) \in A\}). \tag{4.3.3}$$

Например, когда A есть замкнутый интервал $[a, b]$, то приходим к вероятности (4.3.2). Но A также может быть открытым интервалом (a, b), полуоткрытым интервалом $(a, b]$ или $[a, b)$, бесконечным интервалом $(-\infty, b)$ или $(a, +\infty)$, объединением нескольких интервалов или набором целых чисел, скажем, $\{m, m+1, \ldots, m+n\}$. Важным частным случаем является множество A, состоящее из одной точки x. Оно называется *атомом* $\{x\}$.

Различие между точкой x и множеством $\{x\}$ может показаться чисто теоретическим. В любом случае, величина

$$P(X = x) = P(X \in \{x\}) \tag{4.3.4}$$

имеет смысл «вероятности того, что X имеет (или принимает) значение x». Если X — возраст людей в некоторой популяции, $\{X = 18\}$ представляет собой субпопуляцию, состоящую из 18-летних — очень важное множество!

Далее существенную роль для упрощения будет играть гипотеза о том, что Ω — счетное множество. Так как Ω является областью определения для случайной величины X, то множество значений X конечно, если Ω конечно, либо счетно бесконечно, если таково Ω. Действительно, *множество значений* X задается следующим выражением:

$$V_X = \bigcup_{\omega \in \Omega} \{X(\omega)\}, \tag{4.3.5}$$

причем какие-то из чисел, входящих в это объединение, могут совпадать ввиду того, что отображение $\omega \to X(\omega)$ в общем случае не обязано быть взаимно однозначным. В особом случае, когда X — постоянная случайная величина, множество V_X состоит из единственного числа. Запишем все значения из множества V_X в произвольном порядке:

$$\{v_1, v_2, \ldots, v_n, \ldots\}.$$

Эта последовательность может быть как конечной, так и бесконечной. Очевидно, что если $x \notin V_X$, т. е. x не является одним из v_n, то $P(X = x) = 0$. С другой стороны, не запрещается, чтобы некоторые из v_n имели нулевые вероятности. Это означает, что *некоторым точкам выборочного пространства может быть приписана вероятность* 0. Возможно, вы возразите: «Почему же тогда не выбросить эти бесполезные точки из выборочного пространства?» Дело в том, что часто трудно заранее определить, какие именно точки надо выбрасывать. Проще оставить их, так как они безвредны. (В несчетном пространстве Ω каждая отдельная точка ω может иметь нулевую вероятность! Однако мы пока не станем обсуждать это подробнее; см. § 4.5 ниже.)

Введем обозначение

$$p_n = P(X = v_n), \quad v_n \in V_X. \tag{4.3.6}$$

Должно быть ясно, что если известны все p_n, то можно вычислить все вероятности, связанные со случайной величиной X *самой по себе.* Так, вероятности (4.3.2) и (4.3.3) выражаются через p_n следующим образом:

$$P(a \leqslant X \leqslant b) = \sum_{a \leqslant v_n \leqslant b} p_n; \quad P(X \in A) = \sum_{v_n \in A} p_n. \tag{4.3.7}$$

Первое из выражений есть частный случай второго. Последняя из сумм произносится как «сумма таких p_n, для которых соответствующие v_n принадлежат множеству A».

Рассматривая в качестве A бесконечные интервалы вида $(-\infty, x]$ для произвольных действительных x, приходим к следующей функции от x:

$$F_X(x) = P(X \leqslant x) = \sum_{v_n \leqslant x} p_n. \tag{4.3.8}$$

Функция $x \to F_X(x)$, заданная на $\mathbb{R}^1$, называется *функцией распределения* случайной величины X. Ее значение в точке x аккумулирует вероятности всех возможных значений X вплоть до x (включительно); по этой причине к названию функции иногда добавляют прилагательное «кумулятивная». Например, если X — годовой доход (в \$) кормильца, то значение $F_X(10\,000)$ равно вероятности принадлежать к группе людей, зарабатывающих не более \$10 000, а также теоретически включающей всех тех, чей доход отрицателен.

Функция распределения F_X определяется через величины p_n и v_n, как показано в формуле (4.3.8). Обратно, если известна функция F_X, т. е. мы знаем значения $F_X(x)$ при всех x, то можно «восстановить» все p_n и v_n. Мы не станем доказывать здесь это довольно очевидное утверждение. Из соображений удобства, будем говорить, что два множества чисел $\{p_n\}$ и $\{v_n\}$ определяют *вероятностное распределение* случайной величины X. При этом все v_n, для которых $p_n = 0$, можно отбросить. Легко видеть, что если $a < b$, то

$$P(a < X \leqslant b) = P(X \leqslant b) - P(X \leqslant a) = F_X(b) - F_X(a); \tag{4.3.9}$$

но как выразить $P(a \leqslant x \leqslant b)$ или $P(X = x)$ через F_X? (См. задачу 7.)

Теперь вернемся к вероятностям p_n, которые иногда называют «элементарными вероятностями» случайной величины X. В общем случае для них выполняются следующие два свойства:

$$\begin{aligned} &\text{(i)} \quad \forall n: \quad p_n \geqslant 0; \\ &\text{(ii)} \quad \sum_n \quad p_n = 1. \end{aligned} \tag{4.3.10}$$

Сравните их со свойствами (2.4.1). Сумма в (ii) состоит из конечного или бесконечного числа слагаемых в зависимости от конечности или бесконечности множества V_X. Свойство (i) очевидно за исключением уже сделанного замечания, что некоторые из p_n возможно равны 0. Свойство (ii) равносильно утверждению, что величины $\{v_n\}$ из множества V_X исчерпывают все возможные значения случайной величины X, в силу чего их вероятности должны вместе составлять вероятность всего пространства (нашей «вселенной»). Это — верное словесное описание, однако давайте также учиться переводу словесных рассуждений в доказательства на языке символов. Прежде всего запишем:

$$\bigcup_n \{X = v_n\} = \Omega.$$

Так как значения v_n различны, множества $\{X = v_n\}$ не пересекаются. Поэтому в силу счетной аддитивности (см. § 2.3) имеем:

$$\sum_n P(X = v_n) = P(\Omega) = 1.$$

Это и есть свойство (ii).

Прежде чем углубляться в изучение свойств случайных величин, давайте введем новое фундаментальное понятие во всей его общности. Оно происходит от интуитивного представления о среднем значении случайной характеристики.

Определение математического ожидания. Пусть случайная величина X задана на некотором счетном выборочном пространстве Ω. Тогда ее математическим ожиданием*) $E(X)$ называется число, задаваемое формулой

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}), \tag{4.3.11}$$

при условии, что ряд сходится абсолютно, т. е.

$$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) < \infty. \tag{4.3.12}$$

В таком случае говорят, что математическое ожидание случайной величины X *существует*. Процедура «взятия математического ожидания» может быть описана словами следующим образом: возьмите значение случайной величины X в каждой из точек ω, умножьте его на вероятность соответствующей точки, просуммируйте эти произведения

*) Первая буква обозначения происходит от английского слова «expectation» — ожидание. — *Прим. перев.*

по всем ω из Ω. Если интерпретировать вероятность $P(\{\omega\})$ как вес, приписанный точке ω, то математическое значение $E(X)$ имеет смысл взвешенного среднего для функции X. Заметьте, что если записать точки в виде последовательности $\{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$, то

$$E(X) = \sum_n X(\omega_n) P(\{\omega_n\}).$$

Но мы также можем рассматривать сами ω в качестве индексов, что и подразумевается в формуле (4.3.11).

Пример 6. Пусть Ω обозначает участок земли, разделенный на семь частей для продажи. Вот проценты от площади всего участка, приходящиеся на каждую из частей ω, а также стоимости частей:

5 %,	10 %,	10 %,	10 %,	15 %,	20 %,	30 %;
\$800,	\$900,	\$1000,	\$1200,	\$800,	\$900,	\$800.

Определим $X(\omega)$ как цену за акр земли на ω. Тогда величина $E(X)$ представляет собой среднюю цену за акр земли всего участка и определяется выражением

$$(800)\,\frac{5}{100} + (900)\,\frac{10}{100} + (1000)\,\frac{10}{100} + (1200)\,\frac{10}{100} + (800)\,\frac{15}{100} + {}$$
$$+ (900)\,\frac{20}{100} + (800)\,\frac{30}{100} = 890,$$

т. е. составляет \$890 за акр. Величину $E(X)$ также можно подсчитать, объединяя части, имеющие одинаковую цену, а затем суммируя по всем различным ценам:

$$(800)\left(\frac{5}{100} + \frac{15}{100} + \frac{30}{100}\right) + (900)\left(\frac{10}{100} + \frac{20}{100}\right) + (1000)\,\frac{10}{100} + (1200)\,\frac{10}{100} =$$
$$= (800)\,\frac{50}{100} + (900)\,\frac{30}{100} + (1000)\,\frac{10}{100} + (1200)\,\frac{10}{100} = 890.$$

Иногда в термине «математическое ожидание» опускают прилагательное или заменяют сам термин на эквиваленты: ожидаемое значение, среднее (значение), первый момент (последнее из понятий рассматривается в § 6.3). Во всяком случае, не стоит надеяться, что, наблюдая случайную величину X, вы получите значение $E(X)$. Например, пусть при бросании симметричной монеты вы выигрываете \$1 в случае выпадения решки и не выигрываете ничего, когда выпадает орел. Тогда ни в одном из испытаний вы не получаете ожидаемое значение выигрыша \$0.5! Однако, если монета бросается n раз, причем n достаточно велико, то с большой вероятностью вы можете рассчитывать выиграть сумму, близкую к $n/2$ долларам. Этот вывод следует из закона больших чисел, подробно обсуждаемого в § 7.6.

Теперь мы собираемся рассмотреть условие (4.3.12) абсолютной сходимости ряда. Оно, очевидно, автоматически выполняется, когда пространство Ω конечно. Если же Ω счетно бесконечно, то условие (4.3.12) существенно, так как позволяет вычислять математическое ожидание, как угодно переставляя и группируя члены ряда из формулы (4.3.11) без риска получить противоречащие друг другу результаты. Другими словами, абсолютно сходящийся ряд имеет однозначно определенную сумму, не зависящую от порядка суммирования членов ряда. Возможно вам представляется удивительным, что на самом деле могут возникнуть противоречия, если данное условие не выполняется. Если так, то читателю рекомендуется освежить свои знания из математического анализа, относящиеся к теме сходимости и абсолютной сходимости числовых рядов. Эта часть курса анализа нередко поверхностно усваивается студентами (иногда по причине плохого преподавания). Однако она важна для теории вероятностей не только в связи с рассматриваемым вопросом, но и в более широком плане. Можете ли вы, скажем, привести пример, когда ряд в (4.3.11) сходится, а ряд из (4.3.12) расходится? (Имейте в виду, что вероятности p_n должны удовлетворять условию (4.3.10), в то время как значения v_n совершенно произвольны. Таким образом, вопрос несколько труднее, чем просто предъявить пример ряда, не сходящегося абсолютно; см. задачу 21.) В случае расходимости ряда (4.3.12) математическое ожидание считается *неопределенным*. Причина строгости нашего изложения кроется в том, что абсолютно сходящиеся ряды допускают обращение, непозволительное в случае сходящихся рядов, не сходящихся абсолютно[*)]. В самом деле, определение $E(X)$ потеряет смысл, если значение математического ожидания будет изменяться при перестановке некоторых членов ряда из (4.3.11), которая всего лишь означает изменение нумерации точек выборочного пространства. Но это может произойти, если не потребовать выполнения условия (4.3.12)!

Часто оказывается полезным следующий общий метод вычисления математического ожидания $E(X)$. Пусть выборочное пространство Ω разбито на непересекающиеся множества A_n,

$$\Omega = \bigcup_n A_n, \tag{4.3.13}$$

[*)] Их называют условно сходящимися. — *Прим. перев.*

таким образом, что случайная величина X имеет одно и то же значение для всех ω из каждого A_n. Тогда можно записать,что

$$X(\omega) = a_n \quad \text{для} \quad \omega \in A_n, \tag{4.3.14}$$

где a_n не обязательно различны. Отсюда имеем:

$$E(X) = \sum_n P(A_n)a_n = \sum_n P(X = a_n)a_n. \tag{4.3.15}$$

Эта формула получена за счет перегруппировки точек ω в записи (4.3.11): сначала они были объединены в множества A_n, затем выполнено суммирование по всем n. В частности, если $(v_1, v_2, \ldots, v_n, \ldots)$ — возможные значения X, и мы группируем точки выборочного пространства в соответствии со значениями $X(\omega)$, т. е. полагаем

$$A_n = \{\omega \mid X(\omega) = v_n\}, \quad P(A_n) = p_n,$$

то приходим к формуле

$$E(X) = \sum_n p_n v_n, \tag{4.3.16}$$

где ряд автоматически абсолютно сходится в силу условия (4.3.12). Эта форма представления показывает, что математическое ожидание случайной величины X определяется ее вероятностным распределением.

Наконец, полезно отметить, что формула (4.3.16) дает выражение для математического ожидания произвольной функции от X:

$$E(\varphi(X)) = \sum_{\omega\in\Omega} \varphi(X(\omega))P(\{\omega\}),$$

при условии, подобном (4.3.12). Действительно, согласно утверждению 2 (или просто из аналогии), $\varphi(X)$ также является случайной величиной. Отсюда вытекает представление

$$E(\varphi(X)) = \sum_n p_n \varphi(v_n), \tag{4.3.17}$$

где v_n те же самые, что и в формуле (4.3.16), причем значения $\varphi(v_n)$ не обязательно должны быть различными. Таким образом, математическое ожидание $\varphi(X)$ можно выразить через вероятностное распределение X и функцию φ, не находя предварительно самого распределения $\varphi(X)$. Это очень удобно при вычислениях. В частности, для $\varphi(x) = x^r$ имеем следующую формулу для *r-го момента* случайной величины X (см. § 6.3):

$$E(X^r) = \sum_n p_n v_n^r. \tag{4.3.18}$$

4.4. Целочисленные случайные величины

В данном параграфе изучаются только случайные величины, принимающие целые неотрицательные значения. При этом удобно считать областью значений все множество таких чисел:

$$\mathbb{N}^0 = \{0, 1, 2, \ldots, n, \ldots\},$$

поскольку можно приписать нулевые вероятности тем из них, которые не потребуются. В данной ситуации для формул (4.3.6), (4.3.8) и (4.3.11) имеем следующие представления:

$$\begin{aligned} p_n &= P(X = n), \quad n \in \mathbb{N}^0, \\ F_X(x) &= \sum_{0 \leqslant n \leqslant x} p_n, \\ E(X) &= \sum_{n=0}^{\infty} n p_n. \end{aligned} \tag{4.4.1}$$

Ввиду того, что все члены в приведенных рядах неотрицательны, нет различия между сходимостью и абсолютной сходимостью. Более того, поскольку такие ряды либо сходятся к конечной сумме, либо расходятся к $+\infty$, в последнем случае можно положить $E(X) = +\infty$. Это не согласуется с общим определением из предыдущего параграфа, но может рассматриваться как удобное обобщение.

Во многих задачах с практической точки зрения допустимо считать, что при выборе достаточно малой единицы измерения случайные величины принимают только целые значения. Например, денежные суммы могут выражаться не только в долларах, но и в центах, или даже в десятых долях цента, если это необходимо. Если при измерении длины дюйм не является достаточно малой единицей, употребляются сотые или тысячные доли дюйма. Существует единица длины, называемая *ангстремом* (Å) и равная 10^{-7} миллиметра, которую используют для измерения длин электромагнитных волн. На практике, конечно, не встречаются несоизмеримые величины (иррациональные числа); какое-то время назад в некоторых из штатов США было законодательно определено, что π равно 3.14! Но не стоит заходить слишком далеко в подобном округлении.

Рассмотрим несколько примеров на применение формул (4.4.1).

Пример 7. Пусть L — некоторое натуральное число. Положим

$$p_n = \frac{1}{L}, \quad 1 \leqslant n \leqslant L. \tag{4.4.2}$$

Тогда автоматически все другие p_n обязаны равняться нулю, поскольку

$$\sum_{n=1}^{L} p_n = L \cdot \frac{1}{L} = 1,$$

и должны выполняться условия (4.3.10). Далее вычисляем

$$E(X) = \frac{1}{L}\sum_{n=1}^{L} n = \frac{1}{L} \cdot \frac{L(L+1)}{2} = \frac{L+1}{2}\,.$$

Сумма здесь найдена по формуле для *арифметической прогрессии*, которую вы наверняка изучали в школе.

В данном случае говорят, что X имеет *равномерное распределение* на множестве $\{1, 2, \ldots, L\}$. На языке, употреблявшемся в гл. 3, можно сказать, что все L возможных событий

$$\{X = 1\}, \quad \{X = 2\}, \quad \ldots, \quad \{X = L\}$$

являются *одинаково правдоподобными*. Ожидаемое значение X равно среднему арифметическому первых L натуральных чисел. Проиллюстрируем смысл этого среднего. Допустим, что вы вытаскиваете наудачу жетон из ящика, содержащего жетоны достоинством от 1 до 100 центов. Тогда ваша ожидаемая награда есть $E(X) = 50.5$ центов. Представляется ли вам разумным такой результат ?

Пример 8. Предположим, что симметричная монета бросается до тех пор, пока не выпадет решка. Пусть X — количество бросаний до того, как это произойдет. Таким образом, $\{X = n\}$ означает, что выпали $n-1$ орлов прежде, чем появилась первая решка. Из обсуждения в примере 8 из § 2.4 следует, что

$$p_n = P(T = n) = \frac{1}{2^n}\,, \tag{4.4.3}$$

потому, что благоприятный исход представляется только одной последовательностью $\underbrace{\text{Н}\text{Н}\ldots\text{Н}}_{n-1 \text{ раз}}\text{Т}$. Чему равно математическое ожидание X?

Согласно (4.4.1), оно задается формулой

$$\sum_{n=1}^{\infty} \frac{n}{2^n} = ? \tag{4.4.4}$$

Давайте выясним, как найти сумму такого ряда, хотя этот вопрос, строго говоря, не относится в данному курсу. Начнем с суммирования ряда, служащего первоисточником для многих подобных рядов:

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots + x^n + \ldots = \sum_{n=0}^{\infty} x^n \quad \text{при} \quad |x| < 1. \tag{4.4.5}$$

Это — простейший геометрический ряд, который, безусловно, вам не раз встречался. Продифференцируем его почленно:

$$\begin{aligned}\frac{1}{(1-x)^2} &= 1 + 2x + 3x^2 + \ldots + nx^{n-1} + \ldots = \\ &= \sum_{n=0}^{\infty}(n+1)x^n \quad \text{при } |x| < 1.\end{aligned} \tag{4.4.6}$$

Почленное дифференцирование допустимо, поскольку радиус сходимости степенного ряда из (4.4.5) равен 1. (При $|x| < 1$ наблюдается абсолютная и равномерная сходимость*) степенного ряда.) Подставив $x = 1/2$ в формулу (4.4.6), получим

$$4 = \sum_{n=0}^{\infty}(n+1)\left(\frac{1}{2}\right)^n. \tag{4.4.7}$$

Ввиду некоторого отличия этого ряда от ряда из (4.4.4), требуются еще небольшие арифметические преобразования. Один из путей — разбить каждый из членов ряда на два слагаемых:

$$4 = \sum_{n=0}^{\infty}\frac{n}{2^n} + \sum_{n=0}^{\infty}\frac{1}{2^n} = \sum_{n=1}^{\infty}\frac{n}{2^n} + 2,$$

где мы просуммировали второй ряд, подставив $x = 1/2$ в (4.4.5). Итак, сумма ряда в формуле (4.4.4) равна 2. Другой способ нахождения этой суммы состоит в изменении индекса суммирования: $n+1 = \nu$. В результате имеем равенство:

$$4 = \sum_{n=0}^{\infty}\frac{n+1}{2^n} = \sum_{\nu=1}^{\infty}\frac{\nu}{2^{\nu-1}} = 2\sum_{\nu=1}^{\infty}\frac{\nu}{2^\nu},$$

которое, конечно, приводит к тому же самому ответу. Оба использованных технических приема весьма полезны!

Среднее значение $E(X) = 2$ интуитивно представляется правильным. Поскольку вероятность получения решки в отдельном бросании равна 1/2, то два бросания дают $2 \cdot 1/2 = 1$ решку *в среднем.* Данное правдоподобное рассуждение (которое иногда встречается в тестовых работах способных студентов) можно сделать строгим доказательством, но необходимые для этого аргументы намного сложнее, чем вы, возможно, предполагаете. Речь идет о частном случае *равенства Вальда***)

*) Строго говоря, равномерная сходимость есть в любом круге с радиусом $r < 1$. — *Прим. перев.*

**) Названного в честь Абрахама Вальда (1902–1950), выдающегося американского статистика.

или *мартингальной теоремы* (это замечание предназначено для более подготовленного читателя).

Давайте сразу обобщим рассмотренную задачу на случай несимметричной монеты, выпадающей решкой с вероятностью p и орлом с вероятностью $q = 1 - p$. Тогда вероятность (4.4.3) преобразуется в

$$p_n = \underbrace{(q \dots q)}_{n-1} p = q^{n-1} p, \tag{4.4.8}$$

а вместо ряда (4.4.4) получим

$$\sum_{n=1}^{\infty} n q^{n-1} p = p \sum_{n=0}^{\infty} (n+1) q^n = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}\,. \tag{4.4.9}$$

Случайная величина X называется *временем ожидания* появления решки или так называемого «успеха». Распределение $\{q^{n-1}p;\ n = 1, 2, \dots\}$ будем именовать *геометрическим распределением с вероятностью «успеха» p.*

Пример 9. Симметричная монета бросается n раз. Обозначим через S_n количество выпавших решек. Используя обозначения из § 2.4, запишем: $S_n = X_1 + \dots + X_n$. Мы знаем из § 3.2, что

$$p_k = P(S_n = k) = \frac{1}{2^n}\binom{n}{k}, \quad 0 \leqslant k \leqslant n. \tag{4.4.10}$$

Если верить вероятностной теории, то согласно условию (4.3.10) должно выполняться равенство $\sum\limits_{k=0}^{\infty} p_n = 1$. Следовательно,

$$\sum_{k=0}^{n} \frac{1}{2^n}\binom{n}{k} = 1 \quad \text{или} \quad \sum_{k=0}^{n} \binom{n}{k} = 2^n. \tag{4.4.11}$$

Это соотношение уже было установлено ранее (см. (3.3.7)) с помощью вероятностного рассуждения. Его также можно вывести из приведенной ниже формулы (4.4.13), положив в ней $x = 1$. Далее запишем:

$$E(S_n) = \sum_{k=0}^{n} \frac{k}{2^n}\binom{n}{k}. \tag{4.4.12}$$

Здесь опять требуется просуммировать ряд, в данном случае — конечный. Мы найдем сумму двумя разными способами, каждый из которых пригодится нам в дальнейшем для других вычислений. Первый заключается в следующем преобразовании членов ряда:

$$\sum_{k=0}^{n} \frac{k}{2^n}\ \frac{n!}{k!\,(n-k)!} = \frac{n}{2^n} \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{n}{2^n} \sum_{k=1}^{n} \binom{n-1}{k-1}.$$

Здесь мы сократили k в числителе с наибольшим сомножителем в произведении $k!$, отщепили n от $n!$ и отбросили нулевой член ввиду множителя $k = 0$. Теперь изменим индекс суммирования, положив $k - 1 = j$ (как уже делалось в примере 8):

$$\frac{n}{2^n} \sum_{k=1}^{n} \binom{n-1}{k-1} = \frac{n}{2^n} \sum_{j=0}^{n-1} \binom{n-1}{j} = \frac{n}{2^n} \cdot 2^{n-1} = \frac{n}{2},$$

где предпоследний переход совершен с учетом формулы (4.4.11), в которой параметр n заменен на $n - 1$. Итак, в ответе получаем $n/2$.

Рассмотренный способ можно рекомендовать, если вы с легкостью обращаетесь с комбинаторными формулами и такими понятиями, как биномиальные коэффициенты. Однако многие читатели, вероятно, сочтут более простым другой способ, отчасти напоминающий кулинарный рецепт. Начнем с бинома Ньютона, представленного в следующем виде:

$$(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{4.4.13}$$

Обратим внимание, что данная формула представляет собой просто разложение левой части по степеням x и является частным случаем ряда Тейлора так же, как и ряды в формулах (4.4.5) и (4.4.6). Дифференцируя, получаем:

$$n(1 + x)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k-1}. \tag{4.4.14}$$

Подставим $x = 1$:

$$n 2^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k;$$

Разделив на 2^n, приходим снова к ответу $n/2$ для суммы (4.4.12). Таким образом, ожидаемое количество решек при n бросаниях монеты равно $n/2$. Еще раз отметим, что вполне разумно ожидать, что решки будут выпадать примерно в половине испытаний.

Нетрудно обобщить и эту задачу на случай несимметричной монеты. Тогда вероятности (4.4.12) заменяются вероятностями

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leqslant k \leqslant n. \tag{4.4.15}$$

Эта формула уже встречалась нам в § 2.4. Теперь мы можем сказать, что она задает вероятностное распределение случайной величины $S_n = \sum\limits_{i=1}^{n} X_i$. Такое распределение называется *биномиальным* и имеет обозначение $B(n;p)$. Случайные величины X_i (их распределения) называются *бернуллиевскими*. В случае $p = 1/2$ добавляется прилагательное *симметричные*. Далее, формуле (4.4.12) соответствует

$$\sum_{k=0}^{n} \binom{n}{k} kp^k q^{n-k} = np. \tag{4.4.16}$$

Оба способа доказательства по-прежнему работают. Второй — быстрее: положив $x = p/q$ в соотношении (4.4.14), с учетом равенства $p + q = 1$ выводим

$$n\left(1 + \frac{p}{q}\right)^{n-1} = \frac{n}{q^{n-1}} = \sum_{k=0}^{n} \binom{n}{k} k \left(\frac{p}{q}\right)^{k-1};$$

умножая обе части последнего равенства на pq^{n-1}, устанавливаем тождество (4.4.16).

Пример 10. Если в задаче 1 из § 3.4 обозначить через X количество бракованных изделий, то вероятности $P(X = j)$ задаются формулой (3.4.1). Такое распределение называют *гипергеометрическим*.

4.5. Случайные величины, имеющие плотности

В предыдущих параграфах достаточно подробно обсуждались случайные величины, принимающие только счетное множество значений. Однако даже на элементарном уровне возникает немало важных вопросов, для ответа на которые необходимо рассматривать случайные величины, не удовлетворяющие данному ограничению. Это означает, что нам придется изучать несчетные вероятностные пространства. В таком случае неизбежно появляются технические проблемы, связанные с понятием измеримости, которые нельзя удовлетворительно изложить без использования более сложной математики. Как уже отмечалось в гл. 2, данного вида трудности возникают из невозможности назначения вероятности каждому подмножеству несчетного выборочного пространства. Решение проблемы заключается в том, чтобы ограничиться множествами, входящими в достаточно широкий класс, называемый сигма-алгеброй; см. дополнение 1. Не углубляясь более в эту проблематику, рассмотрим частную, но очень важную модель, пригодную для большинства приложений и требующую не очень сложной математической техники. Ею описываются случайные величины, обладающие так называемой «плотностью».

Пусть функция $f\colon u \to f(u)$ определена на всей числовой прямой $\mathbb{R}^1 = (-\infty, \infty)$ и удовлетворяет двум условиям:

$$\begin{aligned} &\text{(i)} \quad \forall u\colon \quad f(u) \geqslant 0; \\ &\text{(ii)} \quad \int\limits_{-\infty}^{\infty} f(u)\,du = 1. \end{aligned} \tag{4.5.1}$$

Такая функция называется *плотностью* на $\mathbb{R}^1$. Интеграл в условии (ii) — интеграл Римана, изучаемый в математическом анализе. Надеемся, читатель помнит, что если функция f непрерывна или *кусочно непрерывна*, то определенный интеграл $\int\limits_a^b f(u)\,du$ существует для любого отрезка $[a, b]$. Тем не менее, чтобы был определен несобственный интеграл по неограниченному интервалу $(-\infty, \infty)$ нужны дополнительные условия, обеспечивающие достаточно быстрое убывание функции $f(u)$ при больших значениях $|u|$. Такие функции называются «интегрируемыми на $\mathbb{R}^1$». Требование, чтобы интеграл по всей прямой был равен 1 менее серьезно, чем может показаться. Действительно, если

$$\int\limits_{-\infty}^{\infty} f(u)\,du = M < \infty,$$

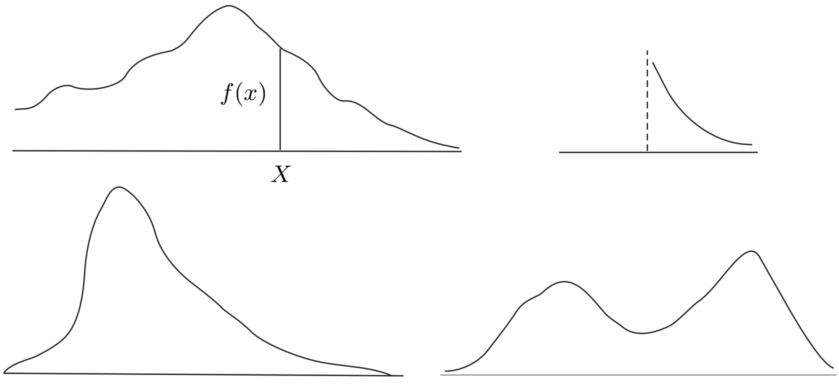
то достаточно разделить обе части на M и использовать функцию f/M вместо f. Приведем несколько возможных графиков плотностей. Какие-то из них — гладкие функции, какие-то — нет.

Из рис. 20 читатель может получить представление о том, какие бывают плотности. Ограничения таковы: кривая обязана лежать всюду выше оси абсцисс, площадь под графиком должна иметь смысл, а общая площадь — равняться 1. Согласитесь, что требования не слишком

Теперь можно определить интересующий нас класс случайных величин на произвольном выборочном пространстве следующим образом. Пусть, как и в § 4.2, X обозначает функцию на Ω: $\omega \to X(\omega)$, но сейчас связанные с ней вероятности для любого отрезка $[a, b]$ зададим с помощью плотности f так:

$$P(a \leqslant X \leqslant b) = \int\limits_a^b f(u)\,du. \tag{4.5.2}$$

Немного обобщим: если A — объединение интервалов, не обязательно непересекающихся, причем некоторые из них могут быть бес-



Графики плотностей

Рис. 20

конечными, то

$$P(X \in A) = \int\limits_A f(u)\,du. \tag{4.5.3}$$

Тогда говорят, что случайная величина X *имеет плотность* f. (В некоторых книгах такие случайные величины называют непрерывными, а величины из § 2 — дискретными. Оба прилагательных не совсем точны, поэтому мы не станем использовать их в этой книге.)

Любое конечное объединение интервалов A представляется в виде объединения *непересекающихся* интервалов (некоторые из них могут примыкать друг к другу):

$$A = \bigcup_{j=1}^{k} [a_j, b_j].$$

Тогда правая часть определения (4.5.3) записывается в виде суммы интегралов:

$$\int\limits_A f(u)\,du = \sum_{j=1}^{k} \int\limits_{a_j}^{b_j} f(u)\,du.$$

Это свойство интегралов очевидно с геометрической точки зрения, при которой они интерпретируются как площади. Далее, если взять $A = (-\infty, x]$, то получаем формулу

$$F(x) = P(X \leqslant x) = \int\limits_{-\infty}^{x} f(u)\,du; \tag{4.5.4}$$

сравните ее с определением (4.3.8). Тем самым установлено, что *функция распределения* F служит *первообразной* для плотности f. В силу известной теоремы из математического анализа при условии *непрерывной плотности* функция f является *производной* от F:

$$F'(x) = f(x). \tag{4.5.5}$$

Таким образом, в данном случае обе функции взаимно определяют друг друга. Когда f не является непрерывной всюду, формула (4.5.5) остается верной для всех тех x, в которых f непрерывна. Это доказывается в анализе.

Отметим, что в приведенном выше определении случайной величины, имеющей плотность, *подразумевается*, что вероятности приписаны любым множествам вида $\{a \leqslant X \leqslant b\}$ и $\{X \in A\}$: они выражаются через функцию плотности формулами (4.5.2) и (4.5.3). Для нас важно обратить внимание читателя на это обстоятельство, но мы не собираемся разрабатывать его детально. (В противном случае, нам придется столкнуться с трудностями (см. дополнение 1), которых сейчас хотелось бы избежать.) Вместо этого давайте подчеркнем близкое сходство только что выписанных формул с соответствующими формулами из § 4.3. Приведем сравнительный список, дополненный определением математического ожидания для случайных величин, обладающих плотностью.

	Счетный случай	Есть плотность
Область значений	$v_n, n = 1, 2, \ldots$	$-\infty < u < +\infty$
Элемент вероятности	p_n	$f(u)\,du = dF(u)$
$P(a \leqslant X \leqslant b)$	$\sum\limits_{a \leqslant v_n \leqslant b} p_n$	$\int\limits_a^b f(u)\,du$
$P(X \leqslant x) = F(x)$	$\sum\limits_{v_n \leqslant x} p_n$	$\int\limits_{-\infty}^x f(u)\,du$
$E(X)$	$\sum\limits_n p_n v_n$	$\int\limits_{-\infty}^{\infty} u f(u)\,du$
Условие существования $E(X)$	$\sum\limits_n p_n \lvert v_n \rvert < \infty$	$\int\limits_{-\infty}^{\infty} \lvert u \rvert f(u)\,du < \infty$

При наличии плотности аналогом формулы (4.3.17) служит интеграл

$$E(\varphi(X)) = \int\limits_{-\infty}^{\infty} \varphi(u) f(u)\,du. \tag{4.5.6}$$

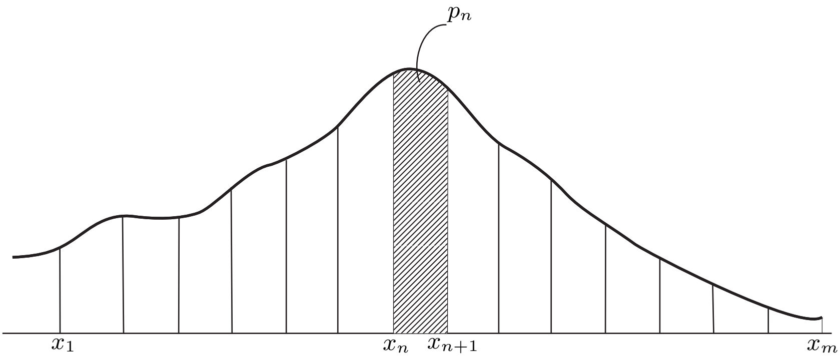


Рис. 21

Можно игнорировать стоящее справа во второй строке таблицы выражение, которое содержит *дифференциал*, если вы не знакомы с этим понятием.

Для более глубокого понимания аналогии рассмотрим рис. 21.

Кривая представляет собой график плотности f. Разобьем ось абсцисс на $m+1$ интервалов, не обязательно равных по длине и не обязательно малых. Обозначим через p_n площадь под графиком на интервале между x_n и x_{n+1}:

$$p_n = \int\limits_{x_n}^{x_{n+1}} f(u)\,du, \quad 0 \leqslant n \leqslant m,$$

где $x_0 = -\infty$, $x_{m+1} = +\infty$. Очевидно, что

$$\forall n: \quad p_n \geqslant 0; \quad \sum_n p_n = 1.$$

Следовательно, числа p_n удовлетворяют условиям (4.3.10). Вместо конечного разбиения, можно устроить счетное, используя подходящую индексацию, например, $\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots$. Действуя таким образом, мы можем получать наборы «элементарных вероятностей» на основе функции плотности бесконечным числом способов. Данный процесс называется *дискретизацией*. Для X с плотностью f рассмотрим такую случайную величину Y, что

$$P(Y = x_n) = p_n,$$

где допустимо заменить x_n на произвольную точку из отрезка $[x_n, x_{n+1}]$. Если плотность f непрерывна, а разбиение достаточно мелкое в том смысле, что образующие его интервалы малы, то из геометрических соображений очевидно, что случайная величина Y является в некотором смысле приближением для X. К примеру,

$$E(Y) = \sum_n p_n x_n$$

служит аппроксимацией для $E(X) = \int\limits_{-\infty}^{\infty} uf(u)\,du$. Помните ли вы, что изучаемые в анализе *римановы суммы* сходятся к интегралу Римана? Там полоски с криволинейным верхом на рис. 21 заменяются на полоски постоянной высоты (прямоугольники), но сама идея приближения совершенно аналогична. На практике реально измеряется именно дискретная аппроксимация, в то время как непрерывная плотность является лишь математической идеализацией. Мы скоро вернемся к этой теме.

Подробно рассмотрев сходство двух классов случайных величин, необходимо указать на фундаментальное различие между ними. Если X имеет плотность, то в силу определения (4.5.2) при $a = b = x$ получаем:

$$P(X = x) = \int\limits_x^x f(u)\,du = 0. \tag{4.5.7}$$

Геометрически это равенство всего лишь выражает тривиальный факт, что отрезок имеет нулевую площадь. Так как x в формуле (4.5.7) — произвольное число, то случайная величина X принимает любое заданное значение с вероятностью нуль. Это прямо противоположно поведению случайных величин со счетным множеством значений, поскольку некоторые из значений должны приниматься с положительными вероятностями. Представляется парадоксальным, что с одной стороны $X(\omega)$ *есть некоторое число* при каждом ω, а с другой стороны *любое фиксированное значение принимается с вероятностью нуль*. Следующий простой пример поможет прояснить ситуацию.

Пример 11. Раскрутим стрелку, закрепленную на оси. Остановившись, она будет указывать на определенный угол θ (отмеряемый, скажем, от горизонтальной прямой). В обычных условиях разумно предполагать, что он *равномерно распределен* между 0° и 360° (сравните с примером 7 из § 4.4). Это означает, что его плотность задается формулой

$$f(u) = \begin{cases} \dfrac{1}{360}, & \text{если } 0 \leqslant u \leqslant 360, \\ 0, & \text{в противном случае.} \end{cases}$$

Отсюда для любых $\theta_1 < \theta_2$ имеем:

$$P(\theta_1 \leqslant \theta \leqslant \theta_2) = \int_{\theta_1}^{\theta_2} \frac{1}{360}\, du = \frac{\theta_2 - \theta_1}{360}\,. \tag{4.5.8}$$

Из формулы следует, что вероятность того, что стрелка остановится внутри угла, образованного двумя фиксированными направлениями, пропорциональна величине этого угла. Если уменьшить (сжать) угол до нуля, то же самое произойдет и с соответствующей вероятностью. Поэтому вероятность события, состоящего в том, что стрелка будет указывать точно на заданное значение θ, равна нулю. С эмпирической точки зрения такое событие не имеет смысла, так как реальная стрелка обладает некоторой шириной. Таким образом, воображаемая стрелка представляет собой воплощение математического понятия «линии без ширины», в чем и заключается корень парадокса.

Полезно более пристально рассмотреть данную ситуацию. Понятно, что вместо того, чтобы вращать стрелку, можно просто выбирать наудачу число из отрезка $[0, 1]$. Для этого достаточно развернуть окружность в отрезок и изменить масштаб. Запишем теперь произвольную точку из $[0, 1]$ в десятичном представлении, например,

$$0.141592653589793\ldots\,. \tag{4.5.9}$$

Точка, записываемая с помощью конечного числа знаков, ничем не отличается от других, так как ее запись можно дополнить бесконечной последовательностью нулей, которые равноправны с другими цифрами. Таким образом, выбор точки из $[0, 1]$ сводится к выбору, одной за другой, цифр ее десятичного представления. Этот вид деятельности хорошо выполняется компьютерами. Шансы получить заданную цифру, скажем, цифру 1, в первом разряде числа (4.5.9), равны $1/10$. Результаты выбора цифр образуют реализацию последовательности независимых испытаний (см. § 2.4). Следовательно, шансы случайно выбрать 15 цифр, формирующих число (4.5.9), составляют

$$\underbrace{\frac{1}{10}\cdot\frac{1}{10}\cdots\frac{1}{10}}_{15\ \text{раз}} = \left(\frac{1}{10}\right)^{15}.$$

Так как 10^9 — это 1 миллиард, то данная вероятность представляется настолько малой, что, согласно высказыванию Эмиля Бореля*), *в зем-*

*) Эмиль Борель (1871–1956) — великий французский математик, один из основоположников теории вероятностей.

ной жизни ею можно пренебречь и фактически считать равной нулю! При этом мы фиксировали только 15 разрядов числа $\pi - 3$, поэтому вообще не возникает вопроса о том, чтобы выбрать наудачу само это число. Даже если вы готовы сколь угодно долго продолжать процесс, вам придется остановиться на некотором шаге и в результате получить число, которое столь же *непредсказуемо заранее*, как и $\pi - 3$. Таким образом, мы снова сталкиваемся со сложностями, порожденными математической абстракцией — системой действительных чисел.

Обобщим этот пример. Пусть $[a, b]$ — произвольный конечный невырожденный отрезок на $\mathbb{R}^1$. Положим

$$f(u) = \begin{cases} \dfrac{1}{b-a}, & \text{если } a \leqslant u \leqslant b, \\ 0, & \text{в противном случае.} \end{cases}$$

Очевидно, эта функция является плотностью. Соответствующее распределение называется *равномерным распределением на отрезке* $[a, b]$. Функция распределения допускает следующее явное представление:

$$F(x) = \frac{[(a \vee x) \wedge b] - a}{b - a}$$

(если вам нравятся подобные «хитрые» формулы).

Пример 12. В круге случайно проводится хорда. Какова вероятность того, что ее длина окажется больше стороны вписанного в круг равностороннего треугольника?

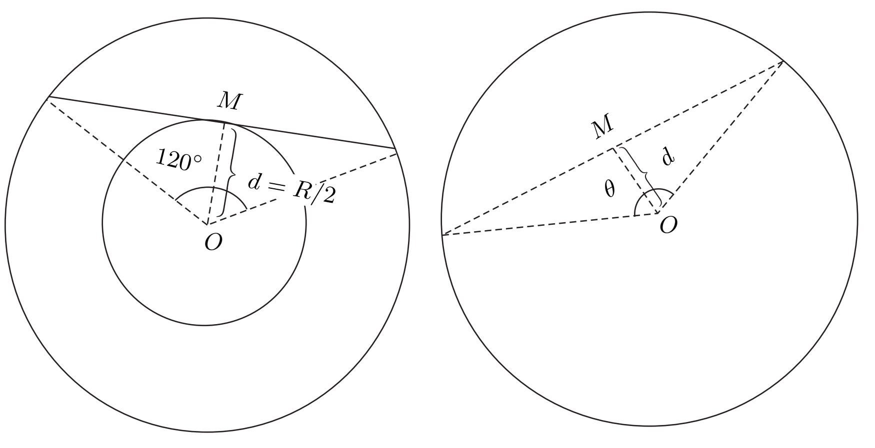


Рис. 22

Давайте нарисуем такой треугольник в круге с центром O и радиусом R. Заметим, что сторона треугольника находится на расстоянии $R/2$ от точки O; середина стороны лежит на концентрической окружности радиуса $R/2$; стороной стягивается угол величиной $120°$ с вершиной в точке O. Вы должны знать, как вычислить длину этой стороны, но она не потребуется. Обозначим через A интересующее нас событие, что длина случайной хорды превзойдет длину стороны треугольника. Длина произвольной хорды однозначно определяется любой из следующих трех характеристик: ее расстоянием d от точки O, местонахождением ее середины M, стягиваемым ею углом θ с вершиной в точке O. Мы хотим предположить последовательно для каждой из этих характеристик, что она равномерно распределена на своем диапазоне, и вычислить вероятности события A в каждом из случаев.

(1). Допустим, что расстояние d равномерно распределено на отрезке $[0, R]$. Это предположение выглядит правдоподобным, если представить, что линейка с постоянной скоростью движется параллельно оси ординат к центру окружности, начиная с касательного положения, а затем внезапно останавливается для проведения хорды. Из геометрических соображений очевидно, что событие A происходит тогда и только тогда, когда $d < R/2$. Отсюда $P(A) = 1/2$.

(2). Допустим, что точка M равномерно распределена внутри круга D, ограниченного окружностью. Такое предположение правдоподобно, когда маленький дротик бросается в круг D, и затем хорда проводится перпендикулярно линии, соединяющей точку попадания с центром O. Обозначим через D' концентрический круг радиуса $R/2$. Тогда событие A наблюдается тогда и только тогда, когда M оказывается внутри D'. Следовательно, $P(A) = P(M \in D') = (\text{площадь } D')/(\text{площадь } D) = 1/4$.

(3). Допустим, что угол θ равномерно распределен между $0°$ и $360°$. Это выглядит правдоподобно, если произвольно зафиксировать один из концов хорды, а другой конец получать в результате пересечения с окружностью прямой, равномерно вращающейся вокруг этой точки до остановки в некоторый момент. Тогда из рисунка понятно, что событие A происходит в том и только том случае, когда угол θ оказывается между $120°$ и $240°$. Поэтому $P(A) = (240 - 120)/360 = 1/3$.

Итак, ответами задачи являются вероятности $1/2$, $1/4$ или $1/3$ в соответствии с различными гипотезами. Следовательно, гипотезы не совместимы друг с другом. Возможны и другие гипотезы, приводящие к ответам, отличным от перечисленных. Удастся ли читателю придумать такую гипотезу? Рассматриваемая задача была известна под именем *парадокса Бертрана* еще в те далекие времена, когда велись дискуссии о том, что такое вероятность. Однако причина парадокса заключается исключительно в том факте, что задача на может считаться корректно сформулированной, если не указана истинная природа присутствующей в ней случайности. Не удивительно, что разные способы рандомизации приводят к разным ответам, что нетрудно установить и экспериментально с помощью описанных механических процедур. Вот житейская аналогия. Представим, что вас спросили, сколько времени занимает путь от дома до школы, не уточняя, идет ли речь о движении пешком, на велосипеде или на автомобиле. Надо ли называть парадоксом то, что ответы будут разными?

Мы завершим этот параграф еще несколькими примерами, в которых участвуют случайные величины, имеющие плотности. Другой важный частный случай — нормальное распределение — обсуждается в гл. 6.

Пример 13. Предположим, что вы стоите на автобусной остановке, расположенной на относительно спокойной проселочной дороге и наблюдаете за проезжающими мимо автомобилями. С помощью часов с секундомером вы можете засечь время, прошедшее с момента начала наблюдений до появления первого автомобиля. Эту случайную величину T обычно называют *временем ожидания*. При определенных условиях разумная гипотеза заключается в том, что плотностью распределения случайной величины T служит зависящая от параметра $\lambda > 0$ функция

$$f(u) = \lambda e^{-\lambda u}, \quad u \geqslant 0. \tag{4.5.10}$$

Не стоит и говорить, что $f(u) = 0$ при $u < 0$. Соответствующая функция распределения, именуемая *экспоненциальным (показательным) законом с параметром* λ, вычисляется интегрированием функции f (см. формулу (4.5.4)):

$$F(x) = \int\limits_{-\infty}^{x} f(u)\,du = \int\limits_{0}^{x} \lambda e^{-\lambda u}\,du = 1 - e^{-\lambda x}.$$

В частности, если положить $x = +\infty$, или, точнее, устремить $x \to \infty$ в этой формуле, легко увидеть, что f удовлетворяет условиям (4.5.1). Поэтому она на самом деле является плотностью. Итак,

$$P(T \leqslant x) = F(x) = 1 - e^{-\lambda x}; \tag{4.5.11}$$

но в данном случае часто более удобной оказывается так называемая «хвостовая вероятность»:

$$P(T > x) = 1 - F(x) = e^{-\lambda x}. \tag{4.5.12}$$

Она может быть получена прямо из формулы (4.5.3) при $A = (x, \infty)$:

$$P(T \in (x, \infty)) = \int\limits_{(x,\infty)} \lambda e^{-\lambda u}\, du = \int\limits_{x}^{\infty} \lambda e^{-\lambda u}\, du = e^{-\lambda x}.$$

Для любого фиксированного x, скажем, 5 (секунд), вероятность $e^{-5\lambda}$ в формуле (4.5.12) убывает с ростом λ. Это означает, что время ожидания имеет тенденцию сокращаться при увеличении λ. На перегруженном транспортом шоссе значение λ на самом деле будет большим. Среднее время ожидания задается следующей формулой:

$$E(T) = \int\limits_0^\infty u\lambda e^{-\lambda u}\, du = \frac{1}{\lambda}\int\limits_0^\infty te^{-t}\, dt = \frac{1}{\lambda}\,. \tag{4.5.13}$$

(Можете ли вы вычислить последний интеграл методом интегрирования по частям, не обращаясь к таблицам?) Этот результат подтверждает наше предыдущее наблюдение, что T в среднем уменьшается, когда λ возрастает.

Экспоненциальное распределение служит очень полезной моделью для различных процессов, имеющих дело с временами ожидания: телефонными вызовами, временами обслуживания, распадом радиоактивных частиц и т. п.; см. § 7.2.

Пример 14. Предположим, что в задаче из предыдущего примера нас интересует не само время ожидания T, а его логарифм (по основанию e):

$$S = \ln T. \tag{4.5.14}$$

Тогда S также является случайной величиной (см. утверждение 2 в § 4.2). Она принимает отрицательные значения, если $T < 1$; равна нулю, если $T = 1$; положительна, если $T > 1$. Каковы связанные с ней вероятности? Интерес могут представлять вероятности вида $P(a \leqslant S \leqslant b)$. Понятно, что для их нахождения достаточно выписать $P(S \leqslant x)$, т. е. функцию распределения F_S случайной величины S. Так как функция $x \to \ln x$ монотонна, то и обратная к ней функция $x \to e^x$ обладает этим свойством. Поэтому

$$S \leqslant x \quad \Leftrightarrow \quad \ln T \leqslant x \quad \Leftrightarrow \quad T \leqslant e^x.$$

Следовательно, в силу (4.5.11) имеем:

$$F_S(x) = P\{S \leqslant x\} = P\{T \leqslant e^x\} = 1 - e^{-\lambda e^x}.$$

Отсюда дифференцированием находим плотность f_S:

$$f_S(x) = F'_S(x) = \lambda e^x e^{-\lambda e^x} = \lambda e^{x-\lambda e^x}.$$

Формула выглядит довольно громоздко, но, как вы только что убедились, легко выводится.

Пример 15. Каждый год река разливается. Предположим, что отметка нижнего уровня воды установлена на высоте 1, а верхний уровень Y имеет функцию распределения

$$F(y) = P(Y \leqslant y) = 1 - \frac{1}{y^2}, \quad 1 \leqslant y < \infty. \tag{4.5.15}$$

Обратите внимание, что $F(1) = 0$, $F(y)$ возрастает по y, $F(y) \to 1$ при $y \to \infty$. Все соответствует смыслу вероятности $P(Y \leqslant y)$. Дифференцированием находим плотность:

$$f(y) = F'(y) = \frac{2}{y^3}, \quad 1 \leqslant y < \infty. \tag{4.5.16}$$

Нет необходимости проверять, что $\int\limits_{-\infty}^{\infty} f(y)\,dy = 1$, поскольку это равносильно тому, что $\lim\limits_{y\to\infty} F(y) = 1$. Вычислим ожидаемое значение Y:

$$E(Y) = \int\limits_1^\infty u \cdot \frac{2}{u^3}\,du = \int\limits_1^\infty \frac{2}{u^2}\,du = 2.$$

Таким образом, в среднем, максимум Y в два раза превосходит минимум.

Что произойдет, если установить отметку нижнего уровня на 0 вместо 1 и выбрать новую единицу измерения высоты, составляющую 1/10 от прежней? Тем самым мы переходим к величине

$$Z = 10(Y - 1). \tag{4.5.17}$$

Тогда, как и в примере 13, можно записать,что

$$Z \leqslant z \quad \Leftrightarrow \quad 10(Y-1) \leqslant z \quad \Leftrightarrow \quad Y \leqslant 1 + \frac{z}{10}, \quad 0 \leqslant z < \infty.$$

Отсюда нетрудно найти

$$F_Z(z) = 1 - \frac{100}{(10+z)^2},$$

$$f_Z(z) = \frac{200}{(10+z)^3}.$$

Вычисление $E(Z)$ на основе f_Z довольно скучно, но просто. Ответ таков: $E(Z) = 10$. Сравнивая его с $E(Y) = 2$, замечаем, что

$$E(Z) = 10(E(Y) - 1). \tag{4.5.18}$$

Значит, *средние* величин Y и Z связаны тем же линейным соотношением, что и сами случайные величины. Представляется ли вам это очевидным? Общее утверждение будет рассмотрено в § 6.1.

4.6. Общий случай

В самом общем случае случайной величиной на выборочном пространстве Ω называется такая функция X, что *для любого действительного числа x определена вероятность* $P(X \leqslant x)$.

Честно говоря, это определение ставит телегу впереди лошади. На самом деле первичной является вероятностная мера P, заданная на классе подмножеств Ω. Этот класс называется *сигма-алгеброй* или *борелевским полем* и обозначается через $\mathfrak{F}$. Тогда, если функция X такова, что для каждого x множество $\{\omega \mid X(\omega) \leqslant x\}$ принадлежит классу $\mathfrak{F}$, то она называется случайной величиной. (Сошлемся на приложение 1, где эти понятия раскрываются полнее. Остальной материал текущего параграфа проще и не содержит таких формальностей.) Другими словами, произвольная функция должна пройти тест, чтобы стать членом клуба. Новой здесь является идея того, что вероятностная мера P задана не на всех подмножествах пространства Ω, а только на множествах из класса $\mathfrak{F}$. Если окажется, что мера определена на всех подмножествах, то, конечно, описанная выше проверка станет номинальной, и любая функция будет случайной величиной. Именно такова ситуация для счетного выборочного пространства Ω, изучавшегося в § 4.1. В общем случае, как мы уже неоднократно отмечали, невозможно задать вероятностную меру на всех подмножествах пространства Ω, и поэтому приходится ограничиваться некоторым классом $\mathfrak{F}$. Так как вероятности приписаны только множествам из $\mathfrak{F}$, а мы хотим рассматривать события вида «$X \leqslant x$», мы обязаны потребовать, чтобы они входили в $\mathfrak{F}$. Отсюда легко понять необходимость проверки. Немного удивительным является то, что эта проверка — все, что нам надо. А именно, при выполнении данного требования удается определить вероятности самых разнообразных множеств, например, таких, как $\{a \leqslant X \leqslant b\}$, $\{X = x\}$, $\{X$ принимает рациональное значение$\}$, и в том числе довольно причудливых множеств, скажем, $\{e^x > X^2 + 1\}$.

Далее, для любого действительного x положим

$$F(x) = P(X \leqslant x) \tag{4.6.1}$$

или, что равносильно, для $a < b$

$$F(b) - F(a) = P(a < X \leqslant b);$$

и назовем F *функцией распределения* случайной величины X. Так мы и поступали ранее, но сейчас у нас уже нет конкретных выражений для F из формул (4.3.8) и (4.5.4):

$$F(x) = \sum_{\nu_n \leqslant x} p_n, \quad F(x) = \int\limits_{-\infty}^{x} f(u)\, du$$

в терминах элементарных вероятностей или плотности. На самом деле в общем случае функция F представляется в виде смеси этих двух типов с еще одним довольно замысловатым (*сингулярным*) типом. Однако для наших целей вполне достаточно определения (4.6.1) без дальнейших уточнений. К сожалению, для работы с общим случаем нужна более сложная математическая техника (на уровне курса «Основные понятия математического анализа»). Поэтому мы не станем углубляться в него, а лишь отметим два простых факта, относящихся к функции распределения F:

(i) F — неубывающая функция, т. е. $x \leqslant x' \Rightarrow F(x) \leqslant F(x')$;

(ii) F имеет пределы на $-\infty$ и $+\infty$, равные 0 и 1 соответственно:

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, \quad F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$$

Свойство (i) справедливо ввиду того, что если $x \leqslant x'$, то

$$\{X \leqslant x\} \subset \{X \leqslant x'\}.$$

Свойство (ii) интуитивно очевидно, так как событие $\{X \leqslant x\}$ становится невозможным при $x \to -\infty$ и становится достоверным при $x \to +\infty$. Это объяснение, возможно, удовлетворит вас, но строгое доказательство несколько сложнее. Оно опирается на счетную аддитивность вероятностной меры P (см. § 2.3). Отметим, что существование предела в (ii) вытекает из присутствующей в силу (i) монотонности и фундаментальной теоремы анализа: ограниченная монотонная последовательность действительных чисел имеет предел.

Оставшаяся часть параграфа посвящена краткому изложению некоторых базовых понятий, относящихся к случайным векторам. Изучение этого материала можно отложить, пока он не потребуется в гл. 6.

Для простоты обозначений мы станем рассматривать только две случайные величины X и Y, но обобщение на случай произвольного конечного числа случайных величин не должно вызвать затруднений. Прежде всего, займемся случаем, когда X и Y имеют счетные множества значений. Пусть X принимает значения $\{x_i\}$, Y принимает значения $\{y_j\}$, и положим

$$P(X = x_i, Y = y_j) = p(x_i, y_j). \tag{4.6.2}$$

Когда x_i и y_j пробегают весь возможный диапазон значений, множество «элементарных вероятностей» $p(x_i, y_j)$ представляет собой *совместное распределение компонент случайного вектора* (X, Y). Чтобы найти распределение X отдельно, позволим y_j принимать любое из возможных значений; при этом

$$P(X = x_i) = \sum_{y_i} p(x_i, y_j) = p(x_i, *), \tag{4.6.3}$$

где последняя из величин определяется суммой, стоящей в середине. Когда x пробегает весь возможный диапазон значений, множество вероятностей $p(x_i, *)$ образует *маргинальное распределение* случайной величины X. Аналогично определяется маргинальное распределение случайной величины Y. Отметим, что эти маргинальные распределения в общем случае не определяют совместное распределение двух случайных величин.

Подобно тому (см. (4.3.17)) как выражалось математическое ожидание произвольной функции от X с помощью вероятностного распределения, выразим теперь математическое ожидание произвольной функции от вектора (X, Y):

$$E(\varphi(X, Y)) = \sum_{x_i} \sum_{y_j} \varphi(x_i, y_j)\, p(x_i, y_j). \tag{4.6.4}$$

Поучительно заметить, что эта формула получается в результате переупорядочения слагаемых в определении математического ожидания от функции $\varphi(X, Y)$, которая интерпретируется как *скалярная* случайная величина; см. (4.3.11):

$$E(\varphi(X, Y)) = \sum_{\omega} \varphi(X(\omega), Y(\omega)) P(\omega).$$

Теперь рассмотрим случай, обобщающий модель из § 4.5, когда существует плотность. Говорят, что случайный вектор (X, Y) имеет (совместную) плотность f, если

$$P(X \leqslant x, Y \leqslant y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u,v)\,du\,dv \tag{4.6.5}$$

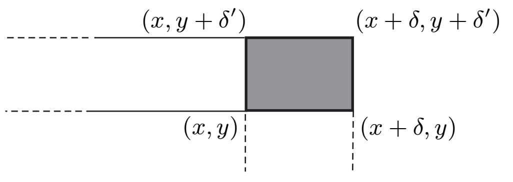
для всех (x,y). Тогда можно показать, что для любого не слишком сложно устроенного (*борелевского*) подмножества S декартовой плоскости верно равенство

$$P((X,Y) \in S) = \iint\limits_{S} f(u,v)\,du\,dv. \tag{4.6.6}$$

К примеру, S может быть многоугольником, кругом, эллипсом или объединением таких фигур. Обратите внимание, что равенство (4.6.5) является весьма частным случаем формулы (4.6.6). Если потребуется, можно принять более емкое условие (4.6.6) в качестве *определения* f как плотности вектора (X,Y). Тем не менее, приведем эвристическое рассуждение, обосновывающее вывод соотношения (4.6.6) из формулы (4.6.5). Обозначим через $R(x,y)$ бесконечный прямоугольник на плоскости со сторонами, параллельными координатным осям, лежащий на юго-запад от точки (x,y). Приведенный ниже рисунок показывает, что для любых $\delta > 0$ и $\delta' > 0$ множество

$$R(x+\delta, y+\delta') - R(x+\delta, y) - R(x, y+\delta') + R(x,y)$$

представляет собой серый прямоугольник.



Тогда, поступая аналогичным образом с соотношением (4.6.5), получаем формулу

$$P(x \leqslant X \leqslant x+\delta,\ y \leqslant Y \leqslant y+\delta') = \int\limits_{x}^{x+\delta} \int\limits_{y}^{y+\delta'} f(u,v)\,du\,dv.$$

Это означает, что равенство (4.6.6) верно для серого прямоугольника. Варьируя x, y, а также δ и δ', видим, что формула справедлива для любого прямоугольника такого вида. Но произвольную (не слишком сложную) фигуру можно аппроксимировать изнутри и снаружи объединением некоторого числа маленьких прямоугольников (или даже квадратов) — факт, известный еще древним грекам. Тогда в пределе получим формулу (4.6.6), как и утверждалось.

Любознательный читатель может удивиться, почему подобные рассуждения не были проведены для одной случайной величины; см. (4.5.3). Ответ в том, что с практической точки зрения на прямой $\mathbb{R}^1$ вряд ли найдутся представляющие интерес множества, отличные от интервалов, точек и их объединений. В одномерном пространстве очень тесно, и наша геометрическая интуиция работает плохо. В этом заключается одна из причин восприятия классической теории меры как сложного предмета.

Плотность f совместного распределения двух случайных величин удовлетворяет следующим условиям:

(i) $f(u, v) \geqslant 0$ для всех (u, v);

(ii) $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(u, v)\, du\, dv = 1.$

Конечно, равенство (ii) подразумевает, что функция f интегрируема на всей плоскости. Часто мы будет также предполагать, что f непрерывна. Приведем формулы, аналогичные соотношению (4.6.3):

$$
\begin{aligned}
P(X \leqslant x) &= \int_{-\infty}^{x} f(u, *)\, du, \quad \text{где} \quad f(u, *) = \int_{-\infty}^{\infty} f(u, v)\, dv;\\
P(Y \leqslant y) &= \int_{-\infty}^{y} f(*, v)\, dv, \quad \text{где} \quad f(*, v) = \int_{-\infty}^{\infty} f(u, v)\, du.
\end{aligned}
\tag{4.6.7}
$$

Функции $u \to f(u, *)$ и $v \to f(*, v)$ называются *маргинальными плотностями* случайных величин X и Y соответственно. Они вычисляются интегрированием совместной плотности по свободной переменной.

Формула (4.6.4) преобразуется в случае существования плотности в следующую: для любой «обыкновенной» (борелевской) функции φ

$$
E(\varphi(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(u, v) f(u, v)\, du\, dv. \tag{4.6.8}
$$

Класс «обыкновенных» функций содержит все ограниченные непрерывные функции от переменных (u, v); индикаторы «не слишком сложных» множеств; функции, непрерывные всюду, за исключением гладких линий разрыва, для которых интеграл в формуле (4.6.8) существует и т. п.

В самом общем случае *(совместная) функция распределения* F случайного вектора (X, Y) определяется формулой

$$F(x, y) = P(X \leqslant x, Y \leqslant y) \quad \text{для всех } (x, y). \tag{4.6.9}$$

Обозначая $\lim\limits_{y \to \infty} F(x, y)$ через $F(x, \infty)$, запишем, что

$$F(x, \infty) = P(X \leqslant x, Y < \infty) = P(X \leqslant x),$$

так как условие «$Y < \infty$» не накладывает никаких ограничений на Y. Таким образом, $x \to F(x, \infty)$ есть *маргинальная функция распределения* случайной величины X. Аналогично определяется маргинальная функция распределения случайной величины Y.

Несмотря на то, что введенные общие понятия образуют фундамент теории, имеющей дело одновременно с несколькими случайными величинами, явное их использование в этой книге будет встречаться довольно редко.

Задачи

1. Пусть X — случайная величина (на счетном выборочном пространстве). Верно ли, что

$$X + X = 2X, \quad X - X = 0?$$

Объясните в деталях.

2. Пусть $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $P(\omega_1) = P(\omega_2) = P(\omega_3) = 1/3$. Определим случайные величины X, Y и Z следующим образом:

$$\begin{array}{lll} X(\omega_1) = 1, & X(\omega_2) = 2, & X(\omega_3) = 3, \\ Y(\omega_1) = 2, & Y(\omega_2) = 3, & Y(\omega_3) = 1, \\ Z(\omega_1) = 3, & Z(\omega_2) = 1, & Z(\omega_3) = 2. \end{array}$$

Покажите, что все три случайные величины имеют одинаковое распределение. Найдите вероятностные распределения величин $X + Y$, $Y + Z$ и $Z + X$.

3. В условиях задачи 2 найдите вероятностные распределения величин

$$X + Y - Z, \quad \sqrt{(X^2 + Y^2)Z}, \quad \frac{Z}{|X - Y|}.$$

4. Возьмите в качестве Ω множество, состоящее из пяти произвольных действительных чисел. Задайте на нем вероятностную меру и случайную величину X, которая принимает значения 1, 2, 3, 4, 5 с вероятностями 1/10, 1/10, 1/5, 1/5, 2/5 соответственно, а также другую случайную величину Y, принимающую значения $\sqrt{2}$, $\sqrt{3}$, π с вероятностями 1/5, 3/10, 1/2. Найдите вероятностное распределение случайной величины XY. [*Указание.* Ответ зависит от вашего выбора и поэтому не однозначен.]

5. Обобщите задачу 4, конструируя Ω, P, X так, что X принимает значения $v_1, v_2, \ldots, v_n$ с вероятностями $p_1, p_2, \ldots, p_n$, где последние удовлетворяют условию (4.3.10).

6. Каков смысл следующих множеств в примере 3 из § 4.1?

$$\{X+Y=7\}, \quad \{X+T\leqslant 7\}, \quad \{X\vee Y>4\}, \quad \{X\neq Y\}.$$

Перечислите все ω в каждом из них.

7*. Пусть X — целочисленная случайная величина и F — ее функция распределения. Убедитесь, что для каждого x и $a<b$

$$P(X=x)=\lim_{\varepsilon\to+0}[F(x+\varepsilon)-F(x-\varepsilon)],$$
$$P(a<X<b)=\lim_{\varepsilon\to+0}[F(b-\varepsilon)-F(a+\varepsilon)].$$

(Эти формулы верны для любых случайных величин, но требуют более сложного доказательства даже в случае, когда Ω счетно.)

8. В условиях примера 4 из § 4.2 предположим, что

$$X=5\,000+X',$$

где случайная величина X' равномерно распределена на множестве целых чисел от 1 до 5 000. Что означает эта гипотеза? Найдите вероятностное распределение и математическое ожидание X при данной гипотезе.

9. Решите задачу 8 в случае, когда

$$X=4\,000+X',$$

где случайная величина X' равномерно распределена от 1 до 10 000.

10*. Решите задачу 8 в случае, когда

$$X=3000+X',$$

где случайная величина X' имеет экспоненциальное распределение со средним 7 000. Найдите математическое ожидание $E(X)$.

11. При $\lambda>0$ определим плотность f так:

$$f(u)=\begin{cases}\dfrac{1}{2}\,\lambda e^{-\lambda u}, & \text{если } u\geqslant 0,\\[2mm] \dfrac{1}{2}\,\lambda e^{+\lambda u}, & \text{если } u<0.\end{cases}$$

Такая функция f называется *двусторонней экспоненциальной.**) Пусть случайная величина X имеет плотность f. Найдите плотность случайной величины $|X|$. [*Указание.* Вычислите сначала функцию распределения.]

12. Предположим, что X — положительная случайная величина с плотностью f. Найдите плотность распределения случайной величины $+\sqrt{X}$.

*) Соответствующее распределение называют *законом Лапласа. — Прим. перев.*

Примените полученный результат для вычисления распределения стороны квадрата, площадь которого равномерно распределена на отрезке $[a, b]$.

13. Пусть X имеет плотность f. Найдите плотность распределения случайной величины

(a) $aX + b$, где a и b — константы;

(b) X^2.

14. Докажите формулу (4.4.5) двумя способами:

(a) раскрывая скобки в произведении $(1 - x)(1 + x + \ldots + x^n)$;

(b) используя разложение в ряд Тейлора.

15. Положим

$$p_n = cq^{n-1}p, \quad 1 \leqslant n \leqslant m,$$

где c — константа, m — натуральное число; сравните с (4.4.8). Задайте константу c так, чтобы выполнялось равенство $\sum\limits_{n=1}^{m} p_n = 1$. (Эта модель описывает распределение времени ожидания «успеха» при условии, что он окажется среди первых m испытаний.)

16. Симметричная монета бросается n раз. Пусть Y_n обозначает разность между количеством выпавших решек и количеством выпавших орлов. Найдите вероятностное распределение величины Y_n и ее математическое ожидание. (Существует простая связь между Y_n и величиной S_n из примера 9 в § 4.4.)

17. Вернемся к задаче 1 из § 3.4. Предположим, что из 550 яблок в мешке 11 являются гнилыми. Выберем случайно 25 яблок. Найдите вероятностное распределение количества X гнилых яблок среди выбранных.

18*. Обобщите задачу 17 на случай произвольных чисел и найдите математическое ожидание случайной величины X. [*Указание.* Для решения задачи надо иметь определенный опыт в комбинаторике. Она легко решается методом, рассмотренным в § 6.1.]

19. Положим

$$P(X = n) = p_n = \frac{1}{n(n+1)}, \quad n \geqslant 1.$$

Образуют ли такие p_n вероятностное распределение? Найдите $P(X \geqslant m)$ и $E(X)$.

20. Представим, что все книги в библиотеке разбросали, а затем наняли обезьяну для того, чтобы расставить их назад по полкам. Можно показать, что хорошим приближением для вероятности обнаружить в точности n книг на своих местах служит величина

$$\frac{e^{-1}}{n!}, \quad n \geqslant 0.$$

Найдите математическое ожидание количества книг, возвращенных на свои места. (Эта часто цитируемая иллюстрация является одним из вариантов задачи о совпадениях, включенной под номером 6 в § 3.4.)

21. Приведите пример, когда ряд $\sum\limits_n p_n v_n$ в формуле (4.3.11) сходится, но не абсолютно. [*Указание.* В этой задаче нет ничего трудного: возьмите, скажем, $p_n = 2^{-n}$, и подберите такие v_n, чтобы $p_n v_n$ являлись членами любого известного вам сходящегося, но не абсолютно, ряда.]

22. Докажите, что если f и g — плотности, то при условии, что $\lambda + \mu = 1$, $\lambda \geqslant 0$, $\mu \geqslant 0$, функция $\lambda f + \mu g$ также является плотностью.

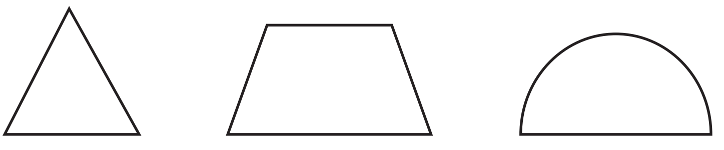
23. Найдите вероятность того, что длина случайно проведенной хорды больше радиуса. Как и в примере 11 из § 4.5, решите задачу отдельно для каждой из трех приведенных там гипотез.

24. Пусть

$$f(u) = ue^{-u}, \quad u \geqslant 0.$$

Убедитесь, что функция f является плотностью. Вычислите $\int\limits_0^\infty uf(u)\,du$.

25. На следующем рисунке изображены равнобедренный треугольник, равнобочная трапеция и полукруг.



Задайте размеры сторон и радиус полукруга так, чтобы фигуры можно было считать графиками плотностей.

26. Предположим, что мишень имеет форму круга с радиусом 10 футов, причем вероятность попадания в любой концентрический круг пропорциональна площади этого круга. Обозначим через R расстояние от точки попадания пули до центра круга. Найдите функцию распределения, плотность и математическое ожидание случайной величины R.

27. Агент 009 попал в ловушку. Он оказался между двумя близко стоящими высокими стенами. Агент размахивал пистолетом так, что направление ствола описывало окружность, касающуюся стен, как показано на рис. 23. Вдруг пистолет выстрелил. Допустим, что направление выстрела образовывало с горизонталью угол, равномерно распределенный между 0° и 90°. Найдите распределение и математическое ожидание высоты точки попадания пули в стену.

28. (*Санкт-Петербургский парадокс.*) Представим, что вы с товарищем играете в игру, представляющую из себя повторное бросание симметричной монеты до выпадения первой решки. Обозначим время ожидания через X. При этом вы согласны заплатить товарищу 2^X центов, когда значение X станет известным (т. е. 2^n центов, если $X = n$). Если вы согласны с тем, что

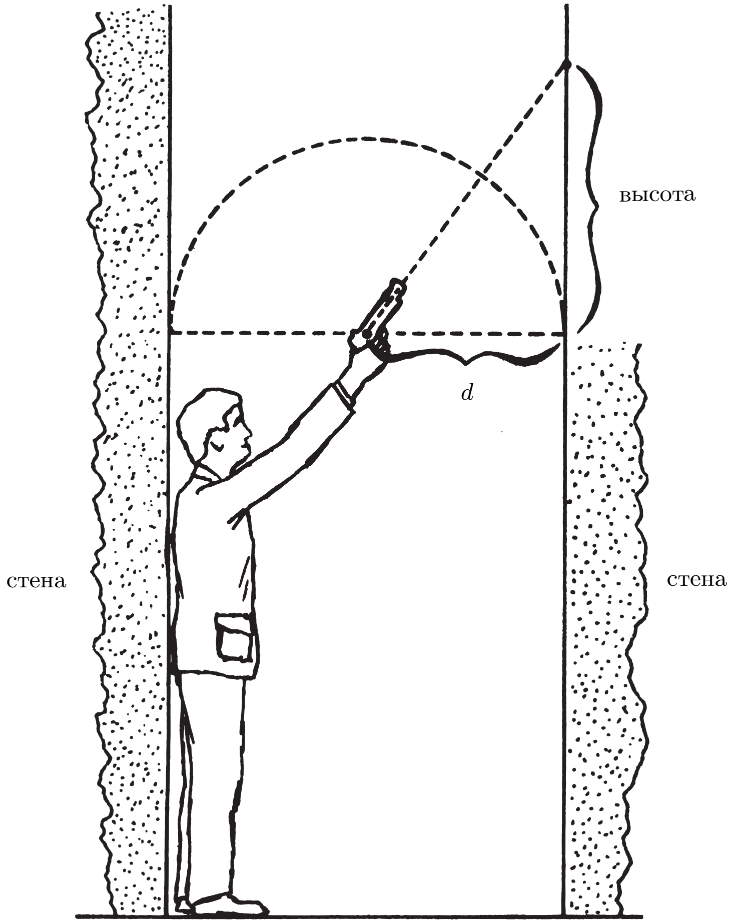


Рис. 23

математическое ожидание $E(2^X)$ является честной ценой, которую он должен заплатить заранее за возможность выигрыша этой случайной суммы, то какова эта цена? Скажите *честно*, на какую цену вы бы сами согласились? (Если вам кажется, что здесь нет парадокса, то вы расходитесь во мнении с такими прославленными математиками, как Д. Бернулли, Ж. Л. Даламбер, С. Д. Пуассон, Э. Борель (мы назвали далеко не всех). Более длинный список приведен в книге [16]. У. Феллер полагает, что парадокс исчезает, если переформулировать задачу на языке более сложной математики. Оставляем за вами право самостоятельно решить, является ли эта задача более интересной с математической или с философской и психологической точек зрения. Тем не менее, см. приложение 3 ниже.)

29. Обычное возражение против схемы из задачи 28 заключается в том, что «время игры должно быть ограничено». Давайте предположим, что допускается не более, чем m бросаний, и ваш товарищ не получает ничего, если решка ни разу не выпала за m бросаний. Рассмотрите случаи $m = 10$ и $m = 100$. Какова теперь справедливая цена за участие в игре? Чувствуете ли вы себя увереннее после такого изменения правил? В данном случае объяснение Феллера теряет свою силу, в то время как психологический элемент остается.

30.* Число μ называется *медианой* распределения случайной величины X, тогда и только тогда, когда $P(X \geqslant \mu) \geqslant 1/2$ и $P(X \leqslant \mu) \geqslant 1/2$. Покажите, что такое число всегда существует, но не обязательно является единственным. Приведем практический пример. После того, как экзаменационные работы были оценены, их расположили в порядке убывания баллов. Тогда одна работа, находящаяся в середине ранжировки, если число работ n — нечетное число, и две работы, если n — четное, соответствует (соответствуют) медиане. Объясните, какая вероятностная модель была здесь использована.

31. Урна содержит n билетов, занумерованных числами от 1 до n. Два билета вынимаются наудачу (без возвращения). Пусть X обозначает меньший, а Y — больший из вынутых номеров. Опишите совместное распределение случайных величин (X, Y), а также их маргинальные распределения. Найдите распределение величины $Y - X$ с помощью совместного распределения.

32. Выберите наудачу два числа из отрезка $[0, 1]$. Определите случайные величины X и Y так же, как в задаче 31, и дайте ответ на тот же вопрос. [*Указание.* Нарисуйте картинку и вычислите площади.]

Приложение 1

Сигма-алгебры и общее определение случайной величины

Когда выборочное пространство Ω несчетно, может оказаться невозможным приписать вероятностную меру всем его подмножествам, как мы поступали в случае счетных пространств в § 2.4. Необходимо ограничиться тем, что мера будет задана только на множествах из некоторого семейства, которое, в свою очередь, должно быть достаточно богатым, чтобы позволять выполнять обычные операции над множествами. Точнее — требуется, чтобы семейство $\mathfrak{F}$ обладало двумя свойствами:

(a) если множество A входит в $\mathfrak{F}$, то его дополнение $A^c = \Omega - A$ также должно принадлежать $\mathfrak{F}$;

(b) если каждое из счетного числа множеств $A_1, A_2, \ldots$ принадлежит $\mathfrak{F}$, то и объединение $\bigcup_n A_n$ обязано входить в $\mathfrak{F}$.

Из законов Де Моргана следует, что объединение в условии (b) можно заменить на пересечение $\bigcap_n A_n$. Таким образом, если производить над членами семейства три указанные операции в любом порядке (см., например, формулу (1.3.1)) счетное число раз, то результат снова будет принадлежать семейству. В этом смысле семейство называют *замкнутым* относительно данных операций, а также относительно других операций, порождаемых ими, скажем, симметричной разности. Такое семейство подмножеств Ω называется *сигма-алгеброй**) на Ω. Вообще говоря, существует много разных сигма-алгебр. Но, например, множество всех подмножеств, которое, очевидно, является сигма-алгеброй, может оказаться слишком большим для того, чтобы на нем удалось определить вероятность. А семейство из двух множеств $\{\varnothing, \Omega\}$, или семейство из четырех множеств $\{\varnothing, A, A^c, \Omega\}$ для заданного множества A — оба в большинстве ситуаций оказываются слишком бедными. Теперь предположим, что выбрана подходящая сигма-алгебра $\mathfrak{F}$, и на ней определена вероятностная мера P. Тогда у нас есть *вероятностная модель* $(\Omega, \mathfrak{F}, P)$,

*) Здесь «сигма» означает «счетный». Знак суммы произошел от заглавной греческой буквы Σ. Обычно в математике «сигма-объект» — понятие, связанное с суммированием счетного числа слагаемых. В оригинале вместо термина «сигма-алгебра» используется устаревший термин «борелевское поле». — *Прим. перев.*

с которой мы можем начать работу. Множества из $\mathfrak{F}$ называются *измеримыми*, как только им приписаны вероятности.

Пусть X — функция на пространстве Ω, принимающая действительные значения. Говорят, что она является *случайной величиной* тогда и только тогда, когда для произвольного действительного числа x множество

$$\{\omega \mid X(\omega) \leqslant x\} \in \mathfrak{F}. \tag{П.1.1}$$

Следовательно, определена вероятность $P(X \leqslant x)$. Как функция от x, она задает функцию распределения F, введенную ранее в формуле (4.6.1). Кроме того, если $a < b$, то множество

$$\{a < X \leqslant b\} = \{X \leqslant b\} - \{X \leqslant a\} \tag{П.1.2}$$

входит в $\mathfrak{F}$, так как сигма-алгебра $\mathfrak{F}$ замкнута относительно разности. А значит, множеству (П.1.2) приписана вероятность, в действительности равная $F(b) - F(a)$.

Когда Ω счетно, и мы берем в качестве $\mathfrak{F}$ множество всех подмножеств Ω, условие (П.1.1), конечно, выполняется для любой функции X. Поэтому в данном случае произвольная функция на Ω является случайной величиной, как и утверждалось в § 4.2. В общем случае условие (П.1.1) необходимо в основном для того, чтобы мы могли определить *математическое ожидание* с помощью опирающейся на него процедуры. Более конкретно: если X — ограниченная случайная величина, то ее математическое ожидание задается следующей формулой:

$$E(X) = \lim_{\delta \to +0} \sum_{n=-\infty}^{\infty} n\,\delta P\{n\,\delta < X \le (n+1)\,\delta\}, \tag{П.1.3}$$

в которой вероятности в сумме корректно определены в силу замечания, относящегося к (П.1.2). Существование предела в формуле (П.1.3) и следующие из этой формулы свойства математического ожидания (помимо обсуждаемых в гл. 5 и 6) составляют часть общей теории, известной как *теория интегрирования Лебега**). Мы отсылаем читателя к стандартным учебникам по данному предмету, вместо того, чтобы вводить математическое ожидание $E(X)$ как интеграл Лебега

$$E(X) = \int\limits_{\Omega} X(\omega) P(d\omega);$$

сравните с дискретным аналогом формулы (4.3.11) в случае счетного Ω.

*) Анри Лебег (1875–1941) — основоположник, совместно с Борелем, современного представления о мере и интегрировании.

Глава 5

Условные вероятности и независимость

5.1. Примеры вычисления условных вероятностей

Мы знаем, что вероятность события A представляет собой отношение суммы весов точек соответствующего множества к весу всего выборочного пространства Ω. Когда Ω конечно, и все точки имеют одинаковый вес (т. е. считаются одинаково правдоподобными), тогда

$$P(A) = \frac{|A|}{|\Omega|}$$

(см. пример 4 из § 2.2). Если Ω счетное, и каждой точке ω приписан вес $P(\omega) = P(\{\omega\})$, то

$$P(A) = \frac{\sum\limits_{\omega \in A} P(\omega)}{\sum\limits_{\omega \in \Omega} P(\omega)} \tag{5.1.1}$$

в силу формулы (2.4.3), поскольку знаменатель здесь равен 1. Во многих задачах интерес представляет пропорциональный вес множества A относительно веса другого множества S. Точнее сказать, отношение веса части A, входящей в S, т. е. $A \cap S$ или AS, к весу S. Формула, аналогичная (5.1.1), выглядит так:

$$\frac{\sum\limits_{\omega \in AS} P(\omega)}{\sum\limits_{\omega \in S} P(\omega)}. \tag{5.1.2}$$

Тем самым мы переключили внимание с пространства Ω на множество S, рассматривая его в качестве нового выборочного пространства («вселенной») вместе с новой мерой — относительной вероятностью. Введем обозначение

$$P(A \mid S) = \frac{P(AS)}{P(S)} \tag{5.1.3}$$

и соответствующую величину назовем *условной вероятностью события A относительно события S*. Могут использоваться и другие слова для описания этого отношения: «при справедливости гипотезы S», «в предположении события S». Конечно, если $P(S) = 0$, то дробь (5.1.3) превращается в бесполезную и не имеющую смысла неопределенность вида $0/0$. Поэтому всегда, когда мы записываем условные вероятности, подобные $P(A \mid S)$, подразумевается выполнение условия $P(S) > 0$, даже если это особо не оговаривается. Обратите внимание, что дробь в формуле (5.1.3) сводится к дроби (5.1.2), когда Ω счетно, но она имеет смысл и в общем случае: необходимо только, чтобы были определены вероятности множеств A и S. Следующие вводные примеры иллюстрируют разнообразные интерпретации и мотивации для использования нового понятия.

Пример 1. Среди студентов, проживающих в студенческом городке, проводится опрос с целью выяснить отношение к определенному кандидату в президенты. Обозначим через D множество студентов, которые его поддерживают. Популяция студентов Ω может быть разбита на разные группы в зависимости от пола, возраста, расы и т. п. Пусть A = женщины, B = чернокожие, C = достигшие избирательного возраста.

Тогда Ω разбивается, как в формуле (1.3.5), на 8 непересекающихся частей ABC, ABC^c, $\ldots$, $A^cB^cC^c$. В результате опроса станут известными количества студентов в них. Множество D, вообще говоря, может иметь непустое пересечение с любой из этих групп. К примеру,

$$P(D \mid A^cBC) = \frac{P(A^cBCD)}{P(A^cBC)}$$

означает долю студентов, поддерживающих данного кандидата, среди чернокожих студентов мужского пола, достигших избирательного возраста;

$$P(D^c \mid A^cC) = \frac{P(A^cCD^c)}{P(A^cC)}$$

означает долю, образуемую студентами, которые не поддерживают данного кандидата, среди юношей, достигших избирательного возраста, и т. п.

Пример 2. Идеальная игральная кость бросается дважды. При условии, что сумма двух выпавших номеров равна 7, найдите вероятность того, что при первом бросании выпала цифра k, $1 \leqslant k \leqslant 6$.

Посмотрите на таблицу в примере 3 из § 4.1. Исходы, для которых сумма равна 7, расположены на «второй главной» диагонали. Всего имеется 6 таких исходов. Среди них есть только один, у которого первая цифра совпадает с k. Поэтому условная вероятность равна $1/6$. Запишем решение с помощью формул. Пусть X_1 и X_2 обозначают, соответственно, номера, выпавшие при первом и втором бросаниях. Тогда, согласно определению (5.3.1), при $A = \{X_1 = k\}$ и $S = \{X_1 + X_2 = 7\}$ запишем:

$$P\{X_1 = k \mid X_1 + X_2 = 7\} = \frac{P\{X_1 = k; X_1 + X_2 = 7\}}{P\{X_1 + X_2 = 7\}} = \frac{1}{6}\,.$$

То, что найденная вероятность оказалась равной *безусловной вероятности* $P\{X_1 = k\}$, является чистой случайностью, связанной с выбором числа 7. Это — единственное значение суммы, при котором допустимы все 6 возможностей для каждого из бросаний. В качестве других примеров имеем:

$$P\{X_1 = k \mid X_1 + X_2 = 6\} = \frac{1}{5}, \quad 1 \leqslant k \leqslant 5,$$

$$P\{X_1 = k \mid X_1 + X_2 = 9\} = \frac{1}{4}, \quad 3 \leqslant k \leqslant 6.$$

Должно быть очевидным, что условная вероятность останется прежней, если поменять местами X_1 и X_2. Почему?

Теперь зададим еще более простой вопрос: «Чему равна условная вероятность события $X_2 = k$ при условии $X_1 = 4$?» Читатель может сразу догадаться, что ответом является 1/6 ввиду того, что первое бросание не оказывает влияние на второе, поэтому условная вероятность $P\{X_2 = k \mid X_1 = 4\}$ должна совпадать с безусловной $P\{X_2 = k\}$. Такое рассуждение, конечно, верно, однако оно использует *независимость* двух испытаний (см. § 2.4). Можно непосредственно получить ответ, опираясь на формулу (5.1.3):

$$P\{X_2 = k \mid X_1 = 4\} = \frac{P\{X_1 = 4; X_2 = k\}}{P\{X_1 = 4\}} = \frac{1/36}{1/6} = \frac{1}{6}. \tag{5.1.4}$$

Наконец, запишем:

$$P\{X_1 + X_2 = 7 \mid X_1 = 4\} = \frac{P\{X_1 = 4; X_1 + X_2 = 7\}}{P\{X_1 = 4\}}. \tag{5.1.5}$$

Даже не глядя на таблицу исходов, понятно, что событие

$$\{X_1 + X_2 = 7;\ X_1 = 4\}$$

совпадает с

$$\{X_1 = 4;\ X_2 = 7 - 4 = 3\}.$$

Следовательно, в действительности равенство (5.1.5) есть частный случай соотношения (5.1.4). Проведенное рассуждение может показаться в данной ситуации слишком окольным, однако оно достаточно характерно для *случайных блужданий*, изучаемых в гл. 8.

Пример 3. Рассмотрим время ожидания X в примере 8 из § 4.4 для случая несимметричной монеты. При условии, что в первых трех бросаниях выпали орлы, чему равна вероятность того, что при двух следующих бросаниях хотя бы один раз монета выпадет решкой?

Таким образом, нас интересует условная вероятность

$$P(X \leqslant 5 \mid X \geqslant 4) = \frac{P(4 \leqslant X \leqslant 5)}{P(X \geqslant 4)}. \tag{5.1.6}$$

Мы знаем, что

$$P(X = n) = q^{n-1}p, \quad n = 1, 2, \ldots; \tag{5.1.7}$$

с учетом этого вычисляем:

$$P(X \geqslant 4) = \sum_{n=4}^{\infty} q^{n-1}p = \frac{q^3 p}{1-q} = q^3 \tag{5.1.8}$$

(каким образом была найдена сумма ряда?). Опять используя вероятности (5.1.7), получаем:

$$P(4 \leqslant X \leqslant 5) = q^3 p + q^4 p.$$

Итак, вероятность (5.1.6) равна $p + qp$. Далее, снова применяя (5.1.7), подсчитываем вероятность выпадения решки по крайней мере в одном их двух испытаний:

$$P(1 \leqslant X \leqslant 2) = p + qp.$$

Сравнивая полученные результаты, заключаем, что три предшествующих «неудачи» не влияют на длину последующего времени ожидания «успеха». Этот вывод, возможно, был очевидным для вас априори, но на самом деле он вытекает из независимости последовательных испытаний. Между прочим, многие бывалые игроки в рулетку полагают, что «если шарик останавливался на «красном» много раз подряд, то правильно будет при следующем вращении поставить на «черное», поскольку «красное» и «черное» в длинной серии испытаний должны быть сбалансированы». Тем не менее, читатель может возразить (имея на своей стороне лорда Кейнса[*]), что если «красное» выпадало, скажем, 10 раз подряд, в отсутствие других свидетельств естественно предположить, что либо из-за перекоса рулетки, либо из-за жульничества крупье «красное» появляется чаще, чем должно. Другими словами, $p > 1/2$ в рассматриваемой выше модели. Следовательно, разумно ставить на «красное». В примере 8 в § 5.2 содержатся похожие рассуждения.

Пример 4. Проведем аналогию между геометрическим распределением, задаваемым вероятностями (5.1.7) (см. также (4.4.8)) и экспоненциальным законом, определяемым формулой (4.5.11). Если случайная величина X имеет первое из распределений, то для любого натурального n

$$P(X > n) = q^n. \tag{5.1.9}$$

[*] Джон Мейнард Кейнс (1883–1946) — английский экономист и писатель.

Это легко доказать суммированием геометрической прогрессии, подобно тому, как это было сделано в формуле (5.1.8). Данное равенство становится очевидным, если вспомнить, что «$X > n$» равносильно событию «во всех первых n испытаниях монета выпадала орлом». Далее, из равенства (5.1.9) выводим для любых натуральных m и n, что

$$P(X > n+m \mid X > m) = \frac{P(X > n+m)}{P(X > m)} = \frac{q^{m+n}}{q^m} = \\ = q^n = P(X > n). \tag{5.1.10}$$

Обозначим через T время ожидания в примере 12 из § 4.5; тогда аналогично имеем для произвольных неотрицательных действительных чисел s и t:

$$P(T > s+t \mid T > s) = \frac{P(T > s+t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \\ = e^{-\lambda t} = P(T > t). \tag{5.1.11}$$

Смысл этой формулы таков: несмотря на то, что мы уже потратили некоторое время на ожидание, распределение оставшегося времени ожидания совпадает с безусловным распределением, как будто прошедшее время не играет никакой роли! Образно это можно назвать *отсутствием памяти* у случайной величины T. Таким свойством обладает только экспоненциальное распределение. Оно лежит в основе теории марковских процессов. Обратите внимание, что хотя геометрическое распределение служит, согласно соотношению (5.1.10), дискретным аналогом экспоненциального закона, строго говоря, оно не имеет свойства «отсутствия памяти». Действительно, равенство (5.1.10) не выполняется, когда m и n не являются натуральными: возьмите, скажем, $m = n = 1/2$.

Пример 5. Рассмотрим семьи, имеющие двух детей, и предположим, что рождение мальчика и девочки одинаково вероятно. Тогда выборочное пространство может быть схематично представлено четырьмя точками:

$$\Omega = \{(bb), (bg), (gb), (gg)\},$$

где b = мальчик, g = девочка[*)]. Порядок в паре отражает очередность рождения. Каждая из точек имеет вероятность $1/4$. Конечно, можно вместо такого Ω использовать пространство, состоящее из $4N$ точек, где N — большое число, в котором данным четырем возможностям отвечают равные количества представителей. Оно более реалистично

[*)] Соответственно английским словам «boy» и «girl». — *Прим. перев.*

отражало бы модель популяции, однако приводимая ниже арифметика будет точно такой же.

Если известно, что в выбранной наудачу из Ω семье есть мальчик, то какова вероятность того, что и второй ребенок — мальчик, т. е. семья имеет тип (bb)? Возможно, читатель быстро даст ответ $1/2$, опираясь на предполагаемую равную вероятность любого пола ребенка. Тем самым, он допустит ошибку, причина которой заключается в неправильном понимании смысла «подчиненного предложения» в вопросе. Приведем подробное объяснение.

Введем обозначения:

$$A = \{\omega \mid \text{в } \omega \text{ есть мальчик}\},$$
$$B = \{\omega \mid \text{в } \omega \text{ — два мальчика}\}.$$

Тогда $B \subset A$, и поэтому $AB = B$. Отсюда

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

Это и есть правильный ответ. Но давайте теперь зададим похожий, но, в действительности, отличающийся вопрос: «При условии, что выбранный наудачу из этих семей ребенок является мальчиком, чему равна вероятность того, что и другой ребенок в этой семье — мальчик?» На этот раз в качестве подходящего представления выборочного пространства следует взять

$$\tilde{\Omega} = \{g_g, g_b, b_g, b_b\},$$

где точками пространства являются уже не сами семьи, а дети из семей. Здесь g_g = у девочки есть сестра, g_b = у девочки есть брат и т. п. Введем обозначения:

$$\tilde{A} = \{\tilde{\omega} \mid \tilde{\omega} \text{ — мальчик}\},$$
$$\tilde{B} = \{\tilde{\omega} \mid \text{у } \tilde{\omega} \text{ есть брат}\}.$$

Поэтому

$$\tilde{A}\tilde{B} = \{\tilde{\omega} \mid \tilde{\omega} = b_b\}.$$

Следовательно,

$$P(\tilde{B} \mid \tilde{A}) = \frac{P(\tilde{A}\tilde{B})}{P(\tilde{A})} = \frac{1/4}{1/2} = \frac{1}{2}\,.$$

Этот пример служит замечательной и ни в коей мере не искусственной иллюстрацией важности понимания того, *что именно является случайной выборкой* в статистическом исследовании.

5.2. Основные формулы

Вообще говоря, большинство вероятностных задач имеет дело с несколькими событиями или случайными величинами. При этом необходимо исследовать их *взаимосвязи* или *совместное поведение.* В некотором смысле все вероятности являются условными потому, что ничего не происходит в вакууме. Как правило, не оговариваются неявные или сами собой разумеющиеся условия, или же мы опускаем условия, когда чувствуем, что они не имеют прямого отношения к нашей ситуации. Например, когда подбрасывается монета, обычно не рассматривается возможность того, что она останется стоять на ребре, а также не уточняется, является ли она американской или канадской. Вероятность победы на выборах определенного кандидата в президенты, конечно, обусловлена тем, что он доживет до конца избирательной компании — предположение, оказавшееся преждевременным в недавней американской истории.

Давайте начнем с нескольких простых, но важных утверждений, включающих в себя условные вероятности.

Утверждение 1. *Для произвольных событий $A_1, A_2, \ldots, A_n$ верно равенство*

$$\begin{gathered} P(A_1A_2\ldots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2) \times \ldots \\ \ldots \times P(A_n \mid A_1A_2\ldots A_{n-1}), \end{gathered} \tag{5.2.1}$$

в предположении, что $P(A_1A_2\ldots A_n) > 0$.

Доказательство. В силу предположения все условные вероятности в формуле (5.2.1) корректно определены, поскольку

$$P(A_1) \geqslant P(A_1A_2) \geqslant \ldots \geqslant P(A_1A_2\ldots A_{n-1}) > 0.$$

Тогда правая часть соотношения (5.2.1) может быть представлена в форме

$$\frac{P(A_1)}{P(\Omega)}\,\frac{P(A_1A_2)}{P(A_1)}\,\frac{P(A_1A_2A_3)}{P(A_1A_2)}\cdots\frac{P(A_1A_2\ldots A_n)}{P(A_1A_2\ldots A_{n-1})},$$

где остается только провести перекрестные сокращения.

Из сопоставления с формулой сложения (2.3.3), справедливой для объединения несовместных событий, равенство (5.2.1) можно назвать *общей формулой умножения* для вероятности пересечения. Обратите внимание, как условные вероятности перемножаются шаг за шагом. Для независимых событий в § 2.4 приведено значительно более простое выражение. В качестве важного применения формулы (5.2.1) рассмотрим ситуацию, когда все случайные величины $X_1, X_2, \ldots, X_n, \ldots$ принимают счетное множество значений; само Ω также предполагается счетным. Тогда для произвольных возможных значений $x_1, x_2, \ldots, x_n, \ldots$, положив

$$A_k = \{X_k = x_k\}, \quad k = 1, 2, \ldots,$$

получим

$$\begin{aligned} P\{X_1 = x_1; X_2 = x_2; \ldots; X_n = x_n\} = \qquad\qquad (5.2.2)\\ = P\{X_1 = x_1\} P\{X_2 = x_2 \mid X_1 = x_1\} P\{X_3 = x_3 \mid X_1 = x_1, X_2 = x_2\} \times \ldots \\ \ldots \times P\{X_n = x_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}\}. \end{aligned}$$

Левую часть называют *совместной вероятностью* величин $X_1, X_2, \ldots, X_n$. Приведенная формула выражает ее через последовательные условные вероятности. Позже будет обсуждаться частный случай этого равенства.

Утверждение 2. *Пусть*

$$\Omega = \sum_n A_n$$

является разбиением выборочного пространства на непересекающиеся части. Тогда для любого множества B справедлива формула

$$P(B) = \sum_n P(A_n) P(B \mid A_n). \qquad (5.2.3)$$

Доказательство. Прежде всего запишем представление

$$B = \Omega B = \left(\sum_n A_n\right) B = \sum_n A_n B,$$

верное в силу простейшей теории множеств, а именно — формулы (1.3.6). Далее выводим равенство

$$P(B) = P\left(\sum_n A_n B\right) = \sum_n P(A_n B),$$

справедливое в силу счетной аддитивности P. Осталось подставить выражение

$$P(A_n B) = P(A_n) P(B \mid A_n),$$

вытекающее из определения (5.1.3). Формула (5.2.3) установлена. Отметьте, что если $P(A_n) = 0$ при некотором n, соответствующий член суммы можно считать нулем, несмотря на то что вероятность $P(B \mid A_n)$ не определена. Доказательство закончено.

На будущее примем соглашение, что $x \cdot 0 = 0$, если x не определен, для того, чтобы избежать впредь подобных замечаний.

Равенство (5.2.3) называют *формулой полной вероятности.* Вот ее полезная интерпретация. Допустим, что событие B может произойти вследствие ряда взаимно исключающих обстоятельств (или причин). Тогда эта формула показывает, каким образом «полная вероятность» события составляется из вероятностей различных обстоятельств и соответствующих условных вероятностей, вычисленных в предположении таких гипотез.

Пусть X и Y — две целочисленные случайные величины, k — целое число. Если применить формулу (5.2.3) к множествам

$$A_n = \{X = n\}, B = \{Y = k\},$$

то получим формулу

$$P(Y = k) = \sum_n P(X = n)P(Y = k \mid X = n), \tag{5.2.4}$$

где суммирование распространяется на все целые числа n, и если $P(X = n) = 0$, то соответствующий член полагается равным 0. Нетрудно обобщить эту формулу на случай, когда X принимает значения из произвольного счетного множества и когда событие «$Y = k$» заменяется, например, на «$a \leqslant Y \leqslant b$» для случайной величины более общего вида, не обязательно принимающей целые значения.

Утверждение 3. *При выполнении допущений из утверждения 2 в используемых там обозначениях имеем:*

$$P(A_n \mid B) = \frac{P(A_n)P(B \mid A_n)}{\sum\limits_n P(A_n)P(B \mid A_n)} \tag{5.2.5}$$

при условии, что $P(B) > 0$.

Доказательство. Знаменатель в формуле (5.2.5) равен $P(B)$ в силу утверждения 2. Поэтому после умножения на него равенство принимает вид

$$P(B)P(A_n \mid B) = P(A_n)P(B \mid A_n).$$

Полученное равенство верно, так как обе его части совпадают с $P(A_nB)$. Доказательство закончено.

Это простое утверждение с легким доказательством, широко известное под именем *теоремы Байеса*, было опубликовано в 1763 г. Оно предназначено для вычисления «перевернутой» вероятности или «вероятности причины» A_n на основе наблюдаемого «эффекта» B. В то время, как вероятность $P(A_n)$ является априорной, вероятность $P(A_n \mid B)$ называют апостериорной. Многочисленные применения этой формулы встречаются во всех областях, изучающих природные явления и человеческое поведение. К примеру, если B обозначает «труп» и A_n суть несколько подозреваемых в убийстве, то теорема может помочь присяжным решить «кто сделал это». (В действительности юриспруденция первоначально была одной из тех областей, в которых весьма широко использовались вероятностные методы.) Если B — это землетрясение, а A_n — разные физические теории, предназначенные для его объяснения, то теорема может помочь ученым произвести выбор между ними. Лаплас*) применял теорему для оценки вероятности того, что «завтра взойдет солнце» (см. пример 9 ниже). В наше время именем Байеса называется целое направление (научная школа) в статистике. Здесь мы только отметим, что Байес безусловно обнаружил замечательную формулу для вычисления «перевернутых» вероятностей, однако ее практическая ценность ограничена тем, что обычно нам не известны априорные вероятности.

Следующие простые примеры предназначены для иллюстрации трех приведенных выше утверждений. Другие утверждения появятся позже по ходу изложения.

Пример 6. Нам уже встречались несколько примеров применения утверждения 1 прежде, в гл. 3. Давайте разберем их еще раз, используя новые понятия.

Чему равна вероятность того, что при бросании шести игральных костей выпадут шесть разных граней? (См. пример 2 в § 3.1.) Грани занумерованы цифрами от 1 до 6. Положим

A_1 = «на первой кости выпадает любой из шести номеров»,

A_2 = «номер на второй кости отличается от выпавшего на первой»,

A_3 = «номер на третьей кости отличен от выпавших на первой и второй»,

и т. д. Тогда в предположении независимости бросаний имеем:

$$P(A_1) = 1, \quad P(A_2 \mid A_1) = \frac{5}{6}, \quad P(A_3 \mid A_1A_2) = \frac{4}{6}, \ldots,$$

$$P(A_6 \mid A_1A_2 \ldots A_5) = \frac{1}{6}.$$

Отсюда, применяя утверждение 1, получаем:

$$P(A_1A_2 \ldots A_6) = \frac{6}{6} \cdot \frac{5}{6} \cdot \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{1}{6} = \frac{6!}{6^6}.$$

*) Пьер Симон Лаплас (1749–1827) — один из величайших математиков всех времен, написавший фундаментальный труд по теории вероятностей около 1815 г.

Задача о днях рождения (задача 5 в § 3.4) теперь представляется практически идентичной только что разобранной, с заменой числа 6 на 365. Обсуждаемый там последовательный метод — просто другая формулировка утверждения 1.

Пример 7. Когда семейной пикник в парке закончился, обнаружилось, что пропала собака. Рассматриваются три гипотезы:

$A =$ «она вернулась домой»,

$B =$ «она по-прежнему грызет кость недалеко от места пикника»,

$C =$ «она убежала из парка и заблудилась в лесу».

Априорные вероятности с учетом привычек собаки оцениваются, соответственно, как $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$. Двух детей отправили на поиски. Одного — на место пикника, другого — на опушку ближайшего леса. Если собака находится вблизи места пикника, то почти наверняка (90 %) она будет найдена. Если она убежала в лес, то шансы обнаружения составляют только 50 % (как при бросании монетки). Какова вероятность того, что собака будет найдена в парке или в лесу?

Здесь события A, B, C являются гипотезами. Положим $D =$ «собака будет найдена в парке или в лесу». Тогда имеем следующие данные:

$$\begin{aligned} P(A) &= \frac{1}{4}, & P(B) &= \frac{1}{2}, & P(C) &= \frac{1}{4}; \\ P(D \mid A) &= 0, & P(D \mid B) &= \frac{90}{100}, & P(D \mid C) &= \frac{50}{100}. \end{aligned}$$

Следовательно, согласно формуле (5.2.3),

$$\begin{aligned} P(D) &= P(A)P(D \mid A) + P(B)P(D \mid B) + P(C)P(D \mid C) = \\ &= \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot \frac{90}{100} + \frac{1}{4} \cdot \frac{50}{100} = \frac{115}{200}. \end{aligned}$$

Какова вероятность того, что собака будет обнаружена дома? Обозначим это событие через D' и предположим, что $P(D' \mid A) = 1$, т. е. если собака побежала домой, то она находится там и будет встречать семью. Ясно, что $P(D' \mid B) = 0$ и $P(D' \mid \mathrm{C}) = 0$. Тогда

$$\begin{aligned} P(D') &= P(A)P(D' \mid A) + P(B)P(D' \mid B) + P(C)P(D' \mid C) = \\ &= \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot 0 + \frac{1}{4} \cdot 0 = \frac{1}{4}. \end{aligned}$$

Какова вероятность того, что собака потеряется? Она равна

$$1 - P(D) - P(D') = \frac{35}{200}.$$

Пример 8. Одна из урн содержит 2 черных и 3 красных шара, а другая — 3 черных и 2 красных шара. Для выбора урны, из которой будут извлекаться шары, бросается симметричная монетка. При этом неизвестно, какая из них какая (где именно больше черных шаров). Предположим, что первым был извлечен черный шар. Затем его положили обратно в урну. Чему равна вероятность того, что второй шар, вынутый из той же самой урны, также окажется черным?

Обозначим урны через U_1 и U_2. Априорные вероятности выбора каждой из них в результате подбрасывания монеты равны 1/2:

$$P(U_1) = \frac{1}{2}, \quad P(U_2) = \frac{1}{2}.$$

Пусть B_1 — событие, состоящее в том, что первый вынутый шар оказался черным, B_2 — событие, означающее, что второй вынутый шар оказался черным. Согласно формуле (5.2.5),

$$P(U_1 \mid B_1) = \frac{1/2 \cdot 2/5}{(1/2 \cdot 2/5) + (1/2 \cdot 3/5)} = \frac{2}{5}, \quad P(U_2 \mid B_1) = \frac{3}{5}.$$

Заметьте, что сумма вероятностей должна равняться 1 (почему?). Ввиду этого, достаточно вычислить только одну из них. Заметьте также, что апостериорные вероятности прямо пропорциональны вероятностям $P(B_1 \mid U_1)$ и $P(B_1 \mid U_2)$. Другими словами, более вероятно, что черный шар был извлечен из урны, которая содержит больше черных шаров, причем вероятности пропорциональны долям черных шаров. Теперь давайте применим формулу (5.2.3) для подсчета вероятности того, что второй извлеченный шар также окажется черным. Здесь A_1 = «B_1 из U_1» и A_2 = «B_1 из U_2» — две альтернативных гипотезы. Так как второе извлечение обусловлено событием B_1, на самом деле вероятности гипотез являются условными:

$$P(A_1) = P(U_1 \mid B_1) = \frac{2}{5}, \quad P(A_2) = P(U_2 \mid B_1) = \frac{3}{5}.$$

С другой стороны, очевидно, что

$$P(B_2 \mid A_1) = \frac{2}{5}, \quad P(B_2 \mid A_2) = \frac{3}{5}.$$

Отсюда находим условную вероятность

$$P(B_2 \mid B_1) = \frac{2}{5} \cdot \frac{2}{5} + \frac{3}{5} \cdot \frac{3}{5} = \frac{13}{25}.$$

Сравним ее с вероятностью

$$P(B_2) = P(U_1)P(B_2 \mid U_1) + P(U_2)P(B_2 \mid U_2) = \frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} = \frac{1}{2}.$$

Заметьте, что $P(B_2 \mid U_1) = P(B_1 \mid U_1)$ (почему?). Мы видим, что знание, что первый извлеченный шар оказался черным, увеличивает вероятность того, что второй шар также будет черным, поскольку увеличивается правдоподобие того, что была выбрана урна с бо́льшим числом черных шаров. Сделаем еще один шаг: при условии, что первые два шара были черного цвета, подсчитаем вероятность того, что и третий шар, извлеченный из той же урны, окажется черным. Аналогично проведенным выше рассуждениям, имеем:

$$P(U_1 \mid B_1B_2) = \frac{1/2 \cdot (2/5)^2}{1/2 \cdot (2/5)^2 + 1/2 \cdot (3/5)^2} = \frac{4}{13}; \quad P(U_2 \mid B_1B_2) = \frac{9}{13};$$

$$P(B_3 \mid B_1B_2) = \frac{4}{13} \cdot \frac{2}{5} + \frac{9}{13} \cdot \frac{3}{5} = \frac{35}{65}.$$

Последняя вероятность превосходит $13/25$, продолжая тем самым процесс роста. Нетрудно убедиться, что можно обобщить результат на любое количество извлечений. При этом

$$P(U_1 \mid B_1B_2 \dots B_n) = \frac{1/2 \cdot (2/5)^n}{1/2 \cdot (2/5)^n + 1/2 \cdot (3/5)^n} = \frac{1}{1 + (3/2)^n},$$

где мы поделили знаменатель на числитель при получении второго равенства. Отсюда следует, что когда n увеличивается, апостериорная вероятность события U_1 становится все меньше. На самом деле, она убывает к нулю, и, как следствие, апостериорная вероятность события U_2 стремится в пределе к 1. Также мы имеем, что

$$\lim_{n\to\infty} P(B_{n+1} \mid B_1B_2 \dots B_n) = \frac{3}{5} = P(B_1 \mid U_2).$$

Этот простой пример имеет важное значение с точки зрения эмпирического понимания вероятности. Заменим наши две урны на монеты, вообще говоря, несимметричные (как все монеты в реальности). Допустим, что вероятность p выпадения орла равна либо $2/5$, либо $3/5$, но мы не знаем, какому именно из двух данных значений. Эти две возможности представляют собой две альтернативные гипотезы, одну из которых необходимо выбрать. Если они имеют априорные вероятности $1/2$, то мы оказываемся в ситуации двух урн. Результат каждого бросания монеты влияет на нашу эмпирическую оценку значения p. Допустим, что по каким-то причинам мы полагаем, что $p = 2/5$. Тогда в случае, если монета выпадет орлом 10 раз подряд, разумно ли по-прежнему считать, что $p = 2/5$, что приводит к вероятности $(2/5)^{10}$ для столь редкого события? Или нам следует заключить, что на самом деле $p = 3/5$, и наблюдаемое событие имеет вероятность $(3/5)^{10}$? Она крайне мала, но все же в $(3/2)^{10}$ раз больше, чем предыдущая. Обычно в задачах, относящихся к теории вероятностей, величина p считается фиксированной, и остальные вычисления базируются на этом предположении. Вопрос заключается в том, какие у нас имеются причины придерживаться представления о неизменности величины p перед лицом опровергающих свидетельств, наблюдаемых в эксперименте? Кейнс строил на данном вопросе свою критику оснований теории вероятностей. С точки зрения аксиоматического подхода, используемого в этой книге, ответ прост: наши формулы верны для каждого произвольного значения p, но на основании аксиом мы, конечно, не можем судить о его истинности или о том, есть ли вообще смысл приписывать исходам какие-либо вероятности. Последнее имеет место, когда речь идет о вероятности обнаружения «больших живых существ где-то в космическом пространстве». (Раньше бы сказали — *на луне*!) Другими словами, характерная особенность математики заключается в том, что она является дедуктивной наукой. Проблемы оценивания или тестирования гипотез об истинности значения p лежат за пределами ее основной области интересов. Конечно, они весьма важны для практики, поэтому для изучения таких проблем и была создана *математическая статистика*. Но здесь мы не станем уделять им много внимания. (Один из авторов книги запомнил мнение Альберта Эйнштейна, высказанное им во время прогулки по Мерсер стрит в Принстоне, на которой он случайно присутствовал, примерно в 1946 или 1947 гг. Вот его суть: в любом направлении науки, имеющем приложения, всегда есть разрыв между теорией и практикой, через который необходимо перебросить мост. Так обстоит дело, например, в геометрии и механике; и теория вероятностей — не исключение.)

Предыдущий пример допускает естественное обобщение на случай, когда p может принимать значения из некоторого интервала. Пожалуй, наиболее известной иллюстрацией служит *задача Лапласа о последовательных восходах солнца*.

Пример 9. Пусть солнце всходило n дней подряд. Какова вероятность того, что оно взойдет еще раз?

Предполагается, что априорная вероятность восхода солнца в любой из дней является неизвестной нам постоянной величиной. Ввиду абсолютного отсутствия предпочтений будем считать, что любое из значений внутри отрезка [0,1] для нее одинаково правдоподобно. Другими словами, данная вероятность рассматривается как случайная величина ξ, равномерно распределенная на отрезке [0,1]. Поэтому плотностью для ξ служит функция $f(p) = 1$ при $0 \leqslant p \leqslant 1$. Это можно записать эвристически как

$$P(p \leqslant \xi \leqslant p + dp) = dp, \quad 0 \leqslant p \leqslant 1; \tag{5.2.6}$$

см. обсуждение в примере 10 из § 4.5. Далее, если истинным значением случайной величины ξ является p, то при этом предположении вероятность наблюдать n последовательных восходов равна p^n, так как они считаются независимыми событиями. Пусть S^n обозначает событие «солнце всходило n раз подряд». Тогда эвристически мы можем записать

$$P(S^n \mid \xi = p) = p^n. \tag{5.2.7}$$

Аналог формулы (5.2.3) должен тогда иметь вид

$$P(S^n) = \sum_{0 \leqslant p \leqslant 1} P(\xi = p)\, P(S^n \mid \xi = p). \tag{5.2.8}$$

Конечно, строго говоря, эта запись не имеет смысла, но если заменить сумму на интеграл и применить равенство (5.2.6), то в результате получим

$$P(S^n) = \int\limits_0^1 P(S^n \mid \xi = p)\, dp = \int\limits_0^1 p^n\, dp = \frac{1}{n+1}\,. \tag{5.2.9}$$

Эта непрерывная версия формулы (5.2.3) на самом деле верна, несмотря на нестрогость приведших к ней рассуждений. Считая данную формулу справедливой, применим ее для n и $n+1$ и, взяв отношение, найдем, что

$$P(S^{n+1} \mid S^n) = \frac{P(S^n S^{n+1})}{P(S^n)} = \frac{P(S^{n+1})}{P(S^n)} = \frac{1/(n+2)}{1/(n+1)} = \frac{n+1}{n+2}\,. \tag{5.2.10}$$

Это и есть ответ Лапласа в задаче о восходах солнца.

Говоря современным языком, Лаплас использовал «урновую модель» для изучения последовательности восходов как случайного процесса. Восход ассоциируется с извлечением черного шара из урны с неизвестным составом. Разнообразные возможные составы ассоциируются со множеством урн, содержащих черные шары в любой из мыслимых пропорций. Наконец, выбор истинного значения пропорции ассоциируется с выбором наудачу точки из [0,1]. Ясно, что перечисленные предположения являются достаточно весомыми. Они вызывают серьезные возражения на разных уровнях. Является ли восход солнца случайным явлением или он детерминирован? Даже если его можно рассматривать как случайное явление, то будет ли он адекватно описываться нашей простой урновой моделью? При допущении, что данная модель в принципе соответствует действительности, остается вопрос, почему априорное распределение истинной вероятности обязано быть равномерным? И как вообще можно установить его вид?

Оставляя эти вопросы в стороне, давайте на минуту вернемся к формуле (5.2.7). Поскольку $P(\xi = p) = 0$ для любого p (см. § 4.5, где приведено соответствующее рассуждение), то формула (5.1.3) *не может* служить определением для условной вероятности (5.2.7). Однако она имеет смысл, изложенный в интерпретации, приведенной перед формулой (5.2.7). В действительности ее можно совершенно корректно определить с помощью более сложного аппарата (производной Радона—Никодима). Если это сделать, то итоговая формула (5.2.9) выводится без применения эвристического равенства (5.2.8). Хотя полное объяснение данных вопросов выходит за уровень этой книги, представляется правильным отметить их здесь в качестве естественного обобщения понятия условной вероятности. Чисто дискретный подход к формуле Лапласа также возможен, но тогда приходится проводить более трудоемкие вычисления (см. задачу 35 ниже).

В завершение текущего параграфа введем понятие условного математического ожидания. Пусть на счетном вероятностном пространстве задана случайная величина Y, принимающая значения $\{y_k\}$. Рассмотрим событие S такое, что $P(S) > 0$. Предположим, что математическое ожидание случайной величины Y существует. Тогда *условное математическое ожидание относительно события* S определяется равенством

$$E(Y \mid S) = \sum_k y_k P(Y = y_k \mid S). \tag{5.2.11}$$

Таким образом, мы в формуле $E(Y) = \sum\limits_k y_k P(Y = y_k)$ просто заменили обычные вероятности на условные. Ряд в соотношении (5.2.11) сходится абсолютно в силу абсолютной сходимости ряда, определяющего $E(Y)$. В частности, если X — другая случайная величина, принимающая значения $\{x_j\}$, то можно взять $S = \{X = x_j\}$ и вычислить $E(Y \mid X = x_j)$. С другой стороны, аналогично формуле (5.2.4) имеем:

$$P(Y = y_k) = \sum_j P(X = x_j)\, P(Y = y_k \mid X = x_j).$$

Умножая слагаемые на y_k, суммируя по k и меняя порядок суммирования, получаем формулу

$$E(Y) = \sum_j P(X = x_j)\, E(Y \mid X = x_j). \tag{5.2.12}$$

Перемена порядка суммирования законна в силу абсолютной сходимости.

Следующие два параграфа содержат довольно специальный материал. Читатель может ознакомиться с начальными частями § 5.3 и § 5.4

вплоть до формулировок теорем 1 и 3 соответственно, чтобы узнать, о чем идет речь. Остальное можно отложить и перейти к § 5.5.

★5.3. Последовательный выбор

В этом параграфе мы продолжим детальное изучение урновой модели. Она относится к простейшим схемам, с которыми удается работать, используя элементарные методы. Тем не менее, в этой модели находят свое представление идеи, важные как для теории, так и для практики, включая аппарат условных ожиданий.

Некая урна содержит b черных и r красных шаров. Шары извлекаются по одному без возвращения. Пусть $X_n = 1$ или 0 в зависимости от того, был n-й вытащенный шар черным или красным. Каждая точка ω выборочного пространства представляется в виде последовательности

$$\{X_1(\omega), X_2(\omega), \ldots, X_{b+r}(\omega)\}, \quad \text{кратко} \quad \{X_n, 1 \leqslant n \leqslant b+r\},$$

см. обсуждение формулы (4.1.3). Данная последовательность называется *стохастическим процессом* (это модное наименование используется для произвольного семейства случайных величин). (Согласно словарю, слово «stochastic» пришло из греческого языка и означает «умеющий угадывать».) В нашем случае семейством является конечная последовательность, индексированная переменной n, меняющейся от 1 до $b+r$. Индекс n можно интерпретировать как *временной параметр*, если представить, что извлечения шаров происходят через единичные временные интервалы. В таком случае, наблюдая последовательные значения величин X_n, говорят об эволюции процесса во времени.

Возможно, вы полагаете, что наша модель есть не что иное, как схема выбора без возвращения, но с упорядочением, обсуждаемая в § 3.2. Вы правы, но дело в том, что сейчас изменилась наша точка зрения, и приведенное выше детальное описание модели предназначено для того, чтобы показать это. Теперь нас интересует, например, не только какое количество черных шаров было получено за определенное число извлечений, как было раньше, но мы также хотим знать, как последовательные извлечения влияют друг на друга, как со временем меняется состав урны и т. п. Другими словами, мы собираемся исследовать взаимную зависимость величин X_n, а для этого нам понадобятся условные вероятности. Давайте начнем с наиболее простых вопросов.

Задача. Шар извлекается из урны и откладывается в сторону. Если считать, что мы не знаем его цвет, то какова вероятность того, что второй вытащенный шар будет черным?

Для краткости обозначим событие $\{X_n = 1\}$ через B_n, а событие $\{X_n = 0\}$ через $R_n = B_n^c$. Тогда, в силу утверждения 2 из § 5.2,

$$P(B_2) = P(B_1)P(B_2 \mid B_1) + P(B_1^c)P(B_2 \mid B_1^c). \tag{5.3.1}$$

Очевидно, что

$$P(B_1) = \frac{b}{b+r}, \quad P(B_1^c) = \frac{r}{b+r}, \tag{5.3.2}$$

в то время, как

$$P(B_2 \mid B_1) = \frac{b-1}{b+r-1}, \quad P(B_1 \mid B_1^c) = \frac{b}{b+r-1},$$

ввиду того что после первого извлечения в урне останется $b+r-1$ шаров, причем среди них черных будет либо $b-1$, либо b в зависимости от того, был ли первый шар черным или нет. Подставляя выписанные формулы в сумму (5.3.1), находим, что

$$P(B_2) = \frac{b}{b+r}\,\frac{b-1}{b+r-1} + \frac{r}{b+r}\,\frac{b}{b+r-1} = \frac{b(b+r-1)}{(b+r)(b+r-1)} = \frac{b}{b+r}.$$

Таким образом, $P(B_2) = P(B_1)$. Другими словами, если принять во внимание обе возможности для цвета первого шара, то вероятности цветов второго шара точно такие же, как если бы вообще не было извлечения. Удивительно это или нет? Любопытные люди захотят узнать, является ли данное совпадение случайным или за ним стоит некоторая теория. Простейший способ проверить это — попытаться сделать еще один или два шага: допустим 2 или 3 шара были извлечены из урны, и их цвета неизвестны, какова вероятность того, что следующий шар будет черным? Нет сомнений, что читатель сможет выполнить необходимые для проверки вычисления. Общий результат кратко формулируется следующим образом.

Теорема 1. *Для каждого* n

$$P(B_n) = \frac{b}{b+r}, \quad 1 \leqslant n \leqslant b+r. \tag{5.3.3}$$

Давайте сделаем здесь паузу и отметим экономичность данной математической формулировки по сравнению с многословным словесным описанием, приведенным выше. *Условие*, что «мы не знаем» цвета $n-1$ предварительно извлеченных шаров, здесь подразумевается по умолчанию и формально выражается в *отсутствии условия* у вероятности $P(B_n)$. Что было бы, если бы мы знали цвета? Тогда речь шла бы о чем-то вроде $P(B_2 \mid B_1)$ или $P(B_3 \mid B_1B_2^c)$, или $P(B_4 \mid B_1B_2^cB_3)$. Их легко подсчитать (почему?), но нас также могут интересовать вероятности типа $P(B_4 \mid B_2)$ или $P(B_4 \mid B_1B_3^c)$, вычислить которые немного труднее (см. задачу 33).

Существует несколько способов доказательства приведенной выше замечательной теоремы. Каждый из них имеет свои достоинства и применяется в других ситуациях. Мы изложим два способа сейчас, а третий в намного более общей форме (теорема 5 из § 5.4) — позже. Существуют и другие подходы, и, может быть, читатель потом подумает о них. Вероятно, первый способ доказательства покажется вам наиболее сложным. Если это так, то пропустите его и сразу переходите ко второму*).

Первый способ. Его можно назвать «лобовым тараном» или «применением грубой силы», так как он предполагает применение тяжелого (хотя и стандартного) оружия из арсенала комбинаторики. Его особенность заключается в том, что он срабатывает, когда уже угадан ответ (как нам после нескольких шагов удалось в рассматриваемом случае). Другими словами, он представляет собой нечто вроде экспериментального подтверждения. Введем новую случайную величину Y_n, равную числу черных шаров, полученных при первых n извлечениях. Она позволяет вычислить пропорцию черных шаров, имеющуюся во время $(n+1)$-го извлечения, поскольку общее количество оставшихся в урне шаров равно $b+r-n$, независимо от исходов первых n извлечений. Поэтому

$$P(B_{n+1} \mid Y_n = j) = \frac{b-j}{b+r-n}, \quad 0 \leqslant j \leqslant b. \tag{5.3.4}$$

В свою очередь, вероятность $P(Y_n = j)$ вычисляется, как в задаче 1 из § 3.4, при $m = b+r$ и $k = b$ по формуле (3.4.1):

$$P(Y_n = j) = \frac{\binom{b}{j}\binom{r}{n-j}}{\binom{b+r}{n}}. \tag{5.3.5}$$

Теперь применим формулу (5.2.4):

$$\begin{aligned} P(B_{n+1}) &= \sum_{j=0}^{b} P(Y_n = j) P(B_{n+1} \mid Y_n = j) = \qquad (5.3.6)\\ &= \sum_{j=0}^{b} \frac{\binom{b}{j}\binom{r}{n-j}}{\binom{b+r}{n}} \, \frac{b-j}{b+r-n}. \end{aligned}$$

Это выражение, конечно, дает искомый ответ, но как, скажите, вычислить такую сумму? На самом деле, это не очень трудная задача. Некоторые превосходные математики сделали карьеру на подобных (и значительно более сложных) вычислениях. Прелесть подобных подсчетов в том, что они выполнимы, если наша догадка верна. Вера в это придает

*) Третий способ заключается в применении метода математической индукции по n.

силы. Давайте выпишем несколько биномиальных коэффициентов в явном виде, сократим и добавим множители с целью преобразования в *новые* биномиальные коэффициенты:

$$\begin{aligned}&\frac{b!}{j!\,(b-j)!}\;\frac{r!}{(n-j)!\,(r-n+j)!}\;\frac{n!\,(b+r-n)!}{(b+r)!}\;\frac{b-j}{b+r-n}=\\&=\frac{b!\,r!}{(b+r)!}\;\frac{(b+r-n-1)!}{(r-n+j)!\,(b-j-1)!}\;\frac{n!}{j!\,(n-j)!}=\\&=\frac{1}{\binom{b+r}{b}}\binom{b+r-n-1}{b-j-1}\binom{n}{j}.\end{aligned}$$

Следовательно,

$$P(B_{n+1})=\frac{1}{\binom{b+r}{b}}\sum_{j=0}^{b-1}\binom{n}{j}\binom{b+r-1-n}{b-1-j}, \tag{5.3.7}$$

где член, отвечающий $j=b$, отброшен, поскольку он дает 0 в соотношении (5.3.6). Новая сумма в (5.3.7) на основе известного тождества для биномиальных коэффициентов сворачивается в $\binom{b+r-1}{b-1}$, см. (3.3.9). Отсюда

$$P(B_{n+1})=\binom{b+r-1}{b-1}\Big/\binom{b+r}{b}=\frac{b}{b+r}\,,$$

как и утверждалось в формуле (5.3.3).

Второй способ. Он является чисто комбинаторным и проводится так же, как в примере из § 3.2. Его достоинство — простота, однако данный способ не удается просто обобщить для использования в другой урновой модели, изучаемой ниже.

Рассмотрим последовательные исходы $n+1$ извлечений: $X_1(\omega)$, $X_2(\omega)$, …, $X_n(\omega)$, $X_{n+1}(\omega)$. Каждая случайная величина $X_j(\omega)$ равна 1 или 0 в зависимости от конкретного вида точки ω; даже количество единиц и нулей среди них зависит от ω при $n+1<b+r$. Две разные последовательности исходов такие, как 0011 и 0101, не будут иметь в общем случае одинаковых вероятностей. Но давайте теперь нанесем номера на шары, скажем, от 1 до b — на черные, от $b+1$ до $b+r$ — на красные. Тогда все шары станут различимыми. Тем самым, мы попадаем в ситуацию выбора без возвращения с упорядочением, обсуждаемую в § 3.2. Полное число возможностей после нумерации задается формулой (3.2.1) с заменой в ней m на $b+r$ и n на $n+1$: $(b+r)_{n+1}$. При этом они все одинаково правдоподобны! Нас интересует случай, когда $(n+1)$-й шар имеет черный цвет. Сколько существует таких возможностей? У нас есть b

вариантов для выбора номера $(n+1)$-го шара, после этого, снова применив формулу (3.2.1), находим, что первые n шаров можно упорядочить $(b+r-1)_n$ способами. Тогда, в силу основного правила из § 3.1, число вариантов, в которых $(n+1)$-й шар является черным, равно $b(b+r-1)_n$. Классическая формула для вычисления вероятности события приводит к ответу

$$P(B_{n+1}) = \frac{b(b+r-1)_n}{(b+r)_{n+1}} = \frac{b}{b+r}\,.$$

Несомненно, данное доказательство, коль скоро оно объяснено, проще предыдущего и требует меньших вычислений. Однако оно предполагает наличие определенной проницательности для того, чтобы догадаться применить указанный метод подсчета. Данное решение привел Пуассон*), при этом его объяснения были еще короче наших. Сформулируем полученное им более общее утверждение.

Теорема 2 (*теорема Пуассона*). *Пусть в урне находятся b черных и r красных шаров. Сначала извлекаются n шаров и откладываются, причем их цвет считается неизвестным. Затем извлекаются еще m шаров. Тогда вероятность того, что среди них будет ровно k черных шаров, оказывается точно такой же, как если бы m шаров были вынуты сразу (без предварительного извлечения n шаров).*

Резюме: на вероятность не оказывает влияния предварительное извлечение шаров, если *мы находимся в неведении о том, каков их цвет.* Ясно, что если цвета извлеченных шаров известны, то в общем случае вероятность изменится. Приведем цитату из книги [16]: «Это исключительно хороший пример..., показывающий, что на вероятность не может повлиять сам *факт осуществления* материального события, но лишь *знание*, которым мы можем обладать, относящееся к осуществлению этого события».

Приведем краткое доказательство Пуассона. Если извлечены $n+m$ шаров, то вероятность комбинации, образуемой n черными и красными шарами, имеющимися в заданной пропорции, за которыми следуют m шаров, среди которых k черных и $m-k$ красных, должна быть той же самой, как в подобной комбинации, где m шаров предшествуют n шарам. Следовательно, вероятность получения k черных шаров в m извлечениях при условии, что n шаров уже были вынуты, должна равняться вероятности того же самого результата, когда шары предварительно не извлекаются.

*) Симеон Дени Пуассон (1781–1840)— французский математик, имя которого носят распределение, случайный процесс, предельная теорема и другие результаты.

Это рассуждение вас убедило? Более подробное комбинаторное доказательство, приведенное выше для случая $m = 1$, можно без труда обобщить и, тем самым, устранить все сомнения. Сомнение — вполне уместная вещь, несмотря на авторитет Пуассона. Обычно, как мы знаем из гл. 3, в истинности комбинаторного доказательств приходится убеждаться путем самостоятельного осмысления.

*5.4. Урновая схема Пойа

Продолжая обсуждение вопросов, затронутых в предыдущем параграфе, сделаем следующий шаг — изучим известное обобщение урновой модели, которое предложил Ж. Пойа [*]. Как и прежде, пусть урна первоначально содержит b черных и r красных шаров, но теперь после извлечения очередного шара будем возвращать его назад вместе с c дополнительными шарами того же цвета, где c — целое число. Если $c < 0$, то добавление c шаров означает извлечение $-c$ шаров. Это можно проводить как с наблюдением, так и без наблюдения цвета вынутого шара. Последнее происходит, например, если выбор производится некоторым автоматом. При $c = 0$ осуществляется обычный выбор с возвращением, а при $c = -1$ мы оказываемся в ситуации, изученной в § 5.3. В общем случае, если c — отрицательное число, то процесс обязательно прервется после некоторого количества извлечений. Если же c — положительное число или 0, то процесс будет продолжаться неограниченно долго. Данную схему можно еще усложнить (вы знаете, что обобщение — конек математиков), если при каждом извлечении мы будем добавлять в урну не только c шаров того же цвета, что у вынутого, но и еще d шаров другого цвета. Но мы не станем рассматривать это обобщение и в дальнейшем, говоря об *урновой модели Пойа*, ограничимся условием $c \geqslant -1$. На самом деле Пойа ввел данную модель, занимаясь одной медицинской проблемой (см. последний абзац этого параграфа).

Пример. Чему равна вероятность того, что в модели Пойа первые три извлеченных шара имеют, по порядку, цвета $\{b, b, r\}$, $\{b, r, b\}$ или $\{r, b, b\}$?

Простое применение утверждения 1 из § 5.2 дает, в обозначениях из § 5.2, равенство

$$\begin{aligned} P(B_1B_2R_3) &= P(B_1)P(B_2 \mid B_1)P(R_3 \mid B_1B_2) = \\ &= \frac{b}{b+r}\;\frac{b+c}{b+r+c}\;\frac{r}{b+r+2c}\,. \end{aligned} \tag{5.4.1}$$

[*] Ж. Пойа — профессор Станфордского университета, один из наиболее известных аналитиков нашего времени, внесший значительный вклад в теорию вероятностей, комбинаторику и их приложения.

Аналогично находим:

$$P(B_1R_2B_3) = \frac{b}{b+r}\,\frac{r}{b+r+c}\,\frac{b+c}{b+r+2c}\,,$$
$$P(R_1B_1B_2) = \frac{r}{b+r}\,\frac{b}{b+r+c}\,\frac{b+c}{b+r+2c}\,.$$

Таким образом, все вероятности одинаковы, т. е. вероятность получить 2 черных шара и 1 красный не зависит от *порядка*, в котором они появляются. Следовательно, вероятность извлечения*) двух черных и одного красного шара в три раза больше, чем правая часть равенства (5.4.1).

Теорема 3. *Вероятность извлечения (с самого начала) любого заданного набора, состоящего из k черных и $n-k$ красных шаров, равняется*

$$\frac{b(b+c)\dots(b+(k-1)c)r(r+c)\dots(r+(n-k-1)c)}{(b+r)(b+r+c)(b+r+2c)\dots(b+r+(n-1)c)}\,, \tag{5.4.2}$$

для всех $n \geqslant 1$, если $c \geqslant 0$; и для $0 \leqslant n \leqslant b+r$, если $c=-1$.

Доказательство. Достаточно применить утверждение 1 из § 5.2, но, все же, *придется потрудиться.* Выше мы уже вывели искомое равенство для случая $k = 2$ и $n = 3$. Если вы попытаетесь самостоятельно исследовать еще несколько случаев, скажем, $n = 4$, $k = 2$ и $n = 5$, $k = 3$, то, вероятно, быстрее разберетесь, как получается общий результат, чем поймете словесные объяснения. Вот главный момент: при m-м извлечении, где $1 \leqslant m \leqslant n$, знаменателем дроби, отвечающей $P(A_m \mid A_1A_2\dots A_{m-1})$ в (5.2.1) служит $b+r+(m-1)c$, поскольку всего к этому времени в урну было добавлено $(m-1)c$ шаров, безотносительно к тому, какой цвет имели извлекаемые шары. Далее, в первый момент, когда был вытащен черный шар, в урне было b черных шаров; при втором извлечении черного шара в урне находилось $b+c$ черных шаров, потому что вместе с ранее извлеченным черным шаром в урну были добавлены еще c черных шаров. Это верно безотносительно к тому, в какой момент (при каком извлечении) был вытащен второй черный шар. Аналогично, при извлечении третьего шара в урне находилось $b+2c$ черных шаров и т. д. Это объясняет происхождение k множителей, содержащих b, в числителе дроби (5.4.2). Теперь рассмотрим красные шары: в первый момент, когда был вытащен красный шар, в урне было r красных шаров; при втором извлечении красного шара в урне находилось $r+c$ красных шаров, потому что после извлечения первого красного шара в урну были добавлены еще c красных шаров и т. д. Это

*) В произвольном порядке. — *Прим. перев.*

объясняет происхождение $n-k$ множителей, содержащих r, в числителе дроби (5.4.2). Искомая вероятность, согласно формуле (5.2.1), получается в результате последовательного перемножения дробей — условных вероятностей. Порядок, в котором появляются множители в числителе, задается конкретным расположением черных и красных шаров в заданном наборе. Однако их произведение — одно и то же при условии, что величины n и k фиксированы. Тем самым, равенство (5.4.2) установлено.

Например, если конкретный набор имеет вид $RBRRB$, то множители в числителе появляются в таком порядке: $rb(r+c)(r+2c)(b+c)$.

Теперь предположим, что задано (фиксировано) только количество черных шаров в наборе, но не сам набор. Тогда верен следующий результат.

Теорема 4. *Вероятность извлечения (с самого начала) k черных шаров равна значению выражения* (5.4.2), *умноженному на* $\binom{n}{k}$. *В терминах обобщенных биномиальных коэффициентов (см. формулу* (5.4.4) *ниже) она имеет вид*

$$\binom{-b/c}{k}\binom{-r/c}{n-k}\Big/\binom{-(b+r)/c}{n}. \tag{5.4.3}$$

Это — обобщение гипергеометрического распределения, см. пример 10 в § 4.5.

Доказательство. Существует $\binom{n}{k}$ вариантов перестановок k черных шаров и $n-k$ красных шаров, см. § 3.2. Каждый отдельный набор, содержащий k черных и $n-k$ красных шаров, в соответствии с формулой (5.4.2) имеет одну и ту же вероятность. Разные перестановки отвечают несовместным событиям. Следовательно, вероятность, указанная в формулировке теоремы, в действительности является суммой $\binom{n}{k}$ вероятностей, каждая из которых равна значению (5.4.2). Осталось только представить ее в виде (5.4.3). Для этого потребуется немного алгебры. Давайте будем считать, что если a — положительное действительное число и j — натуральное число, то по определению:

$$\begin{aligned}\binom{-a}{j} &= \frac{(-a)(-a-1)\ldots(-a-j+1)}{j!} = \\ &= (-1)^j\,\frac{a(a+1)\ldots(a+j-1)}{j!}\,.\end{aligned} \tag{5.4.4}$$

Таким образом, если разделить каждый множитель дроби (5.4.2) на c и ввести обозначения

$$\beta = \frac{b}{c}, \quad \gamma = \frac{r}{c},$$

то, с учетом определения (5.4.4), получим:

$$\frac{\beta(\beta+1)\dots(\beta+k-1)\gamma(\gamma+1)\dots(\gamma+n-k-1)}{(\beta+\gamma)(\beta+\gamma+1)\cdots(\beta+\gamma+n-1)} =$$
$$= \frac{(-1)^k k!\binom{-\beta}{k}(-1)^{n-k}(n-k)!\binom{-\gamma}{n-k}}{(-1)^n n!\binom{-\beta-\gamma}{n}} = \frac{\binom{-\beta}{k}\binom{-\gamma}{n-k}}{\binom{-\beta-\gamma}{n}\binom{n}{k}}.$$

Умножая на $\binom{n}{k}$, приходим к формуле (5.4.3).

Теперь можно получить существенное обобщение теорем 1 и 2 из § 5.3. Как мы увидим, результат так же легко вытекает из фундаментальной формулы (5.4.2), как созревший плод падает с дерева. Правда, потребуется ввести еще несколько новых обозначений.

Вспоминая определение случайной величины X_n из § 5.3, мы можем рассматривать (5.4.2) как формулу, задающую *совместное распределение* n случайных величин $\{X_1, X_2, \dots, X_n\}$. Давайте введем переменные ε_i, $i = 1, \dots, n$, принимающие значения 0 или 1 независимо друг от друга. Однако при этом две такие переменные с одинаковыми индексами, конечно, считаются имеющими определенные значения в границах обсуждения. Например, $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\}$ могут обозначать $\{1, 1, 0, 1\}$ или $\{0, 1, 0, 1\}$, но тогда $\{\varepsilon_1, \varepsilon_2\}$ совпадает с $\{1, 1\}$ в первом случае и с $\{0, 1\}$ — во втором. Теорему 3 можно сформулировать так: если k штук из ε_i в формуле, записанной ниже, равны 1, а остальные $n-k$ равны 0, то вероятность

$$P(X_1 = \textcircled{1}_1, X_2 = \textcircled{1}_2, \dots, X_n = \textcircled{1}_n) \tag{5.4.5}$$

задается выражением (5.4.2). Всего имеется 2^n возможных вариантов значений ε_i в (5.4.5) (почему?), и если выписать все соответствующие им значения выражения (5.4.2), то полученное множество из 2^n вероятностей будет определять совместное распределение величин $\{X_1, X_2, \dots, X_n\}$. Теперь допустим, что $\{n_1, n_2, \dots, n_s\}$ — некоторое подмножество множества $\{1, 2, \dots, n\}$. Совместное распределение случайных величин $\{X_{n_1}, \dots, X_{n_s}\}$ задается набором вероятностей

$$P(X_{n_1} = \textcircled{1}_{n_1}, \dots, X_{n_s} = \textcircled{1}_{n_s}), \tag{5.4.6}$$

где значения ε_i пробегают все 2^s возможных вариантов. Оно называется *маргинальным распределением* по отношению к распределению большего множества величин $\{X_1, \dots, X_n\}$.

Нам потребуются новые обозначения. Пусть $\{n'_1, \dots, n'_t\}$ — это дополнение множества $\{n_1, \dots, n_s\}$ до всего множества $\{1, \dots, n\}$, т. е. те индексы, которые останутся после исключения индексов $\{n_1, \dots, n_s\}$. Конечно, $t = n-s$ и объединение $\{n_1, \dots, n_s, n'_1, \dots, n'_t\}$ представляет собой некоторую перестановку чисел $\{1, \ldots, n\}$. Теперь мы можем выписать следующую формулу, выражающую маргинальные вероятности через совместное распределение большего множества величин:

$$P(X_{n_1} = \textcircled{1}_1, \ldots, X_{n_s} = \textcircled{1}_s) =$$
$$= \sum_{\textcircled{1}'_1, \ldots, \textcircled{1}'_t} P\big(X_{n_1} = \textcircled{1}_1, \ldots, X_{n_s} = \textcircled{1}_s, X_{n'_1} = \textcircled{1}'_1, \ldots, X_{n'_t} = \textcircled{1}'_t\big), \quad (5.4.7)$$

где $\{\varepsilon'_1, \ldots, \varepsilon'_t\}$ — дополнительное множество индексов, и суммирование производится по всем 2^t возможным вариантам значений. Данная формула вытекает из очевидного соотношения между множествами

$$\{X_{n_1} = \varepsilon_1, \ldots, X_{n_s} = \varepsilon_s\} =$$
$$= \sum_{\varepsilon'_1, \ldots, \varepsilon'_t} \{X_{n_1} = \varepsilon_1, \ldots, X_{n_s} = \varepsilon_s, X_{n'_1} = \varepsilon'_1, \ldots, X_{n'_t} = \varepsilon'_t\}$$

и аддитивности вероятностной меры P. (Очевидно, что аналогичная формула верна, когда случайные величины X_i принимают значения, отличные от 0 и 1. В таком случае суммирование должно проводиться по всем возможным значениям.)

Наконец, мы подошли к кульминации (*pièce de résistance*) нашего обсуждения. Она безжалостно проверит вашу готовность к усвоению общих и абстрактных рассуждений. Если вы не сможете «проглотить» ее сейчас — не расстраивайтесь, а вернитесь позже и попробуйте еще раз.

Теорема 5. *Совместное распределение любых s из случайных величин $\{X_1, X_2, \ldots, X_n, \ldots\}$ одинаково.*

Как отмечалось выше, если $c \geqslant 0$, то последовательность, образованная величинами X_n, бесконечна. Напротив, если $c = -1$, то $n \leqslant b + r$.

Доказательство. Что именно утверждается в теореме? Фиксируем число s. Пусть $X_{n_1}, \ldots, X_{n_s}$ — произвольное множество из s случайных величин, входящих в последовательность. Для получения их совместного распределения необходимо рассмотреть все возможные варианты выбора значений этих s величин. Поэтому нам требуется обозначение для произвольного варианта такого вида: скажем, $\varepsilon_1, \ldots, \varepsilon_s$. Запишем вероятность

$$P(X_{n_1} = \varepsilon_1, X_{n_2} = \varepsilon_2, \ldots, X_{n_s} = \varepsilon_s).$$

Мы должны показать, что она имеет одно и то же значение независимо от того, какой именно набор из s индексов был выбран. Другими словами, она должна быть равна

$$P(X_{m_1} = \varepsilon_1, X_{m_2} = \varepsilon_2, \ldots, X_{m_s} = \varepsilon_s),$$

где $\{m_1, \ldots, m_s\}$ — произвольное подмножество размера s. Два множества $\{n_1, \ldots, n_s\}$ и $\{m_1, \ldots, m_s\}$ могут пересекаться (как, например, $\{1, 3, 4\}$ и $\{3, 2, 1\}$). Заметим также, что не требуется, чтобы индексы располагались в порядке возрастания.

Пусть наибольшим из индексов, встретившихся выше, является n. Как и прежде, положим $t = n - s$ и

$$\{n'_1, \ldots, n'_t\} = \{1, \ldots, n\} - \{n_1, \ldots, n_s\},$$
$$\{m'_1, \ldots, m'_t\} = \{1, \ldots, n\} - \{m_1, \ldots, m_s\}.$$

Далее, пусть $\varepsilon'_1, \ldots, \varepsilon'_t$ — произвольный набор значений t переменных. Мы утверждаем, что

$$P(X_{n_1} = \textcircled{1}_1, \ldots, X_{n_s} = \textcircled{1}_s, X_{n'_1} = \textcircled{1}'_1, \ldots, X_{n'_t} = \textcircled{1}'_t) = \qquad (5.4.8)$$
$$= P(X_{m_1} = \textcircled{1}_1, \ldots, X_{m_s} = \textcircled{1}_s, X_{m'_1} = \textcircled{1}_1, \ldots, X_{m_t} = \textcircled{1}'_t).$$

Чтобы понять смысл этого символизма, достаточно заметить, что данная формула непосредственно следует из выражения (5.4.2). Действительно, оба набора, $(n_1, \ldots, n_s, n'_1, \ldots, n'_t)$ и $(m_1, \ldots, m_s, m'_1, \ldots, m'_t)$, суть перестановки полного множества индексов $(1, \ldots, n)$. В свою очередь, набор значений $(\varepsilon_1, \ldots, \varepsilon_s, \varepsilon'_1, \ldots, \varepsilon'_t)$ одинаков в обеих частях равенства (5.4.8). Тем самым, оно всего лишь повторяет утверждение теоремы 3 о том, что любые два заданных набора с одинаковым количеством черных шаров имеют одинаковые вероятности, независимо от перестановки индексов.

Наконец, оставляя $\varepsilon_1, \ldots, \varepsilon_s$ фиксированными и позволяя значениям $\varepsilon'_1, \ldots, \varepsilon'_t$ пробегать все 2^t возможных вариантов, получим 2^t равенств вида (5.4.8). Сложим их и применим равенство (5.4.7): первый раз, как оно записано, а второй — с заменой n_i на m_i. В результате установим, что

$$P(X_{n_1} = \varepsilon_1, \ldots, X_{n_s} = \varepsilon_s) = P(X_{m_1} = \varepsilon_1, \ldots, X_{m_s} = \varepsilon_s),$$

что и требовалось доказать.

В приведенном доказательстве нет никаких фокусов, чего-то действительно трудного. Читатель, возможно, скажет, что «все дело — в обозначениях».

Последовательность случайных величин $\{X_n;\, n = 1, 2, \ldots\}$, обладающая указанным в теореме 5 свойством, называется *перестановочной*. Для нее, в частности, выполняется то, что любые «блоки» заданной длины s вида $X_{s_0+1}, X_{s_0+2}, \ldots, X_{s_0+s}$, где s_0 — любое неотрицательное целое число (и $s_0+s \leqslant b+r$, если $c = -1$), распределены одинаково. Так как индекс обычно интерпретируют как *временной параметр*, то говорят, что распределение таких «блоков» является инвариантным к временному сдвигу. Последовательность случайных величин, обладающая указанным свойством, называется (*строго*)*) *стационарной*. Процессы данного типа широко используются в моделях для описания электрических колебаний, экономических временных рядов, задач теории очередей**) и т. п.

Схему Пойа можно рассматривать как модель для зависящих от удачи исходов (или, говоря попросту, случайных событий), правдоподобие которых имеет тенденцию увеличиваться с каждым наступлением события и уменьшаться, если событие не происходит. Пример такого события дает выбор черного шара из урны в эксперименте Пойа. Пойа ввел урновую модель для описания процесса распространения эпидемий, когда каждый зараженный порождает новые вирусы и, тем самым, увеличивает шансы дальнейшего заражения. Процитируем точно его высказывание (в переводе с французского, выполненного первым из авторов): «Сводя данную проблему к ее простейшему представлению и внося в модель определенную симметрию, удобную для математического изучения, мы приходим к урновой схеме». Под внесенной симметрией имеется в виду добавление в урну красных шаров при извлечении красного шара. Это означает, что каждый наблюдаемый «незараженный» увеличивает шансы не заболеть других «незараженных». Данное допущение не представляется достаточно обоснованным. Это обстоятельство было оставлено без комментариев в книгах некоторых авторов, обсуждавших урновую модель. Прямота профессора Пойа, предупредившего, что допущение введено из соображений математического удобства, может служить примером для ученых, создающих детально разработанные математические теории для описания «грубой» действительности вроде клевания корма курицами (математическая психология) или ползания жуков (математическая биология).

5.5. Независимость и связанные с ней понятия

Крайний и исключительно важный случай при вычислении условной вероятности имеет место, когда условие не влияет на вероятность. Интуитивное понятие независимости возникает из привычного опыта бросания монеты (игральной кости) несколько раз подряд или повторного извлечения с возвращением шаров из урны. Знание исходов предыдущих испытаний не оказывает влияния на «истинные» вероятности исходов в последующих испытаниях, и в этом смысле испытания не зависят друг

*) Или стационарной в узком смысле. — *Прим. перев.*

**) Также называемой теорией массового обслуживания. — *Прим. перев.*

от друга. Мы уже определили понятие независимости событий в § 2.4. Заметим, что определяющая формула (2.4.5) является частным случаем соотношения (5.2.1) при замене условных вероятностей на безусловные. Аналогичная замена в формуле (5.2.2) приводит, как показано ниже, к следующему фундаментальному определению.

Определение независимости случайных величин. Случайные величины $X_1, \ldots, X_n$, принимающие счетное число значений, называются независимыми тогда и только тогда, когда для любых действительных чисел $x_1, \ldots, x_n$ выполняется равенство

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \times \ldots \times P(X_n = x_n). \tag{5.5.1}$$

Оно тривиально, если хотя бы один из сомножителей справа равен нулю, поэтому в определении можно ограничиться только такими x_i, которые принадлежат счетному множеству возможных значений величин X_i.

Несмотря на кажущуюся простоту, условие (5.5.1) на самом деле содержит в себе намного больше, чем представляется на первый взгляд. Чтобы увидеть это, выведем основное обобщение формулы (5.5.1), в котором отдельные значения x_i заменяются на произвольные множества S_i. В приведенных ниже утверждениях 4–6 предполагается, что $X_1, \ldots, X_n$ — независимые случайные величины.

Утверждение 4. *Для произвольных счетных множеств* $S_1, \ldots, S_n$

$$P(X_1 \in S_1, \ldots, X_n \in S_n) = P(X_1 \in S_1) \times \ldots \times P(X_n \in S_n). \tag{5.5.2}$$

Доказательство. Левая часть соотношения (5.5.2) равна

$$\begin{aligned}
&\sum_{x_1 \in S_1} \ldots \sum_{x_n \in S_n} P(X_1 = x_1, \ldots, X_n = x_n) = \\
&\quad = \sum_{x_1 \in S_1} \ldots \sum_{x_n \in S_n} P(X_1 = x_1) \times \ldots \times P(X_n = x_n) = \\
&\quad = \left\{ \sum_{x_1 \in S_1} P(X_1 = x_1) \right\} \times \ldots \times \left\{ \sum_{x_n \in S_n} P(X_n = x_n) \right\}.
\end{aligned}$$

С помощью элементарной алгебры легко проверяется, что данное выражение совпадает с правой частью формулы (5.5.2) (не поленитесь убедиться в этом, если сомневаетесь).

Отметим, что определенная выше независимость множества случайных величин является свойством всего множества в целом. Подобные свойства не обязательно передаются подмножествам; можете ли вы привести простой контрпример? Однако из утверждения 4 вытекает, что любое подмножество из величин $(X_1, \ldots, X_n)$ снова обладает свойством независимости. Чтобы показать, что таково, к примеру, подмножество (X_1, X_2, X_3), когда выше $n > 3$, возьмем $S_i = R^1$ для $i > 3$ и заменим другие S_i на x_i в формуле (5.5.2).

Далее, условие (5.5.2) может быть усилено до его наиболее часто используемой формы.

Утверждение 5. *События*

$$\{X_1 \in S_1\}, \ \ldots, \ \{X_n \in S_n\} \tag{5.5.3}$$

независимы.

Доказательство. Важно напомнить, что для независимости событий требуется не только выполнение соотношения (5.5.2), но также выполнение аналогичных соотношений для всех подмножеств множества $(X_1, \ldots, X_n)$. Однако они выполняются в силу того, что эти подмножества сами являются множествами независимых случайных величин, как мы только что показали.

Прежде чем двигаться дальше, давайте убедимся, что понятие независимости событий из § 2.4 есть частный случай введенного в текущем параграфе понятия независимости случайных величин. С произвольными событиями $\{A_j, 1 \leqslant j \leqslant n\}$ свяжем их индикаторы I_{A_j} (см. § 1.4), где

$$I_{A_j}(\omega) = \begin{cases} 1, & \text{если } \omega \in A_j, \\ 0, & \text{если } \omega \in A_j^c; \end{cases} \quad 1 \leqslant j \leqslant n.$$

Они являются случайными величинами (по крайней мере на счетном выборочном пространстве). Каждая принимает лишь два значения: 0 или 1, причем

$$\{I_{A_j} = 1\} = A_j, \quad \{I_{A_j} = 0\} = A_j^c.$$

Теперь, если применить определение независимости (5.1.1) к случайным величинам $I_{A_1}, \ldots, I_{A_n}$, то оно в точности сведется к равенствам

$$P(\tilde{A}_1 \ldots \tilde{A}_n) = P(\tilde{A}_1) \ldots P(\tilde{A}_n), \tag{5.5.4}$$

где каждое из множеств $\tilde{A}_j$ может быть либо A_j, либо A_j^c, но, конечно, одинаково слева и справа. На самом деле можно показать (задача 36 ниже), что выполнение для всевозможных вариантов $\tilde{A}_j$ условия (5.5.4) в точности совпадает с определением (2.4.5). Таким образом, независимость событий равносильна независимости их индикаторов.

Изучение независимых случайных величин является основной темой в любом из вводных вероятностных курсов. Исторически и эмпирически они связаны с независимыми испытаниями. Мы уже неформально обсуждали данное понятие в § 2.4. Теперь можно сформулировать его в терминах независимых случайных величин: последовательность независимых испытаний — это последовательность независимых случайных величин $(X_1, \ldots, X_n)$, где X_i представляет собой исход i-го испытания. Простые иллюстрации даны в примерах 7 и 8 из § 2.4 (в примере 7 случайные величины отсутствуют, но понятно, как их ввести). Между прочим, данные примеры позволяют установить сам факт *существования* независимых случайны величин, тем самым позволяя нам убедиться, что объект утверждений, содержащихся в текущем параграфе, в действительности существует. На самом деле можно даже построить последовательность независимых случайных величин с произвольными заданными распределениями (см. [4, гл. 3]). (Вероятно, вам будет забавно узнать, что в математике известны случаи, когда определялись и изучались новые объекты, а позже выяснялось, они не существуют!) Это замечание относится к материалу следующих глав, а сейчас мы добавим еще одно общее утверждение для расширения информационного горизонта.

Утверждение 6. *Пусть* $\varphi_1, \ldots, \varphi_n$ — *произвольные действительнозначные функции, заданные на* $(-\infty, \infty)$. *Тогда случайные величины*

$$\varphi_1(X_1),\ \ldots,\ \varphi_n(X_n) \tag{5.5.5}$$

независимы.

Доказательство. Давайте опустим индексы у X и φ и зададим вопрос: как для заданного действительного числа y устроено множество значений x, на котором

$$\varphi(x) = y \quad \text{и} \quad X = x?$$

Так как случайная величина X принимает счетное число значений, это множество должно быть счетным. Обозначим его через S. Оно, конечно, зависит от y, φ и X. Тогда $\{\varphi(X) = y\}$ означает в точности то же самое, что и $\{X \in S\}$. Следовательно, для произвольных $y_1, \ldots, y_n$ события

$$\{\varphi_1(X_1) = y_1\},\ \ldots,\ \{\varphi_n(X_n) = y_n\}$$

как раз такие, как в строке (5.5.3), для соответсвующих множеств $S_1, \ldots, S_n$. Поэтому утверждение 6 вытекает из утверждения 5.

Доказанное утверждение будет неоднократно применяться в гл. 6. На самом деле справедлив следующий более общий результат. Если разбить случайные величины $X_1, \ldots, X_n$ на любое число «блоков» и взять некоторые функции от величин каждого из «блоков», то получившиеся случайные величины будут независимыми. Доказательство не сильно отличается от частного случая, рассмотренного выше, и мы его опускаем.

Что касается случайных величин общего вида*), то они, по определению, независимы тогда и только тогда, когда для произвольных действительных чисел $x_1, \ldots, x_n$ события

$$\{X_1 \leqslant x_1\}, \ \ldots, \ \{X_n \leqslant x_n\} \tag{5.5.6}$$

независимы. В частности,

$$P(X_1 \leqslant x_1, \ \ldots, \ X_n \leqslant x_n) = P(X_1 \leqslant x_1) \times \ldots \times P(X_n \leqslant x_n). \tag{5.5.7}$$

Используя введенное в § 4.6 понятие совместной функции распределения F случайного вектора $(X_1, \ldots, X_n)$, данное равенство можно переписать в виде

$$F(x_1, \ldots, x_n) = F_1(x_1) \times \ldots \times F_n(x_n), \tag{5.5.8}$$

где F_j — маргинальное распределение случайной величины $X_j, 1 \leqslant j \leqslant n$. Таким образом, в случае независимости компонент маргинальные распределения определяют совместное распределение.

Из определения в качестве следствия можно вывести, что события, подобные тем, что записаны в строке (5.5.3), также независимы при условии, что множества $S_1, \ldots, S_n$ не слишком сложно устроены (являются борелевскими). В частности, если существует совместная плотность f, то

$$\begin{aligned} P(X_1 \in S_1, \ \ldots, \ X_n \in S_n) &= \left\{ \int\limits_{S_1} f_1(u)\,du \right\} \times \ldots \times \left\{ \int\limits_{S_n} f_n(u)\,du \right\} = \\ &= \int\limits_{S_1} \ldots \int\limits_{S_n} f_1(u_1) \times \ldots \times f_n(u_n)\,du_1 \ldots du_n, \end{aligned}$$

где $f_1, \ldots, f_n$ обозначают маргинальные плотности. Но, с другой стороны, вероятность в левой части равна

$$\int\limits_{S_1} \ldots \int\limits_{S_n} f(u_1, \ldots, u_n)\,du_1 \ldots du_n,$$

подобно выражению (4.6.6). Сравнение этих двух выражений дает соотношение

$$f(u_1, \ldots, u_n) = f_1(u_1) \ldots f_n(u_n). \tag{5.5.9}$$

*) То есть не обязательно имеющих счетное множество значений. — *Прим. перев.*

Последнее равенство является аналогом формулы (5.5.8) для плотностей.

Итак, мы видим, что стохастическая независимость позволяет разложить на множители совместную вероятность, распределение или плотность. В последующих главах мы покажем, что это свойство позволяет представить в виде произведения математическое ожидание, производящую функцию и другие преобразования случайных величин.

Многочисленные результаты и примеры использования независимых случайных величин появятся в гл. 6 и 7. На самом деле с понятием независимости связано основное содержание классической теории вероятностей. Это понятие настолько важно, что в своей эпохальной монографии «Основные понятия теории вероятностей» А. Н. Колмогоров*) писал: «Возникает ощущение, что именно в независимости заключается, по крайней мере частично, истинная природа вероятностных задач». Здесь мы ограничимся двумя простыми примерами.

Пример 10. Письмо от Паскаля к Ферма (датированное 29 июля 1654 г.) содержит, помимо многих других математических задач, такой текст: «Де Мере сообщил мне, что он обнаружил следующую ошибку в теории чисел. Если игрок ставит на то, что при четырех бросаниях игральной кости хотя бы один раз выпадет 6, то его шансы на выигрыш составляют 671 к 625. Если же он намерен бросать сразу две кости, то шансы выпадения двух шестерок хотя бы единожды за 24 бросания оказываются меньше, чем 1/2. И это несмотря на то, что отношение 24 к 36 (количеству всевозможных пар номеров на двух костях) равно отношению 4 к 6 (числу граней одной кости). Данное наблюдение привело его в негодование и подтолкнуло к тому, что он всем и каждому говорит о несостоятельности математических утверждений и внутренней противоречивости арифметики. Однако вы без труда увидите, что мое объяснение является корректным и опирается на общее для нас с вами понимание принципов».

Эта известная задача (одна из первых зафиксированных в истории теории вероятностей), которая бросала вызов интеллектуальным гигантам прошлых времен, сейчас может быть решена начинающим.

Выпадение грани с номером 6 при четырех бросаниях одной кости означает получение «6» по крайней мере один раз за 4 испытания. Опре-

*) Колмогоров Андрей Николаевич (1903–1987)— выдающийся русский математик, один из основателей современной вероятностной теории.

делим распределения случайных величин $X_n, 1 \leqslant n \leqslant 4$, так:

$$P(X_n = k) = \frac{1}{6}, \quad k = 1, 2, \ldots, 6,$$

и будем считать, что случайные величины X_1, X_2, X_3, X_4 независимы. Положим

$$A_n = \{X_n = 6\}.$$

Тогда интересующее нас событие представляется в виде объединения $A_1 \cup A_2 \cup A_3 \cup A_4$. Проще вычислить вероятность его дополнения $A_1^c A_2^c A_3^c A_4^c$. Испытания предполагаются независимыми, а кость — не имеющей смещения. В качестве частного случая формулы (5.5.4) находим, что

$$P(A_1^c A_2^c A_3^c A_4^c) = P(A_1^c)P(A_2^c)P(A_3^c)P(A_4^c) = \left(\frac{5}{6}\right)^4;$$

следовательно,

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) = 1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} = \frac{671}{1296}.$$

Дробь справа приблизительно равна 0.5177. Поскольку $1296 - 671 = 625$, шансы составляют 671 к 625, как и утверждал Паскаль.

Теперь рассмотрим случай двух костей. Пусть (X_n', X_n'') обозначает исход, полученный при n-м бросании пары костей, и пусть

$$B_n = \{X_n' = 6; X_n'' = 6\}.$$

Тогда $P(B_n^c) = 35/36$ и

$$P(B_1^c B_2^c \ldots B_{24}^c) = \left(\frac{35}{36}\right)^{24},$$

$$P(B_1 \cup B_2 \cup \ldots \cup B_{24}) = 1 - \left(\frac{35}{36}\right)^{24}.$$

Правая часть последней формулы приближенно равна 0.4914, что подтверждает невыгодность такого пари.

Стоит отдать должное Де Мере за его острую наблюдательность. Солидный опыт, накопленный за игровым столом, позволил ему подметить «узкое» неравенство

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) > \frac{1}{2} > P(B_1 \cup B_2 \cup \ldots \cup B_{24}).$$

Его подвела арифметика из-за ошибочности «линейной гипотезы». (Согласно мнению некоторых историков, данная задача появилась еще до Де Мере.)

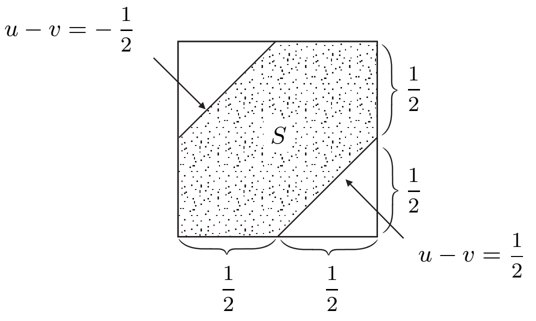


Рис. 24

Пример 11. Если две точки выбираются наудачу из отрезка $[0, 1]$, то какова вероятность того, что расстояние между ними окажется меньше $1/2$?

Мы надеемся, что читатель уже научился расшифровывать подобные формулировки. Вопрос означает следующее: пусть X и Y — две независимые равномерно распределенные на отрезке $[0, 1]$ случайные величины, найдите вероятность $P\,(|X - Y| < 1/2)$. При нашей гипотезе случайный вектор (X, Y) оказывается равномерно распределенным на единичном квадрате U (см. рис. 24). Другими словами, для любого «хорошего» подмножества S квадрата U имеем:

$$P\{(X, Y) \in S\} = \iint\limits_S du\,dv.$$

Это очевидно из обсуждения, приведенного после формулы (4.6.6). Действительно, плотность $f(u, v)$ равна произведению $f_1(u)f_2(v)$, согласно соотношению (5.5.9), а обе плотности f_1 и f_2 равны 1 внутри $[0, 1]$ и 0 снаружи. В нашей задаче множеством S служит множество точек (u, v) из квадрата U, удовлетворяющих неравенству $|u - v| < 1/2$. Вы можете вычислить двойной интеграл по данному множеству, если достаточно хорошо знаете математический анализ, но намного проще сделать это с помощью геометрии. Проведите две прямые $u-v = 1/2$ и $u-v = -1/2$. Тогда S есть область, ограниченная этими прямыми и сторонами квадрата. Дополняющая ее до квадрата область $U - S$ состоит из двух треугольников, площадь каждого из которых равна $1/2\,(1/2)^2 = 1/8$. Следовательно, получаем, что

$$\text{площадь } S = 1 - 2 \cdot \frac{1}{8} = \frac{3}{4}\,,$$

это и есть искомая вероятность.

Пример 12. Пусть $X_1, X_2, \ldots, X_n$ — независимые случайные величины с распределениями $F_1, F_2, \ldots, F_n$, соответственно; распределения задаются формулой (4.5.4). Положим

$$M = \max(X_1, X_2, \ldots, X_n),$$
$$m = \min(X_1, X_2, \ldots, X_n).$$

Требуется найти функции распределения случайных величин M и m.

Для произвольного x, применяя соотношение (5.5.7), получаем:

$$\begin{aligned} F_{\max}(x) &= P(M \leqslant x) = P(X_1 \leqslant x, X_2 \leqslant x, \ldots, X_n \leqslant x) = \\ &= P(X_1 \leqslant x)\, P(X_2 \leqslant x)\, \ldots\, P(X_n \leqslant x) = \\ &= F_1(x)\, F_2(x) \ldots F_n(x). \end{aligned}$$

В частности, если все функции F_j одинаковы, то

$$F_{\max}(x) = [F(x)]^n.$$

В случае минимума удобно ввести «хвостовые распределения» G_j, соответствующие каждой из F_j, следующим образом:

$$G_j(x) = P\{X_j > x\} = 1 - F_j(x).$$

Тогда, используя аналог формулы (5.5.2) на этот раз с $S_j = (x_j, \infty)$, имеем:

$$\begin{aligned} G_{\min}(x) &= P(m > x) = P(X_1 > x, X_2 > x, \ldots, X_n > x) = \\ &= P(X_1 > x)\, P(X_2 > x)\, \ldots\, P(X_n > x) = \\ &= G_1(x)\, G_2(x) \ldots G_n(x). \end{aligned}$$

Следовательно,

$$F_{\min}(x) = 1 - G_1(x)\, G_2(x) \ldots G_n(x).$$

Если все F_j одинаковы, то данное равенство превращается в формулу

$$G_{\min}(x) = [G(x)]^n, \quad F_{\min}(x) = 1 - [G(x)]^n.$$

Вот конкретная иллюстрация. Предположим, что город снабжается водой из трех резервуаров. Пусть величины ежедневных водозаборов из каждого резервуара суть независимые случайные величины, имеющие экспоненциальные плотности $\lambda_1 e^{-\lambda_1 x}$, $\lambda_2 e^{-\lambda_2 x}$, $\lambda_3 e^{-\lambda_3 x}$ соответственно. Допустим, что каждый резервуар может предоставить городу максимум N кубических метров воды в день. Какова вероятность того, что в определенный день городу не хватит воды?

Обозначим через X_1, X_2, X_3 размеры водозаборов в этот день. Тогда искомая вероятность, в силу формулы (4.5.12), равна

$$P(X_1 > N, X_2 > N, X_3 > N) = e^{-\lambda_1 N} e^{-\lambda_2 N} e^{-\lambda_3 N} = e^{-(\lambda_1+\lambda_2+\lambda_3)N}.$$

Оставшаяся часть параграфа посвящена краткому знакомству с новым логическим понятием, которое несколько шире, чем попарная независимость. Оно применяется при статистическом сравнении эмпирических данных, операционном оценивании альтернативных методик и т. п. Некоторые авторы даже пытаются построить философское обоснование статистики на базе этого качественного понятия.

Событие A называется *благоприятным**) для другого события B тогда и только тогда, когда

$$P(AB) \geqslant P(A)P(B). \tag{5.5.10}$$

Символически это обозначается как $A \parallel B$. Таким образом, мы имеем дело с бинарным отношением между событиями, которое включает независимость в качестве частного случая. Замечательный пример такого отношения дает свойство делимости нацело целых чисел; см. § 2.5 и задачу 17 в гл. 2.

Из неравенства (5.5.10) ясно, что отношение $\parallel$ симметрично; оно также рефлексивно, поскольку $P(A) \geqslant P(A)^2$ для любого A. Но оно нетранзитивно, т. е. $A \parallel B$ и $B \parallel C$ не влекут $A \parallel C$. На самом деле, мы покажем в следующем примере, что даже более строгое отношение попарной независимости не является транзитивным.

Пример 13. Рассмотрим семьи, имеющие двух детей, как в примере 5 из § 5.1: $\Omega = \{(bb), (bg), (gb), (gg)\}$. Пусть одна из таких семей выбирается случайно. Рассмотрим следующие три события:

$A = \{$первый ребенок — мальчик$\}$;

$B = \{$дети имеют разный пол$\}$;

$C = \{$первый ребенок — девочка$\}$.

Тогда

$$AB = \{(bg)\}, \quad BC = \{(gb)\}, \quad AC = \varnothing.$$

Элементарные вычисления показывают, что

$$P(AB) = P(A)P(B), \quad P(BC) = P(B)P(C),$$

но $P(AC) = 0 \neq P(A)P(C)$. Поэтому пары $\{A, B\}$ и $\{B, C\}$ состоят из независимых событий, а пара $\{A, C\}$ — нет.

*) В оригинале используется термин «favorable». — *Прим. перев.*

Небольшая модификация этого примера позволит убедиться, что попарная независимость не влечет совместной независимости в случае трех событий. Пусть

$$D = \{\text{второй ребенок — мальчик}\}.$$

Тогда

$$AD = \{(bb)\}, \quad BD = \{(gb)\}, \quad ABD = \varnothing;$$

и поэтому

$$P(ABD) = 0 \neq P(A)P(B)P(D) = \frac{1}{8}\,.$$

Не так давно можно было найти учебники по теории вероятностей и статистике, в которых совместная независимость ошибочно подменялась попарной. Легко «задним умом» предложить каждодневные аналогии для приведенных выше контрпримеров. Скажем, если A дружит с B, а B дружит с C, почему обязательно A должен дружить с C? По-другому: если в каждой паре из трех человек A, B, C люди ладят между собой, то не обязательно они уживутся все вместе.

Эти обыденные иллюстрации предназначены для того, чтобы рассказать нам что-то о пользе и вреде «интуиции». Продвигаясь чуть дальше, давайте запишем еще ряд *ложных заключений*(«$\not\Rightarrow$» читается «не влечет»):

$$\begin{aligned}
&A \parallel C \quad \text{и} \quad B \parallel C \not\Rightarrow (A \cap B) \parallel C;\\
&A \parallel B \quad \text{и} \quad A \parallel C \not\Rightarrow A \parallel (B \cap C);\\
&A \parallel C \quad \text{и} \quad B \parallel C \not\Rightarrow (A \cup B) \parallel C;\\
&A \parallel B \quad \text{и} \quad A \parallel C \not\Rightarrow A \parallel (B \cup C).
\end{aligned} \tag{5.5.11}$$

Вы можете попытаться придумать какие-то словесные объяснения для этих соотношений; несложно также построить строгие, но искусственные примеры, см. задачу 15.

То, что при обращении с условными вероятностями необходимо соблюдать определенную осторожность, вовсе не является чисто теоретическим вопросом. Часто правильность выводов при статистическом анализе экспериментальных данных целиком зависит от критического отношения к использованию базовых принципов. Следующая иллюстрация взята из статьи Colin R. Blyth, «On Simpson's paradox and the sure-thing principle», *Journal of American Statistical Association*, Vol. 67 (1972), p. 364–366.

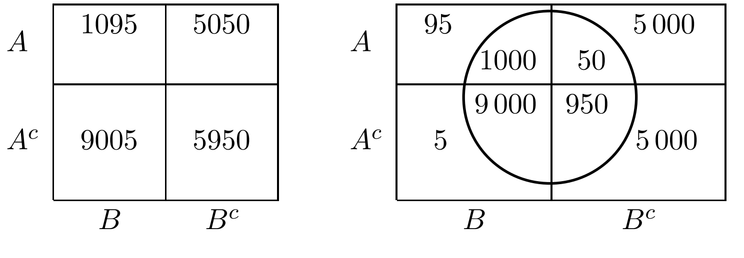
Пример 14. В таблице ниже приведены собранные врачом данные об эффективности нового вида лечения. Из-за того что курс включал в себя продолжительное добавочное лечение после выписки из больницы, врач смог применить его только к немногим пациентам, проживающим за пределами города. В основном ему пришлось иметь дело с городскими жителями.

	Городские жители		Сельские жители	
	Лечившиеся	Не лечившиеся	Лечившиеся	Не лечившиеся
Выздоровевшие	1 000	50	95	5000
Не выздоровевшие	9 000	950	5	5000

Пусть

$$A = \{\text{выздоровевшие}\},$$
$$B = \{\text{лечившиеся}\},$$
$$C = \{\text{городские жители}\}.$$

Выборочное пространство может быть разбито на части сначала относительно A и B; затем относительно A, B и C. Результаты показаны на диаграмме:



Вот разнообразные условные вероятности, точнее, — классификационные пропорции:

$$P(A \mid B) = \frac{1095}{10\,100} = \text{около } 10\,\%; \qquad P(A \mid BC) = \frac{1000}{10000}\,;$$
$$P(A \mid B^c) = \frac{5050}{11\,000} = \text{около } 50\,\%; \qquad P(A \mid B^cC) = \frac{50}{1000}\,;$$
$$P(A \mid BC^c) = \frac{95}{100}\,; \qquad P(A \mid B^cC^c) = \frac{5\,000}{10\,000}\,.$$

Таким образом, если результаты лечения (вопрос «жизни и смерти») оцениваются на базе условных вероятностей из левого столбца, то они выглядят ужасно: лечение уменьшает шансы на выживание в пять раз! Но давайте посмотрим на правый столбец, где содержатся данные для городских и сельских жителей по отдельности:

$$P(A \mid BC) = 10\,\%; \qquad P(A \mid B^cC) = 5\,\%;$$
$$P(A \mid BC^c) = 95\,\%; \qquad P(A \mid B^cC^c) = 50\,\%.$$

В обоих случаях шансы на выживания удваиваются в результате лечения.

Приведем объяснение: в силу некоторых причин (таких, как загрязнение) пациенты из класса C имеют намного меньше шансов поправиться по сравнению с пациентами класса C^c. Но большинство пациентов, применявших новый вид лечения, принадлежали классу C. Естественно, следует ожидать, что лечение должно показать плохую результативность, если оно применяется к серьезно больным пациентам.

Арифметическая загадка легко разрешается при помощи следующих точных формул:

$$\begin{aligned}
P(A \mid B) &= \frac{P(AB)}{P(B)} = \frac{P(ABC) + P(ABC^c)}{P(B)} = \\
&= \frac{P(ABC)}{P(BC)}\ \frac{P(BC)}{P(B)} + \frac{P(ABC^c)}{P(BC^c)}\ \frac{P(BC^c)}{P(B)} = \\
&= P(A \mid BC)\, P(C \mid B) + P(A \mid BC^c)\, P(C^c \mid B) = \\
&= \frac{1\,000}{10\,000}\ \frac{10\,000}{10\,100} + \frac{95}{100}\ \frac{100}{10\,100}\,; \\
P(A \mid B^c) &= P(A \mid B^cC)\, P(C \mid B^c) + P(A \mid B^cC^c)\, P(C^c \mid B^c) = \\
&= \frac{50}{1\,000}\ \frac{1\,000}{11\,000} + \frac{5\,000}{10\,000}\ \frac{10\,000}{11\,000}\,.
\end{aligned}$$

Именно эти «скрытые» коэффициенты $P(C \mid B)$, $P(C^c \mid B)$, $P(C \mid B^c)$, $P(C^c \mid B^c)$ служат причиной парадокса. Небольшая притча поможет прояснить приведенную арифметику. Допустим, что в двух семьях и мужья, и жены работают. Муж в семье 1 зарабатывает больше, чем муж в семье 2, и жена в семье 1 зарабатывает больше, чем жена в семье 2. Ради какого-то повода (или развлечения) муж и жена из семьи 2 выделили по половине своего месячного дохода; но в семье 1 муж внес только 5 % от своего заработка, предоставив жене внести сумму, составляющую 95 % от ее месячного дохода. Не замечали ли вы, что при складчине более бедные пары вносят больше других (или тратят больше денег во время отпуска)?

⋆5.6. Генетические модели

В этом параграфе рассматриваются приложения теории вероятностей к генетике. Обсуждаемая здесь вероятностная модель является одной из простейших и наиболее успешных в естественных науках.

Наследуемые признаки *диплоидных*[*)] организмов, в частности — людей, передаются с помощью генов, образующих пары. В простейшем случае каждый ген в паре может иметь две формы, называемые *аллелями*: A и a. К примеру, A может определять «голубой цвет» глаз у человека, а a — «карий цвет»; или A может приводить к появлению «красных цветков» у садового гороха, а a — «белых». Последний послужил первоначальным объектом исследования Менделя (1822–1884). В рассматриваемом случае имеются три *генотипа*:

$$AA,\ Aa,\ aa,$$

причем генотипы Aa и aA не различаются (природа не упорядочивает пары). Для некоторых признаков ген A может выступать в *доминантной* форме, а ген a — в *рецессивной*, в том смысле, что генотип Aa неотличим от AA на основе внешних проявлений соответствующего признака[**)]. Для других признаков Aa может приводить к промежуточным проявлениям, таким как зеленый оттенок глаз или розовая окраска цветков гороха. Репродуктивные клетки, называемые *гаметами*, формирующиеся в результате расщепления генных пар, содержат только один ген из каждой пары. При спаривании каждый из родителей передает потомку один из своих генов с помощью гаметы. Родители, имеющие чистые типы AA или aa, конечно, передают только A или a соответственно, в то время как обладатели смешанного типа Aa могут передавать либо A, либо a, но не оба сразу. Теперь давайте фиксируем генную пару и предположим, что родительские генотипы AA, Aa, aa присутствуют в пропорции

$$u : 2v : w,$$

где $u > 0$, $v > 0$, $w > 0$, $u + 2v + w = 1$. (Множитель 2 в $2v$ введен для упрощения последующих вычислений.) Предполагается, что общее количество носителей данных трех генотипов очень велико, и что спаривание между особями осуществляется случайным образом. При каждом спаривании каждый из родителей передает потомку один из пары имеющихся у него генов с вероятностью $1/2$, независимо от другого родителя

[*)] Биологический термин, относящийся к особям (клеткам) с двумя гомологичными наборами хромосом. — *Прим. перев.*

[**)] Другими словами, присутствие гена a не отражается в фенотипе. — *Прим. перев.*

и остальных пар родителей. В данных обстоятельствах говорят, что имеет место *случайное спаривание.* Скажем, если горох растет по всему огороду, то эти условия выполняются с достаточной точностью. С другой стороны, если посадки гороха производятся в соответствии с окраской его цветков, то спаривание нельзя рассматривать как чисто случайное.

Стохастическая модель может быть описана следующим образом. Две урны содержат очень большое количество монет трех видов: с A на обеих сторонах, с A на одной из сторон и a — на другой, с a на обеих сторонах. Их доли в каждой из урн относятся между собой как $u : 2v : w$. Из каждой урны извлекают по одной монете таким способом, что выбор любой монеты одинаково правдоподобен. Две вынутые монеты подбрасываются. Стороны монет, оказавшиеся наверху после их падения, определяют генотип потомка. Какова вероятность того, что он будет AA, Aa или aa? При некотором желании, используя частотную интерпретацию вероятности, можно многократно повторить эксперимент и получить оценку для распределения типов. Строго говоря, монеты следует каждый раз класть обратно для того, чтобы во всех испытаниях оставались постоянными вероятности извлечения каждого из типов.

Давайте представим в виде таблицы все случаи, когда в результате спаривания потомок приобретает генотип AA. Понятно, что это возможно только тогда, когда у родителей имеются по крайней мере два гена вида A. Возможные случаи перечислены в первом и втором столбцах таблицы. В третьем столбце приводятся вероятности спаривания между особями, обладающими теми двумя генотипами, которые указаны в первых двух ячейках соответствующей строки; в четвертом столбце содержатся условные вероятности для наследования потомком генотипа AA при условии заданных генотипов родителей; в пятом столбце стоят произведения вероятностей из третьего и четверного столбцов. В силу утверждения 2 из § 5.2, итоговая вероятность образования потомка с генотипом AA получается сложением чисел из пятого столбца.

Генотип отца	Генотип матери	Вероятность образования пары	Вероятность получения потомка AA от такой пары	Вероятность потомка AA
AA	AA	$u \cdot u = u^2$	1	u^2
AA	Aa	$u \cdot 2v = 2uv$	1/2	uv
Aa	AA	$2v \cdot u = 2uv$	1/2	uv
Aa	Aa	$2v \cdot 2v = 4v^2$	1/4	v^2

Итак,

$$P(\text{потомок имеет генотип } AA) = u^2 + uv + uv + v^2 = (u+v)^2.$$

Из симметрии, заменяя u на w, заключаем, что

$$P(\text{потомок имеет генотип } aa) = (v+w)^2.$$

Наконец, перечислим все случаи, когда может появиться потомок типа Aa, в таблице, аналогичной приведенной выше.

Генотип отца	Генотип матери	Вероятность образования пары	Вероятность получения потомка Aa от такой пары	Вероятность потомка Aa
AA	Aa	$u \cdot 2v = 2uv$	$1/2$	uv
Aa	AA	$2v \cdot u = 2uv$	$1/2$	uv
AA	aa	$u \cdot w = uw$	1	uw
aa	AA	$w \cdot u = uw$	1	uw
Aa	aa	$2v \cdot w = 2vw$	$1/2$	vw
aa	Aa	$w \cdot 2v = 2vw$	$1/2$	vw
Aa	Aa	$2v \cdot 2v = 4v^2$	$1/2$	$2v^2$

Складывая числа в последнем столбце, находим:

$$\begin{aligned} P(\text{потомок имеет генотип } Aa) &= 2(uv + uw + vw + v^2) = \\ &= 2(u+v)(v+w). \end{aligned}$$

Введем обозначения:

$$p = u + v, \quad q = v + w, \tag{5.6.1}$$

причем $p > 0$, $q > 0$, $p + q = 1$. Обозначим также через $P_n(\ldots)$ вероятность генотипа потомка в n-м поколении. Тогда полученные выше результаты записываются в следующем виде:

$$P_1(AA) = p^2, \quad P_1(Aa) = 2pq, \quad P_1(aa) = q^2. \tag{5.6.2}$$

Они дают пропорции генотипов родителей для второго поколения. Следовательно, для того, чтобы получить вероятности P_2, надо всего лишь подставить p^2 вместо u, pq вместо v, q^2 вместо w в двух предыдущих формулах. Отсюда

$$\begin{aligned} P_2(AA) &= (p^2 + pq)^2 = p^2, \\ P_2(Aa) &= 2(p^2 + pq)(pq + q^2) = 2pq, \\ P_2(aa) &= (pq + q^2)^2 = q^2. \end{aligned}$$

Вот это да: P_2 совпадает с P_1! Означает ли это, что и P_3 также совпадает с P_1 и т. д.? Это вывод верен, но только с учетом изложенных ниже соображений. Важно, что мы установили равенство $P_1 = P_2$ *для произвольного* P_0 (введенные из предосторожности условия $u > 0$, $v > 0$, $w > 0$ на самом деле могут быть опущены). Двигаясь от поколения к поколению, выводим, что $P_2 = P_3$ даже в случае, если пропорция P_1 окажется не такой, как P_0. Оставшиеся рассуждения очевидны, и сам результат известен под именем теоремы Харди—Вайнберга*).

Теорема. *При случайном спаривании пары генов распределение генотипов, начиная с первого поколения, становится стационарным, независимо от того, каким было первоначальное распределение.*

Давайте припишем числовые значения 2, 1, 0 трем типам AA, Aa, aa в соответствии с количеством генов вида A в паре. Пусть X_n обозначает случайную величину, представляющую собой числовой генотип n-го поколения. Тогда в теореме утверждается, что для $n \geqslant 1$

$$P(X_n = 2) = p^2, \quad P(X_n = 1) = 2pq, \quad P(X_n = 0) = q^2. \tag{5.6.3}$$

Распределение величин X_n стационарно в том смысле, что вероятности не зависят от n. На самом деле можно показать, что случайный процесс $\{X_n, n \geqslant 1\}$ стационарен в описанном в § 5.4 смысле, поскольку он также является цепью Маркова, см. задачу 40.

Результат, содержащийся в формуле (5.6.2), может быть получен даже в более простой модели, чем та, которая обсуждалась выше. Вместо генных пар рассмотрим множество гамет, возникших в результате расщепления пар на индивидуальные гены. Тогда гены вида A и a первоначально присутствуют в пропорции

$$(2u + 2v) : (2v + 2w) = p : q,$$

поскольку в типе AA содержится два A-гена и т. д. Тогда можно представлять себе гаметы в виде множества жетонов в урне, на каждом из которых написана буква A или a, и связать рождение потомка со случайным извлечением (с возвращением) двух гамет, из которых будет образована пара. При этом вероятности получения AA, Aa, aa равны, соответственно,

$$p \cdot p = p^2, \quad p \cdot q + q \cdot p = 2pq, \quad q \cdot q = q^2.$$

Это — тот же самый результат, какой содержится в формуле (5.6.2).

*) Годфри Харолд Харди (1877–1947) — один из ведущих английских математиков. Его основные достижения относятся к теории чисел и классическому математическому анализу.

Новая модель отличается от старой, но приводит к тому же ответу. Возникает искушение считать эти модели идентичными, но единственный логичный подход к этому состоит в том, чтобы подробно разобрать оба случая, как мы и сделали. Вообще говоря, они не эквивалентны. Рассмотрим, например, размножение рыб: сначала самки откладывают миллиарды икринок, а затем приплывают самцы и оплодотворяют их своей спермой. Родители могут никогда не встречаться. В данных обстоятельствах вторая модель отражает картину лучше, особенно если использовать две отдельные урны для икринок и сперматозоидов. (На свете существуют создания, у которых нет разделения полов, и для которых подходит одноурновая модель.) Такую модель можно назвать *моделью нереста*, в отличие от описанной ранее *модели спаривания.* В более сложных случаях, где задействована не только одна пара генов, эти две модели не обязательно приводят к одинаковому результату.

Пример 15. В человеческой генетике известны определенные «плохие» гены, наличие которых приводит к тому, что ребенок рождается калекой или неизлечимо больным. Если a — такой ген, то человек с генотипом aa не доживает до взрослого состояния. Человек с генотипом Aa является *носителем*, но это не проявляется в его внешнем виде из-за того, что a имеет рецессивный характер. Допустим, что вероятность встретить носителя данного гена среди всей популяции равна p независимо от пола. Тогда, если у некоторого человека был больной брат или сестра, умершие в детском возрасте, то он несет на себе груз «плохой наследственности», и его нельзя с генетической точки зрения считать равноправным членом популяции. Вычислим условную вероятность того, что он является носителем. Оба его родителя должны быть носителями, т. е. обладают генотипом Aa, иначе бы они не смогли произвести на свет ребенка с генотипом aa. Поскольку каждый из генов передается с вероятностью $1/2$, вероятности того, что их ребенок будет иметь генотип AA, Aa, aa равны $1/4$, $1/2$, $1/4$ соответственно. Так как человек, о котором идет речь, выжил, то он не может иметь генотип aa. Поэтому условные вероятности того, что его генотип есть AA или Aa таковы:

$$P(AA \mid AA \cup Aa) = \frac{1}{3}, \quad P(Aa \mid AA \cup Aa) = \frac{2}{3}.$$

Если он женится на женщине, о которой известно, что у ее предков не было случаев проявления «плохой наследственности» данного вида, то она обладает генотипом AA или Aa, соответственно, с вероятностями $1 - p$ и p, характерными для всей популяции. Вероятности возможного генотипа их детей приведены в следующей таблице:

Отец	Мать	Вероятность сочетания	Вероятность получения AA	Вероятность получения Aa	Вероятность получения aa
AA	AA	$(1/3)(1-p)$	1	0	0
AA	Aa	$(1/3)p$	1/2	1/2	0
Aa	AA	$(2/3)(1-p)$	1/2	1/2	0
Aa	Aa	$(2/3)p$	1/4	1/2	1/4

Простым вычислением получаем следующее распределение генотипов потомков:

$$P_1(AA) = \frac{2}{3} - \frac{p}{3}, \quad P_1(Aa) = \frac{1}{3} + \frac{p}{6}, \quad P_1(aa) = \frac{p}{6}.$$

Отсюда находим вероятность того, что выживший ребенок является носителем:

$$P_1(Aa \mid AA \cup Aa) = \frac{2+p}{6-p}.$$

Если p пренебрежимо мало, эта вероятность приблизительно равна 1/3. Поэтому с точки зрения выжившего ребенка то, что его отец произвел на свет больного ребенка, вдвое хуже в смысле наследственного риска, чем то, что болен его дядя (или тетя). Можно пойти дальше и вычислить шансы *его* детей, внуков и т. д. Мы оставляем это читателю в качестве упражнения.

Завершая обсуждение данного примера, касающегося серьезных заболеваний, важно подчеркнуть, что простую математическую теорию следует рассматривать только в качестве грубого приближения к действительности, поскольку не были учтены другие генетические факторы.

Задачи

1. Воспользуемся данными из примера 14 в § 5.5. Какова вероятность того, что (a) выживший пациент является городским жителем, (b) выживший лечившийся пациент является сельским жителем?

2. Все болты в машине произведены на одном из заводов A или B, причем оба случая одинаково правдоподобны. Доля дефектных болтов составляет 5 % на заводе A и 1 % на заводе B. Проверяются два болта. Если первый болт оказался качественным, то какова вероятность того, что и второй болт такой же?

3. В электронном приборе имеются лампы двух типов. Прибор не работает тогда и только тогда, когда есть бракованные лампы обоих типов. Вероятность того, что бракованы лампы первого типа, равна 0.1, второго типа — 0.2. Известно, что две лампы бракованы. Какова вероятность того, что, несмотря на это, прибор работает?

4. При условии, что в результате бросания трех идеальных игральных костей на всех костях выпали разные номера, найдите вероятность того, что (a) по крайней мере один из номеров есть «6», (b) сумма номеров равна 8.

5. Рассмотрим семьи с тремя детьми и предположим, что каждый из детей является мальчиком или девочкой с вероятностью 1/2. Если в случайно выбранной такой семье старший ребенок оказался мальчиком, то чему равна вероятность того, что остальные два ребенка — девочки? Тот же самый вопрос, если мальчиком оказался случайно выбранный в семье ребенок.

6. Вместо того, чтобы выбирать случайно семью в задаче 5, представим теперь, что наудачу выбирается ребенок из всех детей во всех семьях. Если он оказался мальчиком, то какова вероятность того, что у него есть две сестры?

7. Выбирается семья, как в задаче 5, а затем в ней случайно выбираются два ребенка. Если они оказались девочками, то какова вероятность того, что у них есть брат?

8. Предположим, что α есть вероятность того, что оба близнеца окажутся мальчиками, а β — что оба окажутся девочками. Допустим также, что если близнецы имеют разный пол, то вероятность того, что первый близнец — девочка, равна 1/2. Если первый близнец оказался девочкой, то какова вероятность того, что второй близнец также девочка?

9. Три стрелка попадают в цель с вероятностями 1/2, 2/3, 3/4 соответственно. Они выстрелили одновременно, причем двое попали. Кто промахнулся? Найдите вероятности.

10. При перелете из Урбаны в Париж мой багаж не прибыл вместе со мной. От пересылался трижды, и вероятности того, что каждая из пересылок не была произведена вовремя, равны 4/10, 2/10, 1/10, соответственно очередности пересылок. Чему равна вероятность того, что оплошала первая из авиалиний?

11. Докажите соотношения, которые можно назвать принципом «так и должно быть»: если

$$P(A \mid C) \geqslant P(B \mid C),$$
$$P(A \mid C^c) \geqslant P(B \mid C^c),$$

то $P(A) \geqslant P(B)$.

12. Покажите, что если $A \parallel B$, то $A^c \parallel B^c$, $A \nparallel B^c$, $A^c \nparallel B$.

13. Убедитесь, что если $A \cap B = \varnothing$, то

(i) $A \parallel C$ и $B \parallel C \Rightarrow (A \cup B) \parallel C$;

(ii) $C \parallel A$ и $C \parallel B \Rightarrow C \parallel (A \cup B)$;

(iii) A и C независимы, B и C независимы $\Rightarrow A \cup B$ и C независимы.

14. Пусть $P(H) > 0$. Покажите, что функция множеств

$$S \to P(S \mid H) \quad \text{для} \quad S \in \Omega \text{ (счетное множество)}$$

является вероятностной мерой.

15.* Постройте примеры для всех утверждений (5.5.11). [*Указание.* Систематический, но утомительный способ сделать это, таков: припишите вероятности $p_1, \ldots, p_s$ каждому из атомов $ABC, \ldots, A^cB^cC^c$ (см. (1.3.5)) и выразите искомые вероятности через них. Работу можно сократить за счет предварительного выбора простых значений для p_i, скажем, положить некоторые из них равными нулю, а остальные взять одинаковыми. Можно также изобрести примеры на основе разных простых свойств небольших наборов целых чисел; см. статью первого из авторов “On mutually favorable events” в *Annals of Mathematical Statistics*, T. 13 (1942), c. 338–349.]

16. Пусть $A_j, 1 \leqslant j \leqslant 5$ — независимые события. Покажите, что события

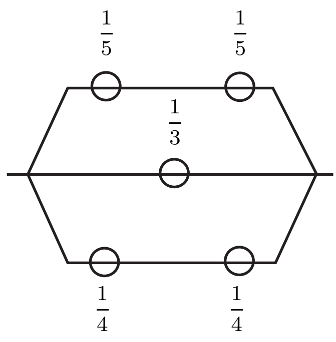
(i) $(A_1 \cup A_2)A_3$ и $A_4^c \cup A_5^c$ независимы,

(ii) $A_1 \cup A_2$, $A_3 \cap A_3$ и A_5^c независимы.

17. Допустим, что в некотором казино стоят «однорукие бандиты» трех видов. Автоматы каждого вида имеются в одинаковом количестве, а вероятности выигрыша на них равны соответственно 1/3, 1/2, 2/3. Один из автоматов раскошелился дважды в четырех играх. Какова вероятность выигрыша в следующей игре?

18. Человек последовательно проходит четыре теста. Вероятность успешного выполнения первого теста равна p, затем вероятность выполнения каждого последующего теста равна p или $p/2$ в зависимости от того, был успех или неудача при прохождении предыдущего. Человек аттестуется, если у него всего оказалось не менее трех успехов. Каковы шансы аттестоваться?

19. Электрическая цепь устроена так, как показано на рисунке, где числа задают вероятности сбоев на соответствующих участках цепи, которые предполагаются независимыми. Какова вероятность того, что цепь пропускает ток?



20. В некотором городе половину времени идет дождь, а прогноз погоды подтверждается с вероятностью $2/3$. Господин Гренкин выходит из дому каждый день и очень обеспокоен возможностью дождя. Поэтому он обязательно берет с собой зонт, если прогноз обещает дождь, и также берет его в одном случае из трех, когда прогноз обещает ясную погоду. Найдите

(a) вероятность того, что он попадет под дождь, оказавшись без зонта,

(b) вероятность того, что он возьмет с собой зонт, а дождя не будет.

Это — два вида ошибок, рассматриваемых в статистической теории Неймана и Пирсона*). [*Указание.* Вычислите вероятности событий {пошел дождь, когда прогноз был «ясно»}, {прогноз был «ясно», зонт не был взят} и т. п.]

21. В телеграфном сигнале «точки» и «тире» появляются в пропорции 3:4. В силу некоторых причин, вызывающих очень сильное искажение сообщения, «точки» заменяются на «тире» с вероятностью $1/4$, в то время как «тире» превращаются в «точки» с вероятностью $1/3$. Если получена «точка», то какова вероятность того, что этот символ был «точкой» при отправлении сообщения?

22. A сказал, что B сообщил ему, что C сказал неправду. Если каждый из этих людей говорит правду с вероятностью p, то какова вероятность того, что C на самом деле солгал? (Трудно поверить, но этот тип вопросов одно время серьезно обсуждался как проблема «достоверности свидетельских показаний». При изложении таких задач на обычном языке они выглядят очень запутанно и требуют большого количества слов для своего объяснения. Чтобы подробно рассмотреть один случай, допустим, что все три человека солгали. Тогда B скажет A, что C сказал правду, так как, предполагается что, B знает, обманывает ли C или нет, но решил сам солгать; A скажет, что B сообщил ему, что C сказал неправду, так как он, не зная, что на самом деле сделал C, намерен солгать о том, что B сказал ему. Это только один из восьми возможных случаев, но остальные рассматриваются аналогично. Намного более прозрачно выглядит формулировка в терминах модели передачи сигналов из задачи 21. C передает «$-$» или «$+$» в соответствии с тем, лжет он или говорит правду; затем B передает сообщение от C верно или неверно, если он лжет или говорит правду; затем A передает сообщение от B аналогичным образом. Здесь не возникает смыслового тупика, даже если мы обобщим задачу на любое количество свидетелей. Вопрос тогда звучит так: если в конце цепочки символов получен «$-$», то какова вероятность того, что он был отправлен первоначально?)

23. Частица стартует из 0 и движется по целым точкам прямой, делая шаг длины 1 налево или направо с вероятностью $1/2$. Шаги предполагаются

*) Мы предполагаем, что читатель сумеет перевести словесные описания в вероятностные термины и получить ответ. В реальной жизни нередко делаются выводы и принимаются решения на основе еще более размытых посылок.

независимыми. Пусть Y_n обозначает ее положение после n шагов. Найдите следующие вероятности:

(a) $P(Y_n \geqslant 0$ для $1 \leqslant n \leqslant 4)$;

(b) $P(|Y_n| \leqslant 2$ для $1 \leqslant n \leqslant 4)$;

(c) $P(Y_n \geqslant 0$ для $1 \leqslant n \leqslant 4 \mid Y_4 = 0)$.

24. В условиях задачи 23 докажите, что если $j < k < n$, то

$$P(Y_n = c \mid Y_j = a, Y_k = b) = P(Y_n = c \mid Y_k = b) = P(Y_{n-k} = c - b),$$

где a, b, c — любые целые числа. Вычислите эти вероятности при $j = 4$, $k = 6$, $n = 10$, $a = 2$, $b = 4$, $c = 6$.

25. Сначала бросается идеальная игральная кость, затем — симметричные монеты в количестве, равном номеру, выпавшему на кости.

(a) Какова вероятность получить в результате k «орлов»?

(b) Если получено 3 «орла», то чему равна вероятность того, что на кости выпал номер n?

26. Из некоторой частицы в ядерной реакции могут образоваться 2 или 3 новые частицы, или же она вовсе не разделится. Вероятности этих возможностей равны p_2, p_3 и p_1. Новые частицы ведут себя так же, причем независимо друг от друга и от предыдущей реакции. Найдите распределения общего числа частиц, образовавшихся в результате двух реакций.

27. Идеальная игральная кость бросается n раз. Пусть M и m обозначают максимум и минимум из выпавших номеров. Найдите $P(m = 2, M = 5)$. [*Указание.* Начните с $P(m \geqslant 2, M \leqslant 5)$.]

28. Пусть X и Y — независимые случайные величины с одинаковым вероятностным распределением $\{p_n, n \geqslant 1\}$. Найдите $P(X \leqslant Y)$ и $P(X = Y)$.

В задачах 29–32 рассматривается пара случайных чисел, выбранных наудачу из отрезка $[0, 1]$.

29. Если наименьшее из чисел меньше $1/4$, то какова вероятность того, что наибольшее из чисел больше $3/4$?

30*. При условии, что наименьшее из чисел меньше x, найдите распределение наибольшего из чисел. [*Указание.* Рассмотрите $P(\min < x, \max < y)$ в двух случаях: $x \leqslant y$ и $x > y$.]

31*. Две выбранные точки разбивают отрезок $[0, 1]$ на три отрезка. Чему равна вероятность того, что из этих отрезков можно образовать треугольник? [*Указание.* Это возможно тогда и только тогда, когда сумма длин любых двух отрезков превосходит длину третьего. Обозначив точки через X и Y, исследуйте сначала случай, когда $X < Y$.]

32. Докажите, что длины трех отрезков из задачи 31 одинаково распределены. [*Указание.* Рассмотрите распределение координаты наименьшей из двух выбранных точек. Затем распределение длины отрезка между точками.

Наконец, используйте соображения симметрии для нахождения распределения расстояния между наибольшей из точек и 1.]

33. В урновой модели Пойа найдите:

(a) $P(R_3 \mid B_1R_2)$; (b) $P(R_3 \mid R_1R_2)$;
(c) $P(R_3 \mid R_2)$; (d) $P(R_1 \mid R_2R_3)$;
(e) $P(R_1 \mid R_2)$; (f) $P(R_1 \mid R_3)$.

34. Рассмотрим две урны U_i, $i = 1, 2$, содержащие r_i красных и b_i черных шаров соответственно. Из U_1 случайно извлекается шар и кладется в U_2, затем из U_2 случайно извлекается шар и кладется в U_1. Какова после этого вероятность извлечения красного шара из U_1? Покажите, что если $b_1 = r_1$, $b_2 = r_2$, то эта вероятность такая же, как если бы не было никаких перекладываний шаров.

35*. Допустим, что априорная вероятность p в задаче о восходе солнца (см. пример 9 в § 5.2) может принимать только значения $k/100, 1 \leqslant k \leqslant 100$, с вероятностью $1/100$ каждое. Найдите $P(S^{n+1} \mid S^n)$. Замените число 100 на N и устремите $N \to \infty$. Что получится в пределе?

36*. Докажите, что события $A_1, \ldots, A_n$ независимы тогда и только тогда, когда

$$P(\tilde{A}_1 \ldots \tilde{A}_n) = P(\tilde{A}_1) \ldots P(\tilde{A}_n),$$

где каждое $\tilde{A}_j$ может быть как A_j, так и A_j^c. [*Указание.* Для вывода этих равенств из независимости используйте индукцию по n и также индукцию по k в $P(A_1^c \ldots A_k^c A_{k+1} \ldots A_n)$; обратное утверждение легко получить индукцией по n.]

37*. Докажите теорему 2 из § 5.3. [*Указание.* Занумеруйте все шары и покажите, что любая фиксированная последовательность шаров имеет ту же самую вероятность расположения шаров на заданных местах, как если бы все шары вынимались по порядку.]

38. Проверьте теорему 5 из § 5.4 непосредственно на парах (X_1, X_2), (X_1, X_3) и (X_2, X_3).

39. Допустим, что три генотипа AA, Aa, aa присутствуют в пропорции $p^2 : 2pq : q^2$, где $p + q = 1$. Если два родителя, выбранных наудачу из популяции, имеют потомка с генотипом AA, то чему равна вероятность того, что их другой ребенок также будет обладать генотипом AA? Дайте ответ на тот же вопрос с заменой AA на Aa.

40. Пусть X_1 и X_2 обозначают генотипы матери и ее ребенка. Предположим, что неизвестный генотип отца распределен так же, как в задаче 39. Используя обозначения из формулы (5.6.3), найдите следующие 9 условных вероятностей:

$$P\{X_2 = k \mid X_1 = j\}, \quad j = 0, 1, 2; \quad k = 0, 1, 2.$$

Они называются *переходными вероятностями* цепи Маркова; см. § 8.3.

41.* Докажите, что если функция φ, заданная на луче $[0,\infty)$, является невозрастающей и удовлетворяет *функциональному уравнению Коши*

$$\varphi(s+t)=\varphi(s)\,\varphi(t),\quad s\geqslant 0,\quad t\geqslant 0;$$

то $\varphi(t)=e^{-\lambda t}$ для некоторого $\lambda\geqslant 0$. Следовательно, положительная случайная величина T обладает свойством

$$P(T>s+t\mid T>s)=P(T>t),\quad s\geqslant 0,\quad t\geqslant 0,$$

тогда и только тогда, когда она имеет экспоненциальное распределение. [*Указание.* $\varphi(0)=1$, $\varphi(1/n)=\alpha^{1/n}$, где $\alpha=\varphi(1)$, $\varphi(m/n)=\alpha^{m/n}$; если $m/n\leqslant t<(m+1)/n$, то $\alpha^{(m+1)/n}\leqslant\varphi(t)\leqslant\alpha^{m/n}$. Отсюда, устремляя $n\to\infty$, получаем доказываемое утверждение.]

42. Игла единичной длины бросается на стол, расчерченный параллельными прямыми, находящимися друг от друга на постоянном расстоянии d, где $d>1$. Обозначим через x расстояние от середины иглы до ближайшей линии, а через θ — угол между иглой и перпендикуляром, опущенным из середины иглы на ближайшую линию. Предположим, что x и θ — независимые случайные величины, причем каждая равномерно распределена на своем диапазоне. Чему равна вероятность того, что игла пересечет линию? Эта задача известна как *задача Бюффона*, и ее решение позволяет предложить эмпирический (Монте-Карло) метод для оценки значения числа π.

Глава 6

Среднее, дисперсия и преобразования случайных величин

6.1. Основные свойства математического ожидания

Математическое ожидание случайной величины, определенное в § 4.3, является одним из наиболее важных понятий теории вероятностей. Как мы увидим, оно играет такую же роль, какую в математическом анализе играет понятие интеграла — мы знаем, что теория интегрирования составляет добрую половину математического анализа. Напомним, что математическое ожидание представляет собой (на счетном выборочном пространстве) взвешенное среднее, где весами служат вероятности, и перепишем формулу (4.3.11) в более простом виде:

$$E(X) = \sum_{\omega} X(\omega)P(\omega). \tag{6.1.1}$$

Если подставить сюда $|X|$ вместо X, то видим, что условие (4.3.12) можно записать в виде

$$E(|X|) < \infty. \tag{6.1.2}$$

Когда выполняется условие (6.1.2), будем говорить, что случайная величина X *интегрируема*. В данном случае говорят также, что «X имеет конечное математическое ожидание (или среднее)» или «ожидание X существует». Последняя формулировка не совсем точна, поскольку, вообще говоря, позволяется, чтобы $E(X)$ было определено и принимало значение $+\infty$, например, когда $X \geqslant 0$ и ряд в (6.1.1) расходится. (См. задачи 27 и 28 в гл. 4.) Мы будем особо оговаривать данную ситуацию, когда она встретится.

Ясно, что если случайная величина X ограничена, т. е. когда существует такая константа M, что

$$|X(\omega)| \leqslant M \quad \text{для всех } \omega \in \Omega,$$

то X является интегрируемой, причем

$$E(|X|) = \sum_{\omega} |X(\omega)|\, P(\omega) \leqslant M \sum_{\omega} P(\omega) = M.$$

В частности, если Ω конечно, то всякая случайная величина ограничена (это не означает, что все случайные величины ограничены одной

и той же константой). Таким образом, класс случайных величин, имеющих конечное математическое ожидание, достаточно широк. Для этого класса отображение
$$X \to E(X) \tag{6.1.3}$$
сопоставляет случайной величине действительное число. К примеру, если X — рост учеников в школе, то $E(X)$ — их средний рост; если X — доход наемных работников, то $E(X)$ — их средний доход; если X — число транспортных средств, проезжающих за день через таможенный пост, то $E(X)$ — средний дневной транспортный поток и т. п.

Для произвольного события A его индикатор I_A (см. § 1.4) является случайной величиной, причем $E(I_A) = P(A)$. В этом смысле математическое ожидание обобщает понятие вероятностной меры.

Напомним, что если X и Y — случайные величины, то $X+Y$ — тоже случайная величина (утверждение 1 в § 4.2). Если X и Y обе имеют конечное математическое ожидание, интуитивно понятно, что и ожидание их суммы $X+Y$ также будет конечным. Благодаря естественной форме нашего определения, эту теорему легко доказать.

Теорема 1. *Если случайные величины X и Y интегрируемы, то таковой является и случайная величина $X+Y$, причем*
$$E(X+Y) = E(X) + E(Y). \tag{6.1.4}$$

Доказательство. Применяя определение (6.1.1) к сумме $X+Y$, имеем:
$$\begin{aligned} E(X+Y) &= \sum_{\omega} (X(\omega) + Y(\omega))\, P(\omega) = \\ &= \sum_{\omega} X(\omega)P(\omega) + \sum_{\omega} Y(\omega)P(\omega) = E(X) + E(Y). \end{aligned}$$

Дело сделано. Вы можете удивиться, почему не потребовалось проверять условие (6.1.2)? Ответ таков: из интегрируемости слагаемых мы автоматически получаем, что ряд, определяющий $E(X+Y)$, абсолютно сходится, как требовалось в § 4.3. Действительно,
$$\begin{aligned} \sum_{\omega} |X(\omega) + Y(\omega)|\, P(\omega) &\leqslant \sum_{\omega} (|X(\omega)| + |Y(\omega)|)\, P(\omega) = \\ &= \sum_{\omega} |X(\omega)|\, P(\omega) + \sum_{\omega} |Y(\omega)|\, P(\omega) < \infty. \end{aligned}$$

Теорема 1 может показаться читателю тривиальной, но она отражает наиболее фундаментальное свойство математического ожидания. Выпишем еще два свойства этой характеристики:
$$E(a) = a, \quad E(aX) = aE(X) \tag{6.1.5}$$

для любой константы a; объединяя формулы (6.1.4) и (6.1.5), получаем, что

$$E(aX + bY) = aE(X) + bE(Y) \tag{6.1.6}$$

для любых двух констант a и b. Это свойство означает, что отображение (6.1.3) является *линейным оператором.* Это — важное понятие в математике; возможно, читатель встречался с ним при изучении линейной алгебры или дифференциальных уравнений.

Используя математическую индукцию, легко обобщить равенство (6.1.4): если $X_1, X_2, \ldots, X_n$ — интегрируемые случайные величины, то такова же и их сумма, причем

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n). \tag{6.1.7}$$

Прежде чем перейти к обсуждению других свойств математического ожидания, давайте применим то, что уже знаем, к некоторым интересным задачам.

Пример 1. В лотерее участвуют 100 билетов. На один из них приходится приз \$10 000, а на остальные — ничего. Если я куплю два билета, то чему равен мой ожидаемый выигрыш?

Если я приобрету только один билет, то мой выигрыш представляется случайной величиной X, которая имеет значение 10 000 в точности на одной точке ω и значение 0 — на всех остальных точках. Поскольку билеты предполагаются одинаково правдоподобными в отношении возможности выигрыша, то

$$X = \begin{cases} 10\,000 & \text{с вероятностью } 1/100, \\ 0 & \text{с вероятностью } 99/100, \end{cases}$$

отсюда

$$E(X) = 10\,000 \cdot \frac{1}{100} + 0 \cdot \frac{99}{100} = 100.$$

Итак, мой ожидаемый выигрыш составляет \$100. Это тривиально. Но теперь, если у я куплю два билета, то мне точно известно, что лишь один из них может оказаться призовым. Тем самым, имеется определенная связь (зависимость) между двумя случайными величинами, представляющими выигрыши билетов. Повлияет ли это на мои ожидания? Задумаемся немного глубже: если я не первый среди участников лотереи, возможно к тому моменту, когда я приобретаю билеты, кто-то уже выиграл приз, но это еще никому не известно. Есть ли различие между шансами ранних и поздних покупателей? Ответы на эти вопросы уже были даны при обсуждении урновой модели в § 5.3. Достаточно отождествить билеты со 100 шарами, среди которых ровно 1 шар — черный. Тогда, если обозначить результат n-го извлечения через X_n, то в силу теоремы 1 из § 5.3 каждая из случайных величин X_n будет иметь точно такое же вероятностное распределение, как рассмотренная выше случайная величина X, и, следовательно, — то же самое математическое ожидание. При $n = 2$ это легко вычисляется непосредственно без обращения за помощью к общей теореме. Тем самым, для произвольных j и k, т. е. для любых двух билетов, купленных в любое время, их общий ожидаемый выигрыш равен

$$E(X_j + X_k) = E(X_j) + E(X_k) = 2E(X) = 200.$$

Вообще мой ожидаемый выигрыш прямо пропорционален количеству купленных билетов — достаточно естественный вывод, но был ли он очевиден с самого начала? В частности, если я скуплю все 100 билетов, то гарантировано получу $100\ E(X) = 10\,000$ долларов. Вероятно, это звучит банально, но зато легко проверяется.

Чтобы углубиться еще на шаг, давайте рассмотрим две совершенно одинаковые лотереи. Вместо покупки двух билетов X и Y на одной из них, я предпочту приобрести по одному билету на каждой. Теперь у меня есть шанс выиграть \$20 000. Делает ли это такую схему более привлекательной для меня? Однако теорема 1 утверждает, что мой ожидаемый выигрыш равен \$200 в любом случае. Как это объяснить? Для ответа на данный вопрос вам следует найти распределение суммы $X+Y$ в каждой из схем и вычислить на его основе $E(X+Y)$. Вы много узнаете, сравнивая результаты.

Пример 2 (***задача собирания купонов***)**.** В мешке лежат N купонов, занумерованных числами от 1 до N. Купоны вынимаются наудачу один за другим с возвращением. Предположим, что мы хотим собрать r разных купонов. Какое в среднем количество извлечений для этого потребуется? Данная задача возникает перед школьниками, собирающими изображения известных бейсболистов, или перед людьми, желающими выиграть швейную машинку, если соберут полный набор купонов, помещаемых в пачки с некоторым моющим средством. В последнем случае возможно мошенничество, заключающееся в том, что какие-то из необходимых купонов встречаются крайне редко. Здесь, конечно, мы будем обсуждать честный подход, при котором все купоны одинаково правдоподобны, а последовательные извлечения независимы.

Данную задачу можно отнести к задачам о времени ожидания: мы ждем r-го появления. Пусть $X_1, X_2, \ldots$ обозначают последовательные времена ожидания нового купона. При этом $X_1 = 1$, так как первый купон — всегда новый. Тогда X_2 — время до появления любого купона, отличного от первого. Поскольку при каждом извлечении всего имеется N купонов, и все, кроме одного, являются новыми, то мы оказывается в условиях примера 8 из § 4.4 с вероятностью «успеха» $p = (N-1)/N$. Поэтому

$$E(X_2) = \frac{N}{N-1}\,.$$

После того как будут собраны два разных купона, время ожидания третьего нового купона будет распределено аналогично, но с вероятностью «успеха» $p = (N-2)/N$. Отсюда

$$E(X_3) = \frac{N}{N-2}\,.$$

Продолжая эти рассуждения, для $1 \leqslant r \leqslant N$ получаем равенство

$$\begin{aligned} E(X_1 + \ldots + X_r) &= \frac{N}{N} + \frac{N}{N-1} + \ldots + \frac{N}{N-r+1} = \\ &= N\left(\frac{1}{N-r+1} + \ldots + \frac{1}{N}\right). \end{aligned}$$

В частности, если $r = N$, то

$$E(X_1 + \ldots + X_N) = N\left(1 + \frac{1}{2} + \ldots + \frac{1}{N}\right). \tag{6.1.8}$$

Если же N — четное число и $r = N/2$, то

$$E(X_1 + \ldots + X_{N/2}) = N\left(\frac{1}{(N/2)+1} + \ldots + \frac{1}{N}\right). \tag{6.1.9}$$

Известная формула из математического анализа утверждает, что

$$1 + \frac{1}{2} + \ldots + \frac{1}{N} = \ln N + C + \varepsilon_N, \tag{6.1.10}$$

где «ln» обозначает натуральный логарифм, т. е. логарифм по основанию e, $C = 0.5772\ldots$ — *постоянная Эйлера* (никто в мире не знает, является она рациональным числом или нет), ε_N стремится к нулю при $N \to \infty$. Для многих целей годится более грубая формула:

$$\lim_{N\to\infty} \frac{1}{\ln N}\left(1 + \frac{1}{2} + \ldots + \frac{1}{N}\right) = 1. \tag{6.1.11}$$

Если подставить ее в равенства (6.1.8) и (6.1.9), то увидим, что для больших N правые части будут примерно равны $N \ln N$ и $N \ln 2 \approx 0.7\,N$ соответственно. (Откуда взялся $\ln 2$ во втором из приближений?) Это означает следующее: количество испытаний (извлечений), необходимое для сбора половины купонов, составляет немногим более половины от их числа, но потребуется «бесконечно» много испытаний для того, чтобы собрать все купоны. Последние несколько купонов получить труднее всего, даже если игра не подстроена.

Еще один устрашающий, но не такой реалистичный, пример применения этой задачи дает изучение эффективности бомбометания во время войны. Результаты поражения в достаточной степени случайны, так как зависят от маскировки, сооружения противником ложных целей, погодных условий и интенсивности заградительного огня. Допустим, что необходимо поразить 100 целей, но бомбы сбрасываются на них случайно, возможно многократно. Тогда потребуется «в среднем» около $100 \ln 100 \approx 460$ бомб для того, чтобы поразить каждую из целей по крайней мере единожды. Таким образом, если противник располагает большим количеством пусковых установок для ответного удара, будет очень трудно уничтожить их все без применения высокоточного оружия. Данный вывод может служить математическим возражением против теории упреждающего удара.

6.2. Случай, когда есть плотность

Возвращаясь к более мирным вопросам, обобщим теорему 1 на случай величин, заданных на произвольном выборочном пространстве. На самом деле, она верна и в общем случае при условии, что математическое ожидание определено надлежащим образом. Некоторый намек на это дает запись математического ожидания в виде абстрактного интеграла:

$$E(X) = \int_{\Omega} X(\omega)\, d\omega, \tag{6.2.1}$$

где «$d\omega$» обозначает «вероятность элемента ω», как это обычно понимается в случае площади или объема в многомерном анализе — так называемый *дифференциал*. В данных обозначениях равенство (6.1.4) записывается в виде

$$\int_{\Omega} (X(\omega) + Y(\omega))\, d\omega = \int_{\Omega} X(\omega)\, d\omega + \int_{\Omega} Y(\omega)\, d\omega, \tag{6.2.2}$$

что является полной аналогией известной формулы из анализа:

$$\int_{I} (f(x) + g(x))\, dx = \int_{I} f(x)\, dx + \int_{I} g(x)\, dx,$$

где I — это некоторый интервал, скажем, $[0, 1]$. Помните ли вы, как доказывается данное равенство? Оно получается с помощью определения интеграла (Римана) на основе его приближения интегральными суммами (Римана). Для вероятностного интеграла из формулы (6.2.1) проводится аналогичная процедура. Он определяется как предел математических ожиданий дискретных случайных величин (введенных в § 4.5), аппроксимирующих непрерывную случайную величину X. Математические ожидания аппроксимирующих величин задаются формулой (6.1.1), и к ним применима теорема 1. Общий результат (6.2.2) получается предельным переходом.

Мы не станем в этой книге вдаваться в подробности данного подхода, так как он требует применения теории меры. Но в случае, когда пара (X, Y) имеет совместную плотность (см. § 4.3), для доказательства теоремы 1 можно применить «ловкий трюк». Используя обозначения из (4.6.7), запишем:

$$E(X) = \int_{-\infty}^{\infty} uf(u, *)\,du, \quad E(Y) = \int_{-\infty}^{\infty} vf(*, v)\,dv. \tag{6.2.3}$$

С другой стороны, если подставить функцию $\varphi(x, y) = x + y$ в формулу (4.6.8), то получим

$$E(X + Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (u + v)f(u, v)\,du\,dv. \tag{6.2.4}$$

Теперь этот двойной интеграл можно разбить на два и каждый из них вычислить с помощью сведения к повторному:

$$\int_{-\infty}^{\infty} u\,du \left[\int_{-\infty}^{\infty} f(u, v)\,dv\right] + \int_{-\infty}^{\infty} v\,dv \left[\int_{-\infty}^{\infty} f(u, v)\,du\right] =$$

$$= \int_{-\infty}^{\infty} uf(u, *)\,du + \int_{-\infty}^{\infty} vf(*, v)\,dv.$$

Сравнение с формулами (6.2.3) позволяет установить справедливость равенства (6.1.4).

Ключом к использованному методу служит формула (6.2.4), которая не была доказана. Обычно приводятся такие разъяснения: «Теперь посмотрите сюда: если X имеет значение u, Y имеет значение v, то $X + Y$ принимает значение $u + v$, а вероятность того, что $X = u$ и $Y = v$ есть

$f(u,v)\,du\,dv$. Понятно?» Разговоры подобного типа можно квалифицировать как «размахивание руками» и «запудривание мозгов». Однако фактически прикладники находят такие разъяснения вполне убедительными. Они их вполне устраивают до тех пор, пока им не приходится вглядеться в них повнимательнее, если такое вообще происходит. В нашем случае рекомендуем читателю решить задачу 40, которая является дискретным аналогом рассуждения, проведенного выше для плотностей, причем абсолютно строгим. Подобные методы будут применяться снова в следующих параграфах.

Пример 3. Вернемся к примеру 5 из § 4.2. Величины S_n обозначают последовательные моменты поступления требований. Положим

$$S_1 = T_1,\ S_2 - S_1 = T_2, \ldots, S_n - S_{n-1} = T_n, \ldots .$$

Таким образом, T_n — *длительности промежутков между поступлениями требований.* Они важны не только для нашего страхового примера, но также для различных моделей, описывающих *длительности простоя* оборудования, используемого время от времени, или *дистанции между автомобилями* при движении транспорта, где T_n измеряют расстояния, а не время. Во многих приложениях именно величины T_n служат объектом статистического анализа. В простейшем случае мы можем предположить, что они экспоненциально распределены, как в примере 13 из § 4.5. Если все T_n обладают плотностью $\lambda e^{-\lambda t}$, то $E(T_n) = 1/\lambda$. Так как

$$S_n = T_1 + \ldots + T_n,$$

то согласно теореме 1 для нашего случая, когда есть плотность, имеем:

$$E(S_n) = E(T_1) + \ldots + E(T_n) = \frac{n}{\lambda}\,.$$

Отметим, что здесь не предполагается, что случайные величины T_n независимы. Другими словами, допускается наличие взаимного влияния между ними. К примеру, несколько несчастных случаев могут быть вызваны одной и той же причиной, как при столкновении 20 машин на шоссе. Более того, величины T_n могут иметь разные параметры λ_n, скажем, из-за ежедневных или сезонных изменений. Если $E(T_n) = 1/\lambda_n$, то

$$E(S_n) = \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_n}\,.$$

Мы завершим этот параграф решением задачи, оставшейся с § 1.4 (см. также задачу 20 из гл. 1 и задачу 6 из § 3.4).

Формула Пуанкаре. Для произвольных событий $A_1, \dots, A_n$ верно равенство

$$P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_j P(A_j) - \sum_{j,k} P(A_j A_k) + \sum_{j,k,l} P(A_j A_k A_l) - \\ - \dots + (-1)^{n-1} P(A_1 \dots A_n), \qquad (6.2.5)$$

где индексы в каждой из сумм различны и меняются от 1 до n.

Доказательство. Пусть $\alpha_j = I_{A_j}$ обозначает индикатор события A_j. Тогда индикатором события $A_1^c \dots A_n^c$ является $\prod_{j=1}^{n}(1-\alpha_j)$. Следовательно, индикатор его дополнения задается формулой

$$I_{A_1 \cup \dots \cup A_n} = 1 - \prod_{j=1}^{n}(1-\alpha_j) = \sum_j \alpha_j - \sum_{j,k} \alpha_j \alpha_k + \\ + \sum_{j,k,l} \alpha_j \alpha_k \alpha_l - \dots + (-1)^{n-1} a_1 \dots a_n.$$

Математическое ожидание индикаторной случайной величины совпадает с вероятностью соответствующего события:

$$E(I_A) = P(A).$$

Если взять математическое ожидание от каждого слагаемого в приведенном выше разложении и использовать формулу (6.1.7), то получим доказываемое равенство (6.2.5). (Анри Пуанкаре называли последним энциклопедистом среди математиков; его вклад в теорию вероятностей — по большей части философский и педагогический. Приведенная выше формула представляет собой вариант «принципа включения — исключения», приписываемого Сильвестру.)

Пример 4 (*задача совпадений* или *случайной встречи*). Две колоды состоят из карточек, занумерованных числами от 1 до n. Обе колоды тщательно перемешаны. Какова вероятность того, что при последовательном сопоставлении карточек из этих двух колод произойдет хотя бы одно совпадение номеров?

Решение. Пусть событие A_j состоит в том, что совпали номера при j-м по счету сравнении, безотносительно результатов остальных сравнений. Всего существует $n!$ перестановок номеров из второго множества относительно номеров первого (последние можно считать упорядоченными по возрастанию от 1 до n). Если номера j-х карточек совпадают, то для остальных номеров имеется $(n-1)!$ возможностей, поэтому

$$P(A_j) = \frac{(n-1)!}{n!} = \frac{1}{n}\,. \qquad (6.2.6)$$

Аналогично, если номера j-й и k-й карточек совпадают, где $j \neq k$, то для остальных номеров существует $(n-2)!$ возможных перестановок. Отсюда

$$P(A_j A_k) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}\,. \tag{6.2.7}$$

Далее, если номера j, k, l-х карточек совпадают, то

$$P(A_j A_k A_l) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}\,,$$

и т. д. Теперь заметим, что имеется ровно $\binom{n}{1}$ слагаемых в первой сумме из правой части равенства (6.2.5), $\binom{n}{2}$ слагаемых во второй сумме, $\binom{n}{3}$ слагаемых в третьей сумме и т. д. Поэтому вся правая часть приобретает вид

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

Известно, что

$$1 - e^{-1} = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1}\frac{1}{n!} + \ldots = \sum_{n-1}^{\infty}\frac{(-1)^{n-1}}{n!}\,.$$

Этот ряд сходится очень быстро; в самом деле, нетрудно убедиться, что

$$\left|1 - e^{-1} - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1}\frac{1}{n!}\right)\right| \leqslant \frac{1}{(n+1)!}\,.$$

Поэтому при $n \geqslant 4$ имеем

$$\frac{1}{(n+1)!} \leqslant \frac{1}{5!} = \frac{1}{120}\,,$$

и вероятность появления по крайней мере одного совпадения отличается от $1 - e^{-1} \approx 0.63$ не более, чем на 0.01. Другими словами, вероятность отсутствия совпадений примерно равна 0.37 для всех $n \geqslant 4$.

А что можно сказать о математическом ожидании числа совпадений? Случайное число совпадений задается формулой

$$N = I_{A_1} + \ldots + I_{A_n}. \tag{6.2.8}$$

(Почему? Обдумайте это *тщательно*, не забывая, что события A_j не являются ни несовместными, ни независимыми.) Следовательно, снова применяя теорему 1, находим, что математическое ожидание

$$E(N) = \sum_{j=1}^{n} E(I_{A_j}) = \sum_{j=1}^{n} P(A_j) = n \cdot \frac{1}{n} = 1,$$

т. е. в точности равно 1 для всех n. Это результат верен, но его можно рассматривать как числовой парадокс.

6.3. Теоремы умножения. Дисперсия и ковариация

Мы уже отмечали, что теорема 1 в действительности является общим результатом теории интегрирования, широко используемым во многих направлениях математики. Напротив, следующее утверждение предполагает стохастическую независимость и специфично для теории вероятностей.

Теорема 2. *Если X и Y — независимые интегрируемые случайные величины, то*

$$E(XY) = E(X)E(Y). \tag{6.3.1}$$

Доказательство. Сначала докажем это равенство для счетного выборочного пространства Ω. Тогда обе величины X и Y имеют счетные множества возможных значений. Пусть $\{x_j\}$ исчерпывают все значения, принимаемые величиной X, а $\{y_k\}$ — величиной Y. Далее, положим

$$A_{jk} = \{w \mid X(\omega) = x_j, Y(\omega) = y_k\},$$

т. е. A_{jk} — множество точек выборочного пространства, на которых $X = x_j$ и $Y = y_k$. Тогда множества A_{jk}, где (j, k) пробегает все множество пар индексов, являются несовместными (почему?), и их объединение дает все выборочное пространство:

$$\Omega = \sum_j \sum_k A_{jk}.$$

Случайная величина XY принимает значение $x_j y_k$ на множестве A_{jk}, но некоторые из этих значений могут быть одинаковыми, скажем, для $\{x_j = 2, y_k = 3\}$ и $\{x_j = 3,\ y_k = 2\}$. Невзирая на это, вычислим математическое ожидание величины XY, умножая вероятность каждого из A_{jk} на значение XY на этом множестве, и суммируя такие произведения:

$$E(XY) = \sum_j \sum_k x_j y_k P(A_{jk}). \tag{6.3.2}$$

Эта формула является частным случаем соотношения (4.3.15) и возникает благодаря перегруппировке членов в определяющем математическое ожидание ряде $\sum\limits_\omega X(\omega)Y(\omega)P(\omega)$. Далее, учитывая предположение о независимости X и Y, запишем, что

$$P(A_{jk}) = P(X = x_j)P(Y = y_k).$$

Подставляя это в правую часть формулы (6.3.2), видим, что двойная сумма распадается в произведение обычных сумм:

$$\sum_j \sum_k x_j y_k P(X = x_j) P(Y = y_k) =$$

$$= \left\{ \sum_j x_j P(X = x_j) \right\} \left\{ \sum_k y_k P(Y = y_k) \right\} = E(X)E(Y).$$

Здесь опять законность перестановки слагаемых обеспечивается благодаря абсолютной сходимости двойного ряда в формуле (6.3.2).

Теперь докажем теорему в предположении, что случайный вектор (X, Y) имеет совместную плотность распределения f. Для этого используем метод, уже примененный в § 6.2. По аналогии с формулой (6.2.4) имеем:

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uvf(u, v)\, du\, dv.$$

Используя обозначения из § 4.6, запишем, согласно (5.5.9), равенство

$$f(u, v) = f(u, *) f(*, v).$$

Тогда двойной интеграл может быть представлен в следующем виде:

$$\int_{-\infty}^{\infty} uf(u, *)\, du \int_{-\infty}^{\infty} vf(*, v)\, dv = E(X)E(Y).$$

Строго говоря, мы должны были сначала провести эти выкладки для случайных величин $|X|$ и $|Y|$ (они также независимы в силу утверждения 6 из § 5.5) и получить неравенство

$$E(|XY|) = E(|X|)E(|Y|) < \infty.$$

Оно показывает, что величина XY интегрируема, и выполненные выше преобразования над двойным рядом или двойным интегралом законны. (Эти надоедливые подробности часто отвлекают от основного доказательства, но они являются той ценой, которую приходится платить за математическую строгость. Преподаватель, так же как и читатель, может лишь поверхностно ознакомится с некоторыми из них по своему собственному усмотрению.)

Обобщение доказанной теоремы на любое конечное число независимых интегрируемых случайных величин получается немедленно:

$$E(X_1 \ldots X_n) = E(X_1) \ldots E(X_n). \tag{6.3.3}$$

Эту формулу можно доказать прямо или по индукции. В последнем случае требуется, чтобы произведение X_1X_2 не зависело от X_3 и т. п. Это верно и ранее отмечалось в § 5.5 (еще одна надоедливая деталь).

В частном случае теоремы 2, когда каждая случайная величина X_j есть индикатор события A_j, формула (6.3.3) сводится к равенству

$$P(A_1 \dots A_n) = P(A_1) \dots P(A_n),$$

которое ясно показывает, что теорема 2 не может выполняться без предположения о независимости. Напротив, в этой же самой ситуации формула, соответствующая (6.1.7), имеет вид:

$$E(I_{A_1} + \dots + I_{A_n}) = P(A_1) + \dots + P(A_n).$$

Здесь на события не накладываются дополнительные условия, подобные несовместности. Левую часть необходимо отчетливо отличать от $P(A_1 \cup \dots \cup A_n)$ или любой другой похожей вероятности. Данный вид путаницы возникает как у первооткрывателей, так и у начинающих. Он известен как *парадокс Кардано**).

Пример 5. Измеряются тонкие железные стержни, имеющие цилиндрическую форму. Допустим, что их средняя длина составляет 10 дюймов, а средняя площадь сечения — 1 квадратный дюйм. Средняя ошибка при измерении длины равна 0.005 дюйма, а при изменении площади сечения — 0.01 квадратного дюйма. Какова средняя ошибка при оценивании веса стержня?

Так как вес равен константе, умноженной на объем, достаточно рассмотреть последний: $V = LA$, где L — длина, A — площадь сечения. Обозначим ошибки через ΔL и ΔA соответственно. Тогда ошибка при оценивании объема V вычисляется по формуле

$$\Delta V = (L + \Delta L)(A + \Delta A) - LA = L\,\Delta A + A\,\Delta L + \Delta L\,\Delta A.$$

Предполагая независимость измерений, находим, что

$$E(\Delta V) = E(L)E(\Delta A) + E(A)E(\Delta L) + E(\Delta A)E(\Delta L) =$$
$$= 10 \cdot \frac{1}{100} + 1 \cdot \frac{1}{200} + \frac{1}{100} \cdot \frac{1}{200} \ = 0.105 \text{ куб. дюймов},$$

если отбросить последний член в силу его относительной малости.

*) Джероламо Кардано (1501–1576) написал одну из первых книг об играх, в которых присутствует случайность.

Определение момента. Для натурального числа r математическое ожидание $E(X^r)$ называется *r-м моментом* (*моментом порядка* r) случайной величины X. Тогда если X^r имеет конечное математическое ожидание, то говорят, что случайная величина X обладает конечным r-м моментом. При $r = 1$ получаем первый момент, который, безусловно, совпадает с обычным математическим ожиданием или средним.

Случай $r = 2$ играет особую роль. Поскольку $X^2 \geqslant 0$, мы будем называть $E(X^2)$ вторым моментом величины X, когда он конечен и когда равен $+\infty$, в соответствии с тем, сходится или расходится определяющий его (на счетном Ω) ряд $\sum\limits_{\omega} X^2(\omega)P(\omega)$.

Когда среднее $E(X)$ конечно, часто бывает удобным иметь дело со случайной величиной

$$X^0 = X - E(X) \tag{6.3.4}$$

вместо X, поскольку первый момент этой случайной величины равен нулю. Мы будем говорить, что X^0 получена из X *центрированием.*

Определение дисперсии и стандартного отклонения. Второй момент центрированной величины X^0 называется *дисперсией* случайной величины X и обозначается $\sigma^2(X)$. Положительный квадратный корень из дисперсии $\sigma(X)$ называется *стандартным отклонением* случайной величины X.

Приведем важнейшие соотношения между $E(X)$, $E(X^2)$ и $\sigma^2(X)$.

Теорема 3. *Если $E(X^2)$ конечно, то таково же и $E(|X|)$. При этом*

$$\sigma^2(X) = E(X^2) - E(X)^2; \tag{6.3.5}$$

следовательно,

$$E(|X|)^2 \leqslant E(X^2). \tag{6.3.6}$$

Доказательство. Так как

$$X^2 - 2|X| + 1 = (|X| - 1)^2 \geqslant 0,$$

то должно (почему?) выполняться неравенство $E(X^2 - 2|X| + 1) \geqslant 0$, и поэтому $E(X^2) + 1 \geqslant 2E(|X|)$ в силу теоремы 1 и соотношений (6.1.5). Это доказывает первое утверждение теоремы. Далее запишем, что

$$\sigma^2(X) = E\{(X - E(X))^2\} = E\{X^2 - 2E(X)X + E(X)^2\} =$$
$$= E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2.$$

Поскольку $\sigma^2(X) \geqslant 0$ согласно первому из равенств в этой формуле, то, подставляя $|X|$ вместо X в соотношение (6.3.5), выводим неравенство (6.3.6).

Каков смысл стандартного отклонения $\sigma(X)$? Начнем с того, что центрированная случайная величина X^0 представляет собой отклонение случайной величины X от ее среднего и может принимать как положительные, так и отрицательные значения. Если нас интересует только абсолютная величина отклонения, то *среднее абсолютное отклонение* задается формулой $E(|X^0|) = E(|X - E(X)|)$. Эта характеристика может быть использована, но ее часто трудно вычислить. Вместо нее рассматривается *средний квадрат отклонения* $E(|X^0|^2)$, который совпадает с дисперсией. Но затем мы должны избавиться от последствий возведения в квадрат*) путем извлечения квадратного корня, что приводит к стандартному отклонению $+\sqrt{E(|X^0|^2)}$. Тем самым, мы получаем оценку среднего отклонения случайной величины (*наблюдаемых значений*) от ее математического ожидания. Чем она меньше, тем лучше наблюдаемые значения группируются вокруг своего среднего, и более компактной или сконцентрированной представляется популяция. Особую важность этих понятий мы увидим позже в гл. 7 при изучении сходимости к нормальному распределению.

Заметим, что для любой константы c случайные величины X и $X+c$ обладают одинаковой дисперсией; в частности, это верно в случае X и X^0. Следующий результат похож на теорему 1, но только внешне.

Теорема 4. *Если случайные величины X и Y независимы и имеют конечные дисперсии, то*

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y). \tag{6.3.7}$$

Доказательство. С учетом предыдущего замечания можно считать, что обе величины X и Y имеют нулевое среднее. Тогда математическое ожидание суммы $X+Y$ также равно 0, а ее дисперсия совпадает с вторым моментом. Далее,

$$E(XY) = E(X)\,E(Y) = 0$$

в силу теоремы 2. Отсюда

$$\begin{aligned} E\{(X+Y)^2\} &= E\{X^2 + 2XY + Y^2\} = \\ &= E(X^2) + 2\,E(XY) + E(Y^2) = E(X^2) + E(Y^2), \end{aligned}$$

согласно теореме 1. Равенство установлено.

Теорема 4 очевидно обобщается на любое конечное число независимых слагаемых. Тем не менее, существует полезная общая формула

*) Надо вернуться от размерности площади к размерности длины. — *Прим. перев.*

для второго момента, справедливая без предположения о независимости. Начнем ее вывод с алгебраического тождества

$$(X_1 + \ldots + X_n)^2 = \sum_{j=1}^{n} X_j^2 + 2 \sum_{1 \leqslant j < k \leqslant n} X_j X_k.$$

Беря математические ожидания от обеих частей и применяя теорему 1, получаем:

$$E\{(X_1 + \ldots + X_n)^2\} = \sum_{j=1}^{n} E(X_j^2) + 2 \sum_{1 \leqslant j < k \leqslant n} E(X_j X_k). \qquad (6.3.8)$$

Когда случайные величины X_j центрированы и являются независимыми, все смешанные члены во второй сумме выше исчезают, и мы в результате приходим к упомянутому ранее обобщению теоремы 4.

Давайте введем две *неопределенности* (вспомогательные переменные) ξ и η и рассмотрим тождество

$$E\{(X\xi + Y\eta)^2\} = E(X^2)\,\xi^2 + 2\,E(XY)\,\xi\eta + E(Y^2)\,\eta^2.$$

В правой части мы видим квадратичную форму относительно (ξ, η), причем она всегда неотрицательна, поскольку слева стоит математическое ожидание случайной величины, которая не принимает отрицательных значений. Согласно известному результату из высшей алгебры для коэффициентов такой формы $a\xi^2 + 2b\xi\eta + c\eta^2$ обязано выполняться неравенство $b^2 \leqslant ac$. В нашем случае оно имеет вид

$$E(XY)^2 \leqslant E(X^2)E(Y^2). \qquad (6.3.9)$$

Его называют *неравенством Коши—Буняковского—Шварца.*

Если случайные переменные X и Y обе обладают конечными дисперсиями, то величину

$$\begin{aligned} E(X^0Y^0) &= E\{(X - E(X))(Y - E(Y))\} = \\ &= E\{XY - X\,E(Y) - Y\,E(X) + E(X)\,E(Y)\} = \\ &= E(XY) - E(X)\,E(Y) - E(Y)\,E(X) + E(X)\,E(Y) = \\ &= E(XY) - E(X)\,E(Y) \end{aligned}$$

называют *ковариацией* между X и Y и обозначают через $\mathrm{Cov}\,(X, Y)$. В свою очередь, величина

$$\rho(X, Y) = \frac{\mathrm{Cov}\,(X, Y)}{\sigma(X)\,\sigma(Y)}$$

носит имя *коэффициента корреляции* между X и Y, при условии, конечно, что знаменатель не обращается в ноль. Если коэффициент корреляции равен нулю, то говорят, что случайные величины *некоррелированы.*

Это свойство вытекает из независимости, но в общем случае является более слабым. В качестве следствия неравенства (6.3.9) получаем, что всегда $-1 \leqslant \rho(X,Y) \leqslant 1$. Как знак, так и абсолютная величина коэффициента корреляции ρ определенным образом характеризуют зависимость между случайными величинами*). См. также задачу 30 в гл. 7.

Пример 6. Классической моделью для применения предыдущих результатов служит схема Бернулли (см. пример 9 в § 4.4). В ней рассматриваются независимые случайные величины, имеющее следующее общее распределение:

$$X = \begin{cases} 1 & \text{с вероятностью} \quad p, \\ 0 & \text{с вероятностью} \quad q = 1 - p. \end{cases} \tag{6.3.10}$$

Мы встречались с ними, когда изучали процесс бросаний монеты в примере 8 из § 2.4, но данная схема применяется также для описания произвольных повторных испытаний, в которых возможны только два исхода: «успех» ($X = 1$) и «неудача» ($X = 0$). Скажем, в примере 1 из § 6.1 вероятность $p = 1/100$, а денежной единицей служит 10 000 («тьма долларов»). Еще одну иллюстрацию дают шансы на «излечение» или «летальный исход» в результате серьезной хирургической операции.

Нетрудно вычислить среднее и дисперсию бернуллиевской случайной величины X:

$$E(X) = p, \quad \sigma^2(X) = p - p^2 = pq.$$

Пусть $\{X_n, n \geqslant 1\}$ обозначают бернуллиевские случайные величины, а

$$S_n = X_1 + \ldots + X_n \tag{6.3.11}$$

представляет собой n-ю частичную сумму этих величин. Она равна общему количеству «успехов» в n испытаниях. Применяя теорему 1, получаем, что

$$E(S_n) = E(X_1) + \ldots + E(X_n) = np. \tag{6.3.12}$$

Это верно, даже если слагаемые были бы зависимыми. Далее, согласно теореме 3,

$$\sigma^2(S_n) = \sigma^2(X_1) + \ldots + \sigma^2(X_n) = npq. \tag{6.3.13}$$

Легкость, с которой были установлены данные результаты, демонстрирует существенный прогресс в математической технике. Напомним, что

*) Математик Эмиль Артин в 1947 г. сказал мне следующее: «Каждый знает, что теория вероятностей и статистика — это одно и то же, а статистика есть не что иное, как теория корреляции. Корреляция же — всего лишь косинус угла. Следовательно, все это тривиально».

равенство (6.3.12) было получено в (4.4.16) путем довольно «хитроумных» вычислений на основе того, что случайная величина S_n имеет биномиальное распределение. Подобный подход срабатывает и в случае равенства (6.3.13). Настоятельно рекомендуем читателю применить его для практики и сравнения. Однако наш новый подход, опирающийся на вычисление средних и дисперсий отдельных слагаемых и не требующий нахождения вероятностного распределения самой суммы, намного проще. В более сложных ситуациях последнее может оказаться весьма трудоемким, если вообще осуществимым. Это объясняет, почему мы посвятили несколько параграфов обсуждению понятий среднего и дисперсии, которых нередко самих по себе вполне достаточно для теоретических и практических целей.

Пример 7. Возвращаясь к задаче о совпадениях из § 6.2, давайте сейчас вычислим стандартное отклонение количества совпадений. Величины I_{A_j} в сумме (6.2.8) не являются независимыми, но мы можем применить формулу (6.3.8) и получить равенство

$$E(N^2) = \sum_{j=1}^{n} E(I_{A_j}^2) + 2 \sum_{1 \leqslant j < k \leqslant n} E(I_{A_j} I_{A_k}).$$

Очевидно, что

$$E(I_{A_j}^2) = P(A_j) = \frac{1}{n}\,,$$

$$E(I_{A_j} I_{A_k}) = P(A_j A_k) = \frac{1}{n(n-1)}\,,$$

в силу соотношений (6.2.6) и (6.2.7). Подставляя их в предыдущую формулу, находим:

$$E(N^2) + n \cdot \frac{1}{n} + 2\binom{n}{2}\frac{1}{n(n-1)} = 1 + 1 = 2.$$

Следовательно,

$$\sigma^2(N) = E(N^2) - E(N)^2 = 2 - 1 = 1.$$

Редко случается, чтобы интересная общая задача имела бы такой простой числовой ответ.

6.4. Полиномиальное распределение

Широкое поле для иллюстрации понятий и применения техники, развитой в предыдущих параграфах, предоставляет *полиномиальное распре-*

деление. Оно является естественным обобщением биномиального распределения и служит моделью для повторных испытаний, в которых возможны несколько разных исходов, а не только «успех» или «неудача». Мы начнем с алгебраической формулы, известной как *полиномиальная теорема*:

$$(x_1 + \ldots + x_r)^n = \sum \frac{n!}{n_1! \ldots n_r!} x_1^{n_1} \ldots x_r^{n_r}, \tag{6.4.1}$$

где суммирование проводится по всем упорядоченным наборам из r целых чисел $n_1, \ldots, n_r$, удовлетворяющих следующим условиям:

$$n_1 \geqslant 0, \quad \ldots, \quad n_r \geqslant 0, \quad n_1 + \ldots + n_r = n. \tag{6.4.2}$$

При $r = 2$ получаем теорему о биноме. В таком случае имеются $n + 1$ упорядоченных пар:

$$(0, n),\ (1, n-1),\ \ldots,\ (k, n-k),\ \ldots, (n, 0)$$

с соответствующими коэффициентами

$$\frac{n!}{0!\, n!},\ \frac{n!}{1!\,(n-1)!},\ \ldots,\ \frac{n!}{k!\,(n-k)!},\ \ldots,\ \frac{n!}{n!\, 0!}$$

или

$$\binom{n}{0},\ \binom{n}{1},\ \ldots,\ \binom{n}{k},\ \ldots,\ \binom{n}{n}.$$

Следовательно, сумма может быть записана в виде

$$\binom{n}{0} x^0 y^n + \binom{n}{1} x^1 y^{n-1} + \ldots + \binom{n}{k} x^k y^{n-k} + \ldots + \binom{n}{n} x^n y^0 =$$
$$= \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

В общем случае n одинаковых скобок $(x_1 + \ldots + x_r)$ из левой части формулы (6.4.1) перемножаются, и соответствующие члены группируются справа. Члены имеют вид $x_1^{n_1} \ldots x_r^{n_r}$ с показателями n_j, удовлетворяющими условиям (6.4.2). Причем слагаемые каждого такого вида встречаются $n!/(n_1! \ldots n_r!)$ раз, поскольку именно таким количеством способов можно разбить n объектов на r различных категорий (x_j) так, что в точности n_j из них будут принадлежать j-й категории. (Как видите, некоторая комбинаторика заключается в самой природе вещей, и без нее нельзя обойтись, даже если вы пропустили большую часть гл. 3.)

Конкретная модель для полиномиального распределения может быть описана следующим образом. Пусть в урне содержатся шары r разных цветов в пропорции

$$p_1 : p_2 : \ldots : p_r, \quad \text{где } p_1 + \ldots + p_r = 1.$$

Извлекаются n шаров один за другим с возвращением. Предполагается, что результаты последовательных извлечений независимы, что может быть достигнуто за счет тщательного перемешивания шаров при встряхивании урны после каждого извлечения. Чему равна вероятность того, что будут получены заданные количества шаров каждого цвета?

Пусть $X_1, \ldots, X_n$ — независимые случайные величины, имеющие такое же распределение, как следующая случайная величина X:

$$X = \begin{cases} 1 & \text{с вероятностью } p_1, \\ 2 & \text{с вероятностью } p_2, \\ \vdots & \\ r & \text{с вероятностью } p_r. \end{cases} \tag{6.4.3}$$

Каково совместное распределение вектора $(X_1, \ldots, X_n)$, т. е. чему равны вероятности

$$P(X_1 = x_1, \ldots, X_n = x_n) \tag{6.4.4}$$

для всевозможных значений x_j от 1 до r? Здесь номера $1, \ldots, r$ соответствуют цветам шаров. Подобное приписывание некоторых числовых значений категориям не является обязательным, но иногда оказывается удобным. Оно также приводит к вопросам, которые не могли бы возникнуть, если бы использовались только цвета: например, чему равна вероятность «$X_1 + \ldots + X_n = m$». Однако представим, что мы заменили шары на лотерейные билеты, на которые приходятся различные выигрышные суммы, или на людей, имеющих разный возраст или доход. Тогда числовая формализация (6.4.3) становится уместной. А могут ли X_j принимать отрицательные или дробные значения? С данной проблемой можно справиться, применяя линейные преобразования (см. пример 15 в § 4.5) при условии, что все возможные значения соизмеримы, скажем, имеют одинаковую точность десятичного представления. Например, если они представлены с точностью в три знака после запятой и принадлежат диапазону $[-10, \infty)$, то вместо X можно использовать

$$X' = 10\,000 + 1\,000X.$$

Значение -9.995 преобразуется в $10\,000 - 9955 = 5$ в новой шкале. Находясь на сверхпрогматичных позициях, можно даже утверждать, что полиномиальное распределение — это все, что нам нужно для описания случайного выбора при условии независимости испытаний, так как в реальности нам никогда не приходится различать между собой, скажем, $10^{10^{10}}$ категорий чего-либо. Однако данный «конечный» подход приводит к отказу от почти всех достижений математики.

Вычислим вероятность (6.4.4). Она равна

$$P(X_1 = x_1) \times \ldots \times P(X_n = x_n)$$

ввиду независимости, причем каждый сомножитель — одно из p_j в формуле (6.4.3). Для получения явного выражения необходимо знать, какое количество чисел среди x_j равны 1, 2, ..., r? Пусть n_1 из них равны 1, n_2 равны 2, ..., n_r равны r. Тогда эти n_j удовлетворяют условиям (6.4.2), и искомая вероятность равна $p_1^{n_1} \ldots p_r^{n_r}$. Удобно ввести новые случайные величины $N_j, 1 \leqslant j \leqslant r$:

$$N_j = \text{количество тех } X_j \text{ среди } (X_1, \ldots, X_n), \text{ которые имеют значение } j.$$

Каждая из N_j принимает значения от 0 до n, но случайные величины $N_1, \ldots, N_r$ являются зависимыми, так как они удовлетворяют очевидному ограничению

$$N_1 + \ldots + N_r = n. \tag{6.4.5}$$

Тем не менее, их совместное распределение может быть выписано явно:

$$P(N_1 = n_1, \ldots, N_r = n_r) = \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r}. \tag{6.4.6}$$

Доказательство этого утверждения ничем не отличается от того, которое было приведено в начале параграфа для формулы (6.4.1), но мы повторим его. Выше было выяснено, что для любого *отдельного* или *точно указанного* набора $(X_1, \ldots, X_n)$, удовлетворяющего условиям $N_1 = n_1, \ldots, N_r = n_r$, вероятность равна $p_1^{n_1} \ldots p_r^{n_r}$. Но всего существует $n!/(n_1! \ldots n_r!)$ различных отдельных наборов, удовлетворяющих тем же самым условиям, получаемых в результате перестановки n множителей, среди которых n_1 множителей вида p_1, n_2 множителей вида p_2 и т. д. Для прояснения приведем простой числовой пример. Возьмем $n = 4$, $r = 3$, $n_1 = 2$, $n_2 = n_3 = 1$. Это означает, что в четырех извлечениях оказалось 2 шара цвета 1 и по одному шару цветов 2 и 3. Всевозможные отдельные наборы указаны в следующем списке:

$$\begin{array}{llll}
(1,1,2,3), & (1,1,3,2), & (1,2,1,3), & (1,3,1,2), \\
(1,2,3,1), & (1,3,2,1), & (2,1,1,3), & (3,1,1,2), \\
(2,1,3,1), & (3,1,2,1), & (2,3,1,1), & (3,2,1,1).
\end{array}$$

Их общее количество есть $12 = 4!/(2!\,1!\,1!)$, и соответствующая вероятность равна $12p_1^2p_2p_3$.

Формула (6.4.6), в которой индексы n_j пробегают всевозможные целые значения, удовлетворяющие условиям (6.4.2), задает *полиномиальное распределение* случайных величин $(N_1, \ldots, N_r)$. Для него используется обозначение $M(n; r; p_1, \ldots, p_{r-1}, p_r)$, где p_j подчиняются условию

$p_1 + \ldots + p_r = 1$. В биномиальном случае при $r = 2$ это распределение обозначают также через $B(n; p)$, см. пример 9 в § 4.4.

Если разделить обе части формулы (6.4.1) на левую часть и положить

$$p_j = \frac{x_j}{x_1 + \ldots + x_r}\,, \quad 1 \leqslant j \leqslant r,$$

то получим

$$1 = \sum \frac{n!}{n_1! \ldots n_r!}\, p_1^{n_1} \ldots p_r^{n_r}. \tag{6.4.7}$$

Тем самым, мы проверили, что сумма всех полиномиальных вероятностей равна 1. Обратно, поскольку равенство (6.4.7) справедливо в силу вероятностной интерпретации, можно из него вывести соотношение (6.4.1), по крайней мере, когда $x_j \geqslant 0$, записав $(p_1 + \ldots + p_r)^n$ вместо левой части формулы (6.4.7). Это еще одна иллюстрация того, что теория вероятностей может придать новый смысл алгебраическим формулам (сравните с заключительной частью § 3.3).

Маргинальные распределения (см. § 4.6) вектора $(N_1, \ldots, N_r)$ можно найти без вычислений с помощью простых рассуждений. Если нас интересует только N_1, то можно «склеить» все остальные категории в новую категорию «не 1» с соответствующей вероятностью $1 - p_1$. В результате полиномиальное распределение свернется в биномиальное $B(n; p_1)$, а именно:

$$P(N_1 = n_1) = \frac{n!}{n!\,(n - n_1)!}\; p_1^{n_1}(1 - p_1)^{n - n_1}.$$

Отсюда легко найти среднее и дисперсию случайной величины N_1, как в примере 6 из § 6.2. Вообще,

$$E(N_j) = np_j, \quad \sigma^2(N_j) = np_j(1 - p_j), \quad 1 \leqslant j \leqslant r. \tag{6.4.8}$$

Далее, если нас интересует пара (N_1, N_2), то аналогичное «склеивание» приводит к распределению $M(n; 3; p_1, p_2, 1 - p_1 - p_2)$, а именно:

$$\begin{aligned} &P(N_1 = n_1, N_2 = n_2) = \\ &\quad = \frac{n!}{n_1!\, n_2!\, (n - n_1 - n_2)!}\, p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2}. \end{aligned} \tag{6.4.9}$$

Теперь можно выразить $E(N_1 N_2)$, используя формулу (4.3.15) или (6.3.2) (без независимости):

$$\begin{aligned} E(N_1 N_2) &= \sum n_1 n_2\, P(N_1 = n_1, N_2 = n_2) = \\ &= \sum \frac{n!}{n_1!\, n_2!\, (n - n_1 - n_2)!}\, n_1 n_2\, p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2}, \end{aligned}$$

где для индексов суммирования n_1 и n_2 выполняются ограничения (6.4.2). Вычислим эту сумму, обобщив метод, использованный ранее в примере 9 из § 4.4 для биномиального случая. Рассмотрим следующую смешанную частную производную:

$$\frac{\partial^2}{\partial x_1 \partial x_2}(x_1+x_2+x_3)^n = n(n-1)(x_1+x_2+x_3)^{n-2} = \\ = \sum \frac{n!}{n_1!\,n_2!\,(n-n_1-n_2)!}\, n_1 n_2\, x_1^{n_1-1} x_2^{n_2-1} x_3^{n-n_1-n_2}.$$

Умножим обе части на x_1x_2 и положим $x_1 = p_1$, $x_2 = p_2$, $x_3 = 1-p_1-p_2$. В результате получим $n(n-1)p_1p_2$ слева и искомую двойную сумму справа. Следовательно, в общем случае для $j \neq k$ имеем:

$$\begin{aligned} E(N_jN_k) &= n(n-1)p_jp_k; \\ \operatorname{Cov}(N_j, N_k) &= E(N_jN_k) - E(N_j)E(N_k) = \\ &= n(n-1)p_jp_k - (np_j)(np_k) = np_jp_k. \end{aligned} \tag{6.4.10}$$

Любопытно проверить формулу (6.3.8) с учетом ограничения (6.4.5):

$$\begin{aligned} n^2 &= E\{(N_1 + \ldots + N_r)^2\} = \\ &= \sum_{j=1}^{r}\{n(n-1)p_j^2 + np_j\} + 2\sum_{1\leqslant j<k\leqslant r} n(n-1)p_jp_k = \\ &= n(n-1)\left(\sum_{j=1}^{r} p_j\right)^2 + n\sum_{j=1}^{r} p_j = n(n-1) + n = n^2. \end{aligned}$$

Есть другой метод вычисления $E(N_jN_k)$, подобный первому методу из примера 6 в § 6.3. Пусть индекс j фиксирован и

$$\xi(x) = \begin{cases} 1, & \text{если } x = j, \\ 0, & \text{если } x \neq j. \end{cases}$$

Функция ξ, как функция действительного переменного x, является индикатором точки $\{j\}$. Теперь введем случайную величину

$$\xi_\nu = \xi(X_\nu) = \begin{cases} 1, & \text{если } X_\nu = j, \\ 0, & \text{если } X_\nu \neq j. \end{cases} \tag{6.4.11}$$

Тем самым, ξ_ν — индикатор события $\{X_\nu = j\}$. Другими словами, ξ_ν реагируют только на те X_ν, которые принимают значение j. Поэтому $N_j = \xi_1 + \ldots + \xi_n$. Тогда

$$\begin{aligned} E(\xi_\nu) &= P(X_\nu = j) = p_j, \\ \sigma^2(\xi_\nu) &= E(\xi_\nu^2) - E(\xi_\nu)^2 = p_j - p_j^2 = p_j(1-p_j). \end{aligned}$$

Заметим, что в силу утверждения 6 из § 5.5 случайные величины $\xi_1, \ldots, \xi_n$ независимы, так как таковыми являются $X_1, \ldots, X_n$. Следовательно, по теоремам 1 и 4

$$\begin{aligned} E(N_j) &= E(\xi_1) + \ldots + E(\xi_n) = np_j, \\ \sigma^2(N_j) &= \sigma^2(\xi_1) + \ldots + \sigma^2(\xi_n) = np_j(1-p_j). \end{aligned} \tag{6.4.12}$$

Далее, пусть $k \neq j$. Определим η и η_ν таким же образом, как ξ и ξ_ν с заменой j на k. Рассмотрим теперь для $1 \leqslant \nu \leqslant n, 1 \leqslant \nu' \leqslant n$:

$$E(\xi_\nu \eta_{\nu'}) = P(X_\nu = j, X_{\nu'} = k) = \begin{cases} p_j p_k & \text{если } \nu \neq \nu', \\ 0 & \text{если } \nu = \nu'. \end{cases} \tag{6.4.13}$$

Наконец, вычислим:

$$E(N_j N_k) = E\left\{\left(\sum_{\nu=1}^{n} \xi_\nu\right)\left(\sum_{\nu'=1}^{n} \eta_{\nu'}\right)\right\} = E\left\{\sum_{\nu=1}^{n} \xi_\nu \eta_\nu + \sum_{1 \leqslant \nu \neq \nu' \leqslant n} \xi_\nu \eta_{\nu'}\right\} =$$

$$= \sum_{\nu=1}^{n} E(\xi_\nu \eta_\nu) + \sum_{1 \leqslant \nu \neq \nu' \leqslant n} E(\xi_\nu \eta_{\nu'}) = n(n-1)\, p_j p_k,$$

с учетом соотношений (6.4.13), поскольку всего имеется $(n)_2 = n(n-1)$ слагаемых в последней из записанных сумм. Этот результат, конечно, совпадает с результатом (6.4.10).

Завершим параграф численной иллюстрацией рассмотренной выше общей задачи.

Пример 8. Бросаются три одинаковых игральных кости. Какова вероятность того, что сумма трех выпавших номеров равна 9? Кости не предполагаются симметричными: пусть для каждой из костей вероятность выпадения на грань с номером j есть p_j, $1 \leqslant j \leqslant 6$.

Перечислим все возможные случаи в терминах случайных величин X_i и N_j:

X_1	X_2	X_3	N_1	N_2	N_3	N_4	N_5	N_6	Количество вариантов	Вероятность
1	2	6	1	1				1	6	$6p_1p_3p_6$
1	3	5	1		1		1		6	$6p_1p_3p_5$
1	4	4	1			2			3	$3p_1p_4^2$
2	2	5		2			1		3	$3p_2^2p_5$
2	3	4		1	1	1			6	$6p_2p_3p_4$
3	3	3			3				1	p_3^3

Отсюда

$$P(X_1 + X_2 + X_3 = 9) = 6\,(p_1p_2p_6 + p_1p_3p_5 + p_2p_3p_4) + \\ + 3\,(p_1p_4^2 + p_2^2p_5) + p_3^3.$$

Если все p_j равны $1/6$, то выражение справа принимает значение

$$\frac{6+6+6+3+3+1}{6^3} = \frac{25}{216}\,.$$

Числитель 25 равен сумме

$$\sum \frac{3!}{n_1! \ldots n_6!}\ ,$$

где n_j удовлетворяют условиям

$$n_1 + n_2 + n_3 + n_4 + n_5 + n_6 = 3,$$
$$n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + 6n_6 = 9.$$

Всего существует 6 решений, приведенных выше с указанием соответствующих значений N_j.

В общем случае для случайных величин X_i в количестве n вероятность

$$P(X_1 + \ldots + X_n = m)$$

находится суммированием правой части формулы (6.4.6) по всем $(n_1, \ldots, n_r)$, удовлетворяющим как ограничениям (6.4.2), так и равенству

$$1 \cdot n_1 + 2 \cdot n_2 + \ldots + r \cdot n_r = m.$$

Понятно, что при вычислении по таким формулам не обойтись без компьютера. К счастью, в большинстве случаев интерес представляют более грубые результаты, такие как

$$P(a_n \leqslant X_1 + \ldots + X_n \leqslant b_n)$$

для больших значений n. Соответствующие асимптотические формулы и предельные теоремы излагаются в гл. 7. Один из необходимых для этого математических инструментов изучается в следующем параграфе.

6.5. Производящая функция и другие преобразования

Производящая функция — это мощный математический инструмент, подлинное изобретение, созданное великим и разносторонним математиком Леонардом Эйлером (1707–1783) для решения задачи подсчета числа разбиений в теории чисел. Пусть X является случайной величиной, принимающей только целые неотрицательные значения и имеющей следующее вероятностное распределение:

$$P(X = j) = a_j, \quad j = 0, 1, 2, \ldots . \tag{6.5.1}$$

Идея заключается в том, чтобы собрать всю информацию об этой случайной величине внутри некоторой компактной капсулы. Для этого вводится вспомогательная переменная z и рассматривается следующий степенной ряд относительно z:

$$g(z) = a_0 + a_1 z + a_2 z^2 + \ldots = \sum_{j=0}^{\infty} a_j z^j. \tag{6.5.2}$$

Этот ряд называется *производящей функцией* числовой последовательности $\{a_j, j \geqslant 0\}$. В нашем случае мы также можем назвать его производящей функцией случайной величины X с распределением (6.5.1). Таким образом, g есть функция от аргумента z, который мы будем считать действительным числом, несмотря на то, что в некоторых более сложных ситуациях предпочтительнее рассматривать z как комплексную переменную. Если вспомнить, что $\sum\limits_j a_j = 1$, то нетрудно показать, что ряд в правой части формулы (6.5.2) сходится при $|z| \leqslant 1$. В самом деле, его можно оценить сверху так:

$$|g(z)| \leqslant \sum_j |a_j|\,|z|^j \leqslant \sum_j a_j = 1 \quad \text{для } |z| \leqslant 1.$$

(Мы надеемся, что информированность читателя касательно рядов простирается дальше признака сходимости Даламбера. Приведенная выше оценка проще и содержательнее.) Известная теорема из математического анализа в данном случае обеспечивает законность почленного дифференцирования ряда внутри интервала $|z| < 1$ для нахождения производных от функции g:

$$\begin{aligned} g'(z) &= a_1 + 2a_2 z + 3a_3 z^2 + \ldots = \sum_{n=1}^{\infty} n a_n z^{n-1}, \\ g''(z) &= 2a_2 + 6a_3 z + \ldots = \sum_{n=2}^{\infty} n(n-1) a_n z^{n-2}. \end{aligned} \tag{6.5.3}$$

В общем случае получаем такое представление:

$$g^{(j)}(z) = \sum_{n=j}^{\infty} n(n-1) \cdot \ldots \cdot (n-j+1) a_n z^{n-j} = \sum_{n=j}^{\infty} \binom{n}{j} j!\, a_n z^{n-j}. \tag{6.5.4}$$

Если положить здесь $z = 0$, то все члены, за исключением константы, пропадут:

$$g^{(j)}(0) = j!\, a_j \quad \text{или} \quad a_j = \frac{1}{j!}\, g^{(j)}(0). \tag{6.5.5}$$

Это показывает, что можно, зная функцию g, найти все a_j. Тем самым, не только вероятностное распределение определяет производящую функцию, но и наоборот. Поэтому в капсуле нет потери информации. В частности, полагая $z = 1$ в формулах для g' и g'', согласно равенству (4.3.18), выводим, что

$$g'(1) = \sum_{n=0}^{\infty} n a_n = E(X), \quad g''(1) = \sum_{n=0}^{\infty} n^2 a_n - \sum_{n=0}^{\infty} n a_n = E(X^2) - E(X)$$

при условии, что ряды сходятся, т. е. соотношения (6.5.3) выполняются при $z = 1$[*)]. Отсюда

$$E(X) = g'(1), \quad E(X^2) = g'(1) + g''(1). \tag{6.5.6}$$

На практике часто достаточным оказывается следующее качественное заключение, подытоживающее обсуждение данного вопроса.

Теорема 5. *Вероятностное распределение случайной величины, принимающей целые неотрицательные значения, однозначно определяется ее производящей функцией.*

Пусть Y — случайная величина с распределением $\{b_k, k \geqslant 0\}$, где $b_k = P(Y = k)$, и пусть h — ее производящая функция:

$$h(z) = \sum_{k=0}^{\infty} b_k z^k.$$

Предположим, что $g(z) = h(z)$ для всех $|z| < 1$. Тогда теорема утверждает, что $a_k = b_k$ для всех $k \geqslant 0$. Следовательно, случайные величины X и Y имеют одинаковые распределения (именно это означает выражение «однозначно определяется»). Явная формула (6.5.5), конечно, влечет данное утверждение, однако можно привести более простое «доказательство». Так как

$$\sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} b_k z^k, \quad |z| < 1;$$

то, полагая $z = 0$ в этом равенстве, сразу устанавливаем, что $a_0 = b_0$. После исключения констант сократим обе стороны на z и, опять взяв

[*)] Это теорема Абеля; сравните с комментариями после формулы (8.4.17).

$z = 0$, получим, что $a_1 = b_1$. Повторение этого процесса устанавливает справедливость теоремы. Вам следует осознать, что мы сейчас воспроизвели ошибочное доказательство, которое иногда встречается в некоторых учебниках математического анализа! Можете ли вы сказать, что в нем не так и как это исправить?

Перейдем к обсуждению важнейшего свойства производящих функций, возникающего при их умножении. Используя введенные выше обозначения, запишем:

$$g(z)h(z) = \left(\sum_{j=0}^{\infty} a_j z^j\right)\left(\sum_{k=0}^{\infty} b_k z^k\right) = \sum_j \sum_k a_j b_k z^{j+k}.$$

Перегруппируем члены этого двойного ряда для представления его в форме обычного ряда. Получим

$$g(z)h(z) = \sum_{l=0}^{\infty} c_l z^l, \tag{6.5.7}$$

где

$$c_l = \sum_{j+k=l} a_j b_k = \sum_{j=0}^{l} a_j b_{l-j}. \tag{6.5.8}$$

Последовательность $\{c_j\}$ называется *сверткой* двух последовательностей $\{a_j\}$ и $\{b_j\}$. Какой смысл имеет коэффициент c_l? Допустим, что случайные величины X и Y независимы. Тогда с учетом определения (6.5.8) имеем:

$$c_l = \sum_{j=0}^{l} P(X=j)\,P(Y=l-j) = \sum_{j=0}^{l} P(X=j, Y=l-j) = P(X+Y=l).$$

Здесь последнее равенство получено с помощью правил из § 5.2 следующим образом. При выполнении равенства $X = j$ событие $X + Y = l$ происходит тогда и только тогда, когда $Y = l - j$. Отсюда, применяя утверждение 2 из § 5.2 (см. (5.2.4)), выводим, что

$$\begin{aligned} P(X+Y=l) &= \sum_{j=0}^{\infty} P(X=j)\,P(X+Y=l \mid X=j) = \\ &= \sum_{j=0}^{\infty} P(X=j)\,P(Y=l-j \mid X=j) = \\ &= \sum_{j=0}^{l} P(X=j)\,P(Y=l-j), \end{aligned}$$

так как Y не зависит от X и не принимает отрицательных значений. Другими словами, мы показали, что для всех $l \geqslant 0$

$$P(X+Y=l)=c_l,$$

поэтому $\{c_l, l \geqslant 0\}$ является вероятностным распределением случайной величины $X+Y$. Следовательно, согласно определению, ее производящая функция представляется рядом из формулы (6.5.7), и ввиду этого равна произведению производящих функций величин X и Y. Очевидная индукция позволяет сформулировать результат в более общей форме.

Теорема 6. *Пусть случайные величины $X_1, \ldots, X_n$ независимы и имеют производящие функции $g_1, \ldots, g_n$ соответственно. Тогда производящая функция суммы $X_1+\ldots+X_n$ равна произведению $g_1 \cdot \ldots \cdot g_n$.*

Эта теорема очень важна, так как она, как мы убедимся в гл. 7, дает метод для изучения сумм независимых случайных величин на основе их производящих функций. В некоторых случаях произведение производящих функций имеет простую форму, позволяющую найти вероятностное распределение путем разложения в степенной ряд или, что эквивалентно, используя равенства (6.5.5). Ниже будут приведены примеры, демонстрирующие данный метод.

Помимо этого, отметим, что степенной ряд в формуле (6.5.2) можно рассматривать для произвольной последовательности действительных чисел $\{a_j\}$. Его сумму называют производящей функцией, определяемой данной последовательностью. Если такой ряд имеет ненулевой радиус сходимости, то предыдущие выкладки, не считая вероятностной интерпретации, остаются в силе. В частности, свертку двух таких последовательностей можно определить формулой (6.5.8), при этом соотношение (6.5.7) будет выполняться. В § 8.4 появятся производящие функции, коэффициенты которых таковы, что ряд может расходится при $z=1$.

Пример 9. Для каждой из бернуллиевских случайных величин $X_1, \ldots, X_n$ (см. § 6.3, пример 6) производящая функция равна

$$g(z)=q+pz,$$

так как $a_0=q$, $a_1=p$ в (6.5.1). Тогда производящей функцией их суммы $S_n=X_1+\ldots+X_n$, где X_i — независимы, служит n-я степень функции g:

$$g(z)^n=(q+pz)^n.$$

Ее разложение в степенной ряд известно как теорема о биноме, а именно:

$$g(z)^n=\sum_{k=0}^{n}\binom{n}{k}q^{n-k}p^k z^k.$$

С другой стороны, согласно определению производящей функции,

$$g(z)^n = \sum_{k=0}^{\infty} P(S_n = k)\, z^k.$$

Сравнение последних двух выражений показывает, что

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leqslant k \leqslant n; \quad P(X_n = k) = 0, \quad k > n.$$

Это — формула Бернулли, которую мы получали ранее, но теперь доказали новым методом.

Пример 10. Время ожидания из § 4.4 обладает следующим распределением: $p_j = q^{j-1}p$, $j \geqslant 1$. Следовательно,

$$g(z) = \sum_{j=1}^{\infty} q^{j-1} p z^j = \frac{p}{q} \sum_{j=1}^{\infty} (qz)^j = \frac{p}{q}\ \frac{qz}{1-qz} = \frac{pz}{1-qz}\,. \tag{6.5.9}$$

Положим $S_n = T_1 + \ldots + T_n$, где случайные величины T_i независимы и каждая имеет производящую функцию g (6.5.9). Тогда величина S_n есть время ожидания n-го «успеха». Ее производящей функцией служит функция g^n, и она может быть разложена в степенной ряд с помощью бинома и формулы (5.4.4):

$$\begin{aligned} g(z)^n &= \left(\frac{pz}{1-qz}\right)^n = p^n z^n \sum_{j=0}^{\infty} \binom{-n}{j} (-qz)^j = \\ &= \sum_{j=0}^{\infty} \frac{n(n+1)\ldots(n+j-1)}{j!}\, p^n q^j z^{n+j} = \sum_{j=0}^{\infty} \binom{n+j-1}{n-1} p^n q^j z^{n+j} = \\ &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n q^{k-n} z^k. \end{aligned}$$

Отсюда для $j \geqslant 0$ находим, что

$$P(S_n = n+j) = \binom{n+j-1}{j} p^n q^j = \binom{-n}{j} p^n (-q)^j.$$

Вероятностное распределение $\left\{\binom{-n}{j} p^n (-q)^j, j \geqslant 0\right\}$ называют *отрицательно биномиальным порядка* n. Приведенное выше обсуждение показывает, что его производящая функция задается формулой

$$\left(\frac{g(z)}{z}\right)^n = \left(\frac{p}{1-qz}\right)^n. \tag{6.5.10}$$

Но $g(z)/z$ — производящая функция для величины $T_1 - 1$ (почему?), которая представляет собой количество «неудач» до первого «успеха».

Тогда функция (6.5.10) является производящей для случайной величины $S_n - n$, равной общему количеству «неудач» до появления n-го «успеха».

Пример 11. В задаче об игральных костях из примера 8, в случае, когда они симметричны, имеем $p_j = 1/6$, $1 \leqslant j \leqslant 6$. Поэтому производящая функция для результата, выпавшего на одной кости, задается формулой

$$g(z) = \frac{1}{6}\,(z + z^2 + z^3 + z^4 + z^5 + z^6) = \frac{z(1-z^6)}{6\,(1-z)}\,.$$

Производящая функция для суммы номеров, выпавших на всех трех костях, есть g^3. Ее можно разложить в степенной ряд следующим образом:

$$\begin{aligned} g(z)^3 &= \frac{z^3}{6^3}\,\frac{(1-z^6)^3}{(1-z)^3} = \frac{z^3}{6^3}\,(1-3z^6+3z^{12}-z^{18})(1-z)^{-3} = \\ &= \frac{z^3}{6^3}\,(1-3z^6+3z^{12}-z^{18})\sum_{k=0}^{\infty}\binom{k+2}{2}z^k. \end{aligned} \tag{6.5.11}$$

Коэффициент при z^9 легко найти, зная, что существуют только два способа получить степень 9 в приведенном выше произведении:

$$\frac{1}{6^3}\left\{1\cdot\binom{6+2}{2} - 3\cdot\binom{0+2}{2}\right\} = \frac{28-3}{6^3} = \frac{25}{6^3}\,.$$

Возможно, на читателя не произвела особого впечатления скорость решения новым методом по сравнению с комбинаторным подсчетом, проведенными в § 6.4, но важно обратить внимание на следующую математическую технику.

1°. Вероятности $\{P(X = j),\ \ j \geqslant 0\}$ свертываются в производящую функцию g.

2°. Преобразуется сама функция g — возводится в n-ю степень.

3°. Вычисляются вероятности $\{P(S_n = k),\ k \geqslant 0\}$ путем разложения функции g^n в степенной ряд.

Отличительной особенностью алгоритмического подхода является то, что шаги осуществляются механически, подобно преобразованиям в формуле (6.5.10). Вам не приходится обдумывать каждый шаг: вы включаете что-то в сеть, нажимаете несколько кнопок или выключателей, и желаемый результат достигнут. Чтобы еще больше ощутить полезность введенного инструмента, представим производящую функцию случайной величины X в форме, в которой сам Эйлер ее бы не узнал (понятие случайной величины появилось позже, незадолго до 1930 г.):

$$g(z) = E(z^X), \tag{6.5.12}$$

т. е. в форме математического ожидания случайной величины z^X. Давайте сначала напомним, почему для каждого фиксированного z функция $\omega \to z^{X(\omega)}$ является случайной величиной. Для счетного выборочного пространства Ω это следует из утверждения 2 из § 4.2. Когда X принимает значение j, z^X принимает значение z^j. Следовательно, согласно формуле (4.3.15), математическое ожидание случайной величины z^X может быть представлено в виде $\sum\limits_{j=0}^{\infty} P(X=j)\, z^j$, что совпадает с определением функции $g(z)$.

Полезность введенной для $g(z)$ формы (6.5.12) можно сразу увидеть из следующего более компактного доказательства теоремы 6, базирующегося на других принципах. Ввиду вышесказанного и свойства экспоненты, производящая функция суммы $X_1 + \ldots + X_n$ записывается как

$$E\big(z^{X_1+\ldots+X_n}\big) = E\big(z^{X_1} z^{X_2 \cdots} z^{X_n}\big).$$

Здесь случайные величины $z^{X_1}, z^{X_2}, \ldots, z^{X_n}$ являются независимыми в силу утверждения 6 из § 5.5. Тогда, согласно теореме 2 из § 6.3, имеем:

$$E\big(z^{X_1} z^{X_2} \ldots z^{X_n}\big) = E\big(z^{X_1}\big) E\big(z^{X_2}\big) \ldots E\big(z^{X_n}\big). \tag{6.5.13}$$

Так как $E(z^{X_j}) = g_j(z)$ для каждого j, то теорема 6 доказана.

К другим достоинствам формы $E(z^X)$ относится возможность ее обобщения. Если случайная величина X может принимать произвольные действительные значения, то такое представление производящей функции по-прежнему будет иметь смысл. Ради простоты рассмотрим только случай, когда $0 < z \leqslant 1$. Каждое такое число z представляется в виде $e^{-\lambda}$ при некотором $0 \leqslant \lambda < \infty$. В самом деле, отображение $z = e^{-\lambda}$ является взаимно однозначным, см. рис. 25. После такой замены переменной наша форма перейдет в

$$E\big(e^{-\lambda X}\big), \quad 0 \leqslant \lambda < \infty. \tag{6.5.14}$$

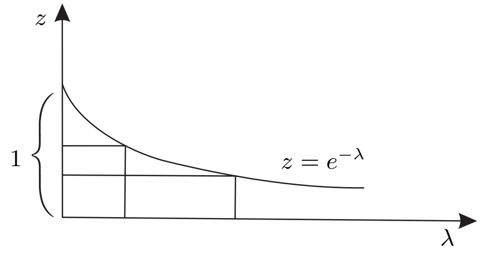


Рис. 25

Если случайная величина X имеет вероятностное распределение (6.5.1),

$$E\big(e^{-\lambda X}\big) = \sum_{j=0}^{\infty} a_j e^{-j\lambda}.$$

Это — не что иное, как наша прежняя функция $g(z)$ с заменой $z = e^{-\lambda}$. В общем случае, если X принимает значения $\{x_j\}$ с вероятностями $\{p_j\}$, то

$$E\big(e^{-\lambda X}\big) = \sum_{j} p_j e^{-\lambda x_j} \tag{6.5.15}$$

при условии, что ряд сходится абсолютно. Это так, когда все значения $x_j \geqslant 0$, поскольку тогда $e^{-\lambda x_j} \leqslant 1$, и ряд мажорируется рядом $\sum\limits_j p_j = 1$. Наконец, если случайная величина X обладает плотностью f, то в силу формулы (4.5.6) с функцией $\varphi(u) = e^{-\lambda u}$, имеем:

$$E\big(e^{-\lambda X}\big) = \int\limits_{-\infty}^{\infty} e^{-\lambda u} f(u)\, du, \tag{6.5.16}$$

при условии, что интеграл сходится. Это так, если $f(u) = 0$ при $u < 0$, другими словами, когда X не принимает отрицательных значений. Тем самым мы с помощью представления (6.5.14) распространили понятие производящей функции на большой класс случайных величин. Наш новый инструмент называется *преобразованием Лапласа* случайной величины X. В аналитическом виде, приведенном в правой части формулы (6.5.16), он широко используется в операционном исчислении, в теории дифференциальных уравнений и инженерных приложениях.

Если в записи (6.5.14) заменить отрицательный действительный параметр $-\lambda$ на чисто мнимый $i\theta$, где $i = \sqrt{-1}$ и θ — действительное число, то получим *преобразование Фурье* $E(e^{i\theta X})$. В теории вероятностей оно также известно под именем *характеристической функции* случайной величины X. Напомним читателю формулу Муавра (которую старшеклассники обычно изучают в курсе тригонометрии): для действительного u

$$e^{iu} = \cos u + i\, \sin u.$$

Ее следствием является формула

$$|e^{iu}|^2 = (\cos u)^2 + (\sin u)^2 = 1.$$

Из нее вытекает, что для любой случайной величины X выполняется равенство $|e^{i\theta X}| = 1$. На основании этого видим, что функция

$$\varphi(\theta) = E(\mathrm{e}^{i\theta X}), \quad -\infty < \theta < \infty, \tag{6.5.17}$$

всегда определена: в самом деле, $|\varphi(\theta)| \leqslant 1$ для всех θ. В этом заключается преимущество данного преобразования перед остальными, обсуждаемыми выше, так как они могут оказаться неопределенными, если соответствующий ряд или интеграл расходится. Но с другой стороны, нам приходится заплатить тем, что необходимо иметь дело с функциями, принимающие комплексные значения, которые обычно не встречаются в элементарных математических курсах. Тем не менее, в гл. 7 мы будем использовать аппарат преобразований Лапласа и Фурье, и для будущих ссылок сформулируем сейчас следующий результат.

Теорема 7. *Теоремы 5 и 6 остаются в силе, если заменить в них производящую функцию на преобразование Лапласа (для неотрицательных случайных величин) или на преобразование Фурье (для произвольных случайных величин).*

В случае теоремы 6 утверждение сразу вытекает из соотношения (6.5.13), если переменную z в нем заменить на $e^{-\lambda}$ или $e^{i\theta}$. Для теоремы 5 аналогия проходит глубже, и требуется более тонкий анализ (см. [4, гл. 6]). Мы просим читателя поверить в ее истинность, принимая во внимание рассмотренную выше *аналогию*, ведущую от $E(z^X)$ к $E(e^{-\lambda X})$ и $E(e^{i\theta X})$. В конце концов, использование аналогий — проверенный временем метод обучения.

Задачи

1. В лотерее штата Массачусетс участвуют 1 миллион билетов. На них приходятся: одна первая премия в размере \$50 000, 9 вторых премий по \$2500 каждая, 90 третьих премий по \$250 каждая, 900 четвертых премий по \$25 каждая. Каков ожидаемый выигрыш при покупке одного билета? Пяти билетов?

2. Допустим, что в лотерее из задачи 1 продано только 80 % билетов. Чему равно математическое ожидание общей суммы выплаченных премий? Если каждый билет стоит 50 центов, то каков ожидаемый средний доход штата?

3. В пяти пригородных жилых блоках был проведен опрос для выяснения расовой принадлежности домовладельцев. В таблице указаны количества домов в каждом из блоков, отдельно для белых и черных домовладельцев:

	1	2	3	4	5
Черные	3	2	4	3	4
Белые	10	10	9	11	10

Если из каждого блока выбираются случайно два дома, то каково ожидаемое число черных домовладельцев в выборке?

4. Шесть костей бросаются один раз. Найдите среднее и дисперсию общего числа выпавших очков. Тот же вопрос для случая n бросаний.

5. Ящик с 1 000 шурупами содержит 1 % шурупов с большими дефектами и 5 % шурупов с небольшими дефектами. Случайно выбираются и проверяются 50 шурупов. Каково ожидаемое число среди них шурупов с большими и небольшими дефектами?

6. Каково среднее число карт пиковой масти на руках у игрока в бридж? Или разных мастей? [*Указание.* Для того, чтобы ответить на второй вопрос, рассмотрите $X_j = 1$ или 0 в зависимости от того, присутствует ли j-я масть среди карт игрока или нет; найдите $E(X_1 + X_2 + X_3 + X_4)$.]

7. Автобус в аэропорту развозит 25 пассажиров по 7 остановкам. Допустим, что каждый из пассажиров одинаково вероятно сходит на любой из остановок, и что они ведут себя независимо друг от друга. Автобус останавливается только тогда, когда кто-нибудь собирается сойти. Какова вероятность того, что никто не сойдет на третьей остановке? Чему равно среднее число остановок, на которых хоть кто-то сойдет? [*Указание.* Положите $X_j = 1$ или 0 в зависимости от того, сходит ли кто-нибудь на j-й остановке или нет.]

8.* Случайно выбираются 500 человек. (a) Какова вероятность того, что среди них окажется несколько человек, у которых день рождения 1 января? (b) Чему равно ожидаемое число людей среди них с данным днем рождения? (c) Каково ожидаемое число дней в году, являющихся днями рождения по крайней мере одного их этих людей? (d) Каково ожидаемое число дней в году, являющихся днями рождения больше, чем одного из этих людей? Для простоты не рассматривайте високосные годы. [*Указание.* Для пп. (b), (c) и (d) действуйте, как в задаче 7.]

9.* Задачи 6, 7 и 8 являются разными вариантами *задач о размещениях*, которые формулируются в общем виде следующим образом. Раскладываются n незанумерованных жетонов по m занумерованным ящикам (см. § 3.3). Каково среднее число ящиков, в которые попадут в точности (или по крайней мере) k жетонов? Можно также спросить, каково среднее число жетонов, которые не будут делить свой ящик ни с одним другим жетоном? Дайте ответы на эти вопросы и перефразируйте их на языке задач 6, 7 и 8.

10.* Используя описанную выше модель размещений, найдите распределение жетонов по ящикам, а именно, вероятности того, что в точности n_j жетонов окажутся в j-м ящике, $1 \leqslant j \leqslant m$. Опишите это распределение на языке задач 6, 7 или 8.

11. Автоматический механизм производит дефектную деталь с вероятностью 2 %. Когда это происходит, выполняется регулировка механизма. Найдите среднее число качественных деталей, производимых между регулировками.

12. Только один из шести внешне похожих ключей открывает определенную дверь. Если пробовать ключи один за другим, то сколько в среднем ключей понадобится испытать, прежде чем дверь будет открыта?

13. Тестируются 100 электрических лампочек. Если вероятность того, что лампочка не загорится, равна p, то чему равны среднее и стандартное отклонение числа незагоревшихся лампочек? (Предполагается, что лампочки стохастически независимы.)

14*. Пятьдесят человек проходят рентгеновский контроль. Допустим, что для четверых из них результат будет «положительным». Чему равно среднее число «отрицательных» результатов, прежде чем будет выявлен первый «положительный»? [*Указание.* Представьте, что эти четверо играют роль разделяющих перегородок между остальными. Тем самым, задача эквивалентна вычислению ожидаемого числа жетонов в первом ящике в условиях (IV′) из § 3.3.]

15. В сумке находятся N купонов с номерами от 1 до N. Они извлекаются один за другим в возвращением. (a) Чему равно среднее число извлечений до того, как купон, вынутый первым, будет извлечен снова? (b)* Чему равно среднее число извлечений до первого дублирования номеров? [*Указание.* В п. (b) вычислите сначала вероятность того, что в n извлечениях не было дублирования.]

16*. В условиях задачи 15 найдите, чему равно математическое ожидание наибольшего номера, полученного в n извлечениях. Тот же вопрос для случая выбора без возвращения. [*Указание.* Найдите $P(\text{maximum} \leqslant k)$.]

17. Для урновой схемы Пойа с параметром $c \geqslant -1$ (см. § 5.4) ответьте на следующие вопросы.

(a) Чему равно ожидаемое число красных шаров в n извлечениях?

(b) Чему равно ожидаемое число красных шаров в урне после n извлечений (и возвращения назад c шаров)?

18. Пусть $p_n \geqslant 0$ и $r_n = \sum\limits_{k=n}^{\infty} p_k$. Покажите, что

$$\sum_{n=1}^{\infty} np_n = \sum_{n=1}^{\infty} r_n,$$

где оба ряда одновременно сходятся или расходятся к $+\infty$. Следовательно, если X есть случайная величина, принимающая целые неотрицательные значения, то

$$E(X) = \sum_{n=1}^{\infty} P(X \geqslant n). \tag{6.6.2}$$

[*Указание.* Запишите $p_n = r_n - r_{n+1}$ и измените порядок суммирования (в некоторых учебниках по математическому анализу этот прием называется методом суммирования Абеля).]

19. Примените формулу (6.6.2) для вычисления математического ожидания времени ожидания в примере 8 из § 4.4. Заметьте, что $P(X \geqslant n) = q^{n-1}$, $n \geqslant 1$.

20. Пусть $X_1, \ldots, X_m$ — независимые неотрицательные целочисленные случайные величины с общим распределением $\{p_n, n \geqslant 0\}$; $r_n = \sum\limits_{k=n}^{\infty} p_k$. Покажите, что

$$E\{\min(X_1, \ldots, X_m)\} = \sum_{n=1}^{\infty} r_n^m.$$

[*Указание.* Используйте задачу 18.]

21. Пусть X — неотрицательная случайная величина с плотностью f. Установите, что если $r(u) = \int\limits_u^{\infty} f(t)\,dt$, то

$$E(X) = \int_0^{\infty} P(X \geqslant u)\,du = \int_0^{\infty} r(u)\,du. \tag{6.6.3}$$

[*Указание.* Эта формула — аналог (6.6.2). Иметь дело с интегралами проще, чем с суммами.]

22. Примените формулу (6.6.3) к случайной величине X, обладающей экспоненциальной плотностью $\lambda e^{-\lambda t}$, $t \geqslant 0$.

23. Продолжительность T телефонных разговоров определенного типа, как выяснилось, удовлетворяет соотношению

$$P(T > t) = a e^{-\lambda t} + (1-a) e^{-\mu t}, \quad t \geqslant 0,$$

где $0 \leqslant a \leqslant 1$, $\lambda > 0$, $\mu > 0$ — константы, оцениваемые статистически. Найдите среднее и дисперсию случайной величины T. [*Указание.* Быстрый способ решения для случая среднего дает применение результата из задачи 21.]

24. Допустим, что время до выхода из строя электронного прибора, измеряемое в часах, имеет экспоненциальную плотность $\lambda e^{-\lambda t}$. При условии, что прибор уже проработал n часов, сколько еще времени он проработает? Сравните это с ожидаемой первоначально продолжительностью работы. Не видите ли вы здесь противоречия?

25. Пусть одновременно тестируются пять приборов, описанных выше. (a) Сколько придется ждать, пока один из них выйдет из строя? (b) Сколько придется ждать, пока испортятся все пять приборов?

26. Средняя относительная ошибка при измерении диаметра круглого диска составляет 0.2 %. На основании результата данного измерения вычисляется площадь диска. Какова средняя относительная ошибка в площади, если не учитывать квадрат относительной ошибки, допущенной при измерении диаметра?

27. Пусть a и b — некоторые константы. Выразите среднее и дисперсию случайной величины $aX + b$ через среднее и дисперсию X. Примените полученный результат к пересчету температуры из шкалы Цельсия в шкалу Фаренгейта: $F = \frac{9}{5}\,C + 32$.

28. Игрок пришел в выводу, что он может всегда обыграть казино, удваивая ставку каждый раз, чтобы компенсировать прошлые потери. Точнее говоря, он решил закончить игру, как только выиграет, в противном случае — он станет удваивать ставку до тех пор, пока не выиграет. Единственным недостатком его стратегии является то, что ему придется прекратить игру, когда у него закончатся деньги. Допустим, что у игрока есть \$150, и он начинает со ставки в \$1. Предположим также, что в каждой из игр его шансы на победу равны 50 %. Какова вероятность того, что он окажется победителем, и сколько он тогда выиграет? Какова вероятность того, что ему придется прекратить игру после очередного проигрыша из-за недостатка денег для удвоения ставки, и сколько он в таком случае проиграет? Чему равно математическое ожидание дохода при такой стратегии? Тот же вопрос, если игрок ставит на кон весь свой оставшийся капитал, когда уже не может удвоить ставку.

29. Выбираются наудачу n точек из отрезка $[0, 1]$. Найдите математическое ожидание максимума, минимума и размаха (= maximum − minimum).

30. Рассмотрим n независимых событий A_j с $P(A_j) = p_j$, $1 \leqslant j \leqslant n$. Пусть N обозначает (случайное) число тех из них, которые произошли. Найдите производящую функцию случайной величины N и вычислите с ее помощью $E(N)$.

31. Пусть $\{p_j, j \geqslant 0\}$ — вероятностное распределение и

$$u_k = \sum_{j=0}^{k} p_j, \quad g(z) = \sum_{k=0}^{\infty} u_k z^k.$$

Докажите, что последний степенной ряд сходится при $|z| < 1$. В качестве частного случая рассмотрим S_n из примера 9 в § 6.5 и $p_j = P\{S_n = j\}$. Каков тогда смысл u_k? Найдите производящую функцию g.

32. Можно определить производящую функцию для случайной величины, принимающей положительные и отрицательные значения. Пусть для простоты

$$P(X = k) = p_k, \quad k = 0, \pm 1, \pm 2, \ldots, \pm N.$$

Тогда функция $g(z) = \sum\limits_{k=-N}^{+N} p_k z^k$ является *дробно-рациональной функцией* от z, т. е. отношением двух многочленов. Найдите g для случая $p_k = 1/(2N + 1)$, что соответствует равномерному распределению на множестве

$$\{-N, -(N-1), \ldots, -1, 0, +1, \ldots, N-1, N\}.$$

Для проверки вычислите математическое ожидание с помощью g'.

33. Пусть $\{X_j, 1 \leqslant j \leqslant n\}$ — независимые случайные величины, такие, что

$$X_j = \begin{cases} 1 & \text{с вероятностью } 1/4, \\ 0 & \text{с вероятностью } 1/2, \\ -1 & \text{с вероятностью } 1/4, \end{cases}$$

и $S_n = \sum\limits_{j=1}^{n} X_j$. Найдите производящую функцию случайной величины S_n в смысле определения из задачи 32 и вычислите с ее помощью $P(S_n = 0)$. В качестве конкретного применения представим, что игроки A и B бросают симметричную монету n раз каждый. Какова вероятность того, что они получат одинаковое число гербов? (Эту задачу также можно решить без использования производящей функции, опираясь на формулу (3.3.9).)

34.* Пусть в задаче 15 о собирании купонов величина T обозначает количество извлечений до того момента, когда будут собраны все купоны. Найдите производящую функцию случайной величины T. Используя ее, найдите математическое ожидание в качестве замечательной проверки результата (6.1.8). [*Указание.* Обозначим через T_j время ожидания между получением $(j-1)$-го и j-го новых номеров. Тогда T_j имеют геометрические распределения с вероятностями $p_j = (N-j+1)/N$, причем они независимы.]

35. Рассмотрим X и g из (6.5.1) и (6.5.2). Получите явные формулы для первых четырех моментов случайной величины X в терминах функции g и ее производных.

36. Обозначим через $L(\lambda)$ преобразование Лапласа случайной величины X, определяемое формулой (6.5.16). Выразите n-й момент случайной величины X в терминах L и ее производных.

37. Найдите преобразования Лапласа, соответствующие плотностям f, приведенным ниже.

(a) $f(u) = 1/c$ на $(0, c)$, где $c > 0$;

(b) $f(u) = 2u/c^2$ на $(0, c)$, где $c > 0$;

(c) $f(u) = (\lambda^n u^{n-1}/(n-1)!)\, e^{-\lambda u}$ на $[0, \infty)$, где $\lambda > 0$, $n \geqslant 1$. (Прежде всего проверьте, что это — плотность! Соответствующее распределение носит название *гамма-распределения* и обозначается через $\Gamma(n; \lambda)$.)

38. Пусть $S_n = T_1 + \ldots + T_n$, где T_j — независимые случайные величины с общей плотностью $\lambda e^{-\lambda t}$. Найдите преобразование Лапласа случайной величины S_n. Сравните результат с ответом из п. (c) задачи 37. Теперь можно использовать теорему 7 для идентификации распределения S_n.

39.* Рассмотрим группу из N налогоплательщиков, которые платят разные суммы налогов со средним m и дисперсией σ^2. Если n из них выбраны случайно, покажите, что среднее и дисперсия их общей суммы налогов равны, соответственно,

$$nm \quad \text{и} \quad \frac{N-n}{N-1}\, n\sigma^2.$$

[*Указание.* Обозначьте отдельные налоговые суммы через $X_1, \ldots, X_n$ и примените формулу (6.3.8). Можно упростить вычисления, заметив, что $E(X_j X_k)$ не зависит от n, поэтому его можно определить, положив $n = N$, однако вполне можно обойтись и без этого приема.]

40. Докажите теорему 1 с помощью метода, использованного при доказательстве теоремы 2. Рассмотрите также случай, когда существуют плотности.

41. Пусть $a(\cdot)$ и $b(\cdot)$ — две плотности. Определим их *свертку* $c(\cdot)$ следующим образом:

$$c(v) = \int\limits_{-\infty}^{\infty} a(u)\, b(v-u)\, du, \quad -\infty < v < \infty;$$

сравните с (6.5.8). Убедитесь, что $c(\cdot)$ является плотностью. (Она часто обозначается через $a * b$.)

42*. В случае $a(u) = \lambda e^{-\lambda u}$ при $u \geqslant 0$ найдите свертку $a(\cdot)$ с самой собой. Найдите по индукции n-кратную свертку

$$\underbrace{a * a * \ldots * a}_{n \text{ раз}}.$$

[*Указание.* Ответ дан в п. (c) задачи 37.]

43. Докажите теорему 4 для неотрицательных целочисленных случайных величин с помощью производящих функций. [*Указание.* Выразите дисперсию через производящую функцию, как в задаче 35, а затем используйте теорему 6.]

44. Докажите аналоги теоремы 6 для преобразований Лапласа и Фурье.

45. Рассмотрим последовательность независимых испытаний, каждое их которых является «успехом» с вероятностью p и «неудачей» с вероятностью $q = 1 - p$. Покажите, что вероятность того, что n-му «успеху» предшествовало в точности j «неудач», равна

$$\binom{n+j-1}{j} p^n q^j.$$

46*. Выведите следующую формулу:

$$\sum_{j+k=l} \binom{m+j-1}{j}\binom{n+k-1}{k} = \binom{m+n+l-1}{l},$$

в которой суммирование проводится по всем $j \geqslant 0$ и $k \geqslant 0$ таким, что $j + k = l$. [*Указание.* Она более узнаваема в виде:

$$\sum_{j+k=l} \binom{-m}{j}\binom{-n}{k} = \binom{-m-n}{l};$$

сравните с (3.3.9). Используйте тождество $(1-z)^{-m}(1-z)^{-n} = (1-z)^{-m-n}$.]

47.* Обобщим задачу о справедливом дележе ставки из примера 6 в § 2.2. Два игрока состязаются в игре, состоящей из серии партий. Вероятности выигрыша в отдельной партии равны p и $q = 1 - p$ для игроков A и B соответственно. Представим, что состязание было прервано, когда игроку A не хватало до выигрыша m побед, а игроку B — n побед. Покажите, что вероятность того, что при продолжении состязания победу в игре одержит игрок A, дается любым из приведенных ниже выражений:

(i) $\displaystyle\sum_{k=m}^{m+n-1} \binom{m+n-1}{k} p^k q^{m+n-1-k}$;

(ii) $\displaystyle\sum_{k=0}^{n-1} \binom{m+k-1}{k} p^m q^k$.

Решение этой задачи впервые было получено Монмаром (1678–1719). [*Указание.* Результат (i) сразу вытекает из формулы Бернулли и основан на идее (см. пример 6 в § 2.2) продолжения состязания еще ровно на $m + n - 1$ партий, даже если A одержит победу прежде, чем все эти партии будут сыграны. Формула (ii) основана на более естественной идее прекращения игры, как только A выиграет m партий прежде, чем B выиграет n партий. Допустим, что это случилось точно после $m + k$ партий. Тогда A должен выиграть последнюю партию, а также $m-1$ партий среди первых $m + k - 1$ партий, где $k \leqslant n - 1$.]

48.* Докажите непосредственно, что выражения из (i) и (ii) в задаче 47 равны между собой. [*Указание.* Можно сделать это, используя индукцию по n для фиксированного m. Но более интересный подход заключается в сравнении идей, лежащих в основе двух решений. Он приводит к представлению (ii) в следующем виде:

$$\sum_{k=0}^{n-1} \binom{m+k-1}{k} p^m q^k (p+q)^{n-1-k} =$$
$$= \sum_{k=0}^{n-1} \binom{m+k-1}{k} p^m q^k \sum_{j=0}^{n-k-1} \binom{n-1-k}{j} p^{n-1-k-j} q^j =$$
$$= \sum_{l=0}^{n-1} p^{m+n-1-l} q^l \sum_{j+k=l} \binom{m+k-1}{k} \binom{n-k-1}{j}.$$

Осталось воспользоваться тождеством из задачи 46. Обратите внимание, что данное равенство связывает биномиальное распределение с отрицательно биномиальным распределением.]

Глава 7

Пуассоновское и нормальное распределения

7.1. Модели, в которых используется пуассоновское распределение

Распределение Пуассона имеет большое значение для теории и практики. Оно также обладает тем полезным качеством, что является достаточно простым математическим объектом. Мы могли бы ввести его в нашей книге на более ранней стадии (читатель был предупрежден об этом в § 4.4). Однако отложенное знакомство придаст ему бóльшую важность. Кроме того, благодаря двум предыдущим главам стало возможным более полное обсуждение этой темы.

Зафиксируем действительное положительное число α и рассмотрим вероятностное распределение $\{a_k, k \in \mathbb{N}^0\}$, где $\mathbb{N}^0$ — множество всех целых неотрицательных чисел, задаваемое формулой:

$$a_k = \frac{e^{-\alpha}}{k!}\,\alpha^k. \tag{7.1.1}$$

Прежде всего, проверим, что

$$\sum_{k=0}^{\infty} a_k = e^{-\alpha}\sum_{k=0}^{\infty}\frac{\alpha^k}{k!} = e^{-\alpha}\cdot e^{\alpha} = 1,$$

где мы использовали разложение e^{α} в ряд Тейлора. Давайте вычислим также математическое ожидание:

$$\sum_{k=0}^{\infty} k a_k = e^{-\alpha}\sum_{k=0}^{\infty} k\,\frac{\alpha^k}{k!} = e^{-\alpha}\alpha\sum_{k=1}^{\infty}\frac{\alpha^{k-1}}{(k-1)!} =$$

$$= e^{-\alpha}\alpha\sum_{k=0}^{\infty}\frac{\alpha^k}{k!} = e^{-\alpha}\alpha e^{+\alpha} = \alpha.$$

(Это небольшое вычисление было проведено подробно по той причине, что первый из авторов обнаружил, что студенты, как правило, не овладевают навыком решения таких задач при изучении «бесконечных рядов» в курсе математического анализа.) Поэтому параметр α в действительности несет определенный смысл. Мы будем называть распределение (7.1.1) *законом Пуассона с параметром* α. Оно будет обозначаться через $\pi(\alpha)$, а отдельные вероятности с индексом k — через $\pi_k(\alpha)$. Тогда если X — случайная величина, имеющая такое распределение, то

$$P(X = k) = \pi_k(\alpha) = \frac{e^{-\alpha}}{k!}\,\alpha^k, \quad k \in \mathbb{N}^0;$$

и

$$E(X) = \alpha. \tag{7.1.2}$$

Далее, вычислим производящую функцию g, определенную в § 6.5. Используя разложение в ряд Тейлора функции $e^{\alpha z}$, находим:

$$g(z) = \sum_{k=0}^{\infty} a_k z^k = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}\, z^k = e^{-\alpha} e^{\alpha z} = e^{\alpha(z-1)}. \tag{7.1.3}$$

Эта функция достаточно проста, и к ней удобно применять инструменты математического анализа. Если продифференцировать ее дважды, то получим:

$$g'(z) = \alpha e^{\alpha(z-1)}, \quad g''(z) = \alpha^2 e^{\alpha(z-1)}.$$

Отсюда, используя соотношения (6.5.6), выводим, что

$$\begin{aligned} E(X) &= g'(1) = \alpha, \\ E(X^2) &= g'(1) + g''(1) = \alpha + \alpha^2, \\ \sigma^2(X) &= \alpha. \end{aligned} \tag{7.1.4}$$

Таким образом, дисперсия равна α, так же, как и среднее (ниже приводится объяснение этого).

Математически распределение Пуассона может быть получено множеством важных способов. Один из них — предельный переход от биномиального распределения. Он исторически известен как предельный закон Пуассона, и мы рассмотрим его первым. Другой способ дает схема сложения независимых экспоненциальных случайных величин. Эта тема составляет основное содержание следующего параграфа.

Вспомним определение биномиального распределения $B(n; p)$ из § 4.4 и запишем:

$$B_k(n; p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leqslant k \leqslant n. \tag{7.1.5}$$

Позволим p меняться вместе с n. Это всего лишь означает, что мы полагаем $p = p_n$ в формуле (7.1.5). Более конкретно, возьмем

$$p_n = \frac{\alpha}{n}, \quad n \geqslant 1. \tag{7.1.6}$$

В результате получим последовательность биномиальных распределений $B(n; \alpha/n)$, вероятности в которых имеют вид

$$B_k\left(n; \frac{\alpha}{n}\right) = \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k}, \quad 0 \leqslant k \leqslant n. \tag{7.1.7}$$

Для краткости обозначим эти вероятности через $b_k(n)$. Теперь зафиксируем k и устремим n к бесконечности. Оказывается, при этом последовательность $\{b_k(n)\}$ сходится при каждом k, и предел может быть найден следующим образом. Для начала возьмем $k = 0$. Тогда

$$\lim_{n\to\infty} b_0(n) = \lim_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^n = e^{-\alpha}. \tag{7.1.8}$$

Эта формула — одна из важнейших для экспоненциальной функции, и вы должны помнить ее из курса математического анализа. Несложно вывести ее из тейлоровского разложения $\ln(1-x) = -\sum\limits_{n=1}^{\infty} x^n/n$:

$$\begin{aligned}\ln\left(1-\frac{\alpha}{n}\right)^n = n\ln\left(1-\frac{\alpha}{n}\right) = n\left\{-\frac{\alpha}{n}-\frac{\alpha^2}{2n^2}-\ldots\right\} = \\ = -\alpha-\frac{\alpha^2}{2n}-\ldots .\end{aligned} \tag{7.1.9}$$

Когда $n\to\infty$, правая часть сходится к $-\alpha = \ln e^{-\alpha}$. Тем самым, формулу (7.1.8) удается получить с помощью взятия логарифма и разложения в ряд Тейлора — метода, широко используемого в прикладной математике. Строгое доказательство должно содержать обоснование того, что членами, вместо которых в конце формулы (7.1.9) стоит многоточие, можно пренебречь; см. задачу 18.

Продолжая, вычислим отношение соседних членов в формуле (7.1.7):

$$\frac{b_{k+1}(n)}{b_k(n)} = \frac{n-k}{k+1}\left(\frac{\alpha}{n}\right)\left(1-\frac{\alpha}{n}\right)^{-1} = \frac{\alpha}{k+1}\left[\left(\frac{n-k}{n}\right)\left(1-\frac{\alpha}{n}\right)^{-1}\right].$$

Оба множителя внутри квадратных скобок здесь сходятся к 1 при $n\to\infty$. Поэтому

$$\lim_{n\to\infty}\frac{b_{k+1}(n)}{b_k(n)} = \frac{\alpha}{k+1}\,. \tag{7.1.10}$$

Начиная с предела (7.1.8) и используя соотношение (7.1.10) для $k = 0, 1, 2, \ldots$, находим, что

$$\begin{aligned}&\lim_{n\to\infty} b_1(n) = \frac{\alpha}{1}\lim_{n\to\infty} b_0(n) = \alpha e^{-\alpha},\\ &\lim_{n\to\infty} b_2(n) = \frac{\alpha}{2}\lim_{n\to\infty} b_1(n) = \frac{\alpha^2}{1\cdot 2}\,e^{-\alpha},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &\lim_{n\to\infty} b_k(n) = \frac{\alpha}{k}\lim_{n\to\infty} b_{k-1}(n) = \frac{\alpha^k}{1\cdot 2\cdot\ldots\cdot k}\,e^{-\alpha}.\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\end{aligned}$$

Эти предельные значения суть последовательные вероятности для распределения $\pi(\alpha)$. Тем самым мы установили теорему Пуассона в ее простейшей форме (*предельный закон Пуассона*):

$$\lim_{n\to\infty} B_k\left(n; \frac{\alpha}{n}\right) = \pi_k(\alpha), \quad k \in \mathbb{N}^0.$$

Данный результат остается в силе, если заменить α/n в левой части на α_n/n, где $\lim\limits_{n\to\infty} \alpha_n = \alpha$. Другими словами, вместо того, чтобы брать $p_n = \alpha/n$ или $np_n = \alpha$, как мы делали в записи (7.1.6), можно рассмотреть $p_n = \alpha_n/n$ или $np_n = \alpha_n$ с условием, что

$$\lim_{n\to\infty} np_n = \lim_{n\to\infty} \alpha_n = \alpha. \tag{7.1.11}$$

Вывод этого обобщения аналогичен проведенному выше за исключением того, что формула (7.1.8) заменяется на следующий более сильный результат: если $\lim\limits_{n\to\infty} \alpha_n = \alpha$, то

$$\lim_{n\to\infty} \left(1 - \frac{\alpha_n}{n}\right)^n = e^{-\alpha}. \tag{7.1.12}$$

С учетом данного обобщения, удается выразить теорему Пуассона в более прагматичной форме следующим образом. Биномиальная вероятность $B_k(n;p)$ при небольших k может быть аппроксимирована с помощью вероятности $\pi_k(\alpha)$, когда n велико по сравнению с $np \approx \alpha$. Напомним, что np есть математическое ожидание в бернуллиевском законе $B(n;p)$ (см. § 4.4). Неудивительно, что аппроксимируемое значение α служит математическим ожиданием предельного распределения Пуассона, как мы видим из формулы (7.1.2). Аналогично, дисперсия биномиального распределения $B(n;p)$ равна $npq = n(\alpha/n)\,[1 - (\alpha/n)]$ для $p = \alpha/n$. При $n \to \infty$ в пределе опять возникает α, как было отмечено после соотношений (7.1.4).

Итак, мы познакомились с законом Пуассона. Предельный переход от биномиальной схемы вполне элементарен, чего нельзя сказать о подходе, который излагается в § 7.3. Но выполняется ли на практике условие (7.1.6) или более слабое условие (7.1.11)? Поразительно, что огромное количество самых разных естественных и созданных человеком случайных явлений прекрасно описываются данной моделью. Приведем четыре примера для иллюстрации того, как эта схема используется в большей или меньшей степени.

Пример 1. Рассмотрим «редкое» событие, т. е. событие, происходящее с малой вероятностью p. К примеру, если кто-то ставит на определенный номер на рулетке, то вероятность выигрыша равна $1/37 \approx 0.027$

в предположении, что все 36 номеров и «зеро» одинаково правдоподобны. (Рулетки в Монте-Карло имеют единственное «зеро», а в Лас-Вегасе — «двойное зеро».) Если делать ставку 37 раз, то можно ожидать одного выигрыша. (Какая теорема говорит об этом?) Но можно также вычислить вероятности не выиграть ни разу, выиграть один раз, два раза и т. д. Точные ответы, конечно, равны трем первым вероятностям распределения $B(37;1/37)$:

$$\left(1-\frac{1}{37}\right)^{37},$$

$$\binom{37}{1}\left(1-\frac{1}{37}\right)^{36}\frac{1}{37}=\left(1-\frac{1}{37}\right)^{36},$$

$$\binom{37}{2}\left(1-\frac{1}{37}\right)^{35}\frac{1}{37^2}=\frac{36}{2\times 37}\left(1-\frac{1}{37}\right)^{35}.$$

Положим

$$c=\left(1-\frac{1}{37}\right)^{37}\approx 0.363.$$

Тогда три числа, приведенных выше, запишутся в виде

$$c,\quad \frac{37}{36}\,c,\quad \frac{37}{36}\times\frac{1}{2}\,c.$$

Если для c использовать аппроксимацию $e^{-1}\approx 0.368$, допуская при этом ошибку в 1.5 %, а также «приравнять» 37/36 к 1, совершая, тем самым, ошибку в 3 %, то придем к трем первым вероятностям закона $\pi(1)$, а именно:

$$e^{-1},\quad e^{-1},\quad \frac{1}{2}\,e^{-1}.$$

Если подобным образом аппроксимировать другие вероятности, то неточность будет увеличиваться, правда какие-то из ошибок будут компенсировать другие. Можно придерживаться и такой стратегии: ставить 111 раз ($111=37\times 3$) на некоторый номер, меняя его время от времени, как обычно и делают участники игры в рулетку. Аналогичный подход дает следующие приближения:

$$\left(1-\frac{1}{37}\right)^{111}=c^3\approx e^{-3},$$

$$\frac{111}{37}\left(1-\frac{1}{37}\right)^{110}=\frac{37}{36}\times 3c^3\approx 3e^{-3},$$

$$\frac{111\times 110}{2}\left(1-\frac{1}{37}\right)^{109}=\frac{111\times 110}{36\times 36}\,\frac{1}{2}\,c^3\approx\frac{9}{2}\,e^{-3}$$

и т. д. Здесь c^3, конечно, хуже аппроксимирует e^{-3}, чем c аппроксимирует e^{-1}. В любом случае должно быть ясно, что мы имеем дело с более или менее грубыми числовыми приближениями без всякого предельного перехода. Сколь бы малым ни было значение p, до тех пор, пока оно фиксировано, как в данном примере, np будет, конечно, стремиться к бесконечности с увеличением n, и обсуждаемая выше предельная схема будет все хуже соответствовать действительности. Тем не менее, достаточно хорошее приближение может быть достигнуто для n и p таких, что np является относительно малым по сравнению с n. Это — как раз тот случай, когда числовая аппроксимация оказывается наиболее простой и точной, и множество приложений связано именно с разнообразными редкими событиями. Когда-то закон Пуассона даже называли «законом малых чисел». В качестве примеров часто приводятся статистические данные о количестве прусских кавалеристов, убитых копытами лошадей, или о количестве детских самоубийств в Пруссии, которые превосходно описываются пуассоновским распределением (см. [16]).

Пример 2. Вернемся к задаче о совпадениях из § 6.2. Если человек, утверждающий, что он обладает сверхестественными способностями, на самом деле — обманщик, который всего лишь пытается угадать номера, написанные на карточках, то будет ли его результат лучше или хуже при увеличении числа карточек? Интуитивно наблюдаются два противоположных эффекта. С одной стороны, становится все больше попыток для угадывания; с другой — становится сложнее угадывать каждый из номеров. Как оказывается (см. § 6.2), оба эффекта уравновешиваются настолько хорошо, что ожидаемое число совпадений равно 1 независимо от количества карточек! Здесь мы попадаем в ситуацию (7.1.6) с $\alpha = 1$. В действительности, можно устроить полное соответствие с биномиальной моделью, допустив возможность повторов при угадывании. Для этого представим, что эксперимент по угадыванию номеров, написанных на карточках, лежащих номерами вниз на столе, проводится с условием забывания уже названных номеров. Тогда вероятность правильно назвать любой из n номеров равна $1/n$ независимо от результатов других угадываний. Вероятность получить при этом в точности k верных ответов задается формулой (7.1.7) с $\alpha = 1$, а пуассоновская аппроксимация годится, когда n велико.

Данный вариант задачи о совпадениях соответствует случайному выбору с возвращением. Он не отражает действительности, если последовательно сопоставляются две колоды с карточками. В таком случае существует взаимная зависимость между отгадываниями, и биномиальное распределение уже не подходит. Однако можно показать, что когда n велико, влияние зависимости мало. Действительно, пусть q_n есть вероятность того, что не будет ни одного правильного ответа при сопоставлении всех n карточек. Мы видели в примере 4 из § 6.2, что

$$q_n \approx e^{-1}$$

служит превосходной аппроксимацией даже для умеренных значений n. Простое комбинаторное рассуждение (задача 19) показывает, что вероятность в точности k совпадений равна

$$\binom{n}{k} \frac{1}{(n)_k} q_{n-k} = \frac{1}{k!} q_{n-k}. \tag{7.1.13}$$

Тогда для фиксированного k это выражение сходится к

$$\frac{1}{k!} e^{-1} = \pi_k(1).$$

Пример 3. При изучении пространственных распределений закон Пуассона обычно возникает в результате подсчета частиц, расположение которых напоминает «однородный хаос». К примеру, можно подсчитывать количества наблюдаемых в микроскоп вирусов внутри каждой из квадратных ячеек сетки, покрывающей область наблюдения. Допустим, что всего имеется N ячеек, среднее число вирусов в одной ячейке равно μ, вирусы свободно передвигаются по всей области наблюдения (см. рис. 26). Тогда их распределение можно аппроксимировать с помощью размещения жетонов по ящикам, описываемого моделью (I') из § 3.3. В нашем случае необходимо разместить μN частиц по N ячейкам, причем каждая из частиц попадает в любую из ячеек с вероятностью $1/N$ независимо от остальных частиц. Тогда вероятность обнаружить в точности k частиц в определенной ячейке задается с помощью биномиального распределения:

$$B_k\left(\mu N; \frac{1}{N}\right) = \binom{\mu N}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{\mu N - k}.$$

Давайте теперь представим, что исследуемый фрагмент является частью намного большей области, заселенной вирусами с такой же средней плотностью μ. На практике данное допущение выполняется с достаточной точностью, скажем, в случае анализа пробы крови, взятой у больного. Тогда, если N велико, законно аппроксимировать приведенную выше вероятность с помощью вероятности $\pi_k(\mu)$. Суть здесь в том, что размер маленьких ячеек, внутри которых производится подсчет вирусов, остается фиксированным, и пространственная однородность позволяет перейти к пределу при возрастании количества ячеек.

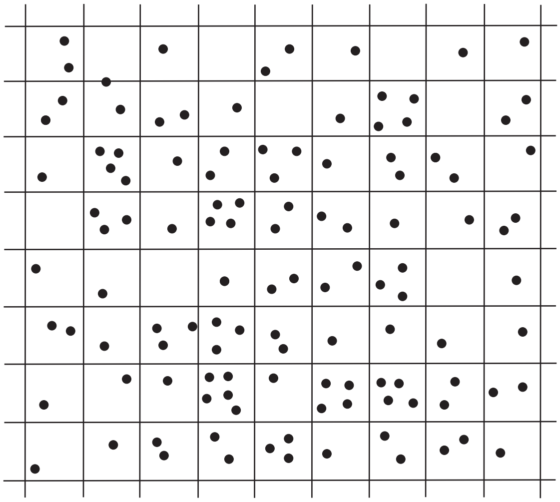


Рис. 26

Мрачный пример пространственного распределения дает подсчет количества самолетов-снарядов, упавших в южной части Лондона во время Второй мировой войны. Вся территория была разбита на $N = 576$ квадратов площадью в 1/4 квадратной мили. Параметр μ был оценен статистически числом 0.9323. Приводимая ниже таблица содержит числа N_k — фактические количества квадратов ровно с k попаданиями и их пуассоновские приближения с $\mu = 0.9323$. Хорошее согласие в данном случае можно объяснить преднамеренной случайностью атак, которая оправдывает применение биномиальной модели.

k	0	1	2	3	4	$\geqslant 5$
N_k	229	211	93	35	7	1
$N\pi_k$	226.74	211.39	98.54	30.62	7.14	1.59

Пример 4. Во многих приложениях роль пространства из предыдущего примера играет время. Если случайные происшествия распределены на некотором временно́м периоде так, что их *интенсивность* (количество происшествий за единицу времени) может считаться примерно постоянной на протяжении всего периода, то применима схема Пуассона, где время выступает в качестве среды для однородного хаоса. Нетрудно повторить рассуждения из примера 3 с заменой пространства на время (последнее, вероятно, даже естественнее разбивать на части). Представим, например, процесс регистрируемых счетчиком Гейгера столкновений космических частиц, происходящих со средней частотой α столкно-

вений в секунду. Тогда вероятность столкновения за малый временной интервал δ представляется в виде $\alpha\delta + o(\delta)$, где «o-малое» обозначает величину, имеющую более высокий порядок малости по сравнению с δ или, проще говоря, «пренебрежимо малую» величину. Теперь разобьем временной отрезок $[0, t]$ на N равных частей. Тогда вероятность регистрации столкновения в каждом из интервалов разбиения равна

$$\frac{\alpha t}{N} + o\left(\frac{t}{N}\right),$$

где роль δ играет t/N. Конечно, $\alpha t/N$ намного меньше 1, если t фиксировано и N велико. Прежде всего, допустим, что для достаточно больших значений N вероятность наблюдать более одного столкновения в любом из интервалов разбиения пренебрежимо мала. Поэтому можно считать, что количества столкновений в каждом из N интервалов принимают только два значения: 0 и 1. Эти количества могут рассматриваться как бернуллиевские случайные величины с вероятностями $\alpha t/N$ «успеха» и $1 - (\alpha t/N)$ «неудачи». Наконец, предположим, что они независимы между собой. Последнее предположение допускает эмпирическую проверку; более глубокий анализ в терминах пуассоновского процесса проводится в следующем параграфе. При сделанных допущениях очевидно, что вероятность наблюдать в точности k столкновений за весь период $[0, t]$ равна $B_k(N; \alpha t/N)$. Действительно, общее количество столкновений во временно́м интервале $[0, t]$ есть сумма определенных выше N независимых бернуллиевских случайных величин. (См. пример 9 в § 4.4.) Так как число N находится в нашем распоряжении и может быть выбрано сколь угодно большим, то в пределе получаем вероятность $\pi_k(\alpha t)$. Таким образом, в данном случае применимость схемы Пуассона базируется на бесконечной делимости времени. Основное предположение, касающееся независимости количеств регистраций на непересекающихся временных интервалах, будет подтверждено в теореме 2 из следующего параграфа.

*7.2. Пуассоновский процесс

Для более глубокого понимания распределения Пуассона построим модель, в которой оно играет главную роль. Эта модель описывает так называемый *пуассоновский процесс*, являющийся одним из фундаментальных случайных процессов.

Рассмотрим последовательность независимых положительных случайных величин, имеющих одинаковую экспоненциальную плотность

$\alpha e^{-\alpha t}$, $\alpha > 0$; см. пример 3 в § 6.2. Обозначим эти случайные величины через $T_1, T_2, \ldots$. Тогда для каждого j

$$P(T_j \leqslant t) = 1 - e^{-\alpha t}, \quad P(T_j > t) = e^{-\alpha t}, \quad t \geqslant 0. \tag{7.2.1}$$

В силу независимости для любых неотрицательных $t_1, \ldots, t_n$

$$\begin{aligned} P(T_1 > t_1, \ldots, T_n > t_n) = P(T_1 > t_1) \ldots P(T_n > t_n) = \\ = e^{-\alpha(t_1 + \ldots + t_n)}. \end{aligned}$$

Тем самым, задано совместное распределение величин T_j, хотя для простоты мы привели «хвостовые» вероятности. Примеры таких случайных величин были рассмотрены выше. Скажем, они могут быть временны́ми промежутками *между автомобилями* на трассе, или между требованиями выплат, получаемыми страховой компанией (см. пример 5 в § 4.2). Они также могут описывать продолжительности телефонных разговоров или времена пребывания атомов на определенном энергетическом уровне. Поскольку

$$E(T_j) = \frac{1}{\alpha}, \tag{7.2.2}$$

понятно, что, чем меньше α, тем больше средняя длина *временны́х промежутков* или *времен ожидания*, или *времен задержки*. К примеру, если T_j — временные промежутки между автомобилями на некотором контрольном пункте, то соответствующая α на проспекте в Лос-Анжелесе должна быть намного больше, чем на дороге в пустыне Невады. В данном случае α называют также *интенсивностью потока* в том смысле, что более плотное движение обладает большей интенсивностью, о которой каждый водитель может судить по своему нервному напряжению.

Теперь положим $S_0 = 0$ и для $n \geqslant 1$ рассмотрим случайные величины

$$S_n = T_1 + \ldots + T_n. \tag{7.2.3}$$

Тогда согласно определению, S_n — время ожидания n-го поступления (прибытия автомобиля). Событие $\{S_n \leqslant t\}$ означает, что n-е прибытие произошло до момента t. (Мы понимаем предлог «до» в смысле «до или в» (момент t). Эту тонкость можно игнорировать в моделях с непрерывным временем, но она важна для моделей, в которых время дискретно.) Эквивалентно $\{S_n \leqslant t\}$ означает, что всего в интервале $[0, t]$ произошло по меньшей мере n прибытий. Данная двойственность точки зрения весьма полезна, поэтому мы обозначим только что описанное количество прибытий через $N(t)$. Тогда наше утверждение можно записать так:

$$\{N(t) \geqslant n\} = \{S_n \leqslant t\}. \tag{7.2.4}$$

Подобно величинам S_n, $N(t)$ также является случайной величиной: $N(t,\omega)$, где в записи опущен аргумент ω, как у $T_j(\omega)$. Надеемся, что вы еще помните наше обсуждение понятия случайной величины как функции от точек ω выборочного пространства. Сейчас самое время вернуться к этой теме. Каков теперь смысл ω? Точно так же, как в примерах из § 4.2, каждая точка ω может рассматриваться как возможная запись информации о транспортном потоке, поступлениях страховых требований, телефонных разговорах или ядерных превращениях. Говоря более точно, $N(t)$ определяется всей последовательностью $\{T_j, j \geqslant 1\}$ и зависит от ω через величины T_j. Действительно, взяв разность обеих частей равенства (7.2.4) при n и $n+1$, получим, что

$$\{N(t)=n\} = \{S_n \leqslant t\} - \{S_{n+1} \leqslant t\} = \{S_n \leqslant t < S_{n+1}\}. \tag{7.2.5}$$

Смысл данного равенства понятен из его непосредственной интерпретации: в интервале $[0,t]$ наблюдаются в точности n поступлений тогда и только тогда, когда n-е поступление происходит до момента t, а $(n+1)$-е происходит после момента t. Поэтому для каждого значения t вероятностное распределение случайной величины $N(t)$ таково:

$$P\{N(t)=n\} = P\{S_n \leqslant t\} - P\{S_{n+1} \leqslant t\}, \quad n \in \mathbb{N}^0. \tag{7.2.6}$$

Обратите внимание на использование нашего соглашения о том, что $S_0 = 0$. Мы собираемся показать, что это распределение на самом деле есть $\pi(\alpha t)$.

Вычислим вероятность $P\{S_n \leqslant t\}$ с помощью преобразования Лапласа случайной величины S_n (см. § 6.5). Первый шаг — найти преобразование Лапласа $L(\lambda)$ для случайной величины T_j, которое корректно определено в силу условия неотрицательности T_j. Используя формулу (6.5.16) с $f(u) = \alpha e^{-\alpha u}$, получаем, что

$$L(\lambda) = \int\limits_0^\infty e^{-\lambda u} \alpha e^{-\alpha u}\, du = \frac{\alpha}{\alpha+\lambda}\,. \tag{7.2.7}$$

Так как случайные величины T_j независимы, применение теоремы 7 из § 6.5 позволяет найти преобразование Лапласа для S_n:

$$L(\lambda)^n = \left(\frac{\alpha}{\alpha+\lambda}\right)^n. \tag{7.2.8}$$

Задачу получения функции распределения или плотности случайной величины S_n из ее преобразования Лапласа называют *проблемой обращения*. Имеются таблицы, содержащие преобразования Лапласа и соответствующие им функции распределения и плотности. В нашем случае ответ был указан в задаче 38 гл. 6. Тем не менее, рассмотрим прием, позволяющий быстро его установить. Он основан на формуле

$$\int_0^\infty e^{-xt}\,dt = \frac{1}{x}\,, \quad x > 0. \tag{7.2.9}$$

Дифференцируя обе части $n-1$ раз, что легко сделать, получим

$$\int_0^\infty (-t)^{n-1} e^{-xt}\,dt = \frac{(-1)^{n-1}(n-1)!}{x^n}\,,$$

или

$$\frac{1}{(n-1)!}\int_0^\infty t^{n-1} e^{-xt}\,dt = \frac{1}{x^n}\,. \tag{7.2.10}$$

Подставляя $\alpha + \lambda$ вместо x в эту формулу и умножая обе части на α^n, выводим, что

$$\int_0^\infty \frac{\alpha^n}{(n-1)!}\; u^{n-1} e^{-\alpha u} e^{-\lambda u}\,du = \left(\frac{\alpha}{\alpha+\lambda}\right)^n.$$

Тогда, положив

$$f_n(u) = \frac{\alpha^n}{(n-1)!}\, u^{n-1} e^{-\alpha u}, \tag{7.2.11}$$

видим, что f_n есть плотность, соответствующая преобразованию Лапласа (7.2.8), т. е. преобразованию Лапласа случайной величины S_n *). Следовательно, можно переписать правую часть формулы (7.2.6) в виде

$$\int_0^t f_n(u)\,du - \int_0^t f_{n+1}(u)\,du. \tag{7.2.12}$$

Чтобы упростить это выражение, вычислим первый интеграл по частям следующим образом:

$$\frac{\alpha^n}{(n-1)!}\int_0^t e^{-\alpha u} u^{n-1}\,du = \frac{\alpha^n}{(n-1)!}\left\{\frac{u^n}{n}\; e^{-\alpha u}\Big|_0^t + \int_0^t \frac{u^n}{n}\; e^{-\alpha u}\alpha\,du\right\} =$$
$$= \frac{\alpha^n}{n!}\; t^n e^{-\alpha t} + \frac{\alpha^{n+1}}{n!}\int_0^t u^n e^{-\alpha u}\,du.$$

---

*) Другой вывод содержится в задаче 42 гл. 6.

Но последний интеграл совпадает с вторым интегралом из (7.2.12). Поэтому разность в точности равна $(\alpha^n/n!)\, t^n e^{-\alpha t} = \pi_n(\alpha t)$. Для фиксированных n и α она представляет собой плотность гамма-распределения $\Gamma(n;\alpha)$; см. п. (c) задачи 37 из гл. 6. Давайте придадим данному утверждению форму теоремы.

Теорема 1. *Общее число поступлений за временной интервал длины $t > 0$ имеет пуассоновское распределение $\pi(\alpha t)$.*

Читатель должен заметить, что теорема утверждает больше, чем было доказано. Дело в том, что в ее формулировке неявно допускается произвольный выбор начального момента, от которого измеряется время наблюдения, т. е. стартовой точки для отсчета времени T_1 ожидания первого поступления. Однако пока результат установлен только для числа поступлений на интервале $[0, t]$. Давайте обозначим число поступлений в течение произвольного интервала $[s, s+t]$ через $N(s, s+t)$. Тогда очевидно, что

$$N(s, s+t) = N(s+t) - N(s)$$

в наших прежних обозначениях, $N(0) = 0$. Но нам требуется еще установить, что распределение $N(s, s+t)$ совпадает с распределением $N(0, t)$. Вопрос звучит так: если начать наблюдение с момента s, останется ли характер потока поступлений таким же, как при наблюдении с момента 0? Правильный ответ — «да», но его обоснование опирается на важнейшее свойство экспоненциального распределения величин T_j. Нестрого говоря, оно заключается в том, что если прервать время ожидания T_j, то время ожидания после прерывания удовлетворяет первоначальному экспоненциальному закону независимо от того, сколько времени уже прошло до прерывания. Это свойство иногда называют «отсутствием памяти». Запишем его формально: для любых $s \geqslant 0$ и $t \geqslant 0$

$$P(T > t + s \mid T > s) = P(T > t) = e^{-\alpha t}; \tag{7.2.13}$$

см. пример 4 в § 5.1. Верно и обратное: если неотрицательная случайная величина T удовлетворяет первому равенству в формуле (7.2.13), то она обязана иметь экспоненциальное распределение; см. задачу 41 гл. 5. Тем самым, «отсутствием памяти» является характеристическим свойством показательного закона, применяемого для описания времен ожидания.

Теперь мы готовы доказать, что поведение потока поступлений, начиная с момента s, такое же, как при отсчете времени от 0. Заданное значение s разбивает последовательность времен между поступлениями, обозначаемых через T_k, на две части.

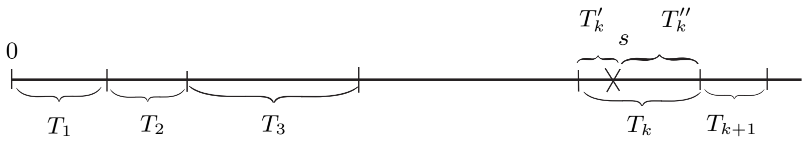


Рис. 27

Согласно рис. 27, вторая часть T_k'' прерванного моментом s времени ожидания T_k обладает тем же самым распределением, что и T_k. При этом она, *очевидно*, не зависит от всех предшествующих времен $T_{k+1}, T_{k+2}, \ldots$. (Это интуитивно понятно, но для формального доказательства необходимо провести некоторые рассуждения, которые мы опускаем.) Отсюда получаем новую последовательность времен ожидания со стартовой точкой s:

$$T_k'', \, T_{k+1}, \, T_{k+2}, \, \ldots \, , \tag{7.2.14}$$

имеющих такие же вероятностные свойства, как первоначальная последовательность, начинающаяся в 0:

$$T_1, \, T_2, \, T_3, \, \ldots \, . \tag{7.2.15}$$

Поэтому наш предыдущий анализ применим к сдвинутому потоку так же, как к первоначальному. В частности, количество поступлений в течение временно́го промежутка $[s, s+t]$ должно быть распределено так же, как количество поступлений в течение интервала $[0, t]$. Это и утверждается в теореме 1.

Тот факт, что величины $N(s, s+t)$ при всех s имеют одинаковое распределение, известен как *временна́я однородность* потока. Напомним, что она установлена в предположении постоянства интенсивности α во времени. На практике данное предположение обычно справедливо только в течение непродолжительных временных периодов. Например, в случае транспортного потока на некотором шоссе оно может выполняться в часы пик или от 2 до 3 ч ночи (дня) с разными значениями α. Для более длительных временных интервалов, скажем, продолжительностью в день, используется среднее значение α за 24 часа. Оно в свою очередь может меняться от года к году или даже от недели к неделе.

До сих пор мы изучали количество поступлений за один временной период, имеющий произвольную длину и начальную точку. Для более полного анализа потока необходимо рассмотреть сразу несколько периодов и изучить зависимость соответствующих количеств поступлений. Другими словами, надо найти совместное распределение случайных величин

$$N(s_1, s_1 + t_1),\ N(s_2, s_2 + t_2),\ N(s_3, s_3 + t_3),\ \ldots. \tag{7.2.16}$$

Ответ содержится в следующей теореме.

Теорема 2. *Если интервалы* $(s_1, s_1{+}t_1), (s_2, s_2{+}t_2), \ldots$ *не пересекаются, то случайные величины в* (7.2.16) *являются независимыми и распределенными по пуассоновским законам* $\pi(\alpha t_1), \pi(\alpha t_2) \ldots$.

Разумно и правильно думать, что если известно совместное распределение количеств поступлений, оказавшихся внутри каждого интервала из произвольного конечного набора непересекающихся временных интервалов, то мы, в принципе, знаем все необходимое. Поэтому теорема 2 обеспечивает нам полный контроль на обсуждаемым процессом.

Доказательство теоремы 2 также опирается на свойство «отсутствия памяти» у случайных величин T_j. Мы объясним здесь главную идею доказательства, не вдаваясь в формальные детали. Вернемся к последовательности (7.2.14), где мы положим $s = s_2$ и заметим, что все случайные величины там не только не зависят друг от друга, но также не зависят от величин, наблюдаемых до момента s, а именно, от случайных величин

$$T_1, \ldots, T_{k-1}, T'_k. \tag{7.2.17}$$

Тот факт, что две части T'_k и T''_k интервала, разбитого точкой s, независимы между собой, вытекает из соотношения (7.2.13). Независимость других случайных величин должна быть интуитивно понятна ввиду того, что они не были затронуты разрывом в точке s. (Здесь снова необходима определенная работа, подтверждающая интуицию.) Далее заметим, что «прошлое поведение» процесса до момента s определяется последовательностью (7.2.17), в то время как «будущее развитие» после момента s определяется последовательностью (7.2.14). Поэтому относительно «настоящего» момента s «прошлое» и «будущее» независимы. В частности, величина $N(s_1, s_1 + t_1)$, относящаяся к «прошлому», не будет зависеть от случайных величин $N(s_2, s_2 + t_2), N(s_3, s_3 + t_3), \ldots$, которые принадлежат «будущему». Повторяя данное рассуждение для $s = s_3, s_4, \ldots$, устанавливаем справедливость утверждения теоремы 2.

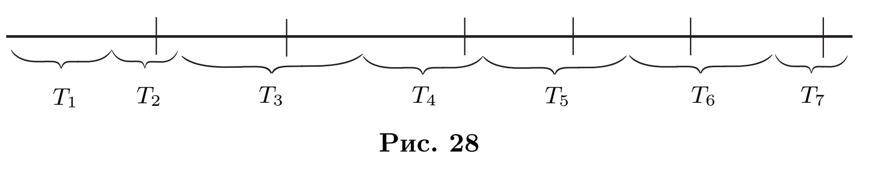


Рис. 28

Теперь все готово для того, чтобы дать общее определение «потока», с которым мы имеем дело.

Определение пуассоновского процесса. Семейство случайных величин $\{X(t)\}$, индексированное действительной переменной t, меняющейся на промежутке $[0, \infty)$, называется *пуассоновским процессом с параметром* (или *средним*) α тогда и только тогда, когда оно удовлетворяет следующим условиям:

(i) $X(0) = 0$;

(ii) приращения $X(s_i + t_i) - X(s_i)$ на произвольных непересекающихся интервалах $(s_i, s_i + t_i)$ из конечного набора — независимые случайные величины;

(iii) для любых $s \geqslant 0$ и $t \geqslant 0$ случайная величина $X(s+t) - X(s)$ имеет распределение Пуассона $\pi(\alpha t)$.

В силу теорем 1 и 2 семейство $\{N(t), t \geqslant 0\}$ удовлетворяет данным условиям и поэтому образует пуассоновский процесс. Обратно, можно показать, что любой пуассоновский процесс имеет тот же вид, что и $N(t)$.

Понятие стохастического процесса уже упоминалось в §§ 5.3–5.4 в связи с урновой моделью Пойа. Последовательность $\{X_n, n \geqslant 1\}$ из теоремы 5 в § 5.4 можно назвать процессом Пойа. В принципе, стохастический процесс — это всего лишь другое название для произвольного семейства случайных величин; но последний термин представляется слишком неясным и всеобъемлющим. То, о чем сейчас идет речь, возвращает нас к основаниям теории вероятности, обсуждаемым в гл. 2, 4 и 5. Там рассматривались такие понятия, как выборочное пространство Ω, состоящее из точек ω, вероятностная мера P, определенная на некоторых подмножествах из Ω, семейство функций $\omega \to X_t(\omega)$, называемых случайными величинами, и связанные с их совместным поведением характеристики: маргинальные и совместные распределения, условные вероятности, математические ожидания и т. п. Все, что мы уже обсуждали (и собираемся обсуждать в будущем) можно считать вопросами теории стохастических процессов, так как она, понимаемая во всей своей общности, включает в себя произвольные случайные величины и подмножества выборочного пространства (через их индикаторы). Однако обычно речь идет о довольно многочисленных и хорошо организованных семействах, управляемых важными и полезными законами. Приведенная выше характеризация пуассоновского процесса дает яркий пример подобного описания.

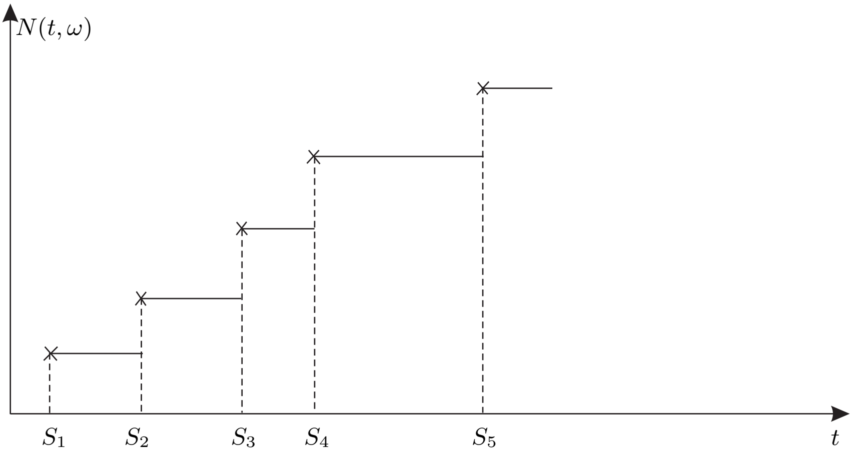


Рис. 29

Согласно определению, отображение $\omega \to N(t,\omega)$ является для каждого значения t случайной величиной, имеющей распределение Пуассона $\pi(\alpha t)$. Существует также двойственная точка зрения, столь же важная при изучении процесса, при которой интерес представляет поведение функций $t \to N(t,\omega)$ при каждом ω. Эти функции называются *выборочными функциями* (*путями* или *траекториями*). К примеру, в случае телефонных разговоров фиксация точки ω означает выбор одной из существующих записей вызовов, поступивших на телефонный узел в течение некоторого 24-часового периода. Записи, конечно, в разные дни выглядят по-разному, поэтому функция $t \to N(t,\omega)$ представляет собой лишь *выбранный вариант* (обозначаемый через ω) последовательности вызовов. Ее график может выглядеть, как график, изображенный на рис. 29. Точки скачков — моменты поступления вызовов $S_n(\omega)$, все скачки имеют величину 1, а длины горизонтальных отрезков отражают продолжительность времен между вызовами. Указанный вид типичен для выборочных функций пуассоновского процесса. Чем выше интенсивность потока, тем гуще располагаются точки скачков.

Последовательность $\{S_n, n \geqslant 1\}$, определенная формулой (7.2.3), также является стохастическим процессом, индексированным переменой n. Выборочная функция $n \to S_n(\omega)$ такого процесса представляет собой возрастающую последовательность положительных чисел $\{S_1(\omega), S_2(\omega), \ldots, S_n(\omega), \ldots\}$. Поэтому ее часто называют *выборочной последовательностью.* Существует взаимосвязь между ней и рассмотренной выше выборочной функцией $N(t,\omega)$. Если поменять местами координатные оси, либо повернуть систему координат на 90° и взглянуть на рис. 29 с другой стороны листа на просвет, то мы получим график $n \to S_n(\omega)$. При этом можно не обращать внимания на вертикальные линии за исключением их нижних концевых точек, которые показывают значения S_n.

Следующие примеры иллюстрируют некоторые свойства распределения Пуассона и пуассоновского процесса.

Пример 5. Рассмотрим количества поступлений в двух непересекающихся интервалах — случайные величины $X_1 = N(s_1, s_1 + t_1)$ и $X_2 = N(s_2, s_2 + t_2)$, аналогичные величинам (7.2.16). Чему равна вероятность того, что общее количество поступлений $X_1 + X_2$ равно n?

В силу теоремы 2 случайные величины X_1 и X_2 независимы и распределены по законам $\pi(\alpha t_1)$ и $\pi(\alpha t_2)$ соответственно. Тогда

$$\begin{aligned} P(X_1 + X_2 = n) &= \sum_{j+k=n} P(X_1 = j) P(X_2 = k) = \\ &= \sum_{j+k=n} \frac{e^{-\alpha t_1}(\alpha t_1)^j}{j!} \, \frac{e^{-\alpha t_2}(\alpha t_2)^k}{k!} = \\ &= \frac{e^{-\alpha(t_1+t_2)}}{n!} \sum_{j=0}^{n} \binom{n}{j} (\alpha t_1)^j (\alpha t_2)^{n-j} = \\ &= \frac{e^{-\alpha(t_1+t_2)}}{n!} \, (\alpha t_1 + \alpha t_2)^n = \pi_n(\alpha t_1 + \alpha t_2). \end{aligned}$$

Другими словами, случайная величина $X_1 + X_2$ также распределена согласно закону Пуассона с параметром $\alpha t_1 + \alpha t_2$. Более общее утверждение выглядит так.

Теорема 3. *Пусть X_j — независимые случайные величины, имеющие пуассоновские распределения $\pi(\alpha_j), 1 \leqslant j \leqslant n$. Тогда случайная величина $X_1 + \ldots + X_n$ распределена по закону $\pi(\alpha_1 + \ldots + \alpha_n)$.*

Доказательство проводится с помощью обычной индукции, но его также можно легко получить методом производящих функций. Если обозначить производящую функцию случайной величины X_i через g_{X_i}, то в силу теоремы 6 из § 6.5 запишем:

$$\begin{aligned} g_{X_1+\ldots+X_n}(z) = g_{X_1}(z)\, g_{X_2} \cdot \ldots \cdot g_{X_n}(z) = e^{\alpha_1(z-1)} e^{\alpha_2(z-1)} \cdot \ldots \cdot e^{\alpha_n(z-1)} = \\ = e^{(\alpha_1+\ldots+\alpha_n)(z-1)}. \end{aligned}$$

Отсюда заключаем, что сумма $X_1 + \ldots + X_n$ имеет производящую функцию, соответствующую распределению $\pi(\alpha_1 + \ldots + \alpha_n)$. Ввиду взаимной однозначности, констатируемой в теореме 7 из § 6.5, она распределена именно по этому закону.

Пример 6. На перекрестке в Америке проводятся наблюдения за тем, в каком из штатов зарегистрированы номера проезжающих мимо автомобилей. Допустим, что поток машин является пуассоновским с интенсивностью α, а вероятности того, что автомобиль зарегистрирован в Калифорнии, Неваде и Аризоне, равны, соответственно, $p_1 = 1/25$, $p_2 = 1/100$, $p_3 = 1/80$. Каково распределение количеств автомобилей из этих трех штатов, проезжающих за единичный временной интервал?

Мы предполагаем, что если подсчитаны n машин, то количества автомобилей из разных штатов удовлетворяют полиномиальному распределению $M(n;\ 50;\ p_1,\ \ldots,\ p_{50})$*), где первые три вероятности p_j приведены выше. Из пуассоновости потока следует, что общее количество машин, проезжающих за единичный период времени, является случайной величиной N с распределением

$$P(N = n) = \frac{e^{-\alpha}}{n!}\ \alpha^n, \quad n = 0, 1, 2, \ldots .$$

Количество машин, зарегистрированных в k-м штате, также является случайной величиной N_k, причем, конечно,

$$N_1 + N_2 + \ldots + N_{50} = N.$$

Задача заключается в вычислении вероятностей

$$P(N_1 = n_1, N_2 = n_2, N_3 = n_3)$$

для произвольных n_1, n_2, n_3. Обозначим через $q = 1 - p_1 - p_2 - p_3$ вероятность того, что машина не относится ни к одному из интересующих нас трех штатов. Для $n \geqslant n_1 + n_2 + n_3$ условная вероятность при гипотезе $N = n$ является полиномиальной. Отсюда

$$P(N_1 = n_1, N_2 = n_2, N_3 = n_3 \mid N = n) = \frac{n!\, p_1^{n_1} p_2^{n_2} p_3^{n_3} q^k}{n_1!\, n_2!\, n_3!\, k!}\ ,$$

где $k = n - n_1 - n_2 - n_3$. Применяя формулу «полной вероятности» (5.2.3), получаем

$$\begin{aligned} P(N_1 = n_1, N_2 = n_2, N_3 = n_3) &= \\ &= \sum_n P(N = n)\, P(N_1 = n_1, N_2 = n_2, N_3 = n_3 \mid N = n) = \\ &= \sum_{k=0}^{\infty} \frac{e^{-\alpha}}{n!}\ \alpha^n\ \frac{n!}{n_1!\, n_2!\, n_3!}\ \frac{p_1^{n_1} p_2^{n_2} p_3^{n_3} q^k}{k!}\ . \end{aligned}$$

*) Здесь 50 — количество штатов США. — *Прим. перев.*

Так как сумма $n_1+n_2+n_3$ фиксирована и $n \geqslant n_1+n_2+n_3$, суммирование выше сводится к суммированию по k, пробегающему все целые неотрицательные значения. Теперь подставим в последнюю формулу выражения

$$e^{-\alpha} = e^{-\alpha(p_1+p_2+p_3)}e^{-\alpha q}, \quad \alpha^n = \alpha^{n_1+n_2+n_3}\alpha^k$$

и вынесем за знак суммы множители, не меняющиеся при суммировании по k. Результат таков:

$$\frac{e^{-\alpha(p_1+p_2+p_3)}}{n_1!\,n_2!\,n_3!}(\alpha p_1)^{n_1}(\alpha p_2)^{n_2}(\alpha p_3)^{n_3}\sum_{k=0}^{\infty}\frac{e^{-\alpha q}}{k!}(\alpha q)^k = \\ = \pi_{n_1}(\alpha p_1)\,\pi_{n_2}(\alpha p_2)\,\pi_{n_3}(\alpha p_3),$$

поскольку последняя сумма равна 1. Таким образом, мы видим, что случайные величины N_1, N_2, N_3 независимы (почему?) и имеют пуассоновские распределения $\pi(\alpha p_1)$, $\pi(\alpha p_2)$, $\pi(\alpha p_3)$.

Замена фиксированного параметра n в полиномиальном распределении $M(n;r;p_1,\ldots,p_r)$ на случайную величину N, имеющую распределение Пуассона, в статистической методологии называется «рандомизированным отбором». Для нашего примера различие заключается в том, что подсчитывается либо фиксированное число машин, либо количество машин в выбранном временном интервале, либо отбор осуществляется каким-нибудь другим методом, допускающим случайные изменения числа машин. Какой из методов предпочтительнее, вообще говоря, зависит от обстоятельств и вида информации, интересующей исследователя.

Конечно, существует обобщение рассмотренного выше примера, которое формулируется следующим образом.

Теорема 4. *При рандомизированном отборе из полиномиальной популяции $M(n;r;p_1,\ldots,p_r)$, где размер выборки — случайная величина N, распределенная по закону Пуассона со средним α, количества $N_1,\ldots,N_r$ отобранных представителей каждой из r разновидностей являются независимыми пуассоновскими случайными величинами со средними $\alpha p_1,\ldots,\alpha p_r$.*

В качестве иллюстрации изящества устройства пуассоновского процесса выведем результат, относящийся к расположению скачков его выборочных функций. Начнем с замечания, что хотя (почти) все выборочные функции имеют бесконечно много скачков на интервале $(0,\infty)$, вероятность того, что скачок произойдет в любой заранее указанный момент времени, равна нулю. Действительно, если момент $t>0$ фиксирован, то при $\delta \to +0$

$$P\{N(t+\delta) - N(t-\delta) \geqslant 1\} = 1 - \pi_0(\alpha, 2\delta) = 1 - e^{-2\alpha\,\delta} \to 0.$$

В частности, количество скачков в интервале (t_1, t_2) имеет одно и то же распределение независимо от того, включаются в него концевые точки t_1 и t_2 или не включаются. Как и прежде, будем обозначать это количество через $N(t_1, t_2)$. Теперь представим, что $N(0, t) = n$ для заданного t, где $n \geqslant 1$, и рассмотрим произвольный подынтервал (t_1, t_2) в промежутке $(0, t)$. Для $0 \leqslant j \leqslant n$ имеем:

$$P\{N(t_1, t_2) = j,\, N(0, t) = n\} = \\ = P\{N(t_1, t_2) = j,\, N(0, t_1) + N(t_2, t) = n - j\}.$$

Положим $s = t_2 - t_1$. Тогда сумма длин интервалов $(0, t_1)$ и (t_2, t) равна $t - s$. С учетом свойства (ii) и теоремы 3 видим, что случайная величина $N(0, t_1) + N(t_2, t)$ распределена по закону $\pi(a(t - s))$ и не зависит от $N(t_1, t_2)$. Поэтому приведенная выше вероятность равна

$$e^{-\alpha s}\, \frac{(\alpha s)^j}{j!}\, e^{-\alpha(t-s)}\, \frac{(\alpha t - \alpha s)^{n-j}}{(n-j)!}\,.$$

Разделив ее на $P\{N(0, t) = n\} = e^{-\alpha t}(\alpha t)^n/n!$, получим условную вероятность

$$P\{N(t_1, t_2) = j \mid N(0, t) = n\} = \binom{n}{j}\left(\frac{s}{t}\right)^j\left(1 - \frac{s}{t}\right)^{n-j}. \qquad (7.2.18)$$

Это — биномиальная вероятность $B_j(n; s/t)$.

Теперь рассмотрим произвольное разбиение интервала $(0, t)$ на конечное число подынтервалов $I_1, \ldots, I_l$ с длинами $s_1, \ldots, s_l$, причем $s_1 + \ldots + s_l = t$. Пусть $n_1, \ldots, n_l$ — произвольные неотрицательные целые числа такие, что $n_1 + \ldots + n_l = n$. Если обозначить через $N(I_k)$ количество скачков процесса в интервале I_k, то можно провести вычисления, похожие на выполненные выше:

$$P\{N(I_k) = n_k, 1 \leqslant k \leqslant l \mid N(0, t) = n\} = \\ = \prod_{k=1}^{l} \frac{e^{-\alpha s_k}(\alpha s_k)^{n_k}}{n_k!}\left(e^{-\alpha t}\,\frac{(\alpha t)^n}{n!}\right)^{-1} = \frac{n!}{n_1! \ldots n_l!}\prod_{k=1}^{l}\left(\frac{s_k}{t}\right)^{n_k}. \quad (7.2.19)$$

Это — полиномиальное распределение, обсуждаемое в § 6.4.

Давайте выберем n точек наудачу в интервале $(0, t)$ и расположим их в порядке возрастания: $0 < \xi_1 \leqslant \xi_2 \leqslant \ldots \leqslant \xi_n < t$. Используя обозначения, введенные выше, под $\tilde{N}(I_k)$ будем понимать количество точек, оказавшихся внутри интервала I_k. Несложно показать, что n-мерное распределение вектора $(\xi_1, \ldots, \xi_n)$ однозначно определяется распределениями $(\tilde{N}(I_1), \ldots, \tilde{N}(I_l))$ для всевозможных разбиений интервала $(0, t)$;

строгое доказательство данного факта предлагается установить в задаче 26. В частности, если n точек выбираются независимо друг от друга, и каждая из них равномерно распределена в интервале $(0, t)$, то из обсуждения в § 6.4 следует, что вероятность $P\{\tilde{N}(I_k) = n_k, 1 \leqslant k \leqslant l\}$ задается правой частью формулы (7.2.19). Таким образом, в предположении, что пуассоновский процесс имеет в точности n скачков в интервале $(0, t)$, *условное распределение* точек скачков таково, как если бы их выбирали наудачу из промежутка $(0, t)$. Это свойство оправдывает термин «однородный хаос».

7.3. От биномиального закона к нормальному

С точки зрения аппроксимации биномиального распределения $B(n; p)$ при больших значениях n, случай, обсуждаемый в § 7.1 и приводящий к пуассоновскому распределению, является в определенном смысле *нетипичным* потому, что p должно быть очень малым, чтобы np оставалось константой или почти константой. Тот факт, что многие случайные явления хорошо согласуются с этим законом, не был известен в ранней истории теории вероятностей. Напомним, что не только радиоактивность в те времена еще не была открыта, но в реальной жизни отсутствовали даже телефоны и автомобильные потоки. С другой стороны, люди подсчитывали количество «орлов» при бросании монет или очков, выпавших на игральных костях, а также уже интенсивно проводились измерения разнообразных физических и биологических величин. Это привело к необходимости изучения биномиального и полиномиального распределений. Но ввиду того, что компьютеры еще не появились, возникла настоятельная потребность найти удобные вычислительные формулы для приближения этих вероятностей. *Типичным* случаем для аппроксимации биномиального распределения

$$B_k(n; p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leqslant k \leqslant n, \tag{7.3.1}$$

является такой, когда значение p фиксировано, а значение n велико. В качестве простейшей иллюстрации допустим, что симметричная монета бросается 100 раз. Чему тогда равна вероятность того, что при этом выпадет в точности 50 «орлов»? Ответ таков:

$$\binom{100}{50} \frac{1}{2^{100}} = \frac{100!}{50!\,50!}\,\frac{1}{2^{100}}\,.$$

Но вряд ли можно считать его удовлетворительным, так как непонятно, какова величина данной вероятности. Как, не используя более сложной математики (которую мы сейчас и собираемся изложить), можно угадать, имеет ли она значение, близкое к 1/2, 1/10 или 1/50?

Очевидно, что *ключевой элемент* в подобных комбинаторных формулах — это факториал $n!$, который встречается в них повсюду. Просмотрите еще раз материал гл. 3. Таким образом, проблема заключается в том, чтобы найти удобную формулу, т. е. некоторую функцию $\chi(n)$ от n, которая является и хорошей аппроксимацией для $n!$, и достаточно простой для вычислений. Но что значит «хорошей»? Поскольку $n!$ возрастает очень быстро с увеличением n (см. небольшую таблицу в § 3.2), нет смысла стремиться сделать малой разность $|n! - \chi(n)|$. (Есть ли различие в том, чтобы иметь миллион или миллион и еще три доллара?) То, что действительно важно, так это отношение $n!/\chi(n)$, которое должно быть близко к 1. Для двух положительных функций ψ и χ от целого аргумента n существует следующее стандартное обозначение:

$$\psi(n) \sim \chi(n), \quad \text{которое означает, что} \quad \lim_{n\to\infty} \frac{\psi(n)}{\chi(n)} = 1. \tag{7.3.2}$$

При этом говорят также, что функции $\psi(n)$ и $\chi(n)$ *асимптотически эквивалентны* при $n \to \infty$. Если это так, то

$$\lim_{n\to\infty} \frac{|\psi(n) - \chi(n)|}{\chi(n)} = 0$$

при условии, что $\chi(n) > 0$ для больших n. Другими словами, разность $|\psi(n) - \chi(n)|$ пренебрежимо мала по сравнению с $\chi(n)$ или $\psi(n)$, хотя на самом деле она может быть большой по величине. Вот тривиальный пример функций, обладающих указанным выше свойством:

$$\psi(n) = 2n^2 + 10n - 100, \quad \chi(n) = 2n^2.$$

Более общее утверждение: полином от n асимптотически эквивалентен члену, имеющему наибольшую степень. Но в нашем случае мы имеем дело с более сложной задачей: найти достаточно простую функцию $\chi(n)$ такую, что

$$\lim_{n\to\infty} \frac{n!}{\chi(n)} = 1.$$

Ответ дает *формула Стирлинга* (см. приложение 2):

$$\chi(n) = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} = n^{n+(1/2)} e^{-n} \sqrt{2\pi}, \tag{7.3.3}$$

или более точно:

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\, e^{\omega(n)}, \text{ где } \frac{1}{12(n + (1/2))} < \omega(n) < \frac{1}{12n}. \tag{7.3.4}$$

Возможно, вы думаете, что функция $\chi(n)$ выглядит хуже, чем $n!$, но она намного проще для вычислений, так как возведение в степень — простая операция. Мы сразу применим ее к встретившейся ранее небольшой задаче. Нет необходимости с самого начала ограничиваться числовыми примерами, поэтому мы рассмотрим выражение

$$\binom{2n}{n}\frac{1}{2^{2n}} = \frac{(2n)!}{n!\,n!}\,\frac{1}{2^{2n}}\,. \tag{7.3.5}$$

Подставляя $\chi(n)$ и $\chi(2n)$ вместо $n!$ и $(2n)!$ соответственно, видим, что это выражение асимптотически эквивалентно

$$\frac{(2n/e)^{2n}\sqrt{4\pi n}}{(n/e)^{2n}\,2\pi n}\,\frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}}\,.$$

В частности, для $n = 50$ получаем искомый ответ $1/\sqrt{50\pi} \approx 0.08$. Попробуйте сделать это, вычисляя логарифм от

$$\frac{(100)_{50}}{50!}\,\frac{1}{2^{100}}\,,$$

и вы еще больше оцените формулу Стирлинга. Исследуем сразу несколько более общее выражение

$$\binom{2n}{n+k}\frac{1}{2^{2n}} = \frac{(2n)!}{(n+k)!\,(n-k)!}\,\frac{1}{2^{2n}}\,, \tag{7.3.6}$$

где k фиксировано. Аналогичное применение формулы (7.3.3) дает

$$\begin{aligned}
&\frac{\left(\dfrac{2n}{e}\right)^{2n}\sqrt{4\pi n}\cdot\dfrac{1}{2^{2n}}}{\left(\dfrac{n+k}{e}\right)^{n+k}\sqrt{2\pi(n+k)}\left(\dfrac{n-k}{e}\right)^{n-k}\sqrt{2\pi(n-k)}} = \\
&= \frac{n^{2n}}{(n+k)^{n+k}(n-k)^{n-k}}\sqrt{\frac{n}{\pi(n+k)(n-k)}} = \\
&= \left(\frac{n}{n+k}\right)^{n+k}\left(\frac{n}{n-k}\right)^{n-k}\sqrt{\frac{n}{\pi(n^2-k^2)}}.
\end{aligned}$$

Очевидно, что последний из сомножителей асимптотически эквивалентен $1/\sqrt{\pi n}$. Что касается других сомножителей, то из соотношений (7.1.8) вытекает, что

$$\begin{aligned}
\lim_{n\to\infty}\left(\frac{n}{n+k}\right)^{n+k} &= \lim_{n\to\infty}\left(1-\frac{k}{n+k}\right)^{n+k} = e^{-k},\\
\lim_{n\to\infty}\left(\frac{n}{n-k}\right)^{n-k} &= \lim_{n\to\infty}\left(1+\frac{k}{n-k}\right)^{n-k} = e^{k}.
\end{aligned} \tag{7.3.7}$$

Поэтому асимптотическое поведение выражения (7.3.6) совпадает с

$$e^{-k}e^{k}\,\frac{1}{\sqrt{\pi n}} = \frac{1}{\sqrt{\pi n}}\,,$$

т. е. является таким же, как и поведение выражения (7.3.5), которое является частным случаем последнего при $k = 0$.

В качестве следствия для фиксированного l выводим, что

$$\lim_{n\to\infty}\sum_{k=-l}^{l}\binom{2n}{n+k}\frac{1}{2^{2n}} = 0, \tag{7.3.8}$$

поскольку, как мы только что показали, каждое слагаемое стремится к нулю при $n \to \infty$, а число слагаемых конечно. Теперь, если вы помните треугольник Паскаля (3.3.5), то знаете, что биномиальные коэффициенты $\binom{2n}{n+k}$, $-n \leqslant k \leqslant n$, достигают максимума $\binom{2n}{n}$ на среднем из слагаемых при $k = 0$ и убывают, когда $|k|$ увеличивается (см. задачу 6). В соответствии с формулой (7.3.8), сумма фиксированного числа слагаемых, окружающих среднее слагаемое, сходится к нулю. Поэтому сумма любого фиксированного числа слагаемых также обязана стремиться к нулю, а именно, для произвольных фиксированных a и b, где $a < b$, имеем:

$$\lim_{n\to\infty}\sum_{j=a}^{b}\binom{2n}{j}\frac{1}{2^{2n}} = 0.$$

Наконец отметим, что данный результат остается верным, если заменить здесь $2n$ на $2n+1$, поскольку отношение соответствующих членов

$$\binom{2n+1}{j}\frac{1}{2^{2n+1}}\Bigg/\binom{2n}{j}\frac{1}{2^{2n}} = \frac{2n+1}{2n+1-j}\cdot\frac{1}{2}$$

сходится к $1/2$, что не оказывает влияния на нулевой предел. Давайте теперь обсудим вероятностный смысл полученных результатов. Обозначим, как обычно, через S_n количество «орлов», выпавших в n бросаниях монеты. Наш результат означает, что для любых фиксированных a и b

$$\lim_{n\to\infty} P(a \leqslant S_n \leqslant b) = 0. \tag{7.3.9}$$

Обратите внимание, что всего имеется $n+1$ возможных значений для S_n, а диапазон $[a, b]$ фиксирован (не зависит от n), т. е. он составляет пренебрежимо малую долю от величины n, когда n велико. Поэтому результат (7.3.9) вряд ли можно считать удивительным, хотя, конечно, он слегка разочаровывает.

Понятно, что если мы желаем «охватить» достаточное количество возможных значений S_n для того, чтобы получить ненулевой предел, то нам придется позволить диапазону суммирования неограниченно расти вместе с n. Поскольку известно, что слагаемые вблизи середины имеют порядок малости $1/\sqrt{n}$, то правдоподобным является предположение, что потребуется сложить порядка $\sqrt{n}$ членов. Говоря точнее, мы ожидаем, что для каждого фиксированного l вероятность

$$P\left(\frac{n}{2} - l\sqrt{n} \leqslant S_n \leqslant \frac{n}{2} + l\sqrt{n}\right) = \sum_{|j-\frac{n}{2}|\leqslant l\sqrt{n}} \binom{n}{j} \frac{1}{2^n} \tag{7.3.10}$$

стремится при $n \to \infty$ к некоторому пределу, лежащему строго между 0 и 1. Здесь диапазон для S_n имеет центр в точке $n/2$ и включает в себя около $2l\sqrt{n}$ членов. Когда n велико, этот диапазон представляет собой по-прежнему очень малую долю от n, но увеличивается достаточно быстро, чтобы обеспечить выполнение нашей цели. Выбор именно $\sqrt{n}$, а не, скажем, $n^{1/3}$ или $n^{3/5}$, весьма важен и определяется достаточно тонкой математической техникой, к изложению которой мы собираемся перейти.

До сих пор в условиях биномиального распределения с вероятностями (7.3.1) изучался случай $p = 1/2$ для того, чтобы продемонстрировать главные выводы на простейшем примере. Однако данное упрощение скрывает роль величин np и npq, встречающихся в общих формулах, приводимых ниже. Мы советуем читателю самостоятельно провести последующие выкладки для случая $p = q = 1/2$, чтобы попрактиковаться и приобрести уверенность в выполнении подобных вычислений.

Теорема 5. *Пусть $0 < p < 1$. Положим $q = 1 - p$ и*

$$x_{nk} = \frac{k - np}{\sqrt{npq}}, \quad 0 \leqslant k \leqslant n. \tag{7.3.11}$$

Понятно, что величины x_{nk} зависят от n и k, но ниже для них используются обозначения x_k.

Пусть A — произвольная положительная константа. Тогда в диапазоне значений k таком, что

$$|x_k| \leqslant A, \tag{7.3.12}$$

имеем

$$\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}}\, e^{-x_k^2/2}. \tag{7.3.13}$$

Сходимость является равномерной для k из указанного выше диапазона.

Доказательство. С учетом определения (7.3.11) запишем:

$$k = np + \sqrt{npq}\, x_k, \quad n - k = nq - \sqrt{npq}\, x_k. \tag{7.3.14}$$

Поэтому в диапазоне из (7.3.12)

$$k \sim np, \quad n-k \sim nq. \tag{7.3.15}$$

Применение формулы Стирлинга (7.3.3) с учетом эквивалентностей (7.3.15) позволяет представить левую часть формулы (7.3.13) в следующем виде:

$$\begin{aligned}\frac{\left(\dfrac{n}{e}\right)^n \sqrt{2\pi n}\, p^k q^{n-k}}{\left(\dfrac{k}{e}\right)^k \sqrt{2\pi k}\left(\dfrac{n-k}{e}\right)^{n-k} \sqrt{2\pi(n-k)}} &= \sqrt{\frac{n}{2\pi k(n-k)}}\,\varphi(n,k) \sim \\ &\sim \frac{1}{\sqrt{2\pi npq}}\,\varphi(n,k).\end{aligned} \tag{7.3.16}$$

Здесь было введено обозначение

$$\varphi(n,k) = \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

Логарифмируя и используя разложение в ряд Тейлора

$$\ln(1+x) = x - \frac{x^2}{2} + \ldots + (-1)^{n-1}\frac{x^n}{n} + \ldots, \quad |x| < 1,$$

согласно равенствам (7.3.14), получаем:

$$\begin{aligned}\ln\left(\frac{np}{k}\right)^k &= k\ln\left(1 - \frac{\sqrt{npq}\,x_k}{k}\right) = \\ &= k\left(-\frac{\sqrt{npq}\,x_k}{k} - \frac{npqx_k^2}{2k^2} - \ldots\right), \\ \ln\left(\frac{nq}{n-k}\right)^{n-k} &= (n-k)\ln\left(1 + \frac{\sqrt{npq}\,x_k}{n-k}\right) = \\ &= (n-k)\left(\frac{\sqrt{npq}\,x_k}{n-k} - \frac{npqx_k^2}{2(n-k)^2} + \ldots\right)\end{aligned} \tag{7.3.17}$$

при условии, что

$$\left|\frac{\sqrt{npq}\,x_k}{k}\right| < 1 \quad \text{и} \quad \left|\frac{\sqrt{npq}\,x_k}{n-k}\right| < 1. \tag{7.3.17$'$}$$

В силу условий (7.3.12) и (7.3.15) эти неравенства выполняются для достаточно больших n. Складывая два приведенных выше разложения, в которых первые члены благополучно сокращаются, игнорируя при этом многоточия и используя знак «$\sim$» вместо «$=$», выводим формулу:

$$\ln\varphi(n,k) \sim -\frac{npqx_k^2}{2k} - \frac{npqx_k^2}{2(n-k)} = -\frac{n^2pqx_k^2}{2k(n-k)}.$$

В приложении 2 приводится строгое доказательство данного соотношения. Снова применяя эквивалентности (7.3.15), видим, что

$$\ln \varphi(n,k) \sim -\frac{n^2pqx_k^2}{2npnq} = -\frac{x_k^2}{2}\,. \tag{7.3.18}$$

С учетом условия (7.3.12) (почему необходимо сделать такую ссылку?) это соотношение эквивалентно $\varphi(n,k) \sim e^{-x_k^2/2}$. Возвращаясь к формуле (7.3.16), устанавливаем соотношение (7.3.13).

Теорема 6 (*Муавра—Лапласа*). *Для произвольных констант a и b, $-\infty < a < b < +\infty$,*

$$\lim_{n\to\infty} P\left(a < \frac{S_n - np}{\sqrt{npq}} \leqslant b\right) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2}\,dx. \tag{7.3.19}$$

Доказательство. Пусть k — одно из возможных значений случайной величины S_n, т. е. равенство $S_n = k$ означает, что $(S_n - np)/\sqrt{npq} = x_k$ согласно преобразованию (7.3.11). Тогда вероятность в левой части формулы (7.3.19) есть всего лишь

$$\sum_{a<x_k\leqslant b} P(S_n = k) = \sum_{a<x_k\leqslant b} \binom{n}{k} p^k q^{n-k}.$$

Заменяя каждое слагаемое на его асимптотический эквивалент из соотношения (7.3.13) и учитывая, что, ввиду определения (7.3.11),

$$x_{k+1} - x_k = \frac{1}{\sqrt{npq}}\,,$$

получаем

$$\frac{1}{\sqrt{2\pi}} \sum_{a<x_k\leqslant b} e^{-x_k^2/2}(x_{k+1} - x_k). \tag{7.3.20}$$

Соответствие между k и x_k является взаимно однозначным. Когда k меняется от 0 до n, x_k пробегает интервал $[-\sqrt{np/q}, \sqrt{nq/p}]$ с шагом $x_{k+1} - x_k = 1/\sqrt{npq}$. Для достаточно больших n этот интервал включает в себя промежуток $(a, b]$, и точки x_k, попавшие внутрь $(a, b]$, формируют его разбиение на подынтервалы, имеющие одинаковую длину $1/\sqrt{npq}$. Допустим, что наименьшее и наибольшее значения k, удовлетворяющие условиям $a < x_k \leqslant b$, суть j и l. Тогда

$$x_{j-1} \leqslant a < x_j < x_{j+1} < \ldots < x_{l-1} < x_l \leqslant b < x_{l+1},$$

и сумму в формуле (7.3.20) можно записать в следующем виде:

$$\sum_{k=j}^{l} \varphi(x_k)(x_{k+1} - x_k); \quad \text{где } \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}. \tag{7.3.21}$$

Это — сумма Римана для определенного интеграла $\int\limits_a^b \varphi(x)\,dx$, несмотря на то что при изложении в стандартных учебниках теории интегрирования по Риману концевые точки a и b обычно также включают в разбиение. На самом деле данное различие не играет никакой роли из-за измельчения разбиения при $n \to \infty$. Поэтому приведенная выше сумма сходится к интегралу, стоящему в правой части равенства (7.3.19).

Доказанное утверждение называют *теоремой Муавра—Лапласа.* Абрахам Де Муавр (1667–1754), известный как последователь Ньютона, привел данный результат в своей *Доктрине шансов* (1714). Кроме того, он раньше Стирлинга (1692–1770) вывел формулу, носящую имя последнего. Лаплас обобщил результат и указал на его важность в своем монументальном труде *Аналитическая теория вероятностей* (1812). Теорема Муавра—Лапласа является первым установленным частным случаем *центральной предельной теоремы*, которая будет обсуждаться в следующем параграфе. Она решает проблему аппроксимации, сформулированную в начале текущего параграфа. Простые примеры ее применения будут даны в конце § 7.5, а также содержатся среди задач к настоящей главе.

7.4. Нормальное распределение

Введем формально вероятностное распределение с плотностью φ из формулы (7.3.21):

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\,du, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

Оно называется *нормальным распределением*, или *распределением Гаусса—Лапласа.* Иногда к названию добавляют уточнение «стандартное» для того, чтобы выделить данное распределение из всего семейства нормальных распределений, получающихся из него с помощью линейных преобразований, см. ниже. Но нам еще следует убедиться, что φ действительно является плотностью в соответствии с определением из § 4.5, а именно, что

$$\int\limits_{-\infty}^{\infty} \varphi(x)\,dx = 1. \tag{7.4.1}$$

Эвристическое доказательство данного факта может быть получено, если положить $a = -\infty$, $b = +\infty$ в равенстве (7.3.19) и заметить, что при этом вероятность слева становится равной 1. Почему такое рассуждение нельзя признать строгим? Потому что в нем участвуют два (или три) предельных перехода, перемена порядка которых может оказаться незаконной. На самом деле, данное доказательство можно сделать строгим (см. приложение 2), но более важно, чтобы вы осознали, что такая «доводка» действительно необходима. Здесь проходит рубеж, где высшая математика отделяется от элементарной, с которой мы в основном имеем дело в данной книге.

Прямое доказательство равенства (7.4.1) также весьма поучительно. Хотя оно содержится в большинстве учебников по математическому анализу, мы воспроизведем его здесь из-за присущей ему оригинальности. Фокус состоит в том, чтобы возвести в квадрат интеграл из левой части формулы (7.4.1) и затем преобразовать его в двойной интеграл:

$$\left(\int_{-\infty}^{\infty} \varphi(x)\,dx\right)\left(\int_{-\infty}^{\infty} \varphi(y)\,dy\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x)\,\varphi(y)\,dx\,dy = \\ = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(1/2)(x^2+y^2)}dx\,dy.$$

Для вычисления последнего интеграла используем переход к полярным координатам:

$$\rho^2 = x^2 + y^2, \quad dx\,dy = \rho\,d\rho\,d\theta,$$

$$\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-(1/2)\rho^2}\rho\,d\rho\,d\theta = \frac{1}{2\pi}\int_0^{2\pi} -e^{-(1/2)\rho^2}\Big|_0^{\infty}\,d\theta = \frac{1}{2\pi}\int_0^{2\pi} 1\,d\theta = 1.$$

Извлечение квадратного корня приводит к равенству (7.4.1).

Плотность φ стандартного нормального распределения обладает множеством замечательных аналитических свойств. В действительности, Гаусс определил ее путем выбора некоторых из них в качестве характеристических для «закона распределения ошибок наблюдений»[*]. Прежде всего заметим, что φ является четной функцией аргумента x,

[*] Карл Фридрих Гаусс (1777–1855), считающийся одним из величайших математиков, также известен своими фундаментальными работами по физике, астрономии и геодезии. Наибольший его вклад в вероятностную науку — это создание теории ошибок наблюдений, иначе называемой методом наименьших квадратов.

т. е. $\varphi(x) = \varphi(-x)$. Отсюда вытекает следующая удобная формула:

$$\int\limits_{-x}^{x} \varphi(u)\, du = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1. \tag{7.4.2}$$

Далее, φ имеет производные любого порядка, причем каждая производная представляет собой произведение функции φ и некоторого многочлена, называемого *полиномом Эрмита*. Существование всех производных делает кривую $x \to \varphi(x)$ очень гладкой. Она обычно описывается как кривая «колоколообразной формы»*). Кроме того, при $|x| \to \infty$ плотность $\varphi(x)$ убывает к 0 очень быстро. Часто оказывается полезной следующая оценка «хвоста» функции Φ:

$$1 - \Phi(x) = \int\limits_{x}^{\infty} \varphi(u)\, du \leqslant \frac{\varphi(x)}{x} = \frac{e^{-x^2/2}}{\sqrt{2\pi}\, x}, \quad x > 0.$$

Для ее вывода заметим, что $-\varphi'(u) = u\varphi(u)$. Отсюда

$$\int\limits_{x}^{\infty} 1 \cdot \varphi(u)\, du \leqslant \int\limits_{x}^{\infty} \frac{u}{x}\, \varphi(u)\, du = \frac{-1}{x} \int\limits_{x}^{\infty} \varphi'(u)\, du = \frac{-1}{x}\, \varphi(u)\Big|_{x}^{\infty} = \frac{\varphi(x)}{x}$$

(еще один ловкий фокус). Из этой оценки следует, что не только распределение Φ обладает моментами всех порядков, но, более того, — интеграл

$$M(\theta) = \int\limits_{-\infty}^{\infty} e^{\theta x} \varphi(x)\, dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{\theta x - (x^2/2)}\, dx \tag{7.4.3}$$

конечен для любого действительного θ, поскольку $e^{-x^2/2}$ убывает намного быстрее, чем возрастает $e^{|\theta x|}$ при $|x| \to \infty$. Функция $M(\theta)$ называется *производящей функцией моментов* для φ или Φ. Заметим, что если заменить θ на чисто мнимое число $i\theta$, то $M(i\theta)$ превратится в характеристическую функцию или преобразованием Фурье для Φ (см. (6.5.17)). Причина, по которой мы не ввели производящую функцию моментов в § 6.5, заключается в том, что интеграл в формуле (7.4.3) обычно расходится, если заменить φ на произвольную плотность. Но для нормальной плотности φ функция $M(\theta)$ выглядит проще, чем $M(i\theta)$, и ее можно использовать с тем же успехом. Вычислим функцию $M(\theta)$. Это нетрудно

*) См. график, изображенный рядом с таблицей значений $\Phi(x)$ в конце книги.

сделать за счет дополнения показателя экспоненты в последнем подынтегральном выражении формулы (7.4.3) до полного квадрата:

$$\theta x - \frac{x^2}{2} = \frac{\theta^2}{2} - \frac{(x-\theta)^2}{2}.$$

Отсюда

$$M(\theta) = e^{\theta^2/2} \int\limits_{-\infty}^{\infty} \varphi(x-\theta)\,dx = e^{\theta^2/2}. \tag{7.4.4}$$

На основе данной формулы можно вычислить все моменты распределения Φ последовательным дифференцированием функции $M(\theta)$ по θ, как в случае производящей функции, который обсуждался в § 6.5. Вот другой (более прямой) путь: разложим $e^{\theta x}$ в формуле (7.4.3) в ряд Тейлора по θ и сравним результат с тейлоровским разложением функции $e^{\theta^2/2}$ из правой части равенства (7.4.4):

$$\int\limits_{-\infty}^{\infty} \left\{1 + \theta x + \frac{(\theta x)^2}{2!} + \ldots + \frac{(\theta x)^n}{n!} + \ldots\right\} \varphi(x)\,dx =$$
$$= 1 + \frac{\theta^2}{2} + \frac{1}{2!}\left(\frac{\theta^2}{2}\right)^2 + \ldots + \frac{1}{n!}\left(\frac{\theta^2}{2}\right)^n + \ldots.$$

Если обозначить n-й момент через $m^{(n)}$:

$$m^{(n)} = \int\limits_{-\infty}^{\infty} x^n \varphi(x)\,dx,$$

то полученное равенство можно записать как

$$\sum_{n=0}^{\infty} \frac{m^{(n)}}{n!}\,\theta^n = \sum_{n=0}^{\infty} \frac{1}{2^n n!}\,\theta^{2n}.$$

Из единственности разложения в степенной ряд (сравните с § 6.5) вытекает, что соответствующие коэффициенты в обеих частях должны совпадать. Поэтому при $n \geqslant 1$

$$m^{(2n-1)} = 0, \quad m^{(2n)} = \frac{(2n)!}{2^n n!}. \tag{7.4.5}$$

Конечно, равенство нулю всех моментов нечетного порядка немедленно следует из четности функции φ.

В общем случае для любых действительных чисел m и $\sigma^2 > 0$ говорят, что случайная величина X имеет *нормальное распределение* $N(m, \sigma^2)$ тогда и только тогда, когда *стандартизованная случайная величина* $X^* = (X-m)/\sigma$ имеет функцию распределения Φ. В частности, при $m = 0$ и $\sigma^2 = 1$ закон $N(0,1)$ является стандартным нормальным распределением.

Плотностью нормального распределения $N(m, \sigma^2)$ служит функция

$$\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-m)^2/(2\sigma^2)} = \frac{1}{\sigma}\,\varphi\left(\frac{x-m}{\sigma}\right). \tag{7.4.6}$$

Эта формула вытекает из общего утверждения, содержащегося в задаче 13 из гл. 4. Производящую функцию моментов M_X для случайной величины X проще всего получить из производящей функции моментов для стандартизованной переменной X^* так:

$$\begin{aligned} M_X(\theta) = E\big(e^{\theta(m+\sigma X^*)}\big) = e^{m\theta} E\big(e^{(\sigma\theta)X^*}\big) = \\ = e^{m\theta} M(\sigma\theta) = e^{m\theta + \sigma^2\theta^2/2}. \end{aligned} \tag{7.4.7}$$

Ниже приводится основное свойство *семейства нормальных распределений.* Сравните его с аналогичной теоремой 3 из § 7.2 для пуассоновского семейства.

Теорема 7. *Пусть X_j — независимые случайные величины, имеющие нормальные распределения*

$$N(m_j, \sigma_j^2), \quad 1 \leqslant j \leqslant n.$$

Тогда сумма $X_1 + \ldots + X_n$ также распределена согласно нормальному закону

$$N\left(\sum_{j=1}^{n} m_j, \sum_{j=1}^{n} \sigma_j^2\right).$$

Доказательство. Достаточно установить справедливость теоремы для $n = 2$, так как общий случай тогда получается с помощью индукции. Это просто сделать, используя производящую функцию моментов. В силу теоремы умножения, подобной теореме 6 из § 6.5, имеем:

$$\begin{aligned} M_{X_1+X_2}(\theta) = M_{X_1}(\theta) M_{X_2}(\theta) = e^{m_1\theta + (\sigma_1^2\theta^2/2)} e^{m_2\theta + (\sigma_2^2\theta^2/2)} = \\ = e^{(m_1+m_2)\theta + (\sigma_1^2+\sigma_2^2)\theta^2/2}. \end{aligned}$$

Ввиду равенства (7.4.7), справа стоит производящая функция моментов для распределения $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. Следовательно, сумма $X_1 + X_2$ нормально распределена, так как распределение однозначно определяется производящей функцией моментов. (Мы не доказывали данное утверждение, но сравните его с теоремой 7, приведенной в конце § 6.5.)

⋆7.5. Центральная предельная теорема

Вернемся к теореме Муавра—Лапласа и дадим для нее более общую формулировку. Напомним определение случайной величины S_n:

$$S_n = X_1 + \ldots + X_n, \quad n \geqslant 1, \tag{7.5.1}$$

где X_j — независимые бернуллиевские случайные величины. Мы знаем, что для каждого j

$$E(X_j) = p, \quad \sigma^2(X_j) = pq;$$

и для каждого n

$$E(S_n) = np, \quad \sigma^2(S_n) = npq$$

(см. пример 6 в § 6.3). Положим

$$X_j^* = \frac{X_j - E(X_j)}{\sigma(X_j)}\,; \quad S_n^* = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j^*. \tag{7.5.2}$$

S_n^* — это случайные величины, встречавшиеся ранее в левой части равенства (7.3.19), которые иногда называют *нормированными* или *стандартизированными суммами*. Для каждого j и n имеем:

$$\begin{aligned} E(X_j^*) = 0, \quad \sigma^2(X_j^*) = 1, \\ E(S_n^*) = 0, \quad \sigma^2(S_n^*) = 1. \end{aligned} \tag{7.5.3}$$

Линейное преобразование случайных величин X_j в X_j^* или S_n в S_n^* приводит к изменению начала координат и масштаба при измерении случайных величин. Цель такого преобразования — получить нулевое среднее и дисперсию 1, как показано в соотношениях (7.5.3). Каждая случайная величина S_n^* принимает значения из множества

$$x_{n,k} = \frac{k - np}{\sqrt{npq}}\,, \quad k = 0, 1, \ldots, n.$$

Это те самые x_{nk}, которые появились в формуле (7.3.11). Вероятностное распределение случайной величины S_n^* имеет следующий вид:

$$P(S_n^* = x_{n,k}) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leqslant k \leqslant n.$$

Удобнее использовать соответствующие функции распределения, обозначаемые через F_n:

$$P(S_n^* \leqslant x) = F_n(x), \quad -\infty < x < \infty.$$

Наконец, для конечного интервала $I = (a, b]$ и функции распределения F будем использовать запись

$$F(I) = F(b) - F(a).$$

(Теперь вы уже должны понимать, почему мы используем полуинтервал $(a, b]$, а не (a, b) или $[a, b]$. Различия нет, когда функция F непрерывна, но F_n, указанные выше, непрерывными не являются. Конечно, в нашем случае в пределе различие исчезает, но в общем случае его нельзя игнорировать.) В результате всех этих продолжительных приготовлений удается представить формулу Муавра—Лапласа в следующем элегантном виде: для произвольного конечного интервала I

$$\lim_{n\to\infty} F_n(I) = \Phi(I). \tag{7.5.4}$$

Таким образом, мы видим, что имеем дело с (понимаемой в некотором смысле) *сходимостью последовательности функций распределения к заданной функции распределения.*

Такая формулировка допускает широчайшие обобщения. Сходящаяся последовательность не обязана состоять из функций распределения нормированных сумм, предел может отличаться от нормального распределения и даже может быть заранее не известен, смысл сходимости также может быть иным. Например, предельная теорема Пуассона, рассматриваемая в § 7.1, подпадает под нашу обобщенную формулировку. Данная проблематика интенсивно изучалась в 1940-х гг. и до сих пор претерпевает дальнейшее развитие. (Подробнее с ней можно познакомиться по книгам [4,10].) Здесь мы ограничимся только одним результатом — так называемой центральной предельной теоремой в классической постановке, которая является, пожалуй, простейшим обобщением теоремы 6 из § 7.3. Даже в этом случае нам понадобиться мощный инструмент из более сложной теории, который мы применим, но не станем подробно объяснять. Само же обобщение состоит в замене бернуллиевских случайных величин на довольно произвольные случайные величины, которые мы сейчас опишем.

Пусть $\{X_j, j \geqslant 1\}$ — последовательность *независимых и одинаково распределенных* случайных величин. Термин «одинаково распределенные» означает, что величины имеют общее распределение, которое не обязательно должно быть указано. Однако предполагается, что среднее и дисперсия каждой из величин X_j конечны. Они обозначаются, соответственно, через m и σ^2, где $0 < \sigma^2 < \infty$. Определим S_n и S_n^* так же, как раньше. Тогда

$$E(S_n) = nm, \quad \sigma^2(S_n) = n\sigma^2, \tag{7.5.5}$$

и равенства (7.5.3) выполняются, как прежде. Пусть опять F_n обозначает распределение нормированной суммы S_n^*. Тогда теорема 8, приводимая ниже, утверждает, что равенство (7.5.4) остается верным даже при наших необременительных предположениях относительно случайных величин X_j. Отметим только некоторые простые случаи: каждая из X_j может описывать бросание игральной кости, вместо монеты, либо может быть равномерно распределенной случайной величиной (выбор точки наудачу), либо экспоненциальной случайной величиной (продолжительность телефонного разговора). Придумайте другие примеры, если хотите.

Теорема 8. *При выполнении указанных выше общих условий для сумм S_n при произвольных $a < b$*

$$\lim_{n\to\infty} P\left(a < \frac{S_n - nm}{\sqrt{n}\,\sigma} \leqslant b\right) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2}\,dx. \tag{7.5.6}$$

Доказательство. Упомянутый выше мощный инструмент — это характеристические функции, рассмотренные в § 6.5. (Производящие функции моментов использовать нельзя, так как они могут не существовать для S_n.) Характеристическую функцию g стандартного нормального распределения Φ найдем, подставив $i\theta$ вместо θ в формулу (7.4.4):

$$g(\theta) = e^{-\theta^2/2}. \tag{7.5.7}$$

С каждой функцией распределения F_n также связана ее характеристическая функция g_n, которая в общем случае выражается через F_n с помощью *интеграла Стильтьеса*. Это понятие выходит за рамки данной книги. К счастью, мы можем обойтись без него в дальнейших рассуждениях, используя соответствующие случайные величины. (Как видите, читателю оставляется возможность самостоятельно изучить то, что осталось «за кадром»!) Теперь мы готовы сформулировать следующий результат.

Теорема 9. *Если для каждого θ*

$$\lim_{n\to\infty} g_n(\theta) = g(\theta) = e^{-\theta^2/2}, \tag{7.5.8}$$

то для каждого x

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^x e^{-u^2/2}\,du; \tag{7.5.9}$$

в частности, выполняется равенство (7.5.4).

Хотя мы и не станем доказывать это утверждение (см. [4, гл. 6]), давайте по крайней мере убедимся в его важности. Согласно теореме 7 из § 6.5, каждая характеристическая функция g_n однозначно определяет функцию распределения F_n, а g определяет Φ. Данная теорема углубляет соответствие между функциями распределения и их преобразованиями (характеристическими функциями) еще на один шаг, так как она говорит, что *предел* последовательности $\{g_n\}$ тоже определяет *предел* последовательности $\{F_n\}$. Поэтому она называется «теоремой непрерывности» для преобразований. В случае нормального распределения Φ данный результат был доказан Пойа. Общий случай рассматривали Поль Леви́ (1886–1972) и Харальд Краме́р (1893–1985); оба являются основоположниками современной теории вероятностей.

Нам также потребуется следующая простая лемма о характеристических функциях.

Лемма. *Если X — случайная величина со средним 0 и дисперсией 1, то тейлоровское разложение ее характеристической функции h в точке $\theta = 0$ имеет следующий вид:*

$$h(\theta) = 1 - \frac{\theta^2}{2}\,(1 + \varepsilon(\theta)), \tag{7.5.10}$$

где $\varepsilon(\theta)$ есть зависящая от h функция такая, что $\lim\limits_{\theta\to 0} \varepsilon(\theta) = 0$.

Доказательство. Одна из наиболее полезных форм теоремы Тейлора (посмотрите ее в вашем учебнике по математическому анализу) утверждает, что если функция h дважды дифференцируема в точке $\theta = 0$, то справедливо представление

$$h(\theta) = h(0) + h'(0)\theta + \frac{h''(0)}{2}\,\theta^2(1 + \varepsilon(\theta)). \tag{7.5.11}$$

Из равенства

$$h(\theta) = E\big(e^{i\theta X}\big)$$

формальным дифференцированием получаем, что

$$h'(\theta) = E\big(e^{i\theta X} iX\big), \quad h''(\theta) = E\big(e^{i\theta X}(iX)^2\big).$$

Отсюда

$$h'(0) = E(iX) = 0, \quad h''(0) = E(-X^2) = -1.$$

Подставляя полученные числовые значения в формулу (7.5.11), приходим к соотношению (7.5.10).

Теорема 8 теперь может быть установлена прямыми выкладками. Рассмотрим характеристическую функцию случайной величины S_n^*:

$$E\big(e^{i\theta S_n^*}\big) = E\big(e^{i\theta(X_1^* + \ldots + X_n^*)/\sqrt{n}}\big).$$

Так как X_j^* — независимые и одинаково распределенные случайные величины (так же, как и X_j), по аналогии с теоремой 6 из § 6.5 находим, что правая часть приведенной выше формулы равна

$$E(e^{i\theta X_1^*/\sqrt{n}})^n = h\left(\frac{\theta}{\sqrt{n}}\right)^n, \tag{7.5.12}$$

где h обозначает характеристическую функцию случайной величины X_1^*. Из леммы следует, что

$$h\left(\frac{\theta}{\sqrt{n}}\right) = 1 - \frac{\theta^2}{2n}\left(1 + \varepsilon\left(\frac{\theta}{\sqrt{n}}\right)\right), \tag{7.5.13}$$

где значение θ фиксировано и $n \to \infty$. Отсюда с учетом равенства (7.1.12) получаем:

$$\lim_{n\to\infty} E\big(e^{i\theta S_n^*}\big) = \lim_{n\to\infty}\left[1 - \frac{\theta^2}{2n}\left(1 + \varepsilon\left(\frac{\theta}{\sqrt{n}}\right)\right)\right]^n = e^{-\theta^2/2}.$$

Это означает, что характеристические функции величин S_n^* сходятся к характеристической функции стандартного нормального закона. Тогда, в силу теоремы 9, функции распределения F_n сходятся к Φ в смысле (7.5.9), откуда вытекает формула (7.5.6).

«Центральной предельной теоремой» обычно называют утверждение о сходимости, в котором нормальное распределение появляется в качестве предела. В частности, она применима к суммам случайных величин из теоремы 8. Исторически такие суммы возникли как ошибки наблюдений, подверженных влиянию случайности. Поэтому теорему можно понимать как всеобъемлющее утверждение о том, что при «нормальных» условиях на слагаемые, стандартизированная сумма распределена согласно *нормальному* закону, называемому также «функцией ошибок». По этой причине результат воспринимался некоторыми как закон природы! Даже в такой узкой постановке теорему 8 удается обобщить в нескольких направлениях: можно отказаться или ослабить требования конечности второго момента, одинаковости распределения слагаемых, их строгой независимости. Наконец, если «нормальные» условия существенно изменить, то центральная предельная теорема перестанет выполняться. На самом деле существуют случайные явления, для которых предельное распределение не является нормальным. Примером этого может служить случай сходимости к закону Пуассона, рассмотренный в § 7.1. Однако имеются и другие предельные законы, тесно связанные с нормальным: так называемые «устойчивые» и «безгранично делимые». Обсуждение этой темы см. [4, гл. 7].

Следует подчеркнуть, что центральная предельная теорема, как она сформулирована в теоремах 6 и 8, является утверждением вида (7.5.4) и не содержит никакой оценки для величины погрешности $|F_n(I) - \Phi(I)|$. Другими словами, она утверждает наличие сходимости, но ничего не говорит о скорости сходимости. Это делает результат бесполезным для точных численных расчетов. Однако при определенных условиях возможно установить верхнюю границу для величины погрешности. К примеру, в случае теоремы Муавра—Лапласа (7.3.19) удается показать, что погрешность не превышает $C/\sqrt{n}$, где C — константа, зависящая от p, но не от a или b (см. [4, § 7.4], где приведен более общий результат). Часто при выполнении приблизительных расчетов эта погрешность просто игнорируется. Так будем ниже поступать и мы.

В отличие от математических обобщений простые практические применения, составляющие каркас «теории больших выборок» в статистике, обычно похожи на кулинарные рецепты. Великая предельная теорема, выражаемая равенством (7.5.6), превращается в нестрогую приближенную формулу, которую можно записать так:

$$P\big(x_1\sigma\sqrt{n} < S_n - mn < x_2\sigma\sqrt{n}\big) \approx \Phi(x_2) - \Phi(x_1).$$

Во многих ситуациях нас интересуют возможные отклонения от среднего на одинаковые расстояния в обе стороны, т. е. случай, когда $x_1 = -x_2$. Тогда с учетом соотношения (7.4.2) последняя формула приобретает вид

$$P\big(|S_n - mn| < x\sigma\sqrt{n}\big) \approx 2\Phi(x) - 1. \tag{7.5.14}$$

Имеются подробные таблицы значений функции Φ и обратной к ней функции Φ^{-1} (небольшая таблица приведена в конце книги). Следующий пример иллюстрирует процедуру применения центральной предельной теоремы.

Пример 7. Некоторая физическая величина измеряется несколько раз для достижения большей точности. Каждое из измерений производится со случайной ошибкой. Разумно предположить, что эта ошибка равномерно распределена между -1 и $+1$ в удобной шкале. Теперь, если вычислить среднее арифметическое всех n выполненных измерений, то какова вероятность того, что оно отличается от истинного значения не более, чем на δ?

Обозначим истинное значение через m. Пусть $X_j, 1 \leqslant j \leqslant n$ — результаты измерения. Тогда гипотеза утверждает, что

$$X_j = m + \xi_j,$$

где ξ_j — случайные величины, равномерно распределенные на отрезке $[-1, +1]$. Поэтому

$$E(\xi_j) = \int\limits_{-1}^{+1} \frac{x}{2}\,dx = 0, \qquad \sigma^2(\xi_j) = E\big(\xi_j^2\big) = \int\limits_{-1}^{+1} \frac{1}{2}\,x^2\,dx = \frac{1}{3}\,,$$

$$E(X_j) = m, \qquad \sigma^2(X_j) = \frac{1}{3}\,.$$

В наших новых обозначениях мы хотим приближенно вычислить $P\{|S_n - mn| < \delta n\}$. Эту вероятность надо представить в форме, стоящей в левой части (7.5.6), и, учитывая приближенное равенство (7.5.14), заменить выражение, содержащее предел, на

$$P\left\{\left|\frac{S_n - mn}{\sqrt{n/3}}\right| < \delta\sqrt{3n}\right\} \approx 2\Phi(\delta\sqrt{3n}) - 1.$$

Например, если $n = 25$ и $\delta = 1/5$, то в ответе получим

$$2\Phi(\sqrt{3}) - 1 \approx 2\Phi(1.73) - 1 \approx 0.92,$$

согласно таблице в конце книги. Таким образом, если произведено 25 измерений, то у нас есть 92 %-я уверенность, что их среднее арифметическое отстоит от истинного значения не дальше, чем на 1/5.

Часто возникает обратная задача: сколько необходимо выполнить измерений, чтобы с вероятностью не менее α (эта величина называется *уровнем значимости*) среднее отклонялось от истинного значения не более, чем на δ? Это означает, что надо найти такое значение x_α, что

$$2\Phi(x_\alpha) - 1 = \alpha \quad \text{или} \quad \Phi(x_\alpha) = \frac{1+\alpha}{2},$$

а затем выбрать n, обеспечивающее выполнение неравенства

$$\delta\sqrt{3n} > x_\alpha.$$

Например, если $\alpha = 0.95$ и $\delta = 1/5$, то из таблицы находим $x_\alpha \approx 1.96$. Отсюда

$$n > \frac{x_\alpha^2}{3\delta^2} \approx 32.$$

Таким образом, 7 или 8 добавочных измерений позволят увеличить нашу степень уверенности с 92 % до 95 %. Стоит их проводить или нет — зависит от стоимости дополнительной работы, а также от важности увеличения вероятности.

Понятно, что в вопросах подобного типа всего участвуют три переменных: δ, α и n. Если две из них заданы, то можно вычислить третью. Так, если величина $n = 25$ фиксирована из-за того, что измерения были выполнены ранее и в распоряжении исследователя имеются только записи результатов, причем для него крайне важна высокая степень доверия к выводам, скажем, $\alpha = 99\,\%$, то придется пойти на компромисс в отношении величины точности δ. Мы оставляет эту задачу в качестве упражнения.

Общепризнанно, что подобные практические применения великой теоремы являются довольно рутинными, но то же самое можно сказать, например, о законах движения Ньютона с точки зрения обыденного уровня.

7.6. Закон больших чисел

В этом параграфе излагаются два результата, тесно связанные с центральной предельной теоремой: закон больших чисел и неравенство Чебышёва.

Покажем, что знаменитый закон больших чисел легко может быть выведен из теоремы 8.

Теорема 10. *Пусть выполнены те же условия, что и в теореме 8. Тогда для произвольной константы* $c > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - m\right| < c\right) = 1. \tag{7.6.1}$$

Доказательство. Так как c фиксировано, то для любой положительной константы l выполняется неравенство

$$l\sigma\sqrt{n} < cn \tag{7.6.2}$$

для всех достаточно больших n. Следовательно, событие

$$\left\{\left|\frac{S_n - mn}{\sigma\sqrt{n}}\right| < l\right\} \quad \text{влечет} \quad \left\{\left|\frac{S_n - mn}{n}\right| < c\right\},$$

и поэтому

$$P\left(\left|\frac{S_n - mn}{o\sqrt{n}}\right| < l\right) \leqslant P\left(\left|\frac{S_n - mn}{n}\right| < c\right) \tag{7.6.3}$$

для больших n. В силу равенства (7.5.6) при $a = -l$ и $b = +l$ левая часть сходится к

$$\frac{1}{\sqrt{2\pi}} \int_{-l}^{l} e^{-x^2/2}\, dx$$

при $n \to \infty$. Для любого заданного $\delta > 0$ можно сначала выбрать l настолько большим, чтобы величина этого интеграла превзошла $1 - \delta$, затем выбрать n настолько большим, чтобы выполнялось (7.6.3). Отсюда следует, что

$$P\left(\left|\frac{S_n}{n} - m\right| < c\right) > 1 - \delta \tag{7.6.4}$$

для всех достаточно больших n. Именно это и утверждается в формуле (7.6.1).

Говоря кратко, закон больших чисел является непосредственным следствием центральной предельной теоремы, потому что произведение сколь угодно большой константы и $\sqrt{n}$ пренебрежимо мало по сравнению с произведением сколь угодно малой константы и n.

В бернуллиевском случае данный закон был впервые установлен Якобом Бернулли, и этот результат можно считать одним из его важнейших достижений*). Его доказательство опиралось на прямые выкладки, проводимые над биномиальными коэффициентами, без использования того преимущества, которые дает формула Стирлинга. В некотором смысле теорема 6 Муавра—Лапласа является непосредственным обобщением данного закона. Изложение этих результатов в порядке, обратном к историческому, приводит к представлению, что закон больших чисел — лишь простое следствие. На самом деле, он является более фундаментальной, хотя и более примитивной, предельной теоремой, которая выполняется при намного менее ограничительных условиях по сравнению с центральной предельной теоремой. К примеру, в формулировке теоремы 8 достаточно предположить только конечность математического ожидания случайных величин X_j, не требуя конечности второго момента. Так как данное обобщение закона больших чисел позволяет иметь дело только со средними, то оно весьма важно. Впервые его доказал А. Я. Хинчин **). На самом деле оно может быть получено с помощью метода, использованного при доказательстве теоремы 8, правда, для этого потребуется существенное усиление теоремы 9, которое представляется чересчур сложным для данной книги (см. теорему 6.4.3 в [4]). Вместо этого мы приведем обобщение теоремы 10 в другом направлении: откажемся от условия одинаковой распределенности величин X_j. Его легко получить с помощью другого знаменитого, но простого, результата, известного как *неравенство Чебышёва****).

Теорема 11. *Допустим, что случайная величина X имеет конечный второй момент. Тогда для произвольной константы $c > 0$*

$$P(|X| \geqslant c) \leqslant \frac{E(X^2)}{c^2}\,. \tag{7.6.5}$$

*) Якоб Бернулли (1654–1705) — швейцарский математик и физик, автор первого известного труда по теории вероятностей — книги *Искусство предположений* (1713), которая содержит обсуждаемую теорему.

**) А. Я. Хинчин (1894–1959) — один из наиболее известных математиков русской вероятностной школы.

***) П. Л. Чебышёв (1821–1894) вместе с А. А. Марковым (1856–1922) и А. М. Ляпуновым (1857–1918) являются основателями русской вероятностной школы.

Доказательство. Мы проведем доказательство для случая, когда X принимает счетное число значений, оставляя случай, когда существует плотность, в качестве упражнения. Идея доказательства годится и для произвольно распределенной случайной величины.

Допустим, что X принимает значения v_i с вероятностями p_i, как в § 4.3. Тогда

$$E(X^2) = \sum_j p_j v_j^2. \tag{7.6.6}$$

Если рассмотреть только те значения v_j, которые удовлетворяют неравенству $|v_j| \geqslant c$, и обозначить через A соответствующее множество индексов j, а именно: $A = \{j\colon |v_j| \geqslant c\}$, то, конечно, $v_j^2 \geqslant c^2$ для $j \in A$. При этом

$$P(|X| \geqslant c) = \sum_{j \in A} p_j.$$

Тогда если просуммировать только по индексам j, входящим в множество A, то получим

$$E(X^2) \geqslant \sum_{j \in A} p_j v_j^2 \geqslant \sum_{j \in A} p_j c^2 = c^2 \sum_{j \in A} p_j = c^2 P(|X| \geqslant c),$$

что совпадает с неравенством (7.6.5).

Теперь мы можем сформулировать обобщение закона больших чисел в следующем виде.

Теорема 12. *Пусть $\{X_j, j \geqslant 1\}$ — последовательность независимых случайных величин таких, что при каждом j*

$$E(X_j) = m_j, \quad \sigma^2(X_j) = \sigma_j^2. \tag{7.6.7}$$

Кроме того, допустим, что существует константа $M < \infty$ такая, что для всех j

$$\sigma_j^2 \leqslant M. \tag{7.6.8}$$

Тогда для каждого фиксированного $c > 0$

$$\lim_{n \to \infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \frac{m_1 + \ldots + m_n}{n}\right| < c\right) = 1. \tag{7.6.9}$$

Доказательство. Если записать

$$X_j^0 = X_j - m_j, \quad S_n^0 = \sum_{j=1}^n X_j^0,$$

то выражение внутри модуля в формуле (7.6.9) — не что иное, как S_n^0/n. Конечно, $E(S_n^0) = 0$, в то время, как

$$E\big((S_n^0)^2\big) = \sigma^2\big(S_n^0\big) = \sum_{j=1}^{n} \sigma^2(X_j^0) = \sum_{j=1}^{n} \sigma_j^2.$$

Данную цепочку равенств легко обосновать с помощью свойств дисперсии, и вы должны без труда узнавать их с первого взгляда. (Если это не так, то вам следует просмотреть еще раз те разделы предыдущих глав, где они обсуждаются.) Далее, условие (7.6.8) влечет, что

$$E\big((S_n^0)^2\big) \leqslant Mn, \quad E\left(\left(\frac{S_n^0}{n}\right)^2\right) \leqslant \frac{M}{n}\,. \tag{7.6.10}$$

Остается только применить теорему 11 к случайной величине $X = S_n^0/n$, чтобы получить оценку

$$P\left(\left|\frac{S_n^0}{n}\right| \geqslant c\right) \leqslant \frac{E((S_n^0/n)^2)}{c^2} \leqslant \frac{M}{c^2 n}\,. \tag{7.6.11}$$

Отсюда следует, что вероятность в левой части стремится к нулю при $n \to \infty$, что эквивалентно утверждению (7.6.9).

На самом деле мы установили даже больше, чем утверждалось: в доказательстве содержится оценка для «скорости сходимости». Точнее: для заданных M, c и δ мы можем сказать, как велико должно быть n, чтобы вероятность в формуле (7.6.9) превысила $1 - \delta$. Отметим также, что теорема 10 является частным случаем теоремы 12: достаточно взять все σ_j^2 равными и положить $M = \sigma_1^2$.

Возможно, читатель согласится с тем, что приведенные выше доказательства теорем 11 и 12 довольно просты по сравнению с усилиями, затраченными в §§ 7.3–7.4. Оглядываясь назад, вероятно, покажется удивительным, что прошли два века прежде, чем Чебышёв открыл *правильное* доказательство теоремы Бернулли. Оно выражает триумф *идеи*, нового образа мышления, но даже сам Чебышёв похоронил свое неравенство под громоздкими и ненужными деталями. Очищенный вариант, продемонстрированный выше, был получен более поздними авторами. Обратим внимание, что метод доказательства применим к любым суммам S_n случайных величин, причем при условии справедливости ключевого неравенства (7.6.10) неважно, зависимы они или нет.

Обсудим теперь смысл закона больших чисел. Его проще всего объяснить на примере простейшей схемы Бернулли, в которой каждая случайная величина X_j принимает значения 1 и 0 с вероятностями p и $q = 1-p$

соответственно, как в теоремах 5 и 6. В этом случае

$$S_n^0 = S_n - np \quad \text{и} \quad E\big(\big(S_n^0\big)^2\big) = \sigma^2(S_n) = npq,$$

поэтому неравенство (7.6.11) превращается в

$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant c\right) \leqslant \frac{pq}{c^2 n} \leqslant \frac{1}{4c^2 n}\,; \tag{7.6.12}$$

заметьте, что $p(1-p) \leqslant 1/4$ при $0 \leqslant p \leqslant 1$. В терминах бросания монеты с вероятностью выпадения «орла», равной p, величина S_n/n представляет собой относительную частоту появлений «орла» в n независимых бросаниях; сравните с примером 3 из § 2.1. Она, конечно, случайна и меняется от эксперимента к эксперименту. Правильно будет рассмотреть выборочную точку ω, имеющую бесконечное количество координат, и записать, что

$$\frac{S_n(\omega)}{n} = \begin{array}{l}\text{относительная частота выпадений «орла» в } n \text{ бросаниях,}\\ \text{связанных с экспериментом, обозначаемым через } \omega.\end{array}$$

Тем самым, каждая точка ω понимается как запись последовательности результатов бросания, кратко называемой *экспериментом*; при этом разные эксперименты представляются разными ω. Последовательность бросаний следует считать *бесконечной* (неограниченной) для того, чтобы позволить параметру n, участвующему в формуле для вычисления частоты, принимать произвольно большие значения (но мы не предполагаем, что частота обязательно имеет предел, см. ниже). Формально ω можно понимать как конкретную реализацию последовательности исходов:

$$\omega = \{X_1(\omega), X_2(\omega), \ldots, X_n(\omega), \ldots\};$$

см. обсуждение в § 4.1. Можно говорить о вероятностях определенных множеств, состоящих из точек ω. На самом деле, мы уже имели дело с вероятностью множества

$$\Lambda_n(c) = \left\{\omega\colon \left|\frac{S_n(\omega)}{n} - p\right| \geqslant c\right\}$$

в неравенстве (7.6.12); см. также, например, формулы (7.3.10) и (7.3.19), где вычисляются подобные вероятности. С помощью оценки сверху из (7.6.12) для заданного значения c и произвольного положительного числа ε удается определить, как велико должно быть n, чтобы $P(\Lambda_n(c)) \leqslant \varepsilon$, а именно,

$$\frac{1}{4c^2 n} \leqslant \varepsilon \quad \text{или} \quad n \geqslant \frac{1}{4c^2 \varepsilon}\,. \tag{7.6.13}$$

Данная граница для n на самом деле завышена из-за того, что неравенство Чебышёва, вообще говоря, является грубой оценкой. Более точная граница может быть получена с помощью теоремы 6 следующим образом. Перепишем вероятность в неравенстве (7.6.12) и применим аппроксимацию (7.3.19) с учетом соотношения (7.4.2):

$$P\left(\left|\frac{S_n - np}{\sqrt{npq}}\right| \geqslant \sqrt{\frac{n}{pq}}\,c\right) \approx 2\left[1 - \Phi\left(\sqrt{\frac{n}{pq}}\,c\right)\right]. \tag{7.6.14}$$

Эта формула *не является* асимптотическим соотношением, на что она, вроде бы, должна претендовать. Причина в том, что в теореме 6 константы a и b должны быть фиксированными, в то время как $\pm c\sqrt{n}/\sqrt{pq}$ меняются вместе с n (и растут достаточно быстро с увеличением n). Можно показать, что разность правой и левой частей приведенной выше формулы имеет порядок малости $A/\sqrt{n}$, где A — константа, зависящая от p, и эту погрешность не следует игнорировать. Однако мы все же поступим так ниже. Теперь положим

$$\eta = \sqrt{\frac{n}{pq}}\,c.$$

Тогда наша задача — найти значение η, обеспечивающее выполнение неравенства

$$2[1 - \Phi(\eta)] \leqslant \varepsilon \quad \text{или} \quad \Phi(\eta) \geqslant 1 - \frac{\varepsilon}{2}, \tag{7.6.15}$$

а затем выразить n через η. Это можно сделать, посмотрев в таблицу значений функции Φ (краткая таблица имеется в конце книги).

Пример 8. Пусть $c = 2\,\%$ и $\varepsilon = 5\,\%$. Тогда неравенство (7.6.15) превращается в условие

$$\Phi(\eta) \geqslant 1 - \frac{5}{200} = 0.975.$$

Из таблицы находим, что оно выполняется при $\eta \geqslant 1.96$. Поэтому

$$n \geqslant \frac{(1.96)^2\,pq}{c^2} = \frac{(1.96)^2 \times 10\,000}{4}\,pq.$$

Правая часть зависит от p, но $p(1-p) \leqslant 1/4$ для всех p, как уже отмечалось. Отсюда получаем $n \geqslant 10\,000 \cdot (1/4) = 2\,500$. Для сравнения, неравенство (7.6.13) дает $n \geqslant 12\,500$; но эта граница была установлена строго, а при нормальной аппроксимации использовались приближенные равенства. Окончательный вывод таков: если монета бросается более 2 500 раз, то у нас есть 95 %-я уверенность, что относительная частота выпадений «орла», наблюдаемая в реальном эксперименте, не будет отличаться от истинного значения p более чем на 2 %.

Подобный результат можно применить двумя способами (их представлял себе уже Бернулли): (i) если мы считаем, что p известно, то можно делать предположения о результате эксперимента; (ii) если мы рассматриваем p как неизвестный параметр, то можно оценить его значение, проводя эксперимент в действительности. Второе применение относится к «обратным задачам» теории вероятностей и лежит в основе так называемого метода Монте-Карло. Вот числовой пример. Известно, что в реально выполненных 10 000 бросаниях монеты выпало 4979 «орлов», подробнее см. [9, с. 21]. Приведенные выше вычисления показывают, что у нас есть 95 %-я уверенность в том, что

$$\left| p - \frac{4979}{10\,000} \right| \leqslant \frac{2}{100} \quad \text{или} \quad 0.4779 \leqslant p \leqslant 0.5179.$$

Возвращаясь к общей ситуации в теореме 10, заметим, что закон больших чисел в указанном виде напоминает определение обычного предела. Для любого $\varepsilon > 0$ найдется $n_0(\varepsilon)$ такое, что для всех $n \geqslant n_0(\varepsilon)$

$$P\left(\left|\frac{S_n}{n} - m\right| < \varepsilon\right) > 1 - \varepsilon. \tag{7.6.16}$$

Без ограничения общности можно считать, что и c, и δ в неравенстве (7.6.4) равны ε (см. задачу 22). Если интерпретировать это неравенство в духе предыдущего примера как утверждение о близости теоретического среднего m и эмпирического среднего S_n/n, то двойная погрешность (неточность), выражающаяся в присутствии двух ε в неравенстве (7.6.16), представляется неизбежной. Причина в том, что в эксперименте нельзя быть ни на 100 % уверенным, ни на 100 % точным, в противном случае он не был бы случайным. Несмотря на это, математикам свойственен идеализм и стремление к совершенству. То, что не встречается в реальном мире, может быть достигнуто в чисто математической схеме. Такую возможность открыл Э. Борель в 1909 г., создавший новую главу в теории вероятностей благодаря своему результату, представленному ниже. Для бернуллиевского случая его знаменитая теорема формулируется так:

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1^{*)}. \tag{7.6.17}$$

Она известна как «усиленный закон больших чисел», который представляет собой существенное обобщение «слабого закона больших чисел», установленного Бернулли. В ней утверждается существование предела

---

*) Обсуждение теоремы Бореля и связанных с ней тем см. в гл. 5 книги [4].

для частоты, равного теоретической вероятности p, для всех точек ω выборочного пространства за исключением множества, имеющего вероятность нуль (но не обязательно пустого множества). Таким образом, предел в формуле (2.1.10) на самом деле существует, но только *для почти всех* ω. Тем самым, эмпирическая теория частот, любимая прикладными учеными, получила свое обоснование с помощью мудреной теоремы. Различие между результатом Бореля и более слабой теоремой Бернулли:

$$\forall\,\varepsilon > 0\colon \quad \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1,$$

достаточно тонкое, и его нельзя строго объяснить без теории меры. Проницательный читатель может заметить, что, хотя в равенстве (7.6.17) утверждается достижение 100 %-й точности со 100 %-й уверенностью, предел частоты невозможно найти эмпирическим путем. Поэтому циник мог бы сказать, что то, в чем мы уверены — всего лишь *идеал*, на что софист возразит, что нам никто не запрещает неограниченно продолжать эксперимент! Даже если так, то следует иметь ввиду, что вероятностная уверенность отличается от абсолютной уверенности в детерминистическом понимании. Приведем аналогию для данного различия, связанную со вторым законом термодинамики из статистической механики. Согласно этому закону, например, если горячее тело находится в контакте с холодным, то *логически* возможно, что тепло будет перетекать от холодного тела к горячему, но вероятность того, что это случится, равна нулю. Подобные исключения допустимы и в теореме Бореля. К примеру, если монета бросается бесконечное число раз, то логически возможно, что всегда будут выпадать «орлы». Такое событие относится к исключениям из формулы (7.6.17), но его вероятность есть $\lim\limits_{n\to\infty} p^n = 0$.

Усиленный закон больших чисел лежит в основании математической теории вероятностей, построенной на понятии частоты; см. § 2.1. Он более понятен, чем слабый закон, и совершенно необходим для некоторых теоретических построений. (В статистической механике его обобщенная форма носит имя *эргодической теоремы.*) Но закоренелый практик, а также представитель радикальной логической школы, называемой интуиционистской, могут считать его идеалистической фикцией. Любопытно процитировать высказывания двух известных авторов.

Феллер:

> «[слабый закон больших чисел] вызывает очень ограниченный интерес, его следует заменить на более точный и полезный усиленный закон больших чисел» [9, с. 52].

Ван дер Варден:

> «[усиленный закон больших чисел] едва ли играет какую-нибудь роль в математической статистике» (с. 98 в *Mathematische Statistik*, 3-е изд., Springer-Verlag, 1971).

Завершим нашу дискуссию, помня о существовании разрыва между наблюдаемыми в реальном мире явлениями и теоретическими моделями, применяемыми для их изучения; см. замечание Эйнштейна, приведенное перед примером 9 из § 5.2. Законы больших чисел, слабый и усиленный, — математические теоремы, выведенные из аксиом. Их применимость к реальным опытам, таким как бросание монетки, неизбежно ограничена и неполна. Разнообразные примеры, приведенные выше, следует воспринимать с учетом этого обстоятельства.

Задачи

1. Допустим, что на 300 страницах книги встречаются 200 опечаток. Используйте пуассоновскую аппроксимацию для вычисления вероятности того, что на определенной странице окажется более одной опечатки.

2. В некоторой школе 4 % учеников пишут левой рукой. Какова вероятность того, что в классе из 25 учеников не будет ни одного левши?

3. Шесть игральных костей бросали 200 раз. Оцените вероятность получения «шести разных граней» ровно k раз, где $k = 0, 1, 2, 3, 4, 5$.

4. В домашней пекарне выпекаются 100 булок с изюмом, для чего используются 2 000 изюмин. Выпишите вероятность того, что в купленной булке окажется от 20 до 30 изюмин.

5. Согласно оценкам, на некотором острове площадью 15 квадратных миль обитают 20 гигантских черепах одного вида и 30 черепах другого. Экологическая экспедиция обнаружила 2 черепахи на площади в 1 квадратную милю, но не зафиксировала их вид. Примените пуассоновское распределение для вычисления вероятностей разных возможностей.

6. Найдите максимальную вероятность (или вероятности) биномиального распределения $B_k(n; p), 0 \leqslant k \leqslant n$. Покажите, что вероятности возрастают вплоть до максимума, а затем — убывают. [*Указание.* Изучите отношения соседних членов распределения.]

7. Найдите максимальную вероятность распределения Пуассона $\pi_k(\alpha)$, $0 \leqslant k < \infty$. Убедитесь в таком же поведении вероятностей, как в задаче 6.

8. Пусть X — случайная величина такая, что $P(X = c + kh) = \pi_k(\alpha)$, где c — действительное число, а h — положительное число. Найдите преобразование Лапласа для X.

9. Найдите свертку двух пуассоновских распределений $\{\pi_k(\alpha)\}$ и $\{\pi_k(\beta)\}$.

10*. Докажите, что если случайная величина X_α имеет распределение Пуассона $\pi(\alpha)$, то

$$\lim_{\alpha\to\infty} P\left\{\frac{X_\alpha-\alpha}{\sqrt{\alpha}}\leqslant u\right\}=\Phi(u)$$

для каждого u. [*Указание.* Покажите, что при $\alpha\to\infty$ преобразование Лапласа $E(e^{-\lambda(X_\alpha-\alpha)/\sqrt{\alpha}})$ сходится к $e^{\lambda^2/2}$, и используйте аналог теоремы 9 из § 7.5.]

11. Допустим, что расстояния между автомобилями, движущимися в одном направлении по некоторому шоссе, экспоненциально распределены со средним значением 100 метров. Какова вероятность того, что на отрезке шоссе длиной в 5 километров находятся от 50 до 60 автомобилей?

12. Предположим, что на некотором шоссе поток автомобилей можно считать пуассоновским с интенсивностью 30 машин в минуту. Выпишите вероятность того, что пройдет более N секунд, пока мимо поста наблюдения проедут n автомобилей. [*Указание.* Используйте формулу (7.2.11).]

13. Идеальная игральная кость бросается 100 раз. Найдите вероятность того, что сумма всех выпавших номеров окажется в пределах от 330 до 380.

14. Требуется найти вероятность p того, что канцелярская кнопка при подбрасывании упадет плоской стороной вниз. Сколько испытаний понадобится для того, чтобы приобрести 95 %-ю уверенность в том, что наблюдаемая относительная частота не отклоняется от p более, чем на $p/10$? [*Указание.* Подбросьте кнопку несколько раз, чтобы получить грубые границы для p.]

15. Два кинотеатра вмещают 1 000 зрителей. Допустим, что каждый зритель выбирает один из кинотеатров случайно и независимо от других зрителей. Сколько мест необходимо иметь в каждом кинотеатре, чтобы вероятность того, что какому-то зрителю не удастся попасть на сеанс из-за отсутствия мест, была не более 1 %?

16. Проводится опрос большого числа избирателей с целью определения для некоторого кандидата процента предпочтения. Допустим, что собирающиеся голосовать за него составляют неизвестную долю p от всех избирателей, и опрашиваемые отвечают независимо друг от друга. Сколько людей необходимо опросить, чтобы предсказать, что $p < 4.5\,\%$ с уверенностью 95 %? (Это — так называемая четырехпроцентная погрешность в предсказании результатов голосования, с учетом того, что число 0.045 считается $\leqslant 0.04$ ввиду правила округления десятичных дробей.)

17. Используем обозначение $\Phi((a,b))$ для разности

$$\Phi(b)-\Phi(a),$$

где $a<b$ и Φ — функция стандартного нормального распределения. Покажите, что $\Phi((0,2))>\Phi((1,3))$ и обобщите соответствующее утверждение на случай любых двух интервалов одинаковой длины. [*Указание.* Функция $e^{-x^2/2}$ убывает, когда $|x|$ возрастает.]

18. Докажите формулу (7.1.8) и затем используйте тот же метод для получения равенства (7.1.12). [*Указание.*

$$|\ln(1-x)+x| \leqslant \frac{1}{2}\sum_{n=2}^{\infty}|x|^n = \frac{x^2}{2(1-|x|)}\,.$$

Поэтому при $|x| \leqslant 1/2$ правая часть не превосходит x^2.]

19. Проверьте справедливость соотношения (7.1.13).

20. Докажите неравенство Чебышёва в случае, когда случайная величина X имеет плотность. [*Указание.*

$$\sigma^2(X) = \int\limits_{-\infty}^{\infty}(x-m)^2 f(x)\,dx \geqslant \int\limits_{|x-m|>c}(x-m)^2 f(x)\,dx.]$$

21. Выведите следующий аналог неравенства Чебышёва, в котором вместо второго момента используется первый абсолютный момент:

$$P(|X-m|>c) \leqslant \frac{1}{c}\,E(|X-m|).$$

22*. Покажите, что $\lim\limits_{n\to\infty} P(|X_n|>\varepsilon) = 0$ для любого ε тогда и только тогда, когда для заданного ε существует число $n_0(\varepsilon)$ такое, что $P(|X_n|>\varepsilon)<\varepsilon$ для $n>n_0(\varepsilon)$. Это также эквивалентно тому, что для заданных δ и ε найдется $n_0(\delta,\varepsilon)$ такое, что $P(|X_n|>\varepsilon)<\delta$ для $n>n_0(\delta,\varepsilon)$. [*Указание.* Возьмите $\varepsilon' = \delta\wedge\varepsilon$ и примените первое утверждение.]

23. Пусть случайная величина X имеет распределение Φ. Покажите, что $|X|$ имеет распределение Ψ, где $\Psi = 2\Phi - 1$*). Распределение Ψ называют «положительным нормальным распределением».

24. Пусть X имеет распределение Φ. Найдите плотность распределения случайной величины X^2 и соответствующую ей функцию распределения. В статистике оно известно под именем «распределения хи-квадрат»**).

[*Указание.* Продифференцируйте $P(X^2<x) = 2/\sqrt{2\pi}\int\limits_0^{\sqrt{x}} e^{-u^2/2}\,du$.]

25*. Используйте задачу 24 для доказательства равенства

$$\int\limits_0^{\infty} x^{-1/2}e^{-x}\,dx = \sqrt{\pi}.$$

Этот интеграл равен $\Gamma(1/2)$, где Γ — *гамма-функция*, определяемая формулой

$$\Gamma(\alpha) = \int\limits_0^{\infty} x^{\alpha-1}e^{-x}\,dx$$

для $\alpha>0$. [*Указание.* Рассмотрите $E(X^2)$ с учетом задачи 24.]

---

*) При $x>0$ и $\Psi(x)=0$ при $x\leqslant 0$. — *Прим. перев.*

**) С одной степенью свободы. — *Прим. перев.*

26*. Пусть $\{\xi_i, 1 \leqslant i \leqslant n\}$ — набор из n случайных величин, удовлетворяющих условию $0 < \xi_1 \leqslant \xi_2 \leqslant \ldots \leqslant \xi_n \leqslant t$. Пусть $(0,t] = \cup_{k=1}^{l} I_k$ — произвольное разбиение промежутка $(0,t]$ на подынтервалы $I_k = (x_{k-1}, x_k]$, где $x_0 = 0$; и пусть $\tilde{N}(I_k)$ обозначает количество случайных величин ξ_i, принадлежащих I_k. Как можно выразить событие $\{\xi_i \leqslant x_k;\ 1 \leqslant k \leqslant l\}$ через $\tilde{N}(I_k), 1 \leqslant k \leqslant l$? Здесь, конечно, $0 < x_1 < x_2 < \ldots < x_n \leqslant t$. Теперь допустите, что $x_k, 1 \leqslant k \leqslant l$, являются произвольными, и ответьте на вопрос снова. [*Указание.* Возьмите $n = 2$ и 3, чтобы понять, что происходит; переобозначьте x_k во второй части.]

27*. Пусть $\{X(t), t \geqslant 0\}$ — пуассоновский процесс с параметром α. Для фиксированного $t > 0$ определим $\delta(t)$ как расстояние от t до последнего скачка, предшествующего моменту t, если имеется хотя бы один скачок, и положим $\delta(t) = t$ в противном случае. Определим $\delta'(t)$ как расстояние от t до скачка, следующего за моментом t. Найдите распределения случайных величин $\delta(t)$ и $\delta'(t)$. [*Указание.* Если $u < t$, то справедливо равенство $P\{\delta(t) > u\} = P\{N(t-u, t) = 0\}$; и для всех $u > 0$ имеем $P\{\delta'(t) > u\} = P\{N(t, t+u) = 0\}$.]

28*. Пусть $\tau(t) = \delta(t) + \delta'(t)$, как в задаче 27. Это есть длина интервала между скачками, который накрывает момент t. Для каждого ω она является одной из случайных величин T_k, определенных в § 7.2. Обладает ли $\tau(t)$ таким же экспоненциальным распределением, как все T_k? (Это — красивый пример, где логика демонстрирует свое преимущество перед интуицией, и его часто называют парадоксом. Ответ несложно получить, опираясь на задачу 27. Дальнейшее обсуждение на несколько более продвинутом уровне, чем в данной книге, содержится в статье Chung «The Poisson process as renewal process,» *Periodica Mathematica Hungarica*, Vol. 2 (1972), pp. 41–48.)

29. Используйте неравенство Чебышёва, чтобы показать, что если X и Y — две произвольные случайные величины, удовлетворяющие условию $E\{(X-Y)^2\} = 0$, то тогда $P(X = Y) = 1$. Другими словами, X и Y совпадают *почти наверное*. [*Указание.* $P(|X - Y| > \varepsilon) = 0$ для любого $\varepsilon > 0$.]

30. Вспомните определение коэффициента корреляции $\rho(X,Y)$ из § 6.3. Покажите, что если $\rho(X,Y) = 1$, то две «стандартизованные» случайные величины
$$\tilde{X} = \frac{X - E(X)}{\sigma(X)}, \quad \tilde{Y} = \frac{Y - E(Y)}{\sigma(Y)}$$
совпадают почти наверное. А что происходит в случае, когда $\rho(X,Y) = -1$? [*Указание.* Вычислите $E\{(\tilde{X} - \tilde{Y})^2\}$ и используйте задачу 29.]

Приложение 2

Формула Стирлинга и теорема Муавра—Лапласа

В данном дополнении приводятся некоторые детали доказательства теоремы 5, касающейся формулы Стирлинга (7.3.3), а также обсуждается ее связь с нормальным интегралом (7.4.1). Начнем с получения оценки.

Лемма. *Если* $|x| \leqslant 2/3$, *то*

$$\ln(1+x) = x - \frac{x^2}{2} + \theta(x), \quad \text{где} \quad |\theta(x)| \leqslant |x|^3.$$

Доказательство. Раскладывая в ряд Тейлора функцию $\ln(1+x)$, имеем:

$$\ln(1+x) = x - \frac{x^2}{2} + \sum_{n=3}^{\infty} (-1)^{n-1} \frac{x^n}{n}.$$

Поэтому $\theta(x)$ совпадает с приведенным выше рядом и

$$\theta(x) \leqslant \sum_{n=3}^{\infty} \frac{|x|^n}{n} \leqslant \frac{1}{3} \sum_{n=3}^{\infty} |x|^n = \frac{|x|^3}{3(1-|x|)}.$$

Для $|x| \leqslant 2/3$ функция $3(1-|x|) \geqslant 1$. Отсюда вытекает утверждение леммы. Выбор константы $2/3$ есть вопрос удобства; аналогичная оценка верна для любой константы, меньшей 1.

Применим лемму, прежде всего, для заполнения пропусков в доказательстве теоремы 5: покажем, что отброшенными членами в разложениях из (7.3.17) действительно можно пренебречь при $n \to \infty$. Когда n достаточно велико, обе величины из неравенств (7.3.17$'$) будут $\leqslant 2/3$. Следовательно, применима лемма, и вклад от «хвостов» двух рядов, представленных там многоточиями, ограничен сверху величиной

$$k \left| \frac{\sqrt{npq}\, x_k}{k} \right|^3 + (n-k) \left| \frac{\sqrt{npq}\, x_k}{n-k} \right|^3.$$

Поскольку $pq < 1$ и $|x_k| \leqslant A$, то она не превосходит

$$\frac{n^{3/2}}{k^2} A^3 + \frac{n^{3/2}}{(n-k)^2} A^3,$$

которая, очевидно, стремится к нулю при $n \to \infty$ в силу эквивалентностей (7.3.15). Тем самым, в пределе «хвосты» исчезают, и мы приходим к формуле (7.3.18), как там и утверждалось.

Докажем следующее соотношение, являющееся решающим шагом к формуле Стирлинга:

$$\lim_{n\to\infty}\left\{\ln n! - \left(n+\frac{1}{2}\right)\ln n + n\right\} = C, \tag{П.2.1}$$

где C — константа, которую мы вычислим позже. Пусть d_n обозначает величину, заключенную в скобки в равенстве (П.2.1). Тогда, проводя простые вычисления, получаем:

$$d_n - d_{n+1} = \left(n+\frac{1}{2}\right)\ln\left(1+\frac{1}{n}\right) - 1.$$

Используя обозначения из леммы, запишем это в виде

$$\left(n+\frac{1}{2}\right)\left(\frac{1}{n} - \frac{1}{2n^2} + \theta\left(\frac{1}{n}\right)\right) - 1 = \left(n+\frac{1}{2}\right)\theta\left(\frac{1}{n}\right) - \frac{1}{4n^2}\,.$$

Отсюда, применяя лемму с $x = 1/n, n \geqslant 2$, выводим неравенство:

$$|d_n - d_{n+1}| \leqslant \left(n+\frac{1}{2}\right)\frac{1}{n^3} + \frac{1}{4n^2} = \frac{2n+1}{2n^3} + \frac{1}{4n^2}\,.$$

Таким образом, ряд $\sum\limits_n |d_n - d_{n+1}|$ сходится в силу признака сравнения. Теперь вспомним, что абсолютно сходящийся ряд обязательно сходится, что означает, что его частичные суммы стремятся к конечному пределу, равному, скажем, C_1. Итак,

$$\lim_{N\to\infty}\sum_{n=1}^{N}(d_n - d_{n+1}) = C_1;$$

но стоящая под знаком предела сумма сворачивается в $d_1 - d_{N+1}$, поэтому

$$\lim_{N\to\infty} d_{N+1} = d_1 - C_1,$$

и мы доказали утверждение (П.2.1) для $C = d_1 - C_1$. Из него следует, что

$$\lim_{n\to\infty}\frac{n!\,e^n}{n^{n+(1/2)}} = e^C,$$

или, положив $K = e^C$,

$$n! \sim Kn^{n+(1/2)}e^{-n}. \tag{П.2.2}$$

Если сравнить эту формулу с соотношением (7.3.3), то увидим, что для получения формулы Стирлинга осталось только доказать, что $K = \sqrt{2\pi}$. Отметим, что даже без нахождения константы K выкладки в теоремах 5 и 6 из § 7.3 будут законными, если всюду заменить $\sqrt{2\pi}$ на K. В частности, формула (7.3.19) при $a = -b$ приобретет вид

$$\lim_{n\to\infty} P\left(\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant b\right) = \frac{1}{K}\int\limits_{-b}^{b} e^{-x^2/2}\,dx. \tag{П.2.3}$$

С другой стороны, можно применить теорему 11 (неравенство Чебышёва) с $X = (S_n - np)/\sqrt{npq}$, $E(X) = 0$, $E(X^2) = 1$ и получить неравенство

$$P\left(\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant b\right) \geqslant 1 - \frac{1}{b^2}\,. \tag{П.2.4}$$

Объединяя два последних результата и учитывая, что вероятность не может превосходить 1, выводим:

$$1 - \frac{1}{b^2} \leqslant \frac{1}{K}\int\limits_{-b}^{b} e^{-x^2/2}\,dx \leqslant 1.$$

Устремляя $b \to \infty$, заключаем, что

$$K = \int\limits_{-\infty}^{\infty} e^{-x^2/2}\,dx. \tag{П.2.5}$$

Поскольку согласно (7.4.1) данный интеграл равен $\sqrt{2\pi}$, тем самым мы доказали, что $K = \sqrt{2\pi}$.

Другой способ вычисления константы K основан на формуле Валлиса, содержащейся во многих учебниках по математическому анализу (см., например, Courant-John, *Introduction to Calculus and Analysis*, Vol. 1, New York: Interscience Publishers, 1965). Если это сделано, то приведенное выше рассуждение дает представление (П.2.5) с $K = \sqrt{2\pi}$, отсюда вытекает формула (7.4.1) для нормального интеграла. Итак, мы придали строгость эвристическому доказательству, как и обещали ранее в абзаце, следующем за формулой (7.4.1), и продемонстрировали близкую связь между двумя результатами, вынесенными в заголовок данного дополнения.

Глава 8

От случайных блужданий к цепям Маркова

8.1. Задача о бродяге и задача о разорении игрока

Простейшее *случайное блуждание* можно описать следующим образом. Частица движется вдоль прямой скачками; каждый скачок (шаг) производится на единицу вправо или влево с вероятностями p и $q = 1 - p$ соответственно, где $0 < p < 1$. Для удобства предположим, что каждый шаг выполняется за единицу времени, поэтому n-й шаг осуществляется в момент n. Кроме того, мы будем считать, что блуждающая частица может оказаться в любой точке с целочисленной координатой на оси. Это множество часто называют «целочисленной решеткой» в $\mathbb{R}^1 = (-\infty, \infty)$ и обозначают через I. Таким образом, частица блуждает вперед и назад по решетке, причем неограниченно долго. Если изобразить график ее текущего положения X_n как функции от времени n, то *траектория* будет иметь вид ломаной, похожей на те, что приведены на рис. 30.

На более выразительном языке частица превращается в бродягу, а прямая — в бесконечную улицу, разделенную на кварталы. За каждую единицу времени, скажем, 5 минут, бродяга проходит один квартал от угла до угла, и на каждом из углов он с вероятностями p и q решает, продолжать ему двигаться вперед или повернуть и идти назад. Тем самым, он осуществляет случайное блуждание, и его следы, оставляемые на дороге, петляют и дублируются. Данное описание сразу приводит к более реалистичной модели, в которой рассматриваются регулярно расположенные горизонтальные и вертикальные улицы, как в Нью-Йорке. В таком случае каждый шаг может быть выполнен в одном из четырех направлений (см. рис. 31). Эта схема соответствует случайному блужданию по целочисленной решетке на плоскости $\mathbb{R}^2$. Мы будем время от времени возвращаться к двумерной модели, но в основном наше обсуждение будет ограничено простейшей одномерной ситуацией.

Теперь дадим математическую формулировку. Пусть ξ_n — это n-й шаг или *смещение*, причем

$$\xi_n = \begin{cases} +1 & \text{с вероятностью } p, \\ -1 & \text{с вероятностью } q; \end{cases} \tag{8.1.1}$$

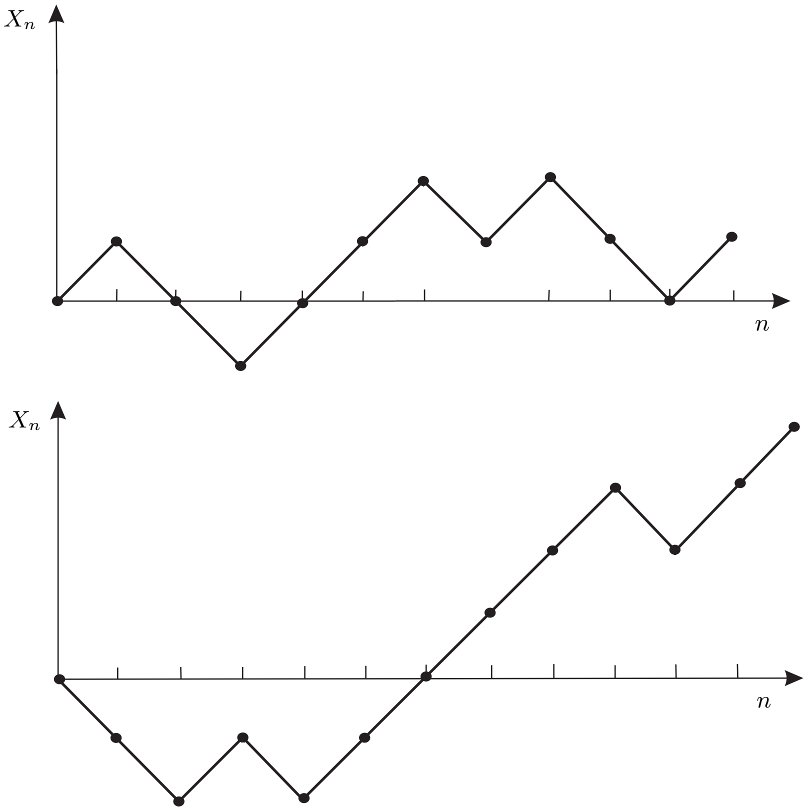


Рис. 30

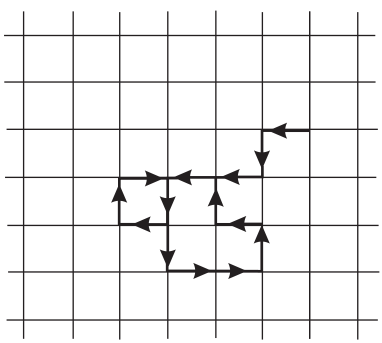


Рис. 31

$\{\xi_n\}$ — независимые случайные величины. Если обозначить начальное положение частицы через X_0, то положение частицы в момент n (т. е. после n шагов) задается формулой

$$X_n = X_0 + \xi_1 + \ldots + \xi_n. \tag{8.1.2}$$

Таким образом, случайное блуждание представляется последовательностью случайных величин $\{X_n, n \geqslant 0\}$, образующих случайный процесс с дискретным временем. На самом деле $X_n - X_0$ — это сумма независимых бернуллиевских случайных величин, которым уделено много внимания в гл. 5, 6 и 7. Мы изменили наши прежние обозначения (см., например, формулу (6.3.11)) на новые (8.1.2) ради удобства дальнейшего использования в § 8.3. Но помимо этого, что здесь нового?

Новой является сама точка зрения. Мы собираемся изучать блуждание целиком, его течение или развитие во времени. Другими словами, каждая траектория частицы или бродяги будет восприниматься как возможное развитие процесса, управляемого вероятностными законами, которым подчиняется движение. До сих пор нас интересовали в основном определенные вероятностные характеристики величины X_n (раньше обозначаемой через S_n), такие, как ее среднее, дисперсия и распределение. Теперь же мы собираемся глубже изучить устройство *последовательности* $\{X_n, n \geqslant 0\}$, задавая вопросы, касающиеся совместного поведения этих величин. Вот некоторые примеры таких вопросов. Попадет ли когда-нибудь движущаяся частица в некоторую заданную точку? Если попадет, то как скоро? Произойдет ли это прежде или после того, как частица посетит некоторые другие точки? Можно также спросить, как часто частица будет попадать в определенную точку или множество; как долго будет она оставаться внутри некоторого множества и т. п. Какие-то из перечисленных вопросов мы уточним ниже и получим на них ответы. Рекомендуем читателю самостоятельно придумать еще несколько подобных вопросов и, возможно, связать их с конкретными моделями, имеющими практическое значение.

Начнем со следующей задачи.

Задача 1. Рассмотрим отрезок $[0, c]$, где $c = a + b$ и $a \geqslant 1$, $b \geqslant 1$. Если частица стартует из точки «a», то какова вероятность того, что она попадет в одну из концевых точек отрезка прежде, чем в другую?

Это — переформулировка известной задачи, обсуждавшейся П. Ферма и Б. Паскалем и решенной в общем виде Монмаром. Два игрока — Петр и Павел — играют серию партий, в каждой из которых Петр побеждает с вероятностью p, а Павел — с вероятностью q. Результаты отдельных партий предполагаются независимыми. К примеру, они могут бросать монету несколько раз или играть в настольный теннис, в котором их уровни мастерства соотносятся как p и q. Проигравший очередную партию платит каждый раз доллар победителю. Если перед началом игры Петр имеет $\$a$, а Павел имеет $\$b$, и они собираются играть до тех пор, пока один из них не разорится (окажется банкротом), то какова вероятность того, что разорится именно Петр?

В такой формулировке положение стартующей из точки a частицы в момент n совпадает с количеством долларов, которые имеются у Петра после n партий. Каждый шаг направо соответствует выигрышу им \$1, а налево — проигрышу \$1. Если частица достигает нуля прежде, чем попадает в точку c, то Петр проигрывает весь свой первоначальный капитал и разоряется. В свою очередь, если частица достигает точки c прежде, чем попадает в 0, то Павел проигрывает весь свой капитал и разоряется. Игра прекращается, как только происходит одно из этих двух событий. Отсюда происходит историческое название — «задача о разорении игрока».

Приведем ее решение. В нем важную роль играет следующее обозначение: для $1 \leqslant j \leqslant c-1$ положим

$$u_j = \text{вероятность того, что частица, выходящая из } j, \text{ попадет в 0 прежде, чем достигнет } c. \tag{8.1.3}$$

Задача состоит в том, чтобы найти u_a, но так как «a» — произвольное число, то нам в действительности нужны все u_j. Идея решения — вывести соотношения, связывающие эти величины, и в результате вычислить их все сразу. Такие соотношения представляют собой следующий набор *разностных уравнений*:

$$u_j = pu_{j+1} + qu_{j-1}, \quad 1 \leqslant j \leqslant c-1, \tag{8.1.4}$$

совместно с *граничными условиями*:

$$u_0 = 1, \quad u_c = 0. \tag{8.1.5}$$

Чтобы объяснить соотношение (8.1.4), представим, что может произойти, когда частица, находящаяся в точке j, совершает один шаг. С вероятностью p она окажется в точке $j+1$, и при этом предположении (условная) вероятность того, что она попадет в 0 прежде, чем в c, равняется u_{j+1}. Аналогично, с вероятностью q она окажется в точке $j-1$, и при этом предположении оговоренная вероятность равняется u_{j-1}. Тогда *полная* вероятность u_j равна сумме двух членов из правой части равенства (8.1.4) в силу утверждения 2 из § 5.2. Такие же рассуждения, проведенные для крайних случаев $j = 1$ и $j = c - 1$, дают указанные в условиях (8.1.5) значения для u_0 и u_c. Они не содержатся среди величин (8.1.3) и, строго говоря, *не совсем* укладываются в приведенные там словесные описания, хотя и получаются из них в результате своего рода экстраполяции.

Оставшаяся работа состоит из чисто алгебраических преобразований. Так как $p + q = 1$, то можно записать левую часть соотношения (8.1.4) в форме $pu_j + qu_j$. После переноса членов равенство приобретает вид:

$$q(u_j - u_{j-1}) = p(u_{j+1} - u_j).$$

Используя обозначения

$$r = \frac{q}{p}, \quad d_j = u_j - u_{j+1},$$

получаем следующее основное рекуррентное соотношение между последовательными *разностями*:

$$d_j = rd_{j-1}. \tag{8.1.6}$$

Итерируя, выводим равенство $d_j = r^j d_0$. Суммируя и последовательно сокращая, находим, что

$$1 = u_0 - u_c = \sum_{j=0}^{c-1}(u_j - u_{j+1}) = \sum_{j=0}^{c-1} d_j = \sum_{j=0}^{c-1} r^j d_0 = \frac{1 - r^c}{1 - r}\, d_0, \tag{8.1.7}$$

при условии, что $r \neq 1$. Далее, аналогично выводим:

$$u_j = u_j - u_c = \sum_{i=j}^{c-1}(u_i - u_{i+1}) = \sum_{i=j}^{c-1} d_j = \sum_{i=j}^{c-1} r^i\, d_0 = \frac{r^j - r^c}{1 - r}\, d_0. \tag{8.1.8}$$

Итак,

$$u_j = \frac{r^j - r^c}{1 - r^c}, \quad 0 \leqslant j \leqslant c. \tag{8.1.9}$$

В случае $r = 1$, отбрасывая крайние правые члены формул (8.1.7) и (8.1.8), получаем, что

$$1 = cd_0, \quad u_j = (c - j)\, d_0 = \frac{c - j}{c}, \quad u_a = \frac{b}{c}. \tag{8.1.10}$$

Итак, половина задачи 1 решена; осталось найти

v_j = вероятность того, что выходящая из точки j частица достигнет c прежде, чем попадет в 0.

Точно такие же рассуждения показывают, что соотношения (8.1.4) остаются верными при замене u_j на v_j и одновременной перестановке граничных условий (8.1.5): $v_0 = 0$, $v_c = 1$. Отсюда можно найти все v_j использованным выше методом. Мы рекомендуем читателю сделать это в качестве превосходного упражнения. Однако существуют и более короткие пути, не требующие подобных усилий.

Один из них, вероятно, легче понять, если рассуждать в игровых терминах. Если заменить p на q (т. е. r на $1/r$) и одновременно заменить j на $c-j$ (когда капитал Петра равен $\$j$, то Павел имеет капитал $\$(c-j)$ и наоборот), то игроки поменяются ролями. При этом u_j перейдет в v_j (но не в v_{c-j}, почему?). Проводя данные изменения в формулах (8.1.9) и (8.1.10), получим, что

$$v_j = \begin{cases} \dfrac{1-r^j}{1-r^c}, & \text{если } p \neq q, \\[2ex] \dfrac{j}{c}, & \text{если } p = q. \end{cases}$$

Теперь приятно заметить, что в обоих случаях выполняется равенство

$$u_j + v_j = 1, \quad 0 \leqslant j \leqslant c. \tag{8.1.11}$$

Тем самым, в качестве побочного результата, мы решили следующую задачу, которая, быть может, приходила вам на ум по ходу предшествующей дискуссии.

Задача 2. Если частица стартует из некоторой внутренней точки отрезка $[0, c]$, то чему равна вероятность того, что когда-нибудь она достигнет границы?

Так как граница состоит из двух концевых точек: 0 и c, то ответ дается равенством (8.1.11), т. е. искомая вероятность равна 1. В терминологии игроков это означает, что один из них обязательно разорится раньше или позже, если игра не имеет временных ограничений. Другими словами, игра не может продолжаться бесконечно. Вы возразите, что ничто не помешает выигрышам Петра и Павла чередоваться сколь угодно долго: такой пример дает последовательность $+1-1+1-1+1-1\ldots$. Объяснение заключается в том, что хотя это событие *логически допустимо*, но вероятность его равна нулю, как мы уже убедились выше. Иначе говоря, оно *почти никогда* не происходит в смысле, рассмотренном в конце § 7.6, и это — все, что мы утверждаем.

Следует отметить, что задача 2 может быть решена без ссылки на задачу 1. В самом деле, понятно, что поставленный в ней вопрос является более широким («качественным»), и ответ на него не обязан основываться на специфических числовых результатах, полученных при решении задачи 1. Несложно показать, что даже если величины ξ_i в сумме (8.1.2) заменить на любые (не равные тождественно нулю) независимые и одинаково распределенные случайные величины (при этом мы получим обобщенное случайное блуждание с шагами произвольной длины), то ответ к задаче 2 будет тем же самым в более широком смысле: частица раньше или позже непременно покинет любой конечный интервал (см., например, [4, теорема 9.2.3]). В нашем частном случае, когда шаги равны ± 1, частица, прежде чем выйти из отрезка $[0, c]$, обязательно должна пройти через одну из концевых точек. Если приведенное выше утверждение, равносильное соотношению (8.1.11), принять на веру (или доказать), то, конечно, мы получим равенство $v_j = 1 - u_j$ без каких-либо добавочных вычислений.

Сформулируем ответ к задаче 2 в следующем виде.

Теорема 1. *При любом случайном блуждании (с произвольным p) частица почти наверное**) *когда-нибудь выходит из любого конечного интервала.*

В качестве следствия мы можем определить случайную величину, обозначающую время ожидания достижения частицей границы. Ее иногда называют *временем до поглощения*, если граничные точки рассматриваются как *поглощающие барьеры*: считается, что частица «прилипает» к ним, как только в них попадает. В терминологии игроков она фигурирует под именем «*продолжительность игры*». Давайте при $1 \leqslant j \leqslant c-1$ положим

$$S_j = \text{первый момент попадания частицы, вышедшей из } j\text{, в точку } 0 \text{ или } c \tag{8.1.12}$$

и обозначим через e_j математическое ожидание $E(S_j)$. Ответ к задаче 2 гарантирует, что время S_j почти наверное конечно, поэтому его можно считать случайной величиной, принимающей положительные значения. (Если бы S_j могла принимать бесконечные значения, то она не являлась бы случайной величиной в смысле определения из § 4.2, поскольку «$+\infty$» — это не число. Однако не все так просто в отношении выборочного пространства, на котором определена S_j, — оно несчетно!) Различные e_j удовлетворяют соотношениям, похожим на соотношения между u_j:

$$\begin{aligned} &e_j = pe_{j+1} + qe_{j-1} + 1, \quad 1 \leqslant j \leqslant c-1, \\ &e_0 = 0, \quad e_c = 0. \end{aligned} \tag{8.1.13}$$

*) В общем случае «почти наверное» означает «с вероятностью 1».

Обоснование выписанных соотношений похоже на обоснование формул (8.1.4) и (8.1.5) за тем исключением, что надо еще объяснить происхождение константы 1 в правой части первого из уравнений. Она появляется из-за того, что на выполнение одного шага из j в $j\pm1$ расходуется единица времени. В равенствах (8.1.13) молчаливо предполагается, что все e_j конечны. Строго доказать это довольно непросто (см. задачу 48 в конце главы).

Решение уравнений (8.1.13) может быть получено либо так же, как прежде, либо более быстрым путем с применением стандартного метода решения разностных уравнений, детально рассматриваемого в задаче 13. Ввиду того, что общее решение выглядит ненамного сложнее, разберем только случай $p = q = 1/2$, который нам понадобится в дальнейшем. Пусть $f_j = e_j - e_{j+1}$. Тогда

$$f_j = f_{j-1} + 2, \quad f_j = f_0 + 2j,$$
$$0 = \sum_{j=0}^{c-1} f_j = c(f_0 + c - 1).$$

Отсюда $f_0 = 1 - c$. Проводя простые вычисления, находим:

$$e_j = \sum_{i=j}^{c-1} f_i = \sum_{i=j}^{c-1} (1 - c + 2i) = j(c - j). \tag{8.1.14}$$

Так как случайное блуждание симметрично, то ожидаемое время до поглощения будет одинаковым, если частица выходит из точки, расположенной на расстоянии j от 0 или от c (во втором случае — на расстоянии $c - j$ от 0). Поэтому заранее ясно, что $e_j = e_{c-j}$, что вполне согласуется с формулой (8.1.14).

8.2. Предельные схемы

Теперь мы готовы вывести важные следствия из предыдущих формул. Прежде всего, устремляя $c \to +\infty$, превратим отрезок $[0, c]$ в полупрямую $[0, \infty)$. Из соотношений (8.1.9) и (8.1.10) вытекает, что

$$\lim_{c\to\infty} u_j = \begin{cases} r^j, & \text{если } r < 1, \\ 1, & \text{если } r \geqslant 1. \end{cases} \tag{8.2.1}$$

На интуитивном уровне данный предельный результат имеет смысл вероятности, с которой частица, выходя из точки j, попадет в 0 прежде, чем «уйдет в $+\infty$». В задаче о разорении игрока это есть вероятность того, что Петр разорится в игре против «бесконечно богатого» Павла, которого невозможно разорить. Иными словами, результат (8.2.1) представляет собой вероятность того, что частица когда-нибудь дойдет из точки j до 0, или того, что Петр, имея в начале игры капитал $\$j$, когда-нибудь обанкротится. Эти интерпретации корректны и дают ответ на следующий вопрос, уточняющий задачу 2.

Задача 3. Если частица стартует из точки $a\,(\geqslant 1)$, то какова вероятность того, что она когда-нибудь окажется в 0?

Ответ таков: 1, если $p \leqslant q$, и $(q/p)^a$, если $p > q$. Заметим, что при $p \leqslant q$ частица движется влево не менее вероятно, чем вправо, поэтому первый вывод нетрудно предвидеть. На самом деле, привлекая усиленный закон больших чисел из § 7.6, в случае $p < q$ можно утверждать даже больше. Используя наши новые обозначения (8.1.2) и то, что $E(\xi_n) = p - q$, видим, что в текущем контексте соотношение (7.6.17) превращается в следующее утверждение: почти наверное

$$\lim_{n\to\infty} \frac{X_n - X_0}{n} = p - q < 0.$$

Оно намного сильнее, чем то, что $\lim\limits_{n\to\infty} X_n = -\infty$. Так как наша частица движется шагами единичной длины, она, уходя в $-\infty$, проходит через *все* точки, находящиеся слева от стартовой. В частности, почти наверное частица попадет из a в 0.

При $p > q$ выводы для игроков представляются довольно любопытными. Если Петр играет намного лучше Павла, то даже если у него есть всего $1, а Павла нельзя разорить, тем не менее он с вероятностью $1 - q/p$ может избежать собственного разорения. На самом деле можно показать, что при осуществлении данного счастливого события Петр выиграет сколь угодно много. Придадим последнему утверждению точный смысл: если X_n означает богатство Петра после n партий, то

$$P\{X_n \to +\infty \mid X_n \neq 0 \quad \text{для всех } n\} = 1.$$

(Это — условная вероятность при условии события $\{X_n \neq 0$ для всех $n\}$.) Является ли данное утверждение интуитивно очевидным? Теорема 1 поможет его объяснить, но, строго говоря, она его не охватывает.

Когда $p = q = 1/2$, приведенное выше рассуждение не проходит. Так как в данном случае есть симметрия между левым и правым направлениями, наше заключение можно выразить в следующей, более сильной, форме.

Теорема 2. *При симметричном случайном блуждании, начинающемся в произвольной точке, частица почти наверное будет попадать в любую точку сколько угодно раз.*

Доказательство. Пусть запись $i \Rightarrow j$ означает, что частица, стартуя из точки i, почти наверное попадает в точку j, где $i \in I$, $j \in I$. Мы уже доказали, что если $i \neq j$, то $i \Rightarrow j$. Поэтому также $j \Rightarrow i$. Но это влечет $j \Rightarrow j$ в силу очевидной диаграммы $j \Rightarrow i \Rightarrow j$. Поэтому также $i \Rightarrow j \Rightarrow j \Rightarrow j \Rightarrow j \ldots$. Тем самым, выходя из точки i, частица будет попадать в j столько раз, сколько мы захотим. Заметьте, что здесь можно взять $j = i$.

Короче говоря, частица *бесконечно часто* попадает во все точки диапазона блуждания I, и такое случайное блуждание называют *рекуррентным* (или *возвратным*). Эти понятия будут перенесены на цепи Маркова в § 8.4.

Для игроков теорема 2 имеет следующую интерпретацию. Если игра является честной, то Петр почти наверное сможет выиграть любую наперед заданную сумму денег при условии, что ему позволяется брать в долг также сколь угодно большую сумму. Таким образом, теорема 2 гарантирует только, что он когда-нибудь выиграет, скажем, \$1 000 000, но ничего не говорит о том, как много придется ему проиграть, пока не осуществится данная цель. Это — не слишком полезная информация, но вполне справедливая с точки зрения Павла! Более реалистический прогноз дает формула (8.1.10), которую можно записать в таком виде:

$$u_a = \frac{b}{a+b}, \quad v_a = \frac{a}{a+b}. \tag{8.2.2}$$

Она показывает, что шансы Петра выиграть желанную сумму b прежде, чем он потеряет свой начальный капитал a, пропорциональны величине этого капитала. Так, если у него есть \$100, то его шансы выиграть \$1 000 000 составляют 100/1 000 100 или примерно 1 к 10 000. Так обычно и обстоит дело при игре в казино, даже если представить, что казино откажется от своего гарантированного преимущества перед отдельными игроками.

Еще один неожиданный результат возникнет, если устремить $c \to +\infty$ в определении e_j. В пределе получим время ожидания момента, когда выходящая из точки j ($\geqslant 1$) частица впервые попадет в 0 (без всяких ограничений на то, как далеко она может уходить вправо от j). Мы видим, что в силу соотношения (8.1.14) данный предел равен бесконечности. Это означает, что даже если у Петра есть всего \$1, и он выступает против бесконечно богатого казино, он может рассчитывать играть очень и очень долго при условии, что игра честная. Данное утверждение звучит фантастически с точки зрения отдельного игрока, но следует понимать, что ввиду закона больших чисел практический смысл математического ожидания проявляется только применительно к «ансамблю». Общеизвестно, что каждый день много мелких игроков выходят из казино с карманами, полными денег. Им повезло — они избежали разорения потому, что у казино не хватило времени обыграть их (несмотря на существование гарантированной прибыли казино)!

Давайте познакомимся с еще одним достаточно оригинальным методом получения формулы (8.2.2). В случае $p = q$ имеем $E(\xi_n) = 0$ для каждого n. Следовательно, из равенства (8.1.2) получаем, что

$$E(X_n) = E(X_0) + E(\xi_1) + \ldots + E(\xi_n) = a. \tag{8.2.3}$$

В игровой терминологии это означает, что ожидаемый капитал Петра остается постоянным на всем протяжении игры при условии, что она честная. Теперь рассмотрим продолжительность игры S_a из формулы (8.1.12). Она представляет собой случайную величину, принимающую целые положительные значения. Учитывая, что равенство (8.2.3) выполняется для каждого целого положительного n, можно допустить, что оно останется справедливым, если вместо n в него подставить S_a. Вообще говоря, подобная подстановка — рискованное дело, но в данном случае она оказывается законной благодаря особому характеру величины S_a, а также самого процесса $\{X_n\}$. Мы не станем сейчас доказывать это (см. дополнение 3), а только используем. Очевидно, что случайная величина X_{S_a} принимает согласно своему определению только два значения: 0 и c. Пусть

$$P(X_{S_a} = 0) = \rho, \quad P(X_{S_a} = c) = 1 - \rho. \tag{8.2.4}$$

Тогда

$$E(X_{S_a}) = \rho \cdot 0 + (1 - \rho) \cdot c = (1 - \rho)c.$$

Отсюда и из равенства $E(X_{S_a}) = a$ находим, что

$$\rho = 1 - \frac{a}{c} = \frac{b}{a + b}\,,$$

что согласуется с соотношениями (8.2.2). Говоря кратко, наши рассуждения показывают, что игра остается справедливой вплоть до момента ее прекращения включительно. Понятно ли это интуитивно?

Теперь перейдем к описанию предельного перехода, ведущего от симметричного случайного блуждания к *броуновскому движению*. Английский ботаник Р. Броун в 1826 г. увидел, что микроскопические частицы, плавающие в жидкости, подвергаются непрестанным ударам молекул и совершают зигзагообразные движения. А. Эйнштейн и М. Смолуховский обнаружили, что, несмотря на их видимую иррегулярность, эти движения можно анализировать с помощью вероятностных законов; оказалось, что смещение частицы за некоторый временной период имеет нормальное распределение. Результат Эйнштейна (1906) заключался в выводе центральной предельной теоремы (см. § 7.4) методом дифференциальных уравнений. Изучение броуновского движения как стохастического процесса было предпринято Винером*) в 1923 г. Он опирался на эвристическую работу Башелье. Вскоре теория приобрела современный вид, благодаря трудам Поля Леви и его последователей. Вместе с пуассоновским процессом (§ 7.2) броуновское движение представляет одну из фундаментальных *разновидностей* случайных процессов, важных как для теории, так и для практики. Несмотря на то, что математический аппарат, используемый в данной книге, недостаточен для всестороннего изучения этой темы, все же возможно дать читателю некоторое представление о том, как броуновское движение возникает из случайного блуждания, и описать некоторые основные свойства этого процесса.

Движение частицы, за которой наблюдал Броун, происходило, конечно, в трехмерном пространстве, но мы станем рассматривать его проекции на координатные оси. Учитывая, что в течение одной секунды частица испытывает огромное количество толчков, уменьшим временную единицу. Но одновременно нам придется уменьшить единицу длины таким образом, чтобы это привело к корректной модели. Пусть δ — новая единица времени, другими словами, — время между последовательными столкновениями. Тогда за время t в прежней шкале частица совершает t/δ шагов. Каждый шаг по-прежнему симметричная бернуллиевская случайная величина, но сейчас мы предположим, что шаги имеют длину $\sqrt{\delta}$, а именно, пусть для всех k

$$P(\xi_k = \sqrt{\delta}) = P(\xi_k = -\sqrt{\delta}) = \frac{1}{2}\,.$$

Отсюда находим, что

$$E(\xi_k) = 0, \quad \sigma^2(\xi_k) = \frac{1}{2}\,(\sqrt{\delta})^2 + \frac{1}{2}\,(-\sqrt{\delta})^2 = \delta.$$

Пусть $X_0 = 0$. Тогда согласно формуле (8.1.2)

$$X_t = \sum_{k=1}^{t/\delta} \xi_k. \tag{8.2.5}$$

*) Норберт Винер (1894–1964) — известный американский математик, основоположник кибернетики.

Если δ намного меньше t, то отношение t/δ велико. Можно считать его целым числом. В силу теоремы 4 из § 6.3

$$E(X_t) = 0, \quad \sigma^2(X_t) = \frac{t}{\delta} \cdot \delta = t. \tag{8.2.6}$$

Более того, если t фиксировано и $\delta \to 0$, то по теореме Муавра—Лапласа (см. теорему 6 в § 7.3), случайная величина X_t будет иметь нормальное распределение $N(0, t)$. Это означает, что если частица совершает шаги $\pm\sqrt{\delta}$ с равными вероятностями за время δ, то переходя к пределу при $\delta \to 0$, мы получим невырожденную модель. Она и называется броуновским движением или *винеровским процессом.* Дадим ее формальное определение.

Определение броуновского движения. Семейство случайных величин $\{X(t)\}$, индексированное непрерывной переменной t, меняющейся на промежутке $[0, \infty)$, называется *броуновским движением* тогда и только тогда, когда выполняются следующие условия:

(i) $X(0) = 0$;

(ii) приращения $X(s_i + t_i) - X(s_i)$ на произвольных непересекающихся интервалах $(s_i, s_i + t_i)$ из конечного набора — независимые случайные величины;

(iii) для любых $s \geqslant 0$ и $t \geqslant 0$ случайная величина $X(s + t) - X(s)$ имеет нормальное распределение $N(0, t)$.

Для произвольной константы a процесс $\{X(t) + a\}$ называется *броуновским движением, выходящим из* точки a.

Мы уже убедились, что процесс, построенный выше с помощью предельного перехода от симметричного случайного блуждания, обладает свойством (iii). Свойство (ii) для него также выполняется ввиду того, что приращения на непересекающихся интервалах находятся суммированием смещений ξ_k в непересекающихся блоках; поэтому в силу замечания, приведенного после утверждения 6 в § 5.5, суммы являются независимыми.

Сравнивая данное определение с определением пуассоновского процесса из § 7.2, замечаем, что единственное отличие заключается в п. (iii). Использованный в § 7.2 способ построения пуассоновского процесса позволяет судить о поведении его траекторий (см. рис. 29). Но для броуновского движения картина не столь очевидна (см. рис. 32). Одним из важнейших открытий Винера является факт непрерывности почти всех его траекторий. Иначе говоря, для почти всех ω функция $t \to X(t, \omega)$ непрерывна по t на промежутке $[0, \infty)$. На практике можно не обра-

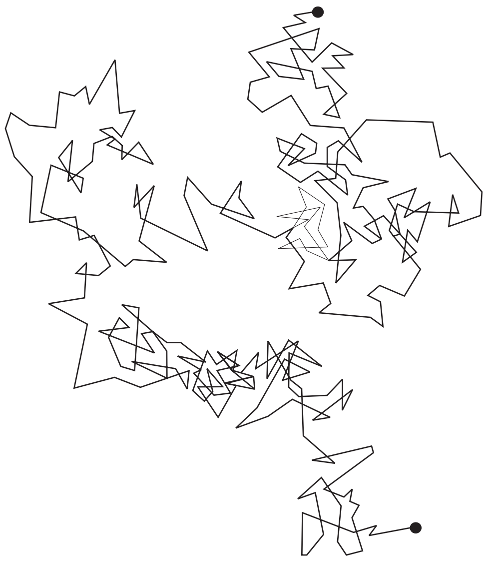


Рис. 32

щать внимания на множество нулевой меры, состоящее из тех точек ω, на которых функции разрывны, и считать все броуновские траектории непрерывными. Это — чрезвычайно важное свойство, которое можно даже присоединить к приведенному выше определению. С другой стороны, Винер также установил, что почти все траектории нигде не дифференцируемы, т. е. к графику нельзя провести касательную ни в какой точке. Это свойство мы упомянули для того, чтобы показать, что в данной области уже нельзя полагаться на интуицию.

Тем не менее, несложно угадать ответы на наши прежние вопросы в случае броуновского движения. На самом деле справедлив следующий аналог теоремы 1: выходя из произвольной точки, траектория будет проходить через любую другую точку бесконечное число раз. Отметим, что ввиду непрерывности траектории этот результат будет вытекать из теоремы «о промежуточном значении» из математического анализа, как только мы установим, что траектория может уходить сколь угодно далеко. Так как каждая допредельная траектория случайного блуждания обладает этим свойством, то, очевидно, что и броуновское движение им обладает. Наконец, давайте покажем, что формула (8.2.2) остается верной также и для броуновского движения, где u_a и v_a обозначают то же самое, что и раньше, но теперь a и c — произвольные действительные числа, $0 < a < c$. Рассмотрим броуновское движение, выходящее из точки a. Тогда в силу свойства (i) имеем $E(X_t) = a$ для всех $t \geqslant 0$, что является непрерывным аналогом формулы (8.2.3). Теперь подставим T_a*) вместо t и получим $E(X_{T_a}) = a$, как и прежде. На этот раз непрерывность траекторий обеспечивает то, что в момент T_a частица окажется точно в положении 0 или c. (Вообще говоря, выражение «достигать границы», используемое в определениях величин u_a, v_a и T_a, следует толковать более аккуратно в случае, когда частица может «перепрыгивать» через границу.) Таким образом, нам остается снова записать соотношения (8.2.4) и получить тот же самый ответ, что и для случайного блуждания.

8.3. Переходные вероятности

Модель случайного блуждания можно существенно обобщить до схемы, называемой *цепью Маркова* (об А. А. Маркове говорится в § 7.6). Как гласит пословица, иногда за деревьями не видно леса. Когда отбрасываются некоторые громоздкие и нехарактерные детали отдельных примеров, возникает общая теория, которая подчас оказывается проще и яснее, а также охватывает широкий круг приложений. Оставшаяся часть главы посвящена изложению основ подобной теории.

Мы будем продолжать использовать язык движения частицы из модели случайного блуждания. Через I обозначим диапазон возможных положений. Он не обязательно будет конечным или бесконечным множеством целых чисел, и, как скоро станет понятно, на I не обязательно должна быть определена какая-то геометрическая или алгебраическая структура (такая, как правое или левое направление, операции сложения и вычитания). Он может быть произвольным счетным множеством, состоящим из некоторых элементов, при условии, что мы обобщим наше определение случайной величины, позволив ей принимать значения в таком множестве. [В § 4.2 случайные величины определялись как функции с числовыми значениями.] Будем называть элементы I *состояниями*, а сам диапазон I — *пространством состояний*. Например, в физической химии состоянием может быть определенный энергетический уровень атома; при опросе общественного мнения — одно из возможных предпочтений опрашиваемых и т. п. Частица будет переходить от одного состояния к другому, а вероятностный закон, управляющий сменами состояний или переходами опишем следующим образом. Допустим, что имеется набор *переходных вероятностей* p_{ij}, где $i \in I$, $j \in I$, таких, что если частица находится в состоянии i в некоторый момент времени, то, *независимо*

*) Определение T_a с $j = a$ содержится в (8.4.2).

от того, в каких состояниях она побывала до этого, вероятность того, что в результате одного шага она попадет в состояние j, равна p_{ij}. Формально: если X_n обозначает положение частицы в момент n, то

$$P\{X_{n+1} = j \mid X_n = i; A\} = P\{X_{n+1} = j \mid X_n = i\} = p_{ij}, \qquad (8.3.1)$$

для произвольного события A, определяемого только случайными величинами $\{X_0, \ldots, X_{n-1}\}$. К примеру, A может быть конкретно заданным «прошлым» вида «$X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}$» или более общим прошлым событием, в котором состояния $i_0, \ldots, i_{n-1}$ заменены на множества состояний: «$X_0 \in J_0, X_1 \in J_1, \ldots, X_{n-1} \in J_{n-1}$». В последнем случае некоторые из этих множеств могут совпадать с пространством I. Тогда соответствующие случайные величины в действительности исчезают из условия. Так, условие «$X_0 \in J_0, X_1 \in I, X_2 \in J_2$» совпадает с «$X_0 \in J_0, X_2 \in J_2$». Первое из равенств (8.3.1) выражает точный смысл фразы «независимо от предшествующего поведения». Оно известно как *марковское свойство.* Второе равенство показывает, что условная вероятность не зависит от n. Это называется *стационарностью переходных вероятностей* или *однородностью по времени.* Вместе они приводят к следующему определению.

Определение цепи Маркова. Стохастический процесс $\{X_n, n \in \mathbb{N}^0\}$*), принимающий значения в некотором счетном множестве I называется *однородной цепью Маркова* или *марковской цепью со стационарными переходными вероятностями* тогда и только тогда, когда выполняются соотношения (8.3.1).

Если первое из равенств (8.3.1) справедливо, а второе — нет, то цепь Маркова называется *неоднородной.* В таком случае переходные вероятности зависят также и от n, и их следует обозначать, скажем, через $p_{ij}(n)$. Мы будем рассматривать только однородные цепи, поэтому при использовании ниже термина «цепь Маркова» или «цепь» будет по умолчанию подразумеваться именно такой случай.

В качестве следствия из определения легко выписываются вероятности последовательных переходов. Если частица находится в состоянии i_0, то независимо от ее прошлого поведения *условную* вероятность того, что она будет проходить через состояния i_1, $i_2, \ldots, i_n$ в указанном порядке в течение следующих n шагов, можно наглядно обозначить с помощью левой части равенства (8.3.2) и вычислить ее, перемножив

*) В некоторых случаях может оказаться более удобным начинать не с $n = 0$, а с $n = 1$.

переходные вероятности в правой части:

$$P\{\ldots.\,i_0 \to i_1 \to i_2 \to \cdots \to i_n\} = p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}; \qquad (8.3.2)$$

пять точек в начале отражают не относящееся к делу и забытое «прошлое». Равенство (8.3.2) получается путем применения соотношений (8.3.1) к общей формуле (5.2.2) для совместной вероятности. Например,

$$\begin{aligned} P\{X_4 = j, X_5 = k, X_6 = l \mid X_3 = i\} = \\ &= P\{X_4 = j \mid X_3 = i\} P\{X_5 = k \mid X_3 = i, X_4 = j\} \times \\ &\quad \times P\{X_6 = l \mid X_3 = i, X_4 = j, X_5 = k\} = \\ &= P\{X_4 = j \mid X_3 = i\}\, P\{X_5 = k \mid X_4 = j\}\, P\{X_6 = l \mid X_5 = k\} = \\ &= p_{ij}\, p_{jk}\, p_{kl}. \end{aligned}$$

Более того, здесь можно добавить в условие любое событие A, определяемое только набором $\{X_0, X_1, X_2\}$, не изменяя результат. Аналогичные вычисления показывают, что при фиксации *в любой момент* положения частицы история ее блуждания до этого момента не оказывает влияния не только на следующий переход, как утверждается в формуле (8.3.1), но также и на любые ее будущие перемещения. Формально: для произвольного события B, определяемого последовательностью $\{X_{n+1}, X_{n+2}, \ldots\}$, имеет место равенство

$$P\{B \mid X_n = i; A\} = P\{B \mid X_n = i\}, \qquad (8.3.3)$$

которое можно рассматривать как обобщение марковского свойства. Но это еще не все! Существует дальнейшее и более глубокое обобщение, связанное с пониманием слов «в любой момент», выделенных выше курсивом. Оно нам потребуется и будет объяснено позже.

Из равенства (8.3.2) понятно, что все вероятности, имеющие отношение к цепи, определяются переходными вероятностями при условии, что начальное положение частицы фиксировано, например, $X_0 = i$. Обобщая модель, рандомизируем начальное положение, полагая

$$P\{X_0 = i\} = p_i, \quad i \in I.$$

Тогда набор вероятностей $\{p_i, i \in I\}$ называется *начальным распределением* цепи. При этом для произвольных состояний $i_0, i_1, \ldots, i_n$ вероятности

$$P\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = p_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n} \qquad (8.3.4)$$

задают совместное распределение случайных величин, участвующих в процессе. Давайте особо рассмотрим случай, когда для каждых $i \in I$ и $j \in I$ выполняется равенство

$$p_{ij} = p_j.$$

Правая часть равенства (8.3.4) тогда превращается в произведение $p_{i_0}p_{i_1}\dots p_{i_n}$, и мы видим, что случайные величины $X_0, X_1, \dots, X_n$ независимы и имеют общее распределение $\{p_j, j \in I\}$. Таким образом, последовательность независимых и одинаково распределенных случайных величин, принимающих счетное множество значений, является частным случаем цепей Маркова, которые охватывают намного более широкий спектр моделей. Основные понятия такой схемы были введены Марковым приблизительно в 1907 г.

Из определения вероятностей p_{ij} очевидно, что

$$\begin{aligned} &\text{(a)} \quad p_{ij} \geqslant 0 && \text{для всех } i \text{ и } j, \\ &\text{(b)} \quad \sum_{j\in I} p_{ij} = 1 && \text{для всех } i. \end{aligned} \tag{8.3.5}$$

На самом деле можно показать, что это — достаточные условия на числа p_{ij} для того, чтобы они были переходными вероятностями однородной цепи Маркова. Другими словами, можно построить марковскую цепь, имеющую такие p_{ij} своими переходными вероятностями. Примеры даны в конце параграфа.

Обозначим через $p_{ij}^{(n)}$ вероятности перехода из состояния i в состояние j в точности за n шагов, т. е.

$$p_{ij}^{(n)} = P\{X_n = j \mid X_0 = i\}. \tag{8.3.6}$$

Там самым, $p_{ij}^{(1)}$ совпадают с нашими прежними p_{ij}, и мы также добавим

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 0, & \text{если } i \neq j, \\ 1, & \text{если } i = j, \end{cases}$$

для удобства. Использованный здесь символ δ_{ij} называют *символом Кронекера.* С ним вы, возможно, встречались в линейной алгебре. Покажем, что для $n \geqslant 1$, $i \in I$, $k \in I$, выполняется равенство

$$p_{ik}^{(n)} = \sum_j p_{ij}p_{jk}^{(n-1)} = \sum_j p_{ij}^{(n-1)}p_{jk}, \tag{8.3.7}$$

где суммирование проводится по всему диапазону I (это сокращение будет часто использоваться ниже). Докажем последнее соотношение. Пусть частица стартует из состояния i; рассмотрим возможности, возникающие после одного шага. Тогда она окажется в состоянии j с вероятностью p_{ij} и при этом условии за $n-1$ оставшихся шагов перейдет в состояние k с вероятностью $p_{jk}^{(n-1)}$, независимо от того, каково было i. Отсюда первое из равенств в формуле (8.3.7) получается суммированием по всем j в соответствии с общей формулой полной вероятности, см. (5.2.3) или (5.2.4). Второе из равенств (8.3.7) доказывается аналогичным образом при рассмотрении сначала перехода за $n-1$ шагов, а затем — перехода за оставшийся шаг.

При $n=2$ формула (8.3.7) превращается в

$$p_{ik}^{(2)} = \sum_j p_{ij} p_{jk}, \tag{8.3.8}$$

вид которой подсказывает, что надо использовать матрицы. Давайте организуем вероятности p_{ij} в форме матрицы

$$\Pi = [p_{ij}], \tag{8.3.9}$$

так что p_{ij} является элементом, стоящим в i-й строке и j-м столбце. Напомним, что элементы матрицы Π удовлетворяют условиям (8.3.5). Такая матрица называется *стохастической.* Произведением $\Pi_1 \times \Pi_2$ двух квадратных матриц служит такая квадратная матрица, что ее элемент из i-й стоки и j-го столбца получается сложением попарных произведений соответствующих элементов i-й строки матрицы Π_1 на элементы j-го столбца матрицы Π_2. В случае, когда обе матрицы Π_1 и Π_2 совпадают с Π, мы получаем в точности правую часть соотношения (8.3.8). Итак, мы имеем:

$$\Pi^2 = \Pi \times \Pi = [p_{ij}^{(2)}],$$

и, проводя индукцию по n с учетом равенства (8.3.7), выводим, что

$$\Pi^n = \Pi \times \Pi^{n-1} = \Pi^{n-1} \times \Pi = [p_{ij}^{(n)}].$$

Другими словами, *n-шаговые переходные вероятности* $p_{ij}^{(n)}$ суть элементы n-й степени матрицы Π. Если I — конечное множество $\{1, 2, \ldots, r\}$, тогда приведенное выше правило умножения совпадает, конечно, с обычным правилом умножения квадратных матриц порядка r. Если I — бесконечное множество, применимо то же правило, но мы должны быть уверены, что получающиеся в результате бесконечные ряды, подобные сумме (8.3.8), сходятся. На самом деле это верно в силу условий (8.3.5). Далее, равенства (8.3.7) можно обобщить следующим образом. Для $n \in \mathbb{N}^0$, $m \in \mathbb{N}^0$ и $i \in I$, $k \in I$ справедливы соотношения

$$p_{ik}^{(n+m)} = \sum_j p_{ij}^{(n)} p_{jk}^{(m)}. \tag{8.3.10}$$

Они известны как *уравнения Колмогорова—Чепмена**). Эти соотношения всего лишь выражают правило возведения матрицы Π в степень:

$$\Pi^{n+m} = \Pi^n \times \Pi^m,$$

*) Синди Чепмен (1888–1970) — английский математик-прикладник.

и могут быть получены либо чисто алгебраически — из соотношений (8.3.7) индукцией по m, либо с помощью вероятностных рассуждений, подобных тем, которые проводились для вывода тех же соотношений. Наконец, запишем тривиальное равенство, выполняющееся для каждого $n \in \mathbb{N}^0$ и $i \in I$:

$$\sum_j p_{ij}^{(n)} = 1. \tag{8.3.11}$$

Матрицу Π^n можно назвать *матрицей переходов за n шагов*. Используя вероятности $p_{ij}^{(n)}$, удается выразить совместные вероятности, когда какие-то из промежуточных состояний не заданы. Это ясно из следующего примера:

$$P\big\{X_4 = j, X_6 = k, X_9 = l \mid X_2 = i\big\} = p_{ij}^{(2)} p_{jk}^{(2)} p_{kl}^{(3)}.$$

Ниже для иллюстрации приводятся несколько примеров однородных марковских цепей и один пример неоднородной.

Пример 1. Пусть $I = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ — множество всех целых чисел;

$$p_{ij} = \begin{cases} p, & \text{если } j = i + 1, \\ q, & \text{если } j = i - 1, \\ 0, & \text{в противном случае}; \end{cases} \tag{8.3.12}$$

$$\Pi = \begin{bmatrix} . & . & . & . & . & . & . & . & . & . \\ . & . & . & q & 0 & p & 0 & 0 & . & . & . \\ . & . & . & 0 & q & 0 & p & 0 & . & . & . \\ . & . & . & 0 & 0 & q & 0 & p & . & . & . \\ . & . & . & . & . & . & . & . & . & . \end{bmatrix},$$

где $p + q = 1$, $p \geqslant 0$, $q \geqslant 0$. Это — *неограниченное случайное блуждание*, обсуждавшееся в § 8.1. В крайних случаях $p = 0$ или $q = 0$ оно, конечно, детерменировано (почти наверное).

Пример 2. Пусть $I = \{0, 1, 2, \ldots\}$ — множество неотрицательных целых чисел, а p_{ij} — такие же, как в примере 1 для $i \neq 0$, но $p_{00} = 1$, что влечет $p_{0j} = 0$ для всех $j \neq 0$. Это — случайное блуждание с одним *поглощающим состоянием* 0. Такая модель годится для задачи 3 из § 8.2. Поглощающее состояние соответствует разорению (банкротству) Петра, тогда как Павел бесконечно богат, соответственно этому I неограничено справа.

Пример 3. Пусть $I = \{0, 1, \ldots, c\}$, $c \geqslant 2$;

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & . & . & . & . & . & . \\ q & 0 & p & 0 & . & . & . & . & . \\ 0 & q & 0 & p & 0 & . & . & . & . \\ . & . & . & . & . & 0 & q & 0 & p \\ . & . & . & . & . & 0 & 0 & 0 & 1 \end{bmatrix}.$$

При $1 \leqslant i \leqslant c-1$ вероятности p_{ij} точно такие же, как в примере 1, но

$$p_{00} = 1, \quad p_{cc} = 1. \tag{8.3.13}$$

Это — случайное блуждание с двумя *поглощающими границами* 0 и c. Такая модель годится для задачи 1 из § 8.1. Здесь Π — квадратная матрица порядка $c+1$.

Пример 4. В примере 3 заменим условия (8.3.13) на такие:

$$p_{01} = 1, \quad p_{c,c-1} = 1;$$

$$\Pi = \begin{bmatrix} 0 & 1 & 0 & 0 & . & . & . \\ q & 0 & p & 0 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & 0 & q & 0 & p \\ . & . & . & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Эта модель представляет собой случайное блуждание с двумя *отражающими границами*: как только частица достигает одной из концевых точек отрезка $[0, c]$, она сразу «отскакивает» назад при следующем шаге. Другими словами, каждому из игроков, как только он разоряется, дают \$1 в утешение. Поэтому игра может продолжаться ради развлечения сколь угодно долго! Мы также можем исключить два состояния 0 и c и рассмотреть $I = \{1, 2, \ldots, c-1\}$,

$$p_{11} = q, \quad p_{12} = p, \quad p_{c-1,c-2} = q, \quad p_{c-1,c-1} = p;$$

$$\Pi = \begin{bmatrix} q & p & 0 & 0 & . & . & . \\ q & 0 & p & 0 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & 0 & q & 0 & p \\ . & . & . & 0 & 0 & q & p \end{bmatrix}.$$

Пример 5. Пусть $p \geqslant 0$, $q \geqslant 0$, $r \geqslant 0$ и $p + q + r = 1$. В примерах 1–4 заменим каждую строку $(\ldots\ q\ 0\ p\ \ldots)$ на $(\ldots\ q\ r\ p\ \ldots)$. Это означает, что в каждом из состояний частица может задерживаться (или что в игре допускаются ничьи) с вероятностью r. При $r = 0$ все сводится к предыдущим примерам.

Пример 6. Пусть $\{\xi_n, n \geqslant 0\}$ — последовательность независимых целочисленных случайных величин таких, что все они, за исключением ξ_0, имеют одинаковое распределение $\{a_k, k \in I\}$, где I — множество всех целых чисел. Определим случайную величину X_n аналогично тому, как это было сделано в формуле (8.1.2):

$$X_n = \sum_{k=0}^{n} \xi_k, \quad n \geqslant 0.$$

Так как $X_{n+1} = X_n + \xi_{n+1}$, а ξ_{n+1} не зависит от $X_0, X_1, \ldots, X_n$, то для произвольного события A, определяемого только величинами $X_0, \ldots, X_{n-1}$, имеем:

$$\begin{aligned} P\{X_{n+1} = j \mid A, X_n = i\} &= P\{\xi_{n+1} = j - i \mid A, X_n = i\} = \\ &= P\{\xi_{n+1} = j - i\} = a_{j-i}. \end{aligned}$$

Поэтому последовательность $\{X_n, n \geqslant 0\}$ образует однородную цепь Маркова с переходной матрицей $[p_{ij}]$, где

$$p_{ij} = a_{j-i}. \tag{8.3.14}$$

Начальным распределением цепи служит распределение случайной величины ξ_0, которое не обязательно должно совпадать с $\{a_k\}$. Говорят, что такая цепь *пространственно однородна* в том смысле, что вероятности p_{ij} зависят только от разности $j - i$. Наоборот, допустим что $\{X_n, n \geqslant 0\}$ — цепь с переходной матрицей, состоящей из элементов (8.3.14). Тогда

$$P\{X_{n+1} - X_n = k \mid X_n = i\} = P\{X_{n+1} = i + k \mid X_n = i\} = p_{i,i+k} = a_k.$$

Отсюда вытекает, что если положить $\xi_{n+1} = X_{n+1} - X_n$, то случайные величины $\{\xi_n, n \geqslant 1\}$ будут независимыми (почему?) и одинаково распределенными с общим распределением $\{a_k\}$. Таким образом, пространственно однородная (а также однородная по времени) марковская цепь идентична последовательности частичных сумм независимых одинаково распределенных целочисленных случайных величин. Изучение последней было одной из наших основных задач в предыдущих главах.

В частности, пример 1 есть частный случай примера 6 при $a_1 = p$, $a_{-1} = q$; можно добавить $a_0 = r$, как в примере 5.

Пример 7. Для каждого $i \in I$ возьмем два неотрицательных числа p_i и q_i, удовлетворяющих условию $p_i + q_i = 1$. Пусть I есть множество всех целых чисел. Положим

$$p_{ij} = \begin{cases} p_i, & \text{если } j = i+1, \\ q_i, & \text{если } j = 1-1, \\ 0 & \text{в противном случае.} \end{cases} \tag{8.3.15}$$

В данной модели частица переходит только в соседние состояния, как в примере 1, но вероятности перехода теперь могут зависеть от состояния. Эту модель можно еще обобщить, как в примере 5, позволив частице задерживаться в каждом состоянии i с вероятностью r_i, где $p_i+q_i+r_i = 1$. Обратите внимание, что данный пример включает в себя также приведенные ранее примеры 2, 3 и 4. Здесь мы имеем дело с цепью, которая уже не представляется в виде суммы независимых*) шагов, как в примере 6. Более подробное обсуждение данной модели см. в [5].

Пример 8 (*модель Эренфестов*). Ее можно считать частным случаем примера 7, в котором $I = \{0, 1, \ldots, c\}$ и

$$p_{i,i+1} = \frac{c-i}{c}\,, \quad p_{i,i-1} = \frac{i}{c}\,. \tag{8.3.16}$$

Такая модель реализуется с помощью урновой схемы следующим образом. Пусть урна содержит c шаров красного или черного цвета. Вынимается случайно шар и заменяется на шар другого цвета. Состоянием цепи считается количество черных шаров в урне. Нетрудно убедиться, что процесс замены шаров может идти сколь угодно долго, и переходные вероятности таковы, как указано выше. П. и Т. Эренфесты использовали похожую модель для изучения обмена энергией между молекулами газа. В действительности их урновая схема была немного иной (см. задачу 14).

Пример 9. Пусть $I = \{0, 1, 2, \ldots\}$ и

$$p_{i,0} = p_i, \quad p_{i,i+1} = 1 - p_i \quad \text{для } i \in I.$$

Здесь p_i — произвольные числа, удовлетворяющие неравенству

$$0 < p_i < 1.$$

Данная модель применяется для изучения рекуррентного явления, представленного состоянием 0. Каждый переход может приводить либо к наступлению явления, либо к увеличению времени ожидания на 1. Легко видеть, что событие «$X_n = k$» означает, что последним моментом $\leqslant n$, когда происходило явление, служит момент $n-k$, где $0 \leqslant k \leqslant n$. Иначе

*) И одинаково распределенных. — *Прим. перев.*

говоря, уже прошел временной период длины k с момента последнего наступления явления. В частном случае, когда все p_i равны p, имеем:

$$P\{X_v \neq 0 \text{ для } 1 \leqslant v \leqslant n-1,\ X_n = 0 \mid X_0 = 0\} = (1-p)^{n-1}p.$$

Таким образом, приходим к геометрическому распределению времени ожидания из примера 8 в § 4.4.

Пример 10. Пусть I — целочисленная решетка в $\mathbb{R}^d$ — евклидовом пространстве d измерений. Она является счетным множеством. Предположим, что, стартуя из произвольной точки решетки, частица за один шаг может переходить только в одну из $2d$ соседних с ней по осям точек решетки с различными (не обязательно одинаковыми) вероятностями. Для $d = 1$ эта модель совпадает с примером 1; для $d = 2$ речь идет о двумерном блуждании бродяги по улицам, упоминавшемся в § 8.1. В последнем случае можно считать состояния цепи парами (i, i'), где i и i' — целые числа. При этом

$$p_{(i,i')(j,j')} = \begin{cases} p_1, & \text{если } j = i+1,\ j' = i', \\ p_2, & \text{если } j = i-1,\ j' = i', \\ p_3, & \text{если } j = i,\ j' = i'+1, \\ p_4, & \text{если } j = i,\ j' = i'-1, \end{cases}$$

где $p_1 + p_2 + p_3 + p_4 = 1$. Если все эти четыре вероятности равны $1/4$, то цепь является симметричным двумерным случайным блужданием. Будет ли по-прежнему частица попадать к каждую точку решетки в вероятностью 1? Будет ли то же самое в случае трех измерений? Ответы на эти вопросы приводятся в следующем параграфе.

Пример 11 (*неоднородная цепь Маркова*). Рассмотрим урновуюсхему Пойа, описанную в § 5.4, при $c \geqslant 1$. Состоянием цепи будем считать количество черных шаров в урне. Поэтому «$X_n = i$» означает, что после n извлечений и добавлений в урне находятся ровно i черных шаров. Понятно, что каждый переход либо увеличивает это количество на c, либо оставляет его неизменным. При этом

$$P\{X_{n+1} = j \mid X_n = i, A\} = \begin{cases} \dfrac{i}{b+r+nc}, & \text{если } j = i+c, \\ 1 - \dfrac{i}{b+r+nc}, & \text{если } j = i, \\ 0 & \text{в противном случае,} \end{cases} \qquad (8.3.17)$$

где A — событие, определяемое исходами первых $n-1$ извлечений. Эти вероятности зависят не только от i и j, но также и от n. Поэтому процесс есть неоднородная цепь Маркова. Можно также позволить c принимать значение -1. Тогда мы будем иметь дело в выбором без возвращения и получим конечную последовательность $\{X_n\colon 0 \leqslant n \leqslant b+r\}$.

Пример 12. Легко определить процесс, который не является марковским. Скажем, в примерах 8 или 11 пусть $X_n = 0$ или 1 в соответствии с тем, является ли вынутый при n-м извлечении шар красным или черным. Тогда понятно, что вероятность «$X_{n+1} = 1$» при условии заданных значений величин $X_1, \ldots, X_n$ не будет, в общем случае, той же самой, как при условии одного фиксированного значения величины X_n. Последняя вероятность на самом деле не очень полезна.

8.4. Структура цепей Маркова

Начнем изучение структуры однородных цепей Маркова с определения бинарного отношения на множестве ее состояний. Будем говорить, что «j достижимо из i» и писать «$i \rightsquigarrow j$» тогда и только тогда, когда существует $n \geqslant 1$ такое, что $p_{ij}^{(n)} > 0$. Будем говорить, что «i сообщается с j» и писать «$i \leftrightsquigarrow j$» тогда и только тогда, когда $i \rightsquigarrow j$ и $j \rightsquigarrow i$. Отношение «$\rightsquigarrow$» транзитивно, т. е. если $i \rightsquigarrow j$ и $j \rightsquigarrow k$, то $i \rightsquigarrow k$. Это вытекает из неравенства

$$p_{ik}^{(n+m)} \geqslant p_{ij}^{(n)} p_{jk}^{(m)}, \tag{8.4.1}$$

которое является алгебраическим следствием формулы (8.3.10), но, быть может, более понятно ввиду своего вероятностного смысла. Действительно, если возможно дойти из i до j за n шагов, а также дойти из j до k за m шагов, то, объединяя эти переходы, возможно добраться из i до k за $n+m$ шагов. Сейчас и в дальнейшем мы будем употреблять такие слова, как «возможно» или «допустимо», в том смысле, что вероятность этого положительна. Важно отметить, что в только что приведенном тривиальном рассуждении использовалось марковское свойство, и без него нельзя обойтись. Отношение «$\leftrightsquigarrow$» очевидно обладает свойствами симметричности и транзитивности, и на его основе состояния цепи разбиваются на непересекающиеся классы следующим образом.

Определение класса. *Классом* называется такое подмножество пространства состояний, что любые два состояния (различные или нет), входящие в класс, сообщаются между собой.

Данный вид классификации, вероятно, известен вам под именем «разбиения на классы эквивалентности». Однако отношение «$\leftrightsquigarrow$» не обязательно является рефлексивным. Иначе говоря, могут существовать состояния, которые не достижимы из самих себя, а поэтому они не сообщаются ни с какими другими состояниями. Их мы просто не будем классифицировать! С другой стороны, класс может состоять из единственного состояния i: в этом случае $p_{ii} = 1$. Такие состояния называются *поглощающими*. Два различных класса не могут пересекаться: если бы у них имелся общий элемент, то они слились бы в единый класс через этот элемент.

Так, в примерах 1, 4, 5, 8 и 9 все состояния попадают в один класс при условии, что $p > 0$ и $q > 0$; то же самое можно сказать о примере 7, если $p_i > 0$ и $q_i > 0$ для всех i. В примере 2 имеются два класса: одиночное поглощающее состояние 0 и все остальные состояния в качестве другого класса. Аналогично, в примере 3 цепь разбивается на три класса. Что касается примера 6, то там ситуация сложнее. Представим, к примеру, что a_k таковы, что $a_k > 0$, если k делится нацело на 5, и $a_k = 0$ в противном случае. Тогда пространство состояний I распадается на пять классов. Два состояния i и j принадлежат одному классу тогда и только тогда, когда $i - j$ делится нацело на 5. Другими словами, классы совпадают с *классами остатков по модулю* 5. Понятно, что в такой ситуации было бы более естественным считать каждый из этих классов *новым* пространством состояний. В самом деле: если частица стартует из некоторого класса, то она (почти наверное) остается в нем навсегда. Тогда зачем обращать внимание на другие классы, в которые она никогда не попадет?

В теории вероятностей, особенно в ее части, относящейся к цепям Маркова, важную роль играет время до осуществления некоторого события. Пусть j — произвольное состояние. Рассмотрим время до первого попадания частицы в это состояние:

$$T_j(\omega) = \min\{n \geqslant 1 \mid X_n(\omega) = j\}, \tag{8.4.2}$$

где правая часть произносится как «наименьшее положительное значение n такое, что $X_n = j$». Может случиться, что для некоторых точек ω выборочного пространства ни одна из $X_n(\omega)$ не принимает значение j, т. е. не существует n, для которого выполняется равенство (8.4.2). Для таких ω величина T_j не определена. Мы примем соглашение, что в данном случае $T_j(\omega) = \infty$. Фразе «событие никогда не происходит» на обычном языке соответствует выражение «события придется ждать вечно» (или «бог знает сколько»). С учетом данного соглашения T_j является случайной величиной, способной принимать значение ∞. Давайте обозначим множество $\{1, 2, \ldots, \infty\}$ через $\mathbb{N}_\infty$. Тогда T_j принимает значения из $\mathbb{N}_\infty$. Тем самым, мы немного обобщили определение из § 4.2.

Выпишем вероятностное распределение случайной величины T_j. Ради упрощения будем использовать обозначение $P_i\{\cdots\}$ для вероятностей, связанных с марковской цепью, в которой частица стартует из состояния i. Для $n \in \mathbb{N}_\infty$ положим

$$f_{ij}^{(n)} = P_i\{T_j = n\}; \tag{8.4.3}$$

$$f_{ij}^{*} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = P_i\{T_j < \infty\}. \tag{8.4.4}$$

Напомним, что $\sum\limits_{n=1}^{\infty}$ на самом деле означает $\sum\limits_{1 \leqslant n < \infty}$. Здесь важно подчеркнуть, что величина ∞ не включается диапазон суммирования. Отсюда

$$P_i\{T_j = \infty\} = f_{ij}^{(\infty)} = 1 - f_{ij}^{*}. \tag{8.4.5}$$

Поэтому $\{f_{ij}^{(n)}, n \in \mathbb{N}_\infty\}$ — вероятностное распределение T_j для цепи, выходящей из i.

Дадим более явные представления для $f_{ij}^{(n)}$ и других вероятностей:

$$\begin{aligned} f_{ij}^{(1)} &= p_{ij} = P_i\{X_1 = j\}, \\ f_{ij}^{(n)} &= P_i\{X_v \neq j \text{ для } 1 \leqslant v \leqslant n-1, X_n = j\},\ n \geqslant 2, \\ f_{ij}^{(\infty)} &= P_i\{X_v \neq j \text{ для всех } v \geqslant 1\}, \\ f_{ij}^{*} &= P_i\{X_v = j \text{ для некоторых } v \geqslant 1\}. \end{aligned} \tag{8.4.6}$$

Заметим, что здесь можно взять $i = j$, а выражение «для некоторых v» означает «для по крайней мере одного значения v».

Случайная величина T_j называется *временем до первого попадания в состояние* j; также используется термин «время достижения состояния j». Важно отметить, что в силу однородности

$$f_{ij}^{(n)} = P\{X_{m+v} \neq j \text{ для } 1 \leqslant v \leqslant n-1,\ X_{m+n} = j \mid X_m = i\} \tag{8.4.7}$$

для любого m, для которого определена эта условная вероятность. Данное свойство будет часто использоваться без акцентирования внимания на нем.

Приведем ключевую формулу, связывающую вероятности $f_{ij}^{(n)}$ и $p_{ij}^{(n)}$.

Теорема 3. *Для любых i, j и $1 \leqslant n < \infty$ верно равенство*

$$p_{ij}^{(n)} = \sum_{v=1}^{n} f_{ij}^{(v)} p_{ij}^{(n-v)}. \tag{8.4.8}$$

Доказательство. Данный результат устанавливается достаточно формальным путем, в котором проявляются основные черты однородных цепей Маркова. Достаточно выписать следующую последовательность равенств:

$$p_{ij}^{(n)} = P_i\{X_n = j\} = P_i\{T_j \leqslant n, X_n = j\} = \sum_{v=1}^{n} P_i\{T_j = v, X_n = j\} =$$
$$= \sum_{v=1}^{n} P_i\{T_j = v\}\, P_i\{X_n = j \mid T_j = v\} =$$
$$= \sum_{v=1}^{n} P_i\{T_j = v\}\, P_i\{X_n = j \mid X_1 \neq j, \ldots, X_{v-1} \neq j, X_v = j\} =$$
$$= \sum_{v=1}^{n} P_i\{T_j = v\}\, P\{X_n = j \mid X_v = j\} =$$
$$= \sum_{v=1}^{n} P_i\{T_j = v\}\, P_j\{X_{n-v} = j\} = \sum_{v=1}^{n} f_{ij}^{(v)} p_{jj}^{(n-v)}.$$

Объясним каждое из этих равенств. Первое является определением вероятности $p_{ij}^{(n)}$. Второе верно, так как событие $\{X_n = j\}$ влечет $\{T_j \leqslant n\}$. Третье — потому что события $\{T_j = v\}$ для $1 \leqslant v \leqslant n$ не пересекаются. Четвертое — ввиду определения условной вероятности. Пятое — в силу смысла $\{T_j = v\}$ [см. (8.4.6)]. Шестое — в силу марковского свойства из (8.3.1), так как события $\{X_1 \neq j, \ldots, X_{v-1} \neq j\}$, а также событие $\{X_0 = i\}$, неявно присутствующее в обозначении P_i, относятся к «истории» процесса до момента v. Седьмое равенство справедливо в силу временно́й однородности перехода из j в j за $n - v$ шагов. Восьмое — всего лишь обозначение. Итак, теорема 3 доказана.

Правда, для соотношения (8.4.8) можно привести (и обычно приводится) более короткое словесное объяснение, но если вы заинтересуетесь деталями и будете прерывать его на каждом шаге вопросами «почему», то оно, по-существу, превратится в краткий пересказ приведенного выше доказательства. Последнее дает образец формализма, встречающегося обычно в общем контексте более сложных разделов теории марковских процессов, поэтому глубокое понимание простейших случаев, подобных нашему, стоит затраченных усилий.

При $i \neq j$ формула (8.4.8) связывает элементы переходной матрицы с индексами (i, j) с диагональными элементами, имеющими индексы (j, j). Существует также двойственная формула, связывающая их с элементами, имеющими индексы (i, i). Она выводится с помощью рассуждений, опирающихся на понятие момента *последнего выхода из* i, рассматриваемого в качестве двойственного по отношению к моменту

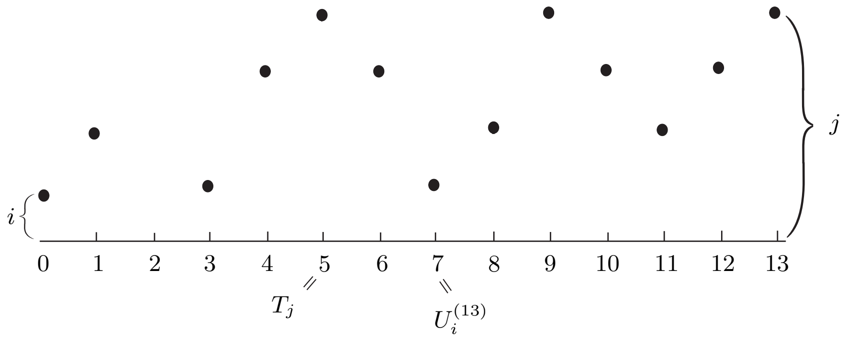


Рис. 33

первого попадания в j. Эта формула несколько более трудна для понимания и, по-видимому, известна лишь немногим специалистам. Мы приведем ее здесь ради симметрии и математической красоты. В действительности она служит полезным инструментом в теории марковских цепей, хотя для наших целей и не является необходимой.

Определим для $n \geqslant 1$

$$U_i^{(n)}(\omega) = \max\{0 \leqslant v \leqslant n \mid X_v(\omega) = i\}; \tag{8.4.9}$$

т. е. $U_i^{(n)}$ — последний момент пребывания в состоянии i до момента n включительно. Эта величина двойственна к T_j, но сложнее ее, так как зависит от n. Далее введем аналоги вероятностей $f_{ij}^{(n)}$ следующим образом:

$$\begin{gathered} g_{ij}^{(1)} = p_{ij}; \\ g_{ij}^{(n)} = P_i\{X_v \neq i \text{ для } 1 \leqslant v \leqslant n-1;\ X_n = j\}, \quad 2 \leqslant n < \infty. \end{gathered} \tag{8.4.10}$$

Так, $g_{ij}^{(n)}$ — это вероятность перемещения из i в j за n шагов *без захода* по пути в i (для $n = 1$ данное ограничение выполняется автоматически, так как $i \neq j$). Сравните ее с $f_{ij}^{(n)}$, которая теперь может быть определена как вероятность перемещения из i в j за n шагов без захода в j. Для обеих вероятностей накладываются запреты на определенные траектории движения, поэтому они называются *вероятностями перехода с запрещением* (более подробное обсуждение см. [5, § 1.9]). Теперь мы готовы к тому, чтобы сформулировать следующий результат.

Теорема 4. *Для* $i \neq j$ *и* $n \geqslant 1$ *верно равенство*

$$p_{ij}^{(n)} = \sum_{v=0}^{n-1} p_{ii}^{(v)} g_{ij}^{(n-v)}. \tag{8.4.11}$$

Доказательство. Будем повторять шаги доказательства теоремы 3 до тех пор, пока это возможно:

$$\begin{aligned}
p_{ij}^{(n)} &= P_i\{X_n = j\} = P_i\{0 \leqslant U_i^{(n)} \leqslant n-1, X_n = j\} = \\
&= \sum_{v=0}^{n-1} P_i\{U_i^{(n)} = v,\ X_n = j\} = \\
&= \sum_{v=0}^{n-1} P_i\{X_v = i, X_u \neq i \text{ для } v+1 \leqslant u \leqslant n-1,\ X_n = j\} = \\
&= \sum_{v=0}^{n-1} P_i\{X_v = i\}\, P_i\{X_u \neq i \text{ для } 1 \leqslant u \leqslant n-v-1,\ X_{n-v} = j\} = \\
&= \sum_{v=0}^{n-1} p_{ii}^{(v)} g_{ij}^{(n-v)}.
\end{aligned}$$

Наибольшее отличие наблюдается в четвертом из равенств, но оно очевидно ввиду смысла величины $U_i^{(n)}$. Обоснование других равенств оставляем читателю.

Введем также обозначение

$$g_{ij}^* = \sum_{n=1}^{\infty} g_{ij}^{(n)}. \tag{8.4.12}$$

Несмотря на то что каждый член этого ряда является вероятностью, заранее не ясно, сходится он или нет (он сходится при условии, что $i \rightsquigarrow j$; см. задачу 33). На самом деле можно показать, что g_{ij}^* равно ожидаемому числу посещений состояния j за время перехода из i в i.

Теоремы 3 и 4 называются, соответственно, *формулами разложения по первому входу или последнему выходу.* Вместе они работают, как две руки, хотя человек способен делать многие вещи, держа одну из рук за спиной, как мы увидим позже. Давайте приведем теперь вспомогательную лемму, предполагающую использование обеих рук.

Лемма. *Отношение $i \rightsquigarrow j$ эквивалентно условию $f_{ij}^* > 0$ или условию $g_{ij}^* > 0$.*

Доказательство. Если $f_{ij}^* = 0$, то $f_{ij}^{(n)} = 0$ для каждого n, и из формулы (8.4.8) вытекает, что $p_{ij}^{(n)} = 0$ для всех n. Следовательно, $i \rightsquigarrow j$ не выполняется. Наоборот, если $f_{ij}^* > 0$, то $f_{ij}^{(n)} > 0$ для некоторого n. Так как $p_{ij}^{(n)} \geqslant f_{ij}^{(n)}$ ввиду смысла этих двух вероятностей, то $p_{ij}^{(n)} > 0$ и поэтому $i \rightsquigarrow j$.

Доказательство для g^*_{ij} *в точности* такое же, если использовать формулу (8.4.11) вместо (8.4.8), что демонстрирует красоту двойственного мышления.

Можно посчитать, что данное доказательство является неоправданно сложным в случае f^*_{ij}, так как небольшое размышление должно убедить нас, что «$i \rightsquigarrow j$» и «$f^*_{ij} > 0$» оба означают, что «можно дойти из i в j за какое-то число шагов» (см. также задачу 31). Однако действительно ли очевидно, что условие «$g^*_{ij} > 0$» означает то же самое? Оно подразумевает, что можно дойти из i в j за какое-то число шагов, не проходя через i. Следовательно, утверждаемая эквивалентность влечет, что если j достижимо из i, то возможно осуществить переход, не возвращаясь в i. Например, представим, что некто способен доехать на автомобиле от Нью-Йорка до Сан-Франциско. Следует ли из этого, что ему удастся доехать, не возвращаясь назад для ремонта, за забытыми вещами или по причине отсрочки поездки? Настолько ли это очевидно, что не требует доказательства?

Эффективным способом обращения с формулами разложения является применение производящих функций, соответствующих последовательностям $\{p^{(n)}_{ij}, n \geqslant 0\}$, $\{f^{(n)}_{ij}, n \geqslant 1\}$, $\{g^{(n)}_{ij}, n \geqslant 1\}$ (см. § 6.5):

$$P_{ij}(z) = \sum_{n=0}^{\infty} p^{(n)}_{ij} z^n, \quad F_{ij}(z) = \sum_{n=1}^{\infty} f^{(n)}_{ij} z^n, \quad G_{ij}(z) = \sum_{n=1}^{\infty} g^{(n)}_{ij} z^n,$$

где $|z| < 1$. Тогда, учитывая равенство (8.4.8) и меняя порядок суммирования, получаем:

$$\begin{aligned} P_{ij}(z) &= \delta_{ij} + \sum_{n=1}^{\infty} \left(\sum_{v=1}^{n} f^{(v)}_{ij} p^{(n-v)}_{jj} \right) z^v z^{n-v} = \\ &= \delta_{ij} + \sum_{v=1}^{\infty} f^{(v)}_{ij} z^v \sum_{n=0}^{\infty} p^{(n-v)}_{jj} z^{n-v} = \delta_{ij} + F_{ij}(z) P_{jj}(z). \qquad (8.4.13) \end{aligned}$$

Изменение порядка суммирования законно, так как оба ряда сходятся абсолютно при $|z| < 1$. Аналогично выводим для $i \neq j$ равенство

$$P_{ij}(z) = P_{ii}(z) G_{ij}(z). \qquad (8.4.14)$$

Сначала применим этот результат для случая $i = j$.

Теорема 5. *Для любого состояния i условие $f^*_{ii} = 1$ выполняется тогда и только тогда, когда*

$$\sum_{n=0}^{\infty} p^{(n)}_{ii} = \infty; \qquad (8.4.15)$$

если же $f_{ii}^* < 1$, *то*

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1 - f_{ii}^*}\,. \tag{8.4.16}$$

Доказательство. Из соотношения (8.4.13) при $i = j$ получаем представление

$$P_{ii}(z) = \frac{1}{1 - F_{ii}(z)}\,. \tag{8.4.17}$$

Полагая здесь $z = 1$ и учитывая, что

$$P_{ii}(1) = \sum_{n=0}^{\infty} P_{ii}^{(n)}, \quad F_{ii}(1) = f_{ii}^*,$$

устанавливаем сразу оба утверждения теоремы. Отметим, что, строго говоря, мы должны были бы устремить $z \to 1-$ в равенстве (8.4.17) (почему?) и использовать следующую теорему из математического анализа: если $c_n \geqslant 0$ и степенной ряд $C(z) = \sum\limits_{n=0}^{\infty} c_n z^n$ сходится при $|z| < 1$, то

$$\lim_{z \to 1-} C(z) = \sum_{n=0}^{\infty} c_n,$$

а если ряд расходится, то данный предел равен бесконечности. Этот важный результат называется *теоремой Абеля* (в честь великого норвежского математика Нильса Абеля). Он нам еще потребуется в дальнейшем.

Дихотомия в теореме 5 приводит к следующему фундаментальному свойству состояний.

Определение возвратности и невозвратности состояния. Состояние i называется *возвратным* тогда и только тогда, когда $f_{ii}^* = 1$, и *невозвратным* в противном случае (т. е. когда $f_{ii}^* < 1$).

Некоторые авторы используют термины «рекуррентный» и «транзиентный» вместо «возвратный» и «невозвратный». Для дальнейшего применения выведем

Следствие (*теоремы 5*). *Если состояние j невозвратно, то*

$$\sum_{n=0}^{\infty} p_{ij}^{(n)} < \infty \quad \text{для каждого } i.$$

В частности,

$$\lim_{n \to \infty} p_{ij}^{(n)} = 0 \quad \text{для каждого } i.$$

Доказательство. При $i = j$ утверждение совпадает с формулой (8.4.16). Если $i \neq j$, то оно следует из соотношения (8.4.13), поскольку

$$P_{ij}(1) = F_{ij}(1)P_{jj}(1) \leqslant P_{jj}(1) < \infty.$$

Нетрудно показать, что два сообщающихся состояния либо оба возвратны, либо оба невозвратны. Тем самым, свойство возвратности распространяется на класс состояний и может рассматриваться как *свойство класса.* Чтобы убедиться в этом, допустим, что $i \leftrightsquigarrow j$. Тогда найдутся числа $m \geqslant 1$ и $m' \geqslant 1$ такие, что $p_{ij}^{(m)} > 0$ и $p_{ji}^{(m')} > 0$. Рассуждения, аналогичные тем, что проводились для обоснования неравенства (8.4.1), объясняют неравенство

$$p_{jj}^{(m'+n+m)} \geqslant p_{ji}^{(m')} p_{ii}^{(n)} p_{ij}^{(m)}.$$

Суммируя по всем $n \geqslant 0$, получаем

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} \geqslant \sum_{n=0}^{\infty} p_{jj}^{(m'+n+m)} \geqslant p_{ji}^{(m')} \left(\sum_{n=0}^{\infty} p_{ii}^{(n)} \right) p_{ij}^{(m)}. \tag{8.4.18}$$

Если состояние i возвратно, то ввиду соотношения (8.4.15) правая часть равна бесконечности. Поэтому и левая часть бесконечна, а в силу теоремы 5 это означает, что состояние j возвратно. Так как i и j можно поменять местами, то наше утверждение для случая возвратных состояний доказано. Утверждение, касающееся невозвратных состояний, вытекает из того, что невозвратность является отрицанием возвратности.

Установленный только что результат красив и полезен, но нам также понадобится его аналог, показывающий, что нельзя перейти от возвратного к невозвратному состоянию. (Обратный переход возможен, как установлено в примере 3 из § 8.3.) Данный аналог является более сильным утверждением и будет доказан ниже двумя разными методами. Первый метод опирается на двойственные теоремы 3 и 4.

Теорема 6. *Если состояние i возвратно и $i \rightsquigarrow j$, то состояние j также возвратно.*

Доказательство. В случае $i = j$ доказывать нечего, поэтому будем считать, что $i \neq j$. Согласно соотношениям (8.4.13) и (8.4.14),

$$P_{ij}(z) = F_{ij}(z)P_{jj}(z), \quad P_{ij}(z) = P_{ii}(z)G_{ij}(z),$$

откуда выводим

$$F_{ij}(z)P_{jj}(z) = P_{ii}(z)G_{ij}(z). \tag{8.4.19}$$

Если устремить $z \to 1-$, как в конце доказательства теоремы 5, то получим, что

$$F_{ij}(1)P_{jj}(1) = P_{ii}(1)G_{ij}(1) = \infty,$$

поскольку $G_{ij}(1) > 0$ в силу леммы и $P_{ii}(1) = \infty$ по теореме 5. Так как $F_{ij}(1) > 0$, согласно лемме заключаем, что $P_{jj}(1) = \infty$. Поэтому состояние j возвратно в силу теоремы 5. Доказательство закончено.

Отметим также, что формула (8.4.19), записанная в виде отношения

$$\frac{P_{ii}(z)}{P_{jj}(z)} = \frac{F_{ij}(z)}{G_{ij}(z)},$$

приводит к другим интересным результатам при $z \to 1-$, которые называются «предельными теоремами об отношениях» (см. [5, § 1.9]).

8.5. Дальнейшее развитие

Для демонстрации глубины понятия возвратности введем еще одну «трансфинитную» вероятность, а именно, — вероятность попадания в заданное состояние *бесконечно часто*:

$$q_{ij} = P_i\{X_n = j \text{ для бесконечного числа значений } n\}. \tag{8.5.1}$$

Мы уже встречались с этим понятием в теореме 2 из § 8.2. Действительно, в наших новых обозначениях эта теорема утверждает, что в симметричном случайном блуждании $q_{ij} = 1$ для произвольных i и j. А все же, каково точное значение слов «бесконечно часто»? Они означают «снова и снова, и так без конца» или более строго: «для любого наперед заданного большого числа, скажем, m, событие произойдет более чем m раз». Это не должно казаться вам чем-то трудно понимаемым, но, вероятно, вы будете удивлены, увидев символьное выражение величины q_{ij}. Оно выглядит так (сравните с формулами из заключительной части § 1.3):

$$q_{ij} = P_i\left\{\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}[X_n = j]\right\}.$$

Для сравнения запишем также, что

$$f_{ij}^{*} = P_i\left\{\bigcup_{n=1}^{\infty}[X_n = j]\right\}.$$

Однако эти пугающие формальности в дальнейшем обсуждении нам не потребуются.

Прежде всего, ввиду смысла данных вероятностей очевидно, что

$$q_{ij} \leqslant f_{ij}^{*}, \tag{8.5.2}$$

потому что, конечно, условие «бесконечно часто» влечет условие «по крайней мере один раз». Следующий результат очень важен.

Теорема 7. *Для любого состояния* i

$$q_{ii} = \begin{cases} 1, & \text{если } i \text{ возвратно,} \\ 0, & \text{если } i \text{ невозвратно.} \end{cases}$$

Доказательство. Положим $X_0 = i$ и $\alpha = f^*_{ii}$. Тогда α есть вероятность по крайней мере одного возвращения в i. В момент первого возвращения частица оказывается в состоянии i, и ее предыдущее поведение не влияет на дальнейшее. Следовательно, начиная с этого момента частица будет двигаться так, как будто она стартовала из i (словно «только что родилась»). Если обозначить через R_m событие, состоящее в том, что произошло «по крайней мере m возвращений», то вышесказанное влечет совпадение условной вероятности $P(R_2 \mid R_1)$ с вероятностью $P(R_1)$. Отсюда

$$P(R_2) = P(R_1 R_2) = P(R_1)P(R_2 \mid R_1) = \alpha \cdot \alpha = \alpha^2.$$

Повторяя данное рассуждение, по индукции выводим для $m \geqslant 1$, что

$$P(R_{m+1}) = P(R_m R_{m+1}) = P(R_m)\, P(R_{m+1} \mid R_m) = \alpha^m \cdot \alpha = \alpha^{m+1}.$$

В качестве следствия имеем, что вероятность бесконечного числа возвращений есть

$$\lim_{m\to\infty} P(R_m) = \lim_{m\to\infty} \alpha^m = \begin{cases} 1, & \text{если } \alpha = 1, \\ 0, & \text{если } \alpha < 1. \end{cases} \tag{8.5.3}$$

Теорема доказана.

Теперь самое время удостовериться в справедливости основного пункта проведенного доказательства:

$$P(R_{m+1} \mid R_m) = \alpha,$$

который был объяснен путем рассмотрения момента m-го возвращения в начальное состояние i и возобновления движения, начиная с этого момента. Рассуждение работает в силу того, что произошедшее до данного момента не сказывается на дальнейшем движении. (Каждый может легко привести ситуацию, в которой дело обстоит иначе — предыдущие возвращения препятствуют появлению новых.) Ввиду марковского свойства данное рассуждение представляется обоснованным, однако оно имеет один существенный изъян. Для определенности возьмем $m = 1$. Тогда моментом первого возвращения служит T_i, определенный формулой (8.4.2), и приведенное выше рассуждение основывается на применении марковского условия (8.3.3) в момент T_i. Но T_i — случайная величина, ее значение зависит от точки ω выборочного пространства. Можем ли мы подставить T_i вместо постоянного момента n в эту формулу? Вероятно, вы полагаете, что, поскольку формула верна для *любого* n, а $T_i(\omega)$ равно *некоторому* n, какой бы ни была ω, подобная замена является корректной. (На самом деле мы производили аналогичную подстановку в § 8.2 без строгого обоснования.) В ошибочности подобного мнения легко убедиться*), но сейчас мы опишем определенный тип случайных величин, для которых данная замена законна.

В модели однородной цепи Маркова $\{X_n, n \in \mathbb{N}^0\}$ случайная величина T называется *марковским моментом*, если для каждого n событие $\{T = n\}$ определяется только величинами $\{X_0, X_1, \ldots, X_n\}$**). Говорят, что событие *предшествует моменту* T, если оно определяется величинами $\{X_0, X_1, \ldots, X_{T-1}\}$, и *происходит после момента* T, если оно определяется величинами $\{X_{T+1}, X_{T+2}, \ldots\}$. (При $T = 0$ не имеет смысла говорить о предшествующих событиях.) Положение частицы в момент T задается, конечно, величиной X_T (обратите внимание, что она является случайной функцией $\omega \to X_{T(\omega)}(\omega)$). Если в качестве T выступает константа n, то эти понятия согласуются с нашими обычными интерпретациями «прошлого» и «будущего», относящимися к «текущему» моменту n. В общем случае они могут зависеть от точки выборочного пространства. В этом нет ничего непривычного, например, во фразах «предродовое наблюдение», «послевоенное устройство» или «день после нашествия саранчи» имеется в виду неопределенная (и поэтому — случайная) дата. Когда игрок решает, что он поставит «на красное» после того, как «черное» выпадет три раза подряд, он имеет дело с X_{T+1}, где значение T зависит от случая. Однако важно отметить, что данные относительные понятия осмысленны исключительно благодаря способу определения момента T. Если бы значение T зависело не только от «прошлого» и «настоящего», но и от «будущего», то «предыстория-T» и «постыстория-T» перемешались бы и стали бесполезными. Если бы игрок мог предвидеть будущее, ему не нужна была бы теория вероятностей! В указанном смысле момент остановки иногда характеризуется как «момент, не зависящий от будущего»; его наступление должно считаться «возможностью» без использования преимуществ ясновидения.

Теперь мы готовы сформулировать следующее обобщение формулы (8.3.3). Для произвольного момента остановки T, любого события A, предшествующего T, и любого события B, происходящего после T, спра-

*) Например, пусть $X_0 = i_0 \neq k$. Возьмем $T = T_k - 1$ в приведенной ниже формуле (8.5.5). Так как $X_{T+1} = k$, равенство не будет выполняться в общем случае.

**) Конечный с вероятностью 1 марковский момент называют *моментом остановки*. — *Прим. перев.*

ведливо равенство

$$P\{B \mid X_T = i; A\} = P\{B \mid X_T = i\}; \tag{8.5.4}$$

в частности, для любых состояний i и j

$$P\{X_{T+1} = j \mid X_T = i; A\} = p_{ij}. \tag{8.5.5}$$

Оно известно как *сильное марковское свойство.* Его, на самом деле, можно получить из очевидно более слабой формы (8.3.3), а также из первоначального определения (8.3.1). В книгах по теории вероятностей нередко приводится слабая, а на самом деле используется сильная форма без упоминания различия. Раз уж мы «вытащили на свет» последнюю, то *примем ее в качестве определения однородной цепи Маркова.* Формальное доказательство см. в [5, § I.13]. Заметим, что сильное марковское свойство нам было необходимо еще во время доказательства теоремы 2 из § 8.2, где оно было намеренно скрыто, чтобы не вызывать преждевременного беспокойства. Сейчас самое время вернуться назад и перечитать доказательство с новым пониманием.

Возвращаясь к теореме 7, проверим, что момент T_i, используемый в доказательстве, в действительности является моментом остановки. Но это следует из явного представления (8.4.6), так как событие

$$\{T_i = n\} = \{X_v \neq i \text{ при } 1 \leqslant v \leqslant n-1, X_n = i\},$$

очевидно, определяется только величинами $\{X_1, \ldots, X_n\}$. Данный аргумент завершает строгое доказательство теоремы 7, к которой мы также добавим следствие.

Следствие (*теоремы 7*). *Для любых состояний i и j*

$$q_{ii} = \begin{cases} f^*_{ij}, & \textit{если } j \textit{ возвратно,} \\ 0, & \textit{если } j \textit{ невозвратно.} \end{cases}$$

Доказательство. Оно немедленно вытекает из теоремы и соотношения

$$q_{ij} = f^*_{ij} q_{jj}. \tag{8.5.6}$$

Действительно, посетить состояние j бесконечно много раз означает попасть в него хотя бы однажды и затем возвращаться бесконечно много раз. Так же, как в доказательстве теоремы 8, использованное здесь рассуждение базируется на сильном марковском свойстве.

Следующий результат показывает силу «мышления в терминах бесконечного».

Теорема 8. *Если состояние i возвратно и $i \rightsquigarrow j$, то*

$$q_{ij} = q_{ji} = 1. \tag{8.5.7}$$

Доказательство. Из утверждения теоремы вытекает, что $i \leftrightsquigarrow j$ и что состояние j возвратно в силу приведенного выше следствия. Тем самым, установив справедливость данного утверждения, мы получим новое доказательство теоремы 6.

Заметим, что для любых двух событий A и B верно включение $A \subset AB \cup B^c$, поэтому

$$P(A) \leqslant P(B^c) + P(AB). \tag{8.5.8}$$

Теперь рассмотрим события

$$A = \{\text{частица посещает } i \text{ бесконечно часто}\},$$
$$B = \{\text{частица попадает в } j \text{ хотя бы однажды}\}.$$

Тогда $P_i(A) = q_{ii} = 1$ в силу теоремы 7 и условия $P_i(B^c) = 1 - f^*_{ij}$. В свою очередь, смысл вероятности $P_i(AB)$ в том, что частица попадет в j за некоторое конечное время и *после этого* посетит i бесконечное число раз, так как «бесконечность минус конечное число есть, по-прежнему, бесконечность». Следовательно, если применить сильное марковское свойство для момента первого попадания в j, то получим

$$P(AB) = f^*_{ij} q_{ji}.$$

Подставляя это выражение в неравенство (8.5.8), выводим, что

$$1 = q_{ii} \leqslant 1 - f^*_{ij} + f^*_{ij} q_{ji}$$

или

$$f^*_{ij} \leqslant f^*_{ij} q_{ji}.$$

Поскольку $f^*_{ij} > 0$, заключаем, что $1 \leqslant q_{ji}$. Отсюда $q_{ji} = 1$. Так как $q_{ji} \leqslant f^*_{ji}$, то $f^*_{ji} = 1$ и поэтому $j \rightsquigarrow i$. Таким образом, i и j сообщаются, и состояние j является возвратным в силу (8.4.18). Зная это, можно поменять ролями i и j в предыдущем рассуждении и вывести, что $q_{ij} = 1$.

Следствие. *В возвратном классе соотношения* (8.5.7) *выполняются для любых двух состояний* i *и* j.

Мы будем называть цепь возвратной, когда пространство состояний цепи представляет собой один возвратный класс; аналогично определяется невозвратная цепь. Положение дел в возвратной цепи, описываемое предыдущим следствием, в точности такое же, как в случае симметричного случайного блуждания (см. теорему 2 в § 8.2). На самом деле последняя модель является частным случаем первой, как мы сейчас покажем.

Применим общие методы, развитые выше, к случайному блужданию, обсуждаемому в § 8.1, точнее — к модели из примера 1 в § 8.3. Начнем с вычисления $p_{ii}^{(n)}$. Это есть вероятность того, что частица окажется снова в начальном положении i в точности через n шагов. Очевидно, $p_{ii}^{(2n-1)} = 0$ при $n \geqslant 1$; и, используя обозначения (8.1.2), по формуле Бернулли (7.3.1):

$$p_{ii}^{(2n)} = P\{\xi_1 + \ldots + \xi_{2n} = 0\} = \binom{2n}{n} p^n q^n, \tag{8.5.9}$$

так как частица должна совершить n шагов вправо и n шагов влево в некотором порядке. Отсюда получаем производящую функцию

$$P_{ii}(z) = \sum_{n=0}^{\infty} \binom{2n}{n} (pqz^2)^n. \tag{8.5.10}$$

Вспоминая определение биномиальных коэффициентов (5.4.4), запишем красивое тождество

$$\binom{-1/2}{n} = \frac{(-1)^n 1 \cdot 3 \ldots (2n-1)}{2^n \cdot n!} = \frac{(-1)^n}{2^{2n}} \binom{2n}{n}, \tag{8.5.11}$$

где второе равенство выводится путем умножения числителя и знаменателя на $2^n \cdot n! = 2 \cdot 4 \cdots (2n)$. Подставляя это тождество в функцию (8.5.10), приходим к явной аналитической формуле:

$$P_{ii}(z) = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-4pqz^2)^n = (1 - 4pqz^2)^{-1/2}, \tag{8.5.12}$$

где средняя часть является разложением в ряд Тейлора бинома, стоящего в правой части.

Отсюда следует, что

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = P_{ii}(1) = \lim_{z \to 1-} P_{ii}(z) = \lim_{z \to 1-} (1 - 4pqz^2)^{-1/2}. \tag{8.5.13}$$

Далее, поскольку $4pq = 4p(1-p) \leqslant 1$ при $0 \leqslant p \leqslant 1$; причем $4pq = 1$ тогда и только тогда, когда $p = 1/2$ (почему?). Поэтому приведенный выше ряд расходится, если $p = 1/2$, и сходится, если $p \neq 1/2$. В силу теоремы 5, состояние i возвратно тогда и только тогда, когда $p = 1/2$. Проведенные вычисления не зависят от значения i благодаря пространственной однородности. Тем самым, мы установили, что при $p = 1/2$ цепь возвратна, в противном случае — невозвратна. Другими словами, случайное блуждание возвратно тогда и только тогда, когда оно симметрично.

Существует другой метод, позволяющий вывести это утверждение прямо из соотношения (8.5.9), без использования производящих функций. Для $p = 1/2$

$$p_{ii}^{(2n)} = \binom{2n}{n} \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}}\,, \tag{8.5.14}$$

в силу следующего за (7.3.5) применения формулы Стирлинга. Опираясь на признак сравнения рядов с положительными членами, выводим отсюда, что ряд из (8.5.13) расходится, так как расходится ряд $\sum\limits_n 1/\sqrt{n}$. Данный метод имеет то преимущество, что его можно использовать для случайных блужданий в пространствах большей размерности. Рассмотрим симметричное случайное блуждание в пространстве $\mathbb{R}^2$ (пример 10 из § 8.3 в случае, когда все четыре вероятности равны 1/4). Возвращение в любое состояние (i, j) за $2n$ шагов означает, что для некоторого k, $0 \leqslant k \leqslant n$, частица выполнила, в некотором порядке, по k шагов на восток и запад, а также — по $n-k$ шагов на север и юг. Вероятность этого, согласно полиномиальной формуле (6.4.6), есть

$$\begin{aligned} p_{(i,j)(i,j)}^{(2n)} &= \frac{1}{4^{2n}} \sum_{k=0}^{n} \frac{(2n)!}{k!\,k!\,(n-k)!\,(n-k)!} = \\ &= \frac{(2n)!}{4^{2n} n!\,n!} \sum_{k=0}^{n} \binom{n}{k}^2 = \frac{(2n)!}{4^{2n} n!\,n!} \binom{2n}{n} = \left[\frac{1}{2^{2n}} \binom{2n}{n}\right]^2, \end{aligned} \tag{8.5.15}$$

где для получения предпоследнего равенства была использована формула, приведенная в задаче 28 гл. 3. То, что данная вероятность оказалась равна в точности квадрату вероятности (8.5.14) — всего лишь приятное совпадение. (Оно не вытекает из очевидной независимости компонент блуждания вдоль двух координатных осей.) Проводя сравнение с формулой (8.5.14), выводим, что

$$\sum_n p_{(i,j)(i,j)}^{(2n)} \sim \sum_n \frac{1}{\pi n} = \infty.$$

Отсюда, применяя теорему 5 еще раз, заключаем, что симметричное случайное блуждание на плоскости, подобно блужданию по прямой, является возвратной цепью Маркова. Похожие, но более громоздкие, рассуждения показывают, что оно невозвратно в пространстве R^d при $d \geqslant 3$, поскольку вероятности, аналогичные вероятности, стоящей в левой части формулы (8.5.15), ограничены сверху величиной $c/n^{d/2}$ (где c — некоторая константа), и ряд $\sum\limits_n 1/n^{d/2}$ сходится при $d \geqslant 3$. Данный результат впервые опубликовал Пойа в 1921 г. Несимметричный случай может быть рассмотрен с помощью нормальной аппроксимации, приведенной в формуле (7.3.13), однако невозвратность такого блуждания вытекает также из усиленного закона больших чисел, как в $\mathbb{R}^1$, см. § 8.2.

В качестве другой иллюстрации выведем явную формулу для $f_{ii}^{(n)}$ в случае $p = 1/2$. Из соотношений (8.4.17) и (8.5.12) получаем, что

$$F_{ii}(z) = 1 - \frac{1}{P_{ii}(z)} = 1 - (1 - z^2)^{1/2}.$$

Тогда еще одно разложение бинома в ряд Тейлора дает

$$F_{ii}(z) = 1 - \sum_{n=0}^{\infty} \binom{1/2}{n} (-z^2)^n = \frac{1}{2}\, z^2 + \sum_{n=2}^{\infty} \frac{1 \cdot 3 \cdots (2n-3)}{2^n \cdot n!}\, z^{2n}.$$

Таким образом, $f_{ii}^{(2n-1)} = 0$ и

$$f_{ii}^{(2n)} = \frac{1}{2^{2n}} \binom{2n}{n} \frac{1}{2n-1}\,, \quad n \geqslant 1. \tag{8.5.16}$$

Данное равенство устанавливается с помощью вычислений, похожих на те, что были выполнены при выводе формулы (8.5.11). В частности, имеем:

n	1	2	3	4	5
$f_{ii}^{(2n)}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{5}{128}$	$\frac{7}{256}$

Сравнение с асимптотикой (8.5.14) показывает, что

$$f_{ii}^{(2n)} \sim \frac{1}{2\sqrt{\pi}\, n^{3/2}}\,.$$

Отсюда $\sum\limits_{n=1}^{\infty} n f_{ii}^{(n)} = \infty$. Это также может быть получено путем вычисления $F'_{ii}(1)$. Таким образом, несмотря на то, что возвращение происходит почти наверное, ожидаемое время до этого момента бесконечно велико. Данный результат, как легко видеть, эквивалентен замечанию из § 8.2 о том, что $e_1 = \infty$.

Подобным методом можно вычислить $f_{ij}^{(n)}$ для любых i и j в случайном блуждании. Однако иногда комбинаторное доказательство короче и более познавательно. К примеру, покажем, что

$$f_{00}^{(2n)} = \frac{1}{2}\, f_{10}^{(2n-1)} + \frac{1}{2}\, f_{-1,0}^{(2n-1)} = f_{10}^{(2n-1)} = f_{01}^{(2n-1)}. \tag{8.5.17}$$

Пусть частица стартует из нуля; рассмотрим разложение ее блуждания по первому шагу, как при выводе соотношения (8.1.4). Тогда, используя симметрию и пространственную однородность, получим остальное. Детали оставляются читателю.

8.6. Стационарное распределение

В этом параграфе рассматривается возвратная марковская цепь. Иначе говоря, мы предполагаем, что пространство состояний представляет собой один возвратный класс.

После того, как частица блуждала в такой цепи достаточно долго, ее можно будет обнаружить в разных состояниях с какими-то вероятностями. Стремятся ли такие вероятности к некоторым предельным значениям? Этот предел физики и инженеры называют «стационарным состоянием» (распределением)*). Они привыкли думать в терминах «ансамбля» или большого числа частиц, движущихся согласно одинаковому вероятностному закону независимо друг от друга, подобно молекулам газа. В нашем случае таким законом служит марковское свойство для модели однородной цепи, обсуждавшее в предыдущих параграфах. По прошествии длительного времени пропорция (доля) частиц, обнаруженных в каждом из состояний, будет близка к стационарной вероятности данного состояния. На самом деле это является частотной интерпретацией понятия вероятности, упоминавшейся в примере 3 из § 2.1, в которой предельные пропорции считаются определением соответствующих вероятностей. Приведем формулировку в наших терминах. Если частица стартует из состояния i, то вероятность множества путей, при которых она оказывается в состоянии j в момент времени n, т. е. вероятности события $\{\omega\colon X_n(\omega) = j\}$, есть $P_i\{X_n = j\} = p_{ij}^{(n)}$. Тем самым, нас интересует асимптотическое поведение $p_{ij}^{(n)}$ при $n \to \infty$. Оказывается, что в определенном смысле более удобной для изучения является величина, представляющая собой среднее значение таких вероятностей за долгий временной период, а именно

$$\frac{1}{n+1}\sum_{v=0}^{n} p_{ij}^{(v)} \quad \text{или} \quad \frac{1}{n}\sum_{v=1}^{n} p_{ij}^{(v)}. \tag{8.6.1}$$

Различие между этими двумя средними пренебрежимо мало для больших n, и мы будем использовать первое. Данная величина имеет следующую удобную интерпретацию. Зафиксируем наше внимание на определенном состоянии j и вообразим, что какой-то считающий прибор записывает количество единиц времени, проведенных частицей в состоянии j. Это можно осуществить, вводя случайные величины, принимающие значение 1 для состояния j и значение 0 для любого другого

---

*) Довольно странно, что они называют распределение состоянием!

состояния:

$$\xi_v(j) = \begin{cases} 1, & \text{если } X_v = j, \\ 0, & \text{если } X_v \neq j. \end{cases}$$

Мы уже использовали такие индикаторы, например, в формуле (6.4.11). Далее положим

$$N_n(j) = \sum_{v=0}^{n} \xi_v(j).$$

Эта случайная величина представляет собой общее *время пребывания* в состоянии j в течение n шагов. Теперь, если обозначить через E_i математическое ожидание, связанное с цепью, в которой частица стартует из i (оно является условным ожиданием; см. конец § 5.2), то

$$E_i(\xi_v(j)) = p_{ij}^{(v)},$$

и поэтому по теореме 1 из § 6.1

$$E_i(N_n(j)) = \sum_{v=0}^{n} E_i(\xi_v(j)) = \sum_{v=0}^{n} p_{ij}^{(v)}. \tag{8.6.2}$$

Итак, величина (8.6.1) оказывается средним ожидаемым временем пребывания в состоянии j.

Для ее изучения рассмотрим, прежде всего, случай $i = j$ и введем *ожидаемое время возвращения* из j в j следующим образом:

$$m_{jj} = E_j(T_j) \sum_{v=1}^{\infty} v f_{jj}^{(v)}, \tag{8.6.3}$$

где случайная величина T_j определяется формулой (8.4.2). Так как j — возвратное состояние, нам известно, что время T_j почти наверное конечно, но его математическое ожидание может быть как конечным, так и бесконечным. Как мы увидим, различие между этими двумя случаями весьма существенно.

Приведем эвристические рассуждения, связывающие формулы (8.6.2) и (8.6.3). Поскольку в среднем на возвращение требуется m_{jj} единиц времени, всего должно произойти примерно n/m_{jj} таких возвращений за время n. Иначе говоря, частица проводит около n/m_{jj} единиц времени в состоянии j в течение первых n шагов, т. е. $E_j(N_n(j)) \approx n/m_{jj}$. Такое же рассуждение показывает, что нет различия в том, стартует ли частица из j или из любого другого состояния i, потому что после первого попадания в j первоначальное i может быть забыто, и мы будем иметь дело только с последовательными возвращениями из j в j. Тем самым, приходим к следующей предельной теореме.

Теорема 9. *Для любых i и j*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{v=0}^{n} p_{ij}^{(v)} = \frac{1}{m_{jj}}\,. \tag{8.6.4}$$

Приведенное выше рассуждение можно сделать строгим с помощью усиленного закона больших чисел в общей форме (см. § 7.5), применяемого ко временам между последовательными возвращениями, которые образуют совокупность независимых и одинаково распределенных случайных величин. К сожалению, технические подробности выходят за уровень сложности, принятый в этой книге. Существует другой подход, опирающийся на мощный аналитический результат, полученный Харди и Литлвудом. (Речь идет о том же самом Харди, что и в теореме Харди—Вайнберга из § 5.6.) Этот результат известен как *тауберова теорема* (в честь А. Таубера, который первым доказал утверждение подобного вида). Ее можно сформулировать следующим образом.

Теорема 10. *Если $A(z) = \sum\limits_{n=0}^{\infty} a_n z^n$, где $a_n \geqslant 0$ для всех n, и ряд сходится при $0 \leqslant z < 1$, то*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{v=0}^{n} a_v = \lim_{z\to 1}(1-z)A(z). \tag{8.6.5}$$

Чтобы лучше понять теорему, представим, что все $a_n = c > 0$. Тогда

$$A(z) = c\sum_{n=0}^{\infty} z^n = \frac{c}{1-z}\,,$$

и равенство (8.6.5) сводится к тривиальному тождеству

$$\frac{1}{n+1}\sum_{v=0}^{n} c = c = (1-z)\ \frac{c}{1-z}\,.$$

Теперь возьмем в качестве $A(z)$ производящую функцию

$$P_{ij}(z) = \sum_{n=0}^{\infty} p_{ij}^{(n)} z^n = F_{ij}(z)P_{ij}(z) = \frac{F_{ij}(z)}{1-F_{jj}(z)}\,,$$

где два последних равенства вытекают из формул (8.4.13) и (8.4.17). Тогда

$$\lim_{z\to 1}(1-z)P_{ij}(z) = F_{ij}(1)\lim_{z\to 1}\frac{1-z}{1-F_{jj}(z)} = \lim_{z\to 1}\frac{1-z}{1-F_{jj}(z)}\,,$$

поскольку $F_{ij}(1) = f_{ij}^* = 1$ в силу теоремы 8 из § 8.5. Последний предел можно вычислить с помощью правила Лопиталя:

$$\lim_{z\to 1} \frac{(1-z)'}{(1-F_{jj}(z))'} = \lim_{z\to 1} -\frac{-1}{F_{jj}'(z)} = \frac{1}{F_{jj}'(1)}\,,$$

где «$'$» обозначает дифференцирование по переменной z. Так как

$$F_{jj}'(z) = \sum_{v=1}^{\infty} v f_{jj}^{(v)} z^{v-1},$$

то

$$F_{jj}'(1) = \sum_{v=1}^{\infty} f_{jj}^{(v)} = m_{jj},$$

и поэтому предел (8.6.4) является частным случаем соотношения (8.6.5).

Теперь, чтобы не усложнять математический аппарат, рассмотрим конечное пространство состояний I. Конечность I немедленно влечет следующее утверждение.

Теорема 11. *Если пространство состояний I конечно и представляет собой один класс (т. е. имеется только конечное число состояний, и все они сообщаются), то цепь необходимо является возвратной.*

Доказательство. Допустим противное. Тогда каждое состояние невозвратно, и поэтому, в силу теоремы 7, почти наверное частица сможет провести в нем только конечное время. Поскольку число состояний конечно, частица сможет провести суммарно во всем пространстве I только конечное время. Но время увеличивается неограниченно, и частица не может никуда уйти. Это противоречие доказывает, что цепь обязана быть возвратной (куда частице податься?).

Пусть $I = \{1, 2, \dots, l\}$, $x = (x_1, \dots, x_l)$ — вектор-строка длины l. Рассмотрим «уравнение стационарности»

$$x = x\Pi \quad \text{или} \quad x(\Delta - \Pi) = 0, \tag{8.6.6}$$

где Δ — единичная матрица, имеющая l строк и l столбцов: $\Delta = (\delta_{ij})$, а Π — переходная матрица (8.3.9). Тем самым, получаем систему из l линейных уравнений с постоянными коэффициентами от l неизвестных. Определитель матрицы $\Delta - \Pi$

$$\begin{vmatrix} 1-p_{11} & -p_{12} & \cdots & -p_{1l} \\ -p_{21} & 1-p_{22} & \cdots & -p_{2l} \\ \hdotsfor{4} \\ -p_{l1} & -p_{l2} & \cdots & 1-p_{ll} \end{vmatrix}$$

равен нулю, поскольку сумма всех элементов в каждой строке есть $1-\sum\limits_j p_{ij} = 0$. Из линейной алгебры известно, что в таком случае система имеет нетривиальное, т. е. отличающееся от нулевого вектора, решение. Понятно, что если x — такое решение, то решением системы также будет и вектор $cx = (cx_1, cx_2, \ldots, cx_l)$ для любой константы c. Приводимая ниже теорема определяет все решения для случая, когда I — единственный конечный класс. В ней используются обозначения

$$\begin{gathered} w_j = \frac{1}{m_{jj}}, \quad j \in I, \\ w = (w_1, w_2, \ldots, w_l), \end{gathered} \tag{8.6.7}$$

а также сокращение $\sum\limits_j$ для $\sum\limits_{j\in I}$.

Теорема 12. *Если пространство I конечно и представляет собой один класс, то*

(i) w является решением уравнения (8.6.6)*;*

(ii) $\sum\limits_i w_j = 1$;

(iii) $w_j > 0$ для всех j;

(iv) любое решение уравнения (8.6.6) *представляется в виде вектора w, умноженного на некоторую константу.*

Доказательство. Из соотношения (8.3.7) следует, что для каждого $v \geqslant 0$

$$p_{ik}^{(v+1)} = \sum_j p_{ij}^{(v)} p_{jk}.$$

Усредняя по v, получим

$$\frac{1}{n+1} \sum_{v=0}^{n} p_{ik}^{(v+1)} = \sum_j \left(\frac{1}{n+1} \sum_{v=0}^{n} p_{ij}^{(v)} \right) p_{jk}.$$

Левая часть отличается от $1/(n+1) \sum\limits_{v=0}^{n} p_{ik}^{(v)}$ на

$$\frac{p_{ik}^{(n+1)} - p_{ik}^{(0)}}{n+1}.$$

Последняя стремится к нулю при $n \to \infty$. Следовательно, в силу теоремы 9, предел левой части равен w_k. Так как пространство I конечно, в правой части можно почленно перейти к пределу при $n \to \infty$. Это приводит к соотношению

$$w_k = \sum_j w_j p_{jk},$$

которое есть не что иное, как $w = w\Pi$. Итак, п. (i) доказан. Далее, итерируя:

$$w = w\Pi = (w\Pi)\Pi = w\Pi^2 = (w\Pi)\Pi^2 = w\Pi^3 = \ldots, \tag{8.6.8}$$

получаем равенство $w = w\Pi^n$ или, в явном выражении для $n \geqslant 1$,

$$w_k = \sum_j w_j p_{jk}^{(n)}. \tag{8.6.9}$$

Теперь используем то, что $\sum\limits_j p_{ij}^{(v)} = 1$ для каждого i и $v \geqslant 1$. Усредняя по v, выводим

$$\frac{1}{n+1}\sum_{v=0}^{n}\sum_j p_{ij}^{(v)} = 1.$$

Отсюда

$$\sum_j w_j = \sum_j \lim_{n\to\infty}\left(\frac{1}{n+1}\sum_{v=0}^{n} p_{ij}^{(v)}\right) = \lim_{n\to\infty}\frac{1}{n+1}\sum_{v=0}^{n}\sum_j p_{ij}^{(v)} = 1,$$

где второе равенство выполняется ввиду того, что пространство I конечно. Тем самым, п. (ii) установлен. Из него вытекает, что по крайней мере одна из координат w_j, скажем, w_i, является положительной. Для любого k верно отношение $i \rightsquigarrow k$, и поэтому найдется такое n, что $p_{ik}^{(n)} > 0$. Используя данное значение n в сотношении (8.6.9), видим, что координата w_k также положительна. Следовательно, выполняется неравенство (iii). Наконец, пусть x — некоторое решение уравнения (8.6.6). Тогда итерируя, как раньше, выводим, что $x = x\Pi^v$ для каждого $v \geqslant 1$, и, усредняя, получаем равенство

$$x = \frac{1}{n+1}\sum_{v=0}^{n} x\Pi^v.$$

В явной записи это выглядит как

$$x_k = \sum_j x_j \left(\frac{1}{n+1}\sum_{v=0}^{n} p_{jk}^{(v)}\right).$$

Переходя к пределу при $n \to \infty$ и используя теорему 9, устанавливаем, что

$$x_k = \left(\sum_j x_j\right) w_k.$$

Следовательно, утверждение п. (iv) верно при $c = \sum\limits_j x_j$. Теорема 12 полностью доказана.

Будем называть $\{w_j, j \in I\}$ *стационарным распределением* цепи Маркова. Оно на самом деле является вероятностным распределением согласно свойству (ii). Следующий результат объясняет значение слова «стационарное».

Теорема 13. *Допустим, что для каждого* j

$$P\{X_0 = j\} = w_j. \tag{8.6.10}$$

Тогда то же самое верно, если заменить X_0 *на любую из случайных величин* $X_n, n \geqslant 1$. *Более того, совместное распределение*

$$P\{X_{n+v} = j_v, 0 \leqslant v \leqslant l\} \tag{8.6.11}$$

для произвольных j_v *одинаково при всех* $n \geqslant 0$.

Доказательство. В силу соотношения (8.6.9),

$$P\{X_n = j\} = \sum_i P\{X_0 = i\} P_i\{X_n = j\} = \sum_i w_i p_{ij}^{(n)} = w_j.$$

Аналогично находим, что вероятность (8.6.11) равна

$$P\{X_n = j_0\}\, p_{j_0 j_1} \ldots p_{j_{l-1} j_l} = w_{j_l} p_{j_0 j_1} \ldots p_{j_{l-1} j_l},$$

т. е. она одна и та же для всех n.

Таким образом, если взять стационарное распределение в качестве начального распределения цепи, то цепь будет стационарным процессом в том смысле, как это понималось в § 5.4. Интуитивно это означает, что если система оказалась в своем стационарном (в смысле постоянства распределений) состоянии, то она будет пребывать в нем сколь угодно долго. Конечно, в системе происходят изменения, но они сбалансированы и кардинально не нарушают равновесия. К примеру, многие экологические системы претерпевали изменения в течение миллионов лет, и их можно было считать достигшими своей стационарной фазы до тех пор, пока человеческое вмешательство грубо не нарушило эволюционный процесс. Однако, если предположить, что это нарушение — всего лишь флуктуация марковской цепи, то она, в соответствии с нашими теоремами, также должна сгладиться с течением времени.

Практическое значение теоремы 12 состоит в том, что она гарантирует существование решения системы уравнений (8.6.6), которое удовлетворяет условиям (ii) и (iii). Чтобы найти данное решение, можно действовать следующим образом. Отбросить одно из l уравнений и разрешить остальные относительно $w_2, \ldots, w_l$, выражая их через w_1. Они будут иметь вид $w_j = c_j w_1, 1 \leqslant j \leqslant l$, где $c_1 = 1$. Тогда искомое решение задается формулами

$$w_j = c_j \left(\sum_{j=1}^{l} c_j \right)^{-1}, \quad 1 \leqslant j \leqslant l.$$

Пример 13. Выключатель имеет два положения: «включен» и «выключен». Обозначим эти два состояния через 1 и 2. В каждую единицу времени состояние может либо сохраниться, либо измениться, причем соответствующие вероятности зависят только от текущего состояния. Поэтому мы имеем дело с однородной цепью Маркова с пространством состояний $I = \{1, 2\}$ и переходной матрицей

$$\Pi = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

где все элементы матрицы предполагаются положительными. Уравнения стационарности имеют вид

$$\begin{aligned} (1 - p_{11}) x_1 - p_{21} x_2 &= 0, \\ -p_{12} x_1 + (1 - p_{22}) x_2 &= 0. \end{aligned}$$

Очевидно, что второе уравнение совпадает с первым, умноженным на -1, и его можно отбросить. Решая первое уравнение, находим

$$x_2 = \frac{1 - p_{11}}{p_{21}}\, x_1 = \frac{p_{12}}{p_{21}}\, x_1.$$

Отсюда

$$w_1 = \frac{p_{21}}{p_{12} + p_{21}}, \quad w_2 = \frac{p_{12}}{p_{12} + p_{21}}.$$

Ввиду теоремы 9 это означает, что в длинной серии наблюдений переключатель будет находится в положениях «включен» и «выключен» в пропорции $p_{21} : p_{12}$.

Пример 14. На карнавале Даниэль выиграл приз за катание на карусельной лошадке (см. рис. 34). Он участвовал в «бесконечном числе» заездов и при каждом звоне колокольчика пересаживался на соседнюю спереди или сзади лошадку с вероятностями p и $q = 1 - p$. Какую долю времени провел он на каждой из лошадок?

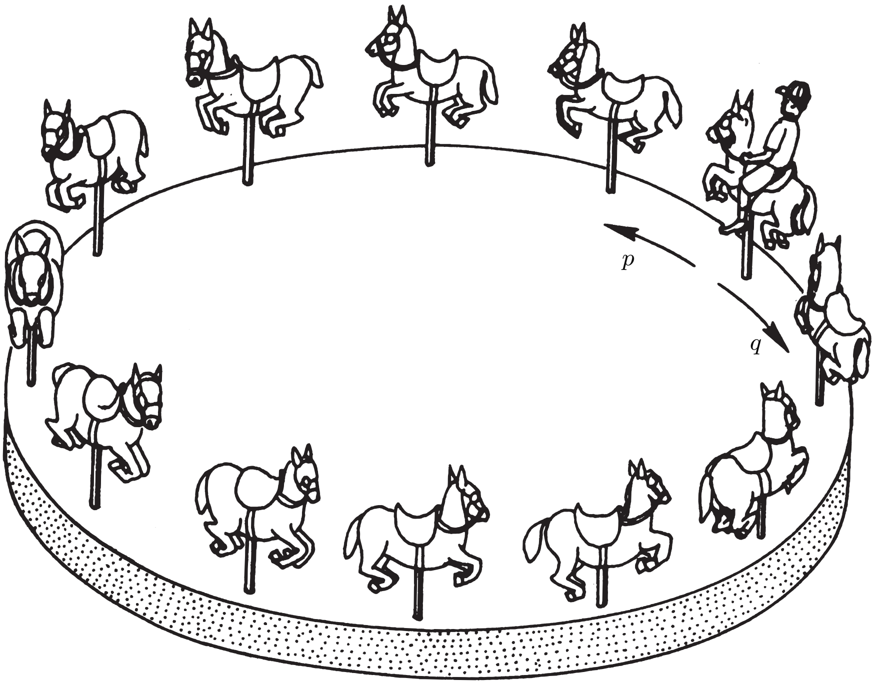


Рис. 34

Это процесс можно описать как «случайное блуждание на окружности». Его переходная матрица выглядит так:

$$\begin{bmatrix} 0 & p & 0 & 0 & \cdots & 0 & 0 & q \\ q & 0 & p & 0 & & 0 & 0 & 0 \\ 0 & q & 0 & p & & 0 & 0 & 0 \\ \hdotsfor{8} \\ 0 & 0 & 0 & 0 & \cdots & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p \\ p & 0 & 0 & 0 & \cdots & 0 & q & 0 \end{bmatrix}$$

Существенной особенностью данной матрицы является то, что сумма элементов каждого столбца (так же, как и каждой строки) равна 1. В наших обозначениях имеем для каждого $j \in I$ соотношение

$$\sum_{i \in I} p_{ij} = 1. \tag{8.6.12}$$

Такие матрицы называются *дважды стохастическими.*

Очевидно, что при условии (8.6.12) вектор $x = (1, 1, \ldots, 1)$, у которого все компоненты равны 1, служит решением системы (8.6.6). Так как в силу п. (iv) теоремы 12 стационарное распределение w должно быть ему пропорционально и, кроме того, удовлетворять условию (iii), то

$$w = \left(\frac{1}{l}, \frac{1}{l}, \ldots, \frac{1}{l}\right),$$

где, как и прежде, l — число состояний в пространстве I. Это означает, что если Даниэль провел 4 часа на карусели с 12 лошадками, то он скакал на каждой из них около 20 минут при условии, что менял лошадок достаточно большое число раз, чтобы можно было применить предельные соотношения (8.6.4).

Для возвратной цепи Маркова с бесконечным пространством состояний теорема 12 заменяется на следующее сильное утверждение о существовании всего лишь двух возможностей:

(a) либо все $w_j > 0$, при этом условия (ii) и (iii) выполняются, как и прежде, и теорема 13 также верна;

(b) либо все $w_j = 0$.

Цепь называют *положительной возвратной* (или *сильно эргодичесой*) в случае (a) и *нулевой возвратной* (или *слабо эргодической*) в случае (b). Примером последней служит симметричное случайное блуждание, обсуждаемое в §§ 8.1–8.2 (см. задачу 38). В монографии [5, § I.7] доказано, что если система (8.6.6) обладает решением $x = (x_1, x_2, \ldots)$, удовлетворяющим условию $0 < \sum\limits_j |x_j| < \infty$, то в действительности все $x_j > 0$, и стационарное распределение имеет вид

$$w_j = \frac{x_j}{\sum\limits_j x_j}, \quad j \in I.$$

Следующий пример иллюстрирует этот результат.

Пример 15. Пусть $I = \{0, 1, 2, \ldots\}$ и $p_{ij} = 0$ при $|i - j| > 1$, а другие p_{ij} — произвольные положительные числа. Тогда для каждого j должно выполняться равенство

$$p_{j,j-1} + p_{jj} + p_{j,j+1} = 1. \tag{8.6.13}$$

Эту модель можно считать частным случаем примера 7 из § 8.3 при $p_{0,-1} = 0$ и соответствующем сокращении пространства состояний. Такую цепь называют *дискретным процессом рождения и гибели*, в котором величина j обозначает размер популяции, а переходы $j \to j + 1$ или

$j \to j-1$ соответствуют отдельному рождению и гибели. Система (8.6.6) превращается в следующую:

$$\begin{aligned} x_0 &= x_0 p_{00} + x_1 p_{10}, \\ x_j &= x_{j-1} p_{j-1,j} + x_j p_{jj} + x_{j+1} p_{j+1,j}, \quad j \geqslant 1. \end{aligned} \tag{8.6.14}$$

Несмотря на то что эта система состоит из бесконечного числа линейных уравнений с постоянными коэффициентами, ее решение можно найти путем последовательного выражения неизвестных $x_1, x_2, \ldots$ через x_0. Действуя описанным образом, получим

$$x_1 = \frac{p_{01}}{p_{10}}\, x_0,$$

$$x_2 = \frac{1}{p_{21}}\{x_1(1-p_{11}) - x_0 p_{01}\} = \frac{p_{01}(1-p_{11}-p_{10})}{p_{21}p_{10}}\, x_0 = \frac{p_{01}p_{12}}{p_{10}p_{21}}\, x_0.$$

Нетрудно догадаться (может быть, после еще нескольких шагов), что в общем случае

$$x_j = c_j x_0, \quad \text{где} \quad c_0 = 1, \quad c_j = \frac{p_{01}p_{12}\cdots p_{j-1,j}}{p_{10}p_{21}\cdots p_{j,j-1}}, \quad j \geqslant 1. \tag{8.6.15}$$

Проверим это с помощью индукции. Допустим, что $p_{j,j-1}x_j = p_{j-1,j}x_{j-1}$. Из равенств (8.6.14) и (8.6.13) выводим, что

$$p_{j+1,j}x_{j+1} = (1-p_{jj})x_j - p_{j-1,j}x_{j-1} = (1-p_{jj}-p_{j,j-1})x_j = p_{j,j+1}x_j.$$

Поэтому данное соотношение выполняется для всех j, и формула (8.6.15) доказана. Тогда

$$\sum_{j=0}^{\infty} x_j = \left(\sum_{j=0}^{\infty} c_j\right) x_0. \tag{8.6.16}$$

Отсюда, при условии возвратности цепи, имеем следующую указанную выше альтернативу (легко видеть, что она выполняется в случае (а)).

Случай (а). Если $\sum\limits_{j=0}^{\infty} c_j < \infty$, то достаточно взять $x_0 = 1$, чтобы получить решение, удовлетворяющее условию $\sum\limits_j x_j < \infty$. Поэтому цепь является положительной возвратной, и ее стационарное распределение имеет вид

$$w_j = \frac{c_j}{\sum\limits_{j=0}^{\infty} c_j}, \quad j \geqslant 0.$$

Случай (b). Если $\sum\limits_{j=0}^{\infty} c_j = \infty$, то при любом выборе значения x_0 в силу равенства (8.6.16) получаем, что либо $\sum\limits_j |x_j| = \infty$, либо $\sum\limits_j |x_j| = 0$. Следовательно, $w_j = 0$ для всех $j \geqslant 0$, и цепь является либо нулевой возвратной, либо невозвратной.

Рассмотренный пример можно модифицировать за счет сокращения пространства состояний до конечного множества, если положить $p_{c,c+1} = 0$ для некоторого $c \geqslant 1$. Частным случаем такой цепи служит модель из примера 8 в § 8.3, к которой мы сейчас вернемся.

Пример 16. Давайте найдем стационарное распределение для модели Эренфестов. Будем действовать точно также, как в примере 15. Это приведет нас к соотношениям (8.6.15), но теперь процесс остановится при $j = c$. Подставляя значения переходных вероятностей (8.3.16), получаем

$$c_j = \frac{c(c-1)\ldots(c-j+1)}{1\cdot 2\ldots j} = \binom{c}{j}, \quad 0 \leqslant j \leqslant c.$$

Из соотношения (3.3.7) имеем, что $\sum\limits_{j=0}^{c} c_j = 2^c$. Отсюда

$$w_j = \frac{1}{2^c}\binom{c}{j}, \quad 0 \leqslant j \leqslant c.$$

Это — известное нам биномиальное распределение $B\,(c;\,1/2)$.

Таким образом, стационарное состояние урны в модели Эренфестов можно моделировать, либо раскрашивая c шаров в красный или черный цвета с вероятностями $1/2$ независимо друг от друга, либо выбирая их наудачу из бесконечно большого резервуара, содержащего красные и черные шары в равных пропорциях.

Далее, вспоминая соотношения (8.6.3) и (8.6.7), видим, что среднее время возвращения задается формулой

$$m_{jj} = 2^c\binom{c}{j}^{-1}, \quad 0 \leqslant j \leqslant c.$$

Для крайних случаев $j = 0$ (нет черных шаров) и $j = c$ (нет красных шаров) оно равно 2^c, что очень велико даже для $c = 100$. Можно установить (см. задачу 42), что ожидаемое время полного изменения цвета содержимого урны на самом деле весьма большое. С другой стороны, цепь является возвратной, поэтому начиная, скажем, с состояния, когда все шары черные, почти наверное в некоторый момент развития процесса Эренфестов все они будут заменены на красные и обратно. Поскольку за один шаг количество черных шаров может измениться только на 1, состав урны обязательно будет проходить через все промежуточные «фазы» снова и снова. Первоначально данная модель предназначалась для демонстрации возвратного физического процесса, имеющего чрезвычайно длинные циклы между возвращениями. «Если ждать достаточно долго, мы снова станем молодыми!»

В заключение опишем без доказательства дальнейшее возможное разложение возвратного класса*). Простейшей иллюстрацией служит классическое случайное блуждание. Его пространство состояний, содержащее все целые числа, может быть разделено на два подкласса — четные и нечетные числа. За один шаг частица обязательно переходит из одного подкласса в другой. Таким образом, чередование двух подклассов является детерминированной особенностью перемещения. В общем случае для каждого возвратного класса C, содержащего по крайней мере два состояния, найдется единственное натуральное число d, называемое *периодом* класса, обладающее следующими свойствами:

(a) для каждого $i \in C$ вероятность $p_{ii}^{(n)} = 0$, если $d \nmid n$**); с другой стороны, $p_{ii}^{(nd)} > 0$ для всех достаточно больших n (насколько больших — зависит от i);

(b) для каждых $i \in C$ и $j \in C$ найдется такое целое число $r, 1 \leqslant r \leqslant d$, что $p_{ij}^{(n)} = 0$, если $d \nmid (n - r)$; с другой стороны, $p_{ij}^{(nd+r)} > 0$ для всех достаточно больших n (насколько больших — зависит от i и j).

Фиксируя состояние i, обозначим через C_r, $1 \leqslant r \leqslant d$, множество всех состояний j, соответствующих одному и тому же числу r в (b). Эти множества не пересекаются, и их объединение есть C. Тогда мы имеем дело со следующими детерминированными циклическими переходами:

$$C_1 \to C_2 \to \ldots \to C_d \to C_1.$$

На рис. 35 приведен пример диаграммы переходов такой цепи с $d = 4$, где переходные вероятности между состояниями указаны с помощью чисел, стоящих над стрелками, соединяющими состояния.

Период d класса C можно найти следующим образом. Возьмем любое $i \in C$ и рассмотрим множество всех $n \geqslant 1$ таких, что $p_{ii}^{(n)} > 0$. Среди общих делителей этого множества найдется наибольший; он и равен d. То, что данное число одинаково для любого выбранного i, является одним из свойств периода. Между прочим, описанное выше разложение

*) На периодические подклассы. — *Прим. перев.*

**) Запись «$d \nmid n$» читается так: «d не делит n» или «n не делится нацело на d».

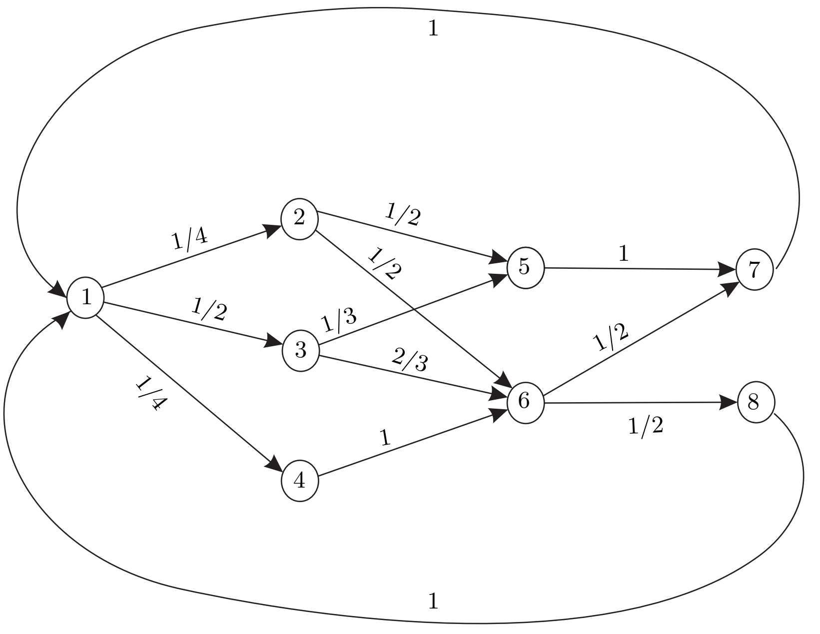


Рис. 35

имеет место для любого стохастически замкнутого класса (определение см. в § 8.7). В частности, неограниченное случайное блуждание обладает периодом 2 независимо от того, возвратно оно или нет.

Когда $d = 1$, класс называется *апериодическим*. Достаточным условием для этого является существование такого целого числа m, что все элементы матрицы Π^m положительны. Тогда из уравнений Колмогорова—Чепмена следует, что то же самое верно для матрицы Π^n при всех $n \geqslant m$. Поэтому свойство (a) влечет равенство $d = 1$. В таком случае фундаментальную предельную теорему, выраженную формулой (8.6.4), можно усилить:

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{m_{jj}}\,; \tag{8.6.17}$$

т. е. предел средних можно заменить индивидуальным пределом. В общем случае, для периода d и состояний i, j, удовлетворяющих условию (b), верна асимптотика

$$\lim_{n\to\infty} p_{ij}^{(nd+r)} = \frac{d}{m_{jj}}\,. \tag{8.6.18}$$

Мы оставляем читателю показать, что если данный предел существует, то он должен быть, в силу соотношения (8.6.4), равен правой части (8.6.18). В действительности, равенство (8.6.18) легко вытекает из формулы (8.6.17), если рассматривать переход за d шагов в качестве нового шага, при котором частица остается внутри фиксированного подкласса. Приведенный выше сильный результат впервые был доказан А. А. Марковым, который изучал только конечные пространства состояний. А. Н. Колмогоров в 1936 г. обобщил его на случай бесконечного пространства. Известны несколько разных доказательств данного результата; одно из них содержится в монографии [5, § I.6].

8.7. Вероятности поглощения

В данном параграфе мы изложим некоторые идеи, касающиеся общего поведения однородной марковской цепи, имеющей как возвратные, так и невозвратные состояния. Пусть R обозначает множество всех возвратных состояний, а T — множество всех невозвратных состояний, тогда $I = R \cup T$. Начнем с полезного определения: множество состояний называется (*стохастически*) *замкнутым* тогда и только тогда, когда частица, стартующая из любого состояния, принадлежащего данному множеству, навсегда остается внутри него. Далее мы будем опускать надоедающую фразу «почти наверное», когда она сама собой разумеется. Перечислим характерные черты глобального движения частицы.

(i) Возвратный класс замкнут. Следовательно, попав в такой класс, частица останется там навсегда.

(ii) Конечное множество невозвратных состояний не замкнуто. Тогда, стартуя из такого множества, в конце концов частица его покинет и будет пребывать где-то вне данного множества.

(iii) Если множество T конечно, то частица в конце концов попадет в один из имеющихся возвратных классов.

(iv) В общем случае частица либо будет поглощена одним из возвратных классов с вероятностью α, либо всегда будет находится внутри множества T с вероятностью $1 - \alpha$, где $0 \leqslant \alpha \leqslant 1$.

Давайте докажем утверждение (i). По теореме 6 частица не может перейти из возвратного ни в какое невозвратное состояние. Она также не может перейти ни в какое возвратное состояние из другого класса из-за того, что два состояния из разных классов не сообщаются по определению (второе не может быть недостижимо из первого в силу теоремы 8). Таким образом, попав в возвратный класс, частица будет двигаться только внутри него. Далее, справедливость утверждения (ii) была установлена в доказательстве теоремы 11, согласно которому частица может находиться лишь ограниченное время в конечном множестве невозвратных состояний. Поэтому в некоторый момент она покинет такое множество. Утверждение (iii) следует из (ii). Оно иллюстрируется примером 3 из § 8.3 (задачей о разорении игрока). Утверждение (iv) представляет очевидную альтернативу для (i) и иллюстрируется примером 1 из § 8.3 при $p > 1/2$ (в данном случае $\alpha = 0$) или примером 9. В последнем случае очевидно, что частица, стартовавшая из $i \geqslant 1$, либо будет поглощена в состоянии 0 с вероятностью f^*_{i0} (в обозначениях (8.4.6)), либо будет бесконечно уходить вправо по невозвратным состояниям $\{i+1, i+2, \ldots\}$ с вероятностью $1 - f^*_{i0}$.

Давайте дополнительно проиллюстрируем некоторые из возможностей на следующем простом числовом примере.

Пример 17. Допустим, что переходная матрица имеет следующий вид:

$$
\begin{array}{c|ccc:cc:cccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & \cdot & \cdot & \cdot \\
\hline
1 & \frac{1}{8} & \frac{3}{8} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \cdot\ \cdot & \cdot \\
2 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} & \frac{1}{6} & 0 & \cdot\ \cdot & \cdot \\
3 & \frac{1}{5} & \frac{3}{10} & 0 & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{10} & \cdot\ \cdot & \cdot \\
\hdashline
4 & & 0 & & \frac{1}{2} & \frac{1}{2} & & 0 & & \\
5 & & & & 1 & 0 & & & & \\
\hdashline
6 & & 0 & & \multicolumn{2}{c:}{0} & & R_2 & & \\
\cdot & & & & & & & & & \\
\cdot & & & & & & & & & \\
\cdot & & & & & & & & &
\end{array}
\tag{8.7.1}
$$

Пространство состояний может быть как конечно, так и бесконечно, в зависимости от вида части R_2, которая сама по себе может быть переходной матрицей любой возвратной цепи Маркова, подобной тем, что рассматривались в примерах 4 или 8 из § 8.3, а также в примере 1 с $p = 1/2$.

Здесь $T = \{1, 2, 3\}, R_1 = \{4, 5\}$ и R_2 — два возвратных класса. Из теории переходов между состояниями следует, что четыре заполненных нулями блока матрицы останутся таковыми при возведении матрицы в любую степень. Попытайтесь убедиться в этом путем нескольких формальных умножений. С другой стороны, некоторые из одиночных нулей могут превратиться в положительные числа в процессе умножения. В данной цепи существуют два невозвратных класса: $\{1, 3\}$ и $\{2\}$. Возможно перейти из первого во второй, но не наоборот. (Это не важно; на самом деле незамкнутый невозвратный класс — не очень полезный объект. Он был назван классом в § 8.4 исключительно в силу обстоятельств!) Из всех трех невозвратных состояний можно попасть как в класс R_1, так и в класс R_2, но несложно было бы добавить к цепи еще и другие состояния, из которых можно попасть только в один из этих классов. Задача вычисления разнообразных вероятностей поглощения будет решена на основе общей процедуры, излагаемой ниже.

Пусть $i \in T$ — невозвратное состояние, а C — некоторый возвратный класс. Положим для $n \geqslant 1$

$$y_i^{(n)} = \sum_{j \in C} p_{ij}^{(n)} = P_i\{X_n \in C\}. \tag{8.7.2}$$

Это есть вероятность того, что частица окажется в C в момент n, при условии, что она стартовала из i. Так как класс C замкнут, то частица также будет находиться в нем в момент $n+1$. Поэтому $y_i^{(n)} \leqslant y_i^{(n+1)}$. Тогда, согласно теореме о монотонной последовательности из математического анализа, при $n \to \infty$ существует предел

$$y_i = \lim_{n \to \infty} y_i^{(n)} = P_i\{X_n \in C \text{ для некоторого } n \geqslant 1\}$$

(почему верно второе равенство?), который равен вероятности поглощения.

Теорема 14. *Определенные выше вероятности $\{y_i\}$ удовлетворяют системе уравнений*

$$x_i = \sum_{j \in T} p_{ij} x_j + \sum_{j \in C} p_{ij}, \quad i \in T. \tag{8.7.3}$$

Если множество невозвратных состояний T конечно, то $\{y_i\}$ являются компонентами единственного решения системы. Оно вычисляется с помощью стандартных методов линейной алгебры.

Доказательство. Пусть частица стартует из состояния i. Рассмотрим ее положение j после одного шага. Если $j \in T$, то в силу марковского свойства условная вероятность поглощения становится равной y_j; если $j \in C$, то частица уже поглощена; если $j \in (I - T) - C$, то она никогда не будет поглощена классом C. Принимая в расчет все возможности, получаем

$$y_i = \sum_{j \in T} p_{ij} y_j + \sum_{j \in C} p_{ij} \cdot 1 + \sum_{j \in (I-T)-C} p_{ij} \cdot 0.$$

Это доказывает первое утверждение теоремы. Допустим теперь, что T есть конечное множество $\{1, 2, \ldots, t\}$. Тогда систему (8.7.3) можно записать в матричной форме:

$$(\Delta_T - \Pi_T)x = y^{(1)}, \tag{8.7.4}$$

где Δ_T — единичная матрица размера $T \times T$; Π_T — ограничение переходной матрицы Π размера $T \times T$, $y^{(1)}$ — вектор-столбец с компонентами, заданными формулой (8.7.2) при $n = 1$. В соответствии с известным результатом из линейной алгебры, данное уравнение имеет единственное решение тогда и только тогда, когда матрица $\Delta_T - \Pi_T$ невырождена, т. е. когда существует обратная матрица $(\Delta_T - \Pi_T)^{-1}$. В таком случае решение имеет вид

$$(\Delta_T - \Pi_T)^{-1} y^{(1)}. \tag{8.7.5}$$

Допустим противное. Тогда согласно тому же результату существует ненулевое решение соответствующей однородной системы. Иначе говоря, найдется вектор-столбец $v = (v_1, \ldots, v_t) \neq (0, \ldots, 0)$, удовлетворяющий условию

$$(\Delta_T - \Pi_T)v = 0 \quad \text{или} \quad v = \Pi_T v.$$

Итерируя, получаем:

$$v = \Pi_T(\Pi_T v) = \Pi_T^2 v = \Pi_T^2(\Pi_T v) = \Pi_T^3 v = \ldots,$$

и поэтому для любого $n \geqslant 1$ имеем:

$$v = \Pi_T^n v$$

(сравните с формулой (8.6.8), но обратите внимание на различие между умножением слева на строку и справа на столбец). Полученное равенство означает, что

$$v_i = \sum_{j \in T} p_{ij}^{(n)} v_j, \quad i \in T.$$

Устремляя $n \to \infty$ и используя следствие теоремы 5, видим, что каждый член суммы сходится к нулю. Отсюда заключаем, что $v_i = 0$ для всех $i \in T$, а это противоречит нашему предположению. Данное противоречие позволяет сделать вывод, что матрица $\Delta_T - \Pi_T$ невырождена. Поэтому существует единственное решение системы (8.7.4), задаваемое выражением (8.7.5). Теорема доказана.

В примере 17 уравнения (8.7.3) для вероятностей поглощения в классе R_1 приобретают следующий вид:

$$\begin{aligned} x_1 &= \frac{1}{8}\,x_1 + \frac{3}{8}\,x_2 + \frac{1}{4}\,x_3 + \frac{1}{4}\,, \\ x_2 &= \qquad\quad \frac{1}{2}\,x_2 \qquad\qquad + \frac{1}{3}\,, \\ x_3 &= \frac{1}{5}\,x_1 + \frac{3}{10}\,x_2 \qquad\quad + \frac{2}{5}\,. \end{aligned}$$

Из второго уравнения сразу находим x_2, а затем вычисляем x_1, x_3 из остальных:

$$x_1 = \frac{26}{33}\,, \quad x_2 = \frac{2}{3}\,, \quad x_3 = \frac{25}{33}\,.$$

Для каждого i вероятности поглощения в классах R_1 и R_2 в сумме дают 1. Отсюда вероятности поглощения в R_2 равны $1 - x_1$, $1 - x_2$, $1 - x_3$. Они составляют единственное решение системы уравнений для R_2, которая отличается от приведенной выше тем, что константы в столбце справа заменяются на 0, 1/6, 1/10. Возможно, вы захотите проверить это. (Полезная привычка — проводить двойную проверку подобных утверждений хотя бы единожды в новой ситуации.)

Поучительно описать, как задача поглощения в возвратном классе сводится к задаче поглощения в некотором состоянии. Действительно, все состояния возвратного класса можно объединить в одно новое состояние, так как нас не интересуют переходы внутри класса. Ни из какого состояния класса нельзя выйти наружу, в то время как вероятность попадания в класс за один шаг из произвольного невозвратного состояния i в точности равна $y_i^{(1)}$. В результате матрица (8.7.1) преобразуется в следующую:

$$\begin{bmatrix} \frac{1}{8} & \frac{3}{8} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{5} & \frac{3}{10} & 0 & \frac{2}{5} & \frac{1}{10} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

в которой последние два состояния $\{4\}$ и $\{5\}$ заняли место классов R_1 и R_2. Вероятностями поглощения теперь являются f^*_{i4} и f^*_{i5} в обозначениях (8.4.6). Обе системы уравнений, конечно, остаются прежними.

В случае, когда множество невозвратных состояний T конечно и имеются в точности два поглощающих состояния, существует другой интересный метод. Как и раньше, пусть $T = \{1, 2, \ldots, t\}$. Будем считать, что поглощающими состояниями являются 0 и $t+1$, поэтому $I = \{0, 1, \ldots, t+1\}$. Метод основан на нахождении такого решения системы уравнений $(\Delta - \Pi)x = 0$, не все компоненты которого одинаковы, т. е. вектора $v = (v_0, v_1, \ldots, v_{t+1})$, удовлетворяющего системе уравнений

$$v_i = \sum_{j=0}^{t+1} p_{ij} v_j, \quad i = 0, 1, \ldots, t+1. \tag{8.7.6}$$

Заметим, что два уравнения, соответствующие $i = 0$ и $i = t+1$, автоматически выполняются для любого v, поскольку $p_{0j} = \delta_{0j}$ и $p_{t+1,j} = \delta_{t+1,j}$. Вектор, все компоненты v_i которого равны 1, всегда является решением системы, но нас интересуют другие решения. Итерируя, выводим равенства

$$v_i = \sum_{j=0}^{t+1} p_{ij}^{(n)} v_j$$

для всех $n \geqslant 1$. Устремляя $n \to \infty$ и замечая, что

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \qquad \text{при } 1 \leqslant j \leqslant t,$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = f_{ij}^* \quad \text{при } j = 0 \quad \text{и} \quad j = t+1,$$

получаем

$$v_i = f_{i0}^* v_0 + f_{i,t+1}^* v_{t+1}. \tag{8.7.7}$$

Вспомним также соотношение

$$1 = f_{i0}^* + f_{i,t+1}^*. \tag{8.7.8}$$

Мы утверждаем, что $v_0 \neq v_{t+1}$. В противном случае из двух последних уравнений будет следовать, что $v_i = v_0$ для всех i, а это противоречит предположению о различии компонент вектора v. Отсюда находим вероятности поглощения:

$$f_{i0}^* = \frac{v_i - v_{t+1}}{v_0 - v_{t+1}}; \quad f_{i,t+1}^* = \frac{v_0 - v_i}{v_0 - v_{t+1}}. \tag{8.7.9}$$

Пример 18. Давайте вернемся к задаче 1 из § 8.1, где $t = c - 1$. Если $p \neq q$, то $v_i = (q/p)^i$ является решением системы (8.7.6), имеющим различные компоненты. Это легко проверяется, но читатель вправе потребовать объяснений: каким-таким образом мы открыли данное решение? В нашем случае ответ прост (однако он основан на знании разностных уравнений из § 8.1): попробуйте найти решение в форме λ^i и увидеть, каково должно быть λ. Теперь, если подставить такие v_i в соотношения (8.7.9), то получим, что f^*_{i0} равны u_i из формулы (8.1.9).

Если же $p = q = 1/2$, то $v_i = i$ образуют отличное от константы решение системы (8.7.6), поскольку

$$i = \frac{1}{2}\,(i+1) + \frac{1}{2}\,(i-1). \tag{8.7.10}$$

Это приводит к тому же самому ответу, который дается формулами (8.1.10). Новое решение имеет отношение к идее мартингала (см. приложение 3). Вот еще один похожий пример.

Пример 19. Следующая модель случайного размножения была введена С. Райтом в его работах по генетике (см., например, монографию [15] для детального ознакомления). У гаплоидного организма гены передаются по одному, а не парами, как в диплоидном случае, рассмотренном в § 5.6. Допустим, что $2N$ генов типов A и a (аллелей) выбираются из каждого поколения. Число A-генов — состояние марковской цепи с пространством состояний $I = \{0, 1, \ldots, 2N\}$ и переходными вероятностями

$$p_{ij} = \binom{2N}{j}\left(\frac{i}{2N}\right)^{i}\left(1 - \frac{i}{2N}\right)^{2N-j}. \tag{8.7.11}$$

Таким образом, если число A-генов в каком-то поколении равно i, то можно представить, что существует бесконечная совокупность генов обоих типов, в которой гены типа A присутствуют в пропорции $i : (2N - i)$ относительно типа a. Производится $2N$ независимых извлечений из этой совокупности, в результате которых возникают гены следующего поколения. Тем самым, мы имеем дело с $2N$ независимыми бернуллиевскими испытаниями с вероятностью «успеха» $i/2N$, что приводит к биномиальному распределению $B(2N;\, i/2N)$ с вероятностями (8.7.11). Отсюда следует (см. формулы(4.4.16) или (6.3.6)), что ожидаемое число A-генов равно

$$\sum_{j=0}^{2N} j\,p_{ij} = 2N\,\frac{i}{2N} = i. \tag{8.7.12}$$

Это означает, что ожидаемое число A-генов в следующем поколении равно имеющемуся (но случайному) числу таких генов в текущем поколении. Поэтому оно остается постоянным во всех поколениях. Ситуация здесь та же самая, что и в модели справедливой игры, обсуждаемой в § 8.2 после формулы (8.2.3). Использованный там недоказанный прием также годится в нашем случае и ведет к тому же самому (с точностью до обозначений) выводу. Однако теперь мы можем применить доказанную формулу (8.7.9) и сразу получить, что

$$f_{i0}^* = \frac{2N - i}{2N}, \quad f_{i,2N}^* = \frac{i}{2N}.$$

Это и есть вероятности превращения всей популяции в совокупность генетически одинаковых организмов, имеющих гены типа a или A соответственно.

Наш заключительный пример посвящен однородной марковской цепи специального, но важного вида. Другой специфический пример — модель из теории очередей*) — очерчен с подробными указаниями в задачах 29–31.

Пример 20. Элементарная частица в результате ядерной реакции может распасться на несколько новых элементарных частиц; ребенок мужского пола, являющийся носителем фамилии, когда вырастет, станет отцом какого-то количества (в частности, нуля) сыновей. Эти процессы могут повторяться многократно, если не произойдет вырождения. Они служат примерами применения определяемой ниже модели *ветвящегося процесса* (см. рис. 36).

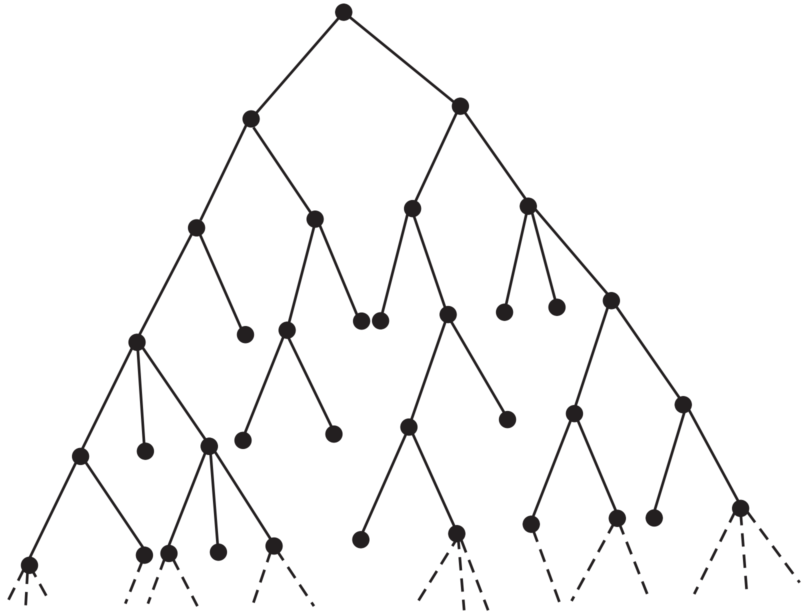


Рис. 36

*) Называемой также теорией массового обслуживания. — *Прим. перев.*

Без ограничения общности будем считать, что в начале процесса есть всего одна частица: $X_0 = 1$. Она дает X_1 потомков в первом поколении, где

$$P(X_1 = j) = a_j, \quad j = 0, 1, 2, \ldots . \tag{8.7.13}$$

Если $X_1 \neq 0$, то каждая из частиц первого поколения производит некоторое количество потомков, имеющее одинаковое для всех частиц распределение (8.7.13). При этом предполагается, что частицы размножаются независимо друг от друга. Каково распределение количества частиц во втором поколении?

Обозначим через g производящую функцию случайной величины X_1:

$$g(z) = \sum_{j=0}^{\infty} a_j z^j.$$

Допустим, что число частиц в первом поколении равно j. Обозначим количества их потомков через $Z_1, \ldots, Z_j$. Согласно нашему предположению, они являются независимыми и одинаково распределенным случайными величинами с производящей функцией g. Общее число частиц во втором поколении есть $X_2 = Z_1 + \ldots + Z_j$. В силу теоремы 6 из § 6.5, производящей функцией для X_2 является g^j. Вспоминая формулу (6.5.12) и определение условного математического ожидания (5.2.11), запишем эту производящую функцию в виде

$$E(z^{X_2} \mid X_1 = j) = g(z)^j, \tag{8.7.14}$$

отсюда, с учетом соотношения (5.2.12), выводим, что

$$E(z^{X_2}) = \sum_{j=0}^{\infty} P(X_1 = j)\, E(z^{X_2} \mid X_1 = j) = \sum_{j=0}^{\infty} a_j g(z)^j = g(g(z)).$$

Пусть g_n — это производящая функция X_n, в частности $g_1 = g$. Тогда мы установили выше, что $g_2 = g(g_1)$. В точности такое же рассуждение приводит к соотношению $g_n = g(g_{n-1}) = g \circ g \circ \cdots \circ g$ (здесь g появляется n раз), где «$\circ$» обозначает суперпозицию функций. Другими словами, g_n является n-кратной суперпозицией функции g с самой собой. Используя определение g_n, можно представить ее также в следующем виде:

$$g_n(z) = E(z^{X_n}) = \sum_{k=0}^{\infty} P(X_n = k) z^k. \tag{8.7.15}$$

Ввиду того, что распределение количества потомков в каждом последующем поколении полностью определяется количеством частиц в текущем поколении и не зависит от прошлого течения процесса, понятно, что последовательность $\{X_n, n \geqslant 0\}$ обладает марковским свойством. Она является однородной цепью Маркова благодаря тому, что закон размножения частиц не меняется от поколения к поколению. Из формулы (8.7.14) находим переходные вероятности цепи:

$$p_{jk} = \text{коэффициент при } z^k \text{ в разложении } g(z)^j \text{ в степенной ряд.} \tag{8.7.16}$$

Для исключения тривиальностей будем считать выполняющимися условия

$$0 < a_0 < a_0 + a_1 < 1. \tag{8.7.17}$$

Пространством состояний цепи служит множество всех целых неотрицательных чисел (почему?). Указанные условия обеспечивают то, что состояние 0 достижимо из любого другого состояния (почему?) и является поглощающим. Поэтому все другие состояния оказываются невозвратными, но их бесконечно много. Общее поведение частицы не описывается п. (iii) из начала параграфа, но подпадает под п. (iv). (Отметим, что слово «частица» используется там в другом смысле.) В действительности мы сейчас перейдем к определению величины α, которая в данной модели называется *вероятностью вырождения*.

Полагая в соотношении (8.7.15) $z = 0$, видим, что $g_n(0) = p_{10}^{(n)}$. С другой стороны, из нашего общего обсуждения задачи о поглощении известно, что

$$\alpha = \lim_{n\to\infty} p_{10}^{(n)} = \lim_{n\to\infty} g_n(0). \tag{8.7.18}$$

Так как $g_n(0) = g(g_{n-1}(0))$, устремляя $n \to \infty$, получаем

$$\alpha = g(\alpha). \tag{8.7.19}$$

Таким образом, искомая вероятность является корнем уравнения

$$\varphi(z) = 0, \quad \text{где } \varphi(z) = g(z) - z.$$

Мы будем кратко называть его корнем функции φ. Поскольку $g(1) = 1$, то $z = 1$ — один из корней. Далее,

$$\varphi''(z) = g''(z) = \sum_{j=2}^{\infty} j(j-1)a_j z^{j-2} > 0$$

при $z > 0$ в предположении выполнения неравенств (8.7.17). Поэтому производная φ' является возрастающей функцией. Вспомним *теорему Ролля* из математического анализа: между двумя корнями дифференцируемой функции существует по крайней мере один корень ее производной. Следовательно, функция φ не может иметь больше двух корней на отрезке $[0, 1]$, так как в противном случае производная φ' обладала бы больше чем одним корнем, что невозможно ввиду возрастания φ'. Таким образом, функция φ может иметь максимум один отличный от 1 корень на отрезке $[0, 1]$, и мы должны рассмотреть два случая[*)].

Случай 1. Функция φ не имеет корней на промежутке $[0, 1)$. Тогда, поскольку $\varphi(0) = a_0 > 0$, то должно выполняться неравенство $\varphi(z) > 0$ для всех z из $[0, 1)$, так как непрерывная функция не может принимать положительные и отрицательные значения в некотором интервале и нигде не обращаться в 0. Поэтому

$$\varphi(1) - \varphi(z) < \varphi(1) = 0, \quad 0 \leqslant z < 1;$$

следовательно,

$$\varphi'(1) = \lim_{z \to 1-} \frac{\varphi(1) - \varphi(z)}{1 - z} \leqslant 0,$$

что равносильно неравенству $g'(1) \leqslant 1$.

Случай 2. Функция φ имеет единственный корень r на промежутке $[0, 1)$. Тогда, в силу теоремы Ролля, производная φ' должна иметь корень s в $[r, 1)$, т. е. $\varphi'(s) = g'(s) - 1 = 0$. Так как g' — возрастающая функция, то

$$g'(1) > g'(s) = 1.$$

Подведем итоги: уравнение $g(z) = z$ тогда и только тогда имеет положительный корень строго меньший 1, когда $g'(1) > 1$.

В случае 1 имеем $\alpha = 1$, поскольку $0 \leqslant \alpha \leqslant 1$, и α является корнем ввиду соотношения (8.7.19). Таким образом, почти наверное популяция выродится.

В случае 2 покажем, что α совпадает с корнем $r < 1$. Учитывая, что g — возрастающая функция, имеем $g(0) < g(r) = r$. Сделаем индуктивное предположение, что $g_{n-1}(0) < r$. Тогда

$$g_n(0) = g(g_{n-1}(0)) < g(r) = r$$

ввиду возрастания функции g. Значит, $g_n(0) < r$ для всех n, и поэтому $\alpha \leqslant r$, в силу соотношений (8.7.18). Но тогда α должно быть равно r, так как оба эти числа являются корнями уравнения на промежутке $[0, 1)$.

Что происходит в случае 2, если популяция избежит вырождения? В соответствии с общим поведением, описанном в п. (iv), размер по-

---

[*)] Обычно рисуют картинки с этими случаями. Приглашаем читателя сделать это и выяснить, помогут ли они ему лучше понять последующие рассуждения.

пуляции будет всегда пребывать в множестве невозвратных состояний $\{1, 2, \ldots\}$ с вероятностью $1-\alpha$. Будет ли он беспорядочно колебаться вверх и вниз от малых значений до больших и обратно? Ответ на этот вопрос дает описание поведения в п. (ii), согласно которому численность популяции обязана покинуть любое конечное множество $\{1, 2, \ldots, \ell\}$ рано или поздно, сколь велико бы ни было ℓ. Поэтому в действительности популяция должна стать бесконечно большой (не обязательно монотонно, а в пределе), т. е.

$$P\{\lim_{n\to\infty} X_n = +\infty \mid X_n \neq 0 \text{ для всех } n\} = 1.$$

Данный вывод называют «бум»-синдромом. То же самое происходит с игроком, имеющим превосходство над бесконечно богатым противником (см. § 8.2): если он не разоряется, то становится в конце концов бесконечно богатым. Теория вероятностей содержит множество таких экстремальных результатов; некоторые из них известны как *законы нуля или единицы* (тем самым, реализуется принцип «все или ничего»).

В нашем случае имеется простое объяснение полученного выше результата. Давайте вычислим математическое ожидание размера популяции в n-м поколении. Пусть $\mu = E(X_1)$ обозначает среднее число потомков каждой отдельной частицы. Заметим, что $\mu = g'(1)$. Поэтому $\mu \leqslant 1$ в случае 1 и $\mu > 1$ в случае 2. Предположим, что $\mu < \infty$. Тогда, если количество частиц в $(n-1)$-м поколении равно j, ожидаемое число частиц в n-м поколении будет равно μj (почему?). В терминах условных математических ожиданий это можно записать так:

$$E\{X_n \mid X_{n-1} = j\} = \mu j.$$

Из формулы (5.2.12) следует, что

$$E(X_n) = \sum_{j=0}^{\infty} \mu j\, P(X_{n-1} = j) = \mu\, E(X_{n-1}).$$

Отсюда, итерируя, выводим формулу

$$E(X_n) = \mu^n E(X_0) = \mu^n.$$

Устремляя $n \to \infty$, получаем, что

$$\lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} \mu^n = \begin{cases} 0, & \text{если } \mu < 1, \\ 1, & \text{если } \mu = 1, \\ \infty, & \text{если } \mu > 1. \end{cases}$$

Этот результат подтверждает наше заключение о неизбежности вырождения в случае 1. На самом деле интуитивно очевидно, что если $\mu < 1$, то популяция не обладает свойством самовоспроизводства на уровне средних. Случай $\mu = 1$, возможно, не столь очевиден. Нетрудно заметить, что для него наблюдается странная ситуация, когда $E(X_n) = 1$ для всех n, но $P\left(\lim\limits_{n\to\infty} X_n = 0\right) = 1$ ввиду принадлежности к случаю 1. Если же $\mu > 1$, грубая интерпретация полученного результата состоит в том, что популяция обязательно станет бесконечно большой. Однако мы установили, что в случае 2 в качестве ужасной альтернативы возможно вырождение с положительной вероятностью. Результат также интересен тем, что он связывает простые вычисления с более глубокой теорией. Последнее замечание предназначено для приглашения читателя к дальнейшему знакомству с вероятностью и ее смыслом.

Задачи

1. Пусть случайные величины X_n заданы соотношениями (8.1.2) с $X_0 = 0$. Найдите следующие вероятности:
   (a) $P\{X_n \geqslant 0$ для $n = 1, 2, 3, 4\}$;
   (b) $P\{X_n \neq 0$ для $n = 1, 2, 3, 4\}$;
   (c) $P\{X_n \leqslant 2$ для $n = 1, 2, 3, 4\}$;
   (d) $P\{|X_n| \leqslant 2$ для $n = 1, 2, 3, 4\}$.

2. Пусть $Y_n = X_{2n}$, где случайные величины X_n такие же, как в задаче 1. Покажите, что последовательность $\{Y_n, n \geqslant 0\}$ является цепью Маркова и найдите ее переходную матрицу. Сделайте то же самое для последовательности $\{Z_n, n \geqslant 0\}$, где $Z_n = X_{2n+1}$. Каково начальное распределение последней цепи?

3. Монета бросается бесконечное число раз. Пусть H_n и T_n обозначают, соответственно, количества «орлов» и «решек», выпавших в первых n испытаниях. Возьмем $X_n = H_n$, $Y_n = H_n - T_n$. Образуют ли они цепи Маркова? Если да, то найдите переходные матрицы.

4*. Для случайных величин, определенных в задаче 3, положим $Z_n = |H_n - T_n|$. Будет ли эта последовательность цепью Маркова? [*Указание.* Вычислите, например, $P\{Y_{2n} = 2i \mid Z_{2n} = 2i\}$ с помощью формулы Бернулли, затем $P\{Z_{2n+1} = 2i \pm 1 \mid Z_{2n} = 2i, Y_{2n} > 0\}$.]

5. Допустим, что переходные матрицы имеют вид:

$$\text{(a)} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\[2mm] \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} p_1 & q_1 & 0 \\ 0 & p_2 & q_2 \\ q_3 & 0 & p_3 \end{pmatrix}.$$

Найдите $f_{11}^{(n)}$, $f_{12}^{(n)}$, $g_{12}^{(n)}$ для $n = 1, 2, 3$ (эти величины определены формулами (8.4.6) и (8.4.10)).

6. В модели, описывающей *процесс дрессировки* крыс, разработанной Истсом, грызун считается находящимся либо в состоянии 1, если он разучил определенный трюк (достал лакомство или избежал удара током), либо в состоянии 2, если грызун еще не научился его выполнять. Допустим, что как только трюк освоен, он осваивается навсегда. Если же он еще не освоен, то грызун обучается ему в ходе очередного урока с вероятностью α. Выпишите переходную матрицу и вычислите вероятности $p_{21}^{(n)}$, $f_{21}^{(n)}$ для всех $n \geqslant 1$, а также m_{21} (см. (8.6.3)).

7. Убедитесь, что нетрудно придумать переходную матрицу, в которой имеется любое заданное число возвратных и невозвратных классов, где каждый из классов содержит заданное число состояний, при условии, что (a) пространство I бесконечно, (b) I конечно, но не все состояния невозвратны.

8. Для произвольной переходной матрицы Π покажите, что ее легко увеличить за счет добавления новых состояний, из которых будут достижимы старые, но нельзя добавить новое состояние, которое сообщается со всеми старыми.

9. В игре «удвоение или проигрыш» вы в каждой партии ставите на кон весь имеющийся капитал. При этом шансы удвоить и потерять весь капитал одинаковы. Пусть игра началась с \$1, и вы решили сыграть не более n партий (возможно, игра закончится раньше, если вы разоритесь). Опишите соответствующую цепь Маркова с помощью ее переходной матрицы.

10. Некий человек играет в «орла и решку». Но монета несимметрична и выпадает «орлом» с вероятностью 0.48. Человек решил, что закончит игру, как только окажется в выигрыше. Какова вероятность того, что игра никогда не закончится?

11. У молодого человека есть две подружки, живущие в разных концах города: первая — на севере, вторая — на юге. Когда он хочет поехать на выходные в гости к одной из них, то выбирает с вероятностью p девушку, живущую на севере города, а с вероятностью $1 - p$ — девушку, живущую на юге. Между двумя визитами он остается на выходные дома. Опишите цепь Маркова, имеющую три состояния (местопребывание во время выходных): «на севере города», «дома» и «на юге города». Найдите частоты каждого из состояний в длинной серии недель. (Это — простейший случай примера 4 из § 8.3. Приведем любопытный парадокс, связанный с данной схемой. Допустим, что молодой человек решил положиться на волю случая при выборе направления поездки: он знал, что от ближайшей к нему автобусной остановки автобусы отправляются каждые 15 минут на север и на юг, поэтому он считал, что, садясь на первый подошедший автобус, будет случайно выбирать направление с вероятностью $p = 1/2$. Но через некоторое время молодой человек обнаружил, что гостит на севере вдвое чаще, чем на юге. Как такое могло произойти? Данный парадокс дает важный урок прикладным статистикам, заключающийся в том, что данные на самом деле могут быть вовсе не такими, как ожидается на первый взгляд. Представим, что молодой человек приходит на автобусную остановку в случайный момент между 6 и 8 часами вечера. Укажите конкретное расписание движения автобусов, при котором он будет отправляться на север с вероятностью $p = 2/3$.)

12. Решите задачу 1 из § 8.1 в случае, когда на каждом шаге существует положительная вероятность r того, что частица останется в текущей позиции.

13.* Получите решение системы (8.1.13) при $p \neq q$ следующим методом. Сначала найдите две такие константы λ_1 и λ_2, что $x_j = \lambda^j$ удовлетворяют соотношениям $x_j = px_{j+1} + qx_{j-1}$. Тогда общее решение данной системы имеет вид $A\lambda_1^j + B\lambda_2^j$, где A и B — некоторые константы. Затем найдите частное решение соотношений $x_j = px_{j+1} + qx_{j-1} + 1$, подставляя в них $x_j = Cj$ и определяя константу C. Тогда общим решением последней системы является $A\lambda_1^j + B\lambda_2^j + Cj$. Наконец, найдите A и B из граничных условий в (8.1.13).

14. На самом деле модель Эренфестов выглядела следующим образом. В каждой из двух урн содержится по N шаров. Наудачу выбирается шар из всех $2N$ шаров и перекладывается в другую урну. Пусть X_n обозначает количество шаров в фиксированной урне после n извлечений. Покажите, что случайные величины X_n образуют цепь Маркова с переходными вероятностями (8.3.16) при $c = 2N$.

15. Схема, похожая на модель из задачи 14, использовалась Даниилом Бернулли (сыном Иоганна, который, в свою очередь, был младшим братом Якоба Бернулли) и Лапласом для изучения потока несжимаемой жидкости, текущей из одного сосуда в другой. Пусть в первой урне содержится N красных шаров, во второй урне — N черных шаров. В каждой из урн наудачу выбирается шар и перекладывается в другую урну. Найдите переходные вероятности для количества красных шаров в заданной урне.

16. В определенных испытаниях, подобных решению задач из домашнего задания, полученный успех ведет к увеличению шансов на достижение следующего успеха ввиду возрастания опыта и уверенности в себе. В других испытаниях может иметь место обратная тенденция. В любом случае предположим, что последействие затрагивает только два испытания, следующих одно за другим. Поэтому успехи и неудачи образуют марковскую цепь с двумя состояниями $\{s, f\}$. Пусть $p_{ss} = \alpha$, $p_{ff} = \beta$, где α и β — произвольные числа между 0 и 1. Найдите частоту успехов в длинной серии испытаний.

17. Следующая модель используется для изучения процесса *распространения заболевания*. Допустим, что среди N человек есть больные гриппом. Сделаем некоторые предположения:

(a) при встрече больного человека со здоровым шансы последнего получить инфекцию равны α;

(b) все контакты происходят только между двумя людьми;

(c) все возможные объединения в пары равновероятны;

(d) контакты происходят один за другим через одинаковые временные интервалы.

Определите марковскую цепь, описывающую процесс распространения заболевания, и выпишите ее переходную матрицу. (Вы подавлены всеми этими упрощающими реальность предположениями? Вся прикладная математика построена на удачном выборе и применении подобных упрощенных моделей.)

18. Время горения электрической лампочки измеряется в днях, доли дня не учитываются. Если лампочка перегорает, то она заменяется на новую в начале следующего дня. Допустим, что исправная в начале дня лампочка (возможно, только что установленная) с вероятностью p не перегорает по крайней мере в течение этого дня, при этом ее «возраст» увеличивается на 1. Предположим также, что сроки службы последовательно заменяемых лампочек независимы. Пусть $X_0 = 0$ и X_n обозначает «возраст» той лампочки, которая используется в начале $(n+1)$-го дня. (Отсчет ведется, начиная с первого дня, поэтому $X_1 = 1$ или 0 в зависимости от того, останется ли на месте первоначальная лампочка, или же она будет заменена на новую в начале второго дня.) Последовательность $\{X_n, n \geqslant 0\}$ дает пример так называемого *процесса восстановления*. Покажите, что она является возвратной марковской цепью. Найдите ее переходные вероятности и стационарное распределение. (Заметим, что продолжительность службы лампочки на самом деле описывается непрерывной случайной величиной. Из-за этого потребовалось так много слов для аккуратной формулировки модели с дискретным временем и удаления двусмысленностей. Возможно, было бы проще и яснее представить задачу в терминах «орлов» и «решек», выпадающих при бросании монеты (как?), но тогда она потеряла бы практическую содержательность!)

19. Найдите стационарное распределение для случайного блуждания с двумя отражающими границами (пример 4 из § 8.3).

20. Следующая цепь Маркова использовалась Б. Коэном как модель в социологическом исследовании «приверженности». У нее четыре состояния: S_1 = решительный противник; S_2 = нерешительный противник; S_3 = нерешительный сторонник; S_4 = решительный сторонник. Обнаружилось, что в ходе группового эксперимента испытуемые после каждого мероприятия меняли свои взгляды в соответствии со следующей переходной матрицей:

	S_1	S_2	S_3	S_4
S_1	1	0	0	0
S_2	0.06	0.76	0.18	0
S_3	0	0.27	0.69	0.04
S_4	0	0	0	1

Найдите вероятности перемещения в конце концов из «неустойчивых» состояний S_2 и S_3 в «устойчивые» состояния S_1 и S_4.

21. Рассмотрим генетическую модель, похожую на модель из примера 19 в § 8.7, с $I = \{0, 1, \ldots, 2N\}$ и

$$p_{ij} = \binom{2i}{j}\binom{2N-2i}{N-j}\bigg/\binom{2N}{N}.$$

Как бы вы описали изменение генотипов от поколения к поколению с помощью некоторой урновой схемы? Найдите вероятности поглощения. [*Указание.* Вычислите $\sum\limits_{j=0}^{2N} j\,p_{ij}$ путем упрощения биномиальных коэффициентов или на основе теоремы 1 из § 6.1.]

22. Найдите вероятность вырождения для ветвящегося процесса из примера 20 в § 8.7, у которого a_0, a_1 и a_2 положительны, а остальные a_j равны 0.

23. Допустим, что частицы в первом поколении ветвящегося процесса размножаются с вероятностями $\{b_j, j \geqslant 0\}$, которые не обязательно совпадают с вероятностями (8.7.13) закона, управляющего размножением первоначальной частицы. Каково тогда распределение числа частиц во втором поколении?

24. Последовательность электрических импульсов измеряется с помощью некоторого прибора, который записывает наибольшее напряжение, проходившее через него вплоть до текущего момента времени. Допустим, что величина импульсов равномерно распределена на множестве $\{1, 2, \ldots, \ell\}$. Определите соответствующую цепь Маркова и найдите ее переходную матрицу. Каково ожидание времени до того момента, когда прибор зарегистрирует максимальное значение ℓ? [*Указание.* Рассуждайте как при выводе формул (8.1.13) для ожидаемого времени поглощения в состоянии ℓ; примените индукцию после вычисления e_{l-1} и e_{l-2}.]

25. При тщательном изучении рукописи каждый читатель обнаруживает по крайней мере одну ошибку. Допустим, что если всего имеется j ошибок, то после проверки останется от 0 до $j-1$ ошибок с одинаковыми вероятностями. Найдите среднее число читателей, требуемых для выявления всех ошибок. [*Указание.* $e_j = j^{-1}(e_1 + \ldots + e_{j-1}) + 1$; теперь упростите разности $e_j - e_{j-1}$.]

26. Колоду из m карт можно перемешивать (тасовать) разными способами. Пусть пространством состояний служит множество из $m!$ всевозможных упорядочений карт. Каждый отдельный способ перемешивания переводит текущее состояние (упорядочение) в другое. Если разные способы выбираются случайно, то это приводит к различным переходным вероятностям. Следуя совету (а) из § 3.4, полезному для решения комбинаторных задач, давайте начнем с $m = 3$ и двух следующих способов перемешивания:

(i) верхняя карта с вероятностью p перемещается вниз колоды,

(ii) с вероятностью p меняются местами верхняя и средняя карты.

Выпишите переходную матрицу. Покажите, что она является дважды стохастической, и все состояния сообщаются. Убедитесь, что если используется только один из способов перемешивания, то не все состояния будут сообщающимися.

27. Изменим точку зрения в задаче 26, фиксируя внимание на некоторой отдельной карте, скажем, на даме пик, если наши три карты — это король, дама и валет пик. Пусть X_n обозначает ее положение после n перемешиваний. Покажите, что в таком случае мы также имеем дело с марковской цепью, обладающей дважды стохастической переходной матрицей.

28.* Теперь обобщим задачи 26 и 27: докажите, что для произвольного m и любых рандомизированных перемешиваний переходные матрицы в обеих формулировках являются дважды стохастическими. [*Указание.* Если рассматривать каждый способ перемешивания как некоторую перестановку, то она имеет обратную. Таким образом, если он переводит упорядочение j в k, он также переводит некоторое упорядочение i в j. Для фиксированного j соответствие $i = i(k)$ является взаимно однозначным, и $p_{ij} = p_{jk}$. Это доказывает результат для общего случая задачи 26. Далее, рассмотрим два упорядочения j_1 и j_2 с фиксированной картой, лежащей, скажем, сверху колоды. Каждый способ упорядочения, переводящий j_1 в упорядочение, где данная карта располагается на втором месте сверху, делает то же самое с j_2. Следовательно, сумма вероятностей таких способов одинакова для j_1 или j_2 и представляет собой переходную вероятность $1 \to 2$ для перемещения данной карты.]

29.* Клиенты прибывают на пункт обслуживания и становятся в очередь, если он занят. Как только обслуживание одного клиента заканчивается, сразу начинается обслуживание другого, если кто-то есть в очереди, либо откладывается до момента прибытия следующего клиента, если очереди нет. Предположим, что время обслуживания является константой (например, когда речь идет об магнитной записи или автомате для сушки рук), тогда его можно считать единицей времени. Допустим, что поступления клиентов образуют пуассоновский процесс с параметром α в данной временной шкале. Пусть для каждого $n \geqslant 1$ случайная величина X_n обозначает число клиентов в очереди в тот момент, когда заканчивается обслуживание n-го клиента; $\{Y_n, n \geqslant 1\}$ — независимые случайные величины с пуассоновским распределением $\pi(\alpha)$ (см. § 7.1). Покажите, что

$$X_{n+1} = (X_n - 1)^+ + Y_n, \quad n \geqslant 1,$$

где $x^+ = x$, если $x > 0$, и $x^+ = 0$, если $x \leqslant 0$. Выведите отсюда, что последовательность $\{X_n, n \geqslant 1\}$ является цепью Маркова на множестве

$\{0, 1, 2, \ldots\}$ со следующей переходной матрицей:

$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 & \cdots \\ c_0 & c_1 & c_2 & c_3 & \cdots \\ 0 & c_0 & c_1 & c_2 & \cdots \\ 0 & 0 & c_0 & c_1 & \cdots \\ \multicolumn{5}{c}{\dotfill} \end{bmatrix},$$

где $c_j = \pi_j(\alpha)$. [*Указание.* Этот процесс относится к моделям *теории очередей*, а последовательность $\{X_n, n \geqslant 1\}$ называют *вложенной цепью Маркова.* В момент, когда заканчивается обслуживание n-го клиента, имеется две возможности.

(i) Очередь не пуста; тогда сразу начинается обслуживание $(n+1)$-го клиента, и в продолжение этого обслуживания (единицы времени) прибывают Y_n новых клиентов. Поэтому, когда оно заканчивается, длина очереди оказывается равной $X_n - 1 + Y_n$.

(ii) Очередь отсутствует; тогда пункт обслуживания остается свободным вплоть до прибытия $(n + 1)$-го клиента. Как только он появляется, немедленно начинается его обслуживание, за время которого прибывают Y_n новых клиентов. Поэтому в момент окончания обслуживания длина очереди будет равна Y_n. Случайные величины Y_n независимы и имеют распределение $\pi(\alpha)$ в силу теорем 1 и 2 из § 7.2.]

30.* Обобщим схему из задачи 29. Будем считать, что время обслуживания — это такая случайная величина S, что $P\{S = k\} = b_k, k \geqslant 1$. Последовательные времена обслуживания независимы и одинаково распределены. Покажите, что утверждения из задачи 29 выполняются для

$$c_j = \sum_{k=1}^{\infty} b_k\, \pi_j(k\alpha).$$

31.* В условиях задач 29 или 30 положим $\mu = \sum\limits_{j=0}^{\infty} j\, c_j$. Докажите, что цепь Маркова является невозвратной, возвратной нулевой или положительной возвратной в зависимости от выполнения одного из условий $\mu < 1$, $\mu = 1$ или $\mu > 1$. (Этот результат установил Линдли. Разобьем доказательство на шаги, которые можно выполнить на основе материала гл. 8. В обозначениях из § 8.4 положим:

$$F_{10}(z) = f(z), \quad g(z) = \sum_{j=0}^{\infty} c_j z^j.$$

(a) $F_{j,j-1}(z) = f(z)$ для всех $j \geqslant 1$; потому что, например,

$$f_{j,j-1}^{(4)} = P\{Y_n \geqslant 1, Y_n + Y_{n+1} \geqslant 2, Y_n + Y_{n+1} + Y_{n+2} \geqslant 3, \\ Y_n + Y_{n+1} + Y_{n+2} + Y_{n+3} = 3 \mid X_n = j\}.$$

(b) $F_{j0}(z) = f(z)^i$ для $j \geqslant 1$, поскольку длина очереди может убывать только на 1 за один шаг.

(c) $f_{10}^{(1)} = c_0$, $f_{10}^{(v)} = \sum_{j=1}^{\infty} c_j f_{j0}^{(v-1)}$ для $v \geqslant 2$; отсюда

$$f(z) = c_0 z + \sum_{j=1}^{\infty} c_j z F_{j0}(z) = zg(f(z)).$$

(d) Аналогично устанавливается, что $F_{00}(z) = zg(f(z))$.

(e) Если $f(1) = \rho$, то ρ есть наименьший корень уравнения $\rho = g(\rho)$ в $[0, 1]$; отсюда, согласно примеру 4 из § 8.7, имеем, что $F_{00}(1) = f(1) < 1$ или $F_{00}(1) = 1$, в зависимости от того, будет ли $g'(1) > 1$ или $g'(1) \leqslant 1$.

(f) $f'(1) = f'(1)\, g'(1) + g(1)$; тогда если $g'(1) \leqslant 1$, то в обозначениях (8.6.3) имеем $m_{00} = F_{00}'(1) = f'(1) = \infty$ или $m_{00} < \infty$ в зависимости от того, будет ли $g'(1) = 1$ или $g'(1) < 1$. Что и требовалось доказать. Более сложные модели теории очередей рассматриваются, например, в книге [15].)

32* Фирма желает использовать s одинаковых приборов. Приборы выходят из строя в соответствии с известным вероятностным законом. В начале каждой недели фирма заказывает новые приборы для замены сломавшихся с целью довести общее количество исправных приборов до s. Требуется неделя на доставку каждого нового заказа. Пусть X_n — число работающих приборов в начале n-й недели, а Y_n обозначает количество приборов, сломавшихся в течение n-й недели. Установите рекурсивную формулу

$$X_{n+1} = s - Y_n$$

и покажите, что случайные величины $\{X_n, n \geqslant 1\}$ образуют цепь Маркова. Допустим, что распределение количества поломок является равномерным, т. е.

$$P\{Y_n = j \mid X_n = i\} = \frac{1}{i+1}, \quad j = 0, 1, \ldots, i.$$

Найдите переходную матрицу цепи, ее стационарное распределение и ожидаемое количество приборов, находящихся в рабочем состоянии в стационарном режиме.

33* В условиях задачи 32 допустим, что количество поломок имеет биномиальное распределение:

$$P\{Y_n = j \mid X_n = i\} = \binom{i}{j} p^j (1-p)^{i-j}, \quad j = 0, 1, \ldots, i,$$

при некоторой вероятности p. Дайте ответ на те же самые вопросы. (Эти две задачи о *замене приборов* были сообщены Д. Иглхартом.)

34. Матрица $[p_{ij}], i \in I, j \in I$ называется *субстохастической*, если для каждого i выполняется условие $\sum_{j \in I} p_{ij} \leqslant 1$. Покажите, что любая степень такой матрицы также будет субстохастической.

35. Установите, что множество состояний C стохастически замкнуто в том и только том случае, когда для каждого $i \in C$ выполняется условие $\sum\limits_{j \in C} p_{ij} = 1$.

36. Покажите, что

$$\max_{0 \leqslant n < \infty} P_i\{X_n = j\} \leqslant P_i\left\{\bigcup_{n=0}^{\infty} [X_n = f]\right\} \leqslant \sum_{n=0}^{\infty} P_i\{X_n = j\}.$$

Отсюда выведите, что $i \rightsquigarrow j$ тогда и только тогда, когда $f_{ij}^* > 0$.

37. Докажите, что если $q_{ij} > 0$, то $\sum\limits_{n=0}^{\infty} p_{ij}^{(n)} = \infty$.

38*. Убедитесь, что если $j \rightsquigarrow i$, то $g_{ij}^* < \infty$. Приведите пример, когда $g_{ij}^* = \infty$. [*Указание.* Покажите, что $g_{ij}^{(n)} f_{ji}^{(v)} \leqslant f_{ii}^{(n+v)}$, и выберите v так, чтобы $f_{ji}^{(v)} > 0$.]

39. Докажите, что если существует такое j, что $i \rightsquigarrow j$, но не $j \rightsquigarrow i$, то состояние i невозвратно. [*Указание.* Примените теорему 9 или рассуждайте так же, как при доказательстве теоремы 9, чтобы установить неравенство $q_{ii} \leqslant p_{ij}^{(n)} \cdot 0 + (1 - p_{ij}^{(n)}) \cdot 1$ для каждого n.]

40. Определим при $n \geqslant 1$ для произвольных i, j и k из пространства состояний I вероятности

$${}_kp_{ij}^{(n)} = P_i\{X_v \neq k \text{ для } 1 \leqslant v \leqslant n-1, X_n = j\}.$$

Покажите, что если $k = j$, то они сводятся к $f_{ij}^{(n)}$. С другой стороны, если $k = i$, то они превращаются в $g_{ij}^{(n)}$. В общем случае докажите, что

$$\sum_{\ell \neq k} {}_kp_{i\ell}^{(n)} \, {}_kp_{\ell j}^{(m)} = {}_kp_{ij}^{(n+m)}.$$

Такие вероятности называют *вероятностями перехода с запрещением*, потому что прохождение через состояние k не допускается.

41. Докажите, что если общее число состояний равно r и $i \rightsquigarrow j$, то найдется такое n, что $1 \leqslant n \leqslant r$ и $p_{ij}^{(n)} > 0$. [*Указание.* Любую последовательность состояний, ведущую из i в j, где какие-то k встречаются дважды, можно укоротить.]

42. Обобщим определение (8.6.3) следующим образом:

$$m_{ij} = E_i(T_j) = \sum_{v=1}^{\infty} v f_{ij}^{(v)}.$$

Докажите, что $m_{ij} + m_{ji} \geqslant m_{ii}$ для любых двух состояний i и j. В частности, в примере 16 из § 8.6 имеем $m_{0c} \geqslant 2^{c-1}$.

43. Докажите, что симметричное случайное блуждание является возвратным нулевым. [*Указание.* $P_{ij}^{(n)} = P\{\xi_1 + \ldots + \xi_n = j - i\}$; используйте асимптотику (7.3.7) и следующую за ней оценку.]

44. Для произвольного состояния i определим *время пребывания* в i так: $S = \max\{n \geqslant 1 \mid X_v = i$ при всех $v = 1, 2, \ldots, n\}$. Найдите распределение случайной величины S.

45*. Задана цепь Маркова $\{X_n, n \geqslant 1\}$, в которой нет поглощающих состояний. Определим новый процесс следующим образом. Пусть n_1 — это наименьшее значение n такое, что $X_n \neq X_1$, n_2 — наименьшее значение $> n_1$ такое, что $X_n \neq X_{n_1}$, n_3 — наименьшее значение $> n_2$ такое, что $X_n \neq X_{n_2}$ и т. д. Теперь положим $Y_v = X_{nv}$. Покажите, что последовательность $\{Y_v, v \geqslant 1\}$ также является марковской цепью и выразите ее переходную матрицу через переходную матрицу цепи $\{X_n, n \geqslant 1\}$. Докажите, что если некоторое состояние возвратно в одной из цепей, то оно также возвратно и в другой.

46. В обозначениях из задачи 3 положим $X_n = H_n^{(2)} = \sum\limits_{v=1}^{n} H_v$. Покажите, что хотя сами случайные величины $\{X_n\}$ не образуют цепь Маркова, но если определить процесс, значениями Y_n в момент n которого служат упорядоченные пары состояний (X_{n-1}, X_n), то последовательность $\{Y_n, n \geqslant 1\}$ будет марковской цепью. Как устроено ее пространство состояний и какова переходная матрица? Процесс $\{H_n^{(2)},\ n \geqslant 0\}$ иногда называют *цепью Маркова порядка* 2. Как обобщить это понятие на более высокие порядки?

47. Есть аналог марковского свойства для *обратного течения времени*. Пусть $\{X_n\}$ — однородная цепь Маркова. Для $n \geqslant 1$ событие B определяется только случайными величинами $X_{n+1}, X_{n+2}, \ldots$. Покажите, что для любых двух состояний i и j

$$P\{X_{n-1} = j \mid X_n = i, B\} = P\{X_{n-1} = j \mid X_n = i\},$$

но эта вероятность может зависеть от n. Однако, если $\{X_n\}$ — стационарная цепь, как в теореме 13, то докажите, что данная вероятность равна

$$\tilde{p}_{ij} = \frac{w_j p_{ji}}{w_i}$$

и поэтому не зависит n. Проверьте, что $[\tilde{p}_{ij}]$ является переходной матрицей. Однородную цепь Маркова с такой переходной матрицей называют *обращенной* по отношению к первоначальной цепи.

48. Докажите, что $E\,(S_j) < \infty$ для каждого j, что усиливает теорему 1. [*Указание.* При выходе из любого j в $[1, c-1]$ потребуется не более $c-1$ шагов, чтобы покинуть данный отрезок. Пусть W обозначает время ожидания этого события. Тогда $P\,(W \geqslant c-1) \leqslant 1 - p^{c-1} = \delta$, скажем. Повторяя данное рассуждение (как именно?), получим, что $P\,(W \geqslant n(c-1)) \leqslant \delta^n$ для любого $n \geqslant 1$. Положим $V = W\,(c-1)^{-1}$. Тогда $P\,(V = n) \leqslant \delta^n$, и поэтому $E(V) \leqslant \sum\limits_{n=1}^{\infty} n\delta^n < \infty$. Хитро? Такова математика.]

Приложение 3

Мартингалы

Пусть каждая из случайных величин X_n имеет конечное математическое ожидание. Для простоты предположим, что величины принимают целые значения. Рекомендуем вспомнить понятие условного математического ожидания, определенное в конце § 5.2. Потребуем, чтобы для каждого события A, порожденного величинами $X_0, \ldots, X_{n-1}$, и любого возможного для X_n значения i выполнялось равенство

$$E\{X_{n+1} \mid A; X_n = i\} = i; \tag{П.3.1}$$

тогда процесс $\{X_n, n \geqslant 0\}$ называется *мартингалом.* Приведенное определение похоже видом условия на определение марковской цепи (8.3.1), но само равенство представляет собой нечто новое. Его можно более выразительно представить в следующей форме:

$$E\{X_{n+1} \mid X_0, X_1, \ldots, X_n\} = X_n.$$

Это означает, что для произвольных фиксированных значений величин $X_0, X_1, \ldots, X_n$ условное математическое ожидание случайной величины X_{n+1} равно значению X_n безотносительно к значениям остальных случайных величин. Иллюстациями данного свойства могут служить симметричное случайное блуждание или генетическая модель из примера 19 в § 8.7. В первой из моделей обозначим через X_n текущее положение блуждающей частицы. В результате следующего шага она попадает либо в $X_n + 1$ с вероятностью $1/2$, либо в $X_n - 1$ с той же самой вероятностью. Следовательно, каково бы ни было значение X_n, справедливо следующее равенство:

$$E\{X_{n+1} \mid X_n\} = \frac{1}{2}(X_n + 1) + \frac{1}{2}(X_n - 1) = X_n;$$

более того, оно остается верным при добавлении в условие случайных величин $X_0, X_1, \ldots, X_{n-1}$, содержащих информацию о траектории блуждания. Тем самым, выполняется мартингальное свойство (П.3.1). С точки зрения игрока оно означает, что игра является честной, поскольку на каждом шаге возможные выигрыш и проигрыш уравновешиваются, благодаря чему математическое ожидание будущего капитала совпадает с его текущим размером. Аналогичное утверждение имеет место и для количества генов типа A в генетической модели. Его можно обобщить: если выполняется условие (8.7.6) и v — функция $i \to v(i), i \in I$, то процесс $\{v(X_n), n \geqslant 0\}$ является мартингалом. Наконец, нетрудно убедиться, что нормированный размер популяции $\{X_n/\mu^n, n \geqslant 0\}$ в примере 20 из § 8.7 есть мартингал.

Если взять в качестве A в формуле (П.3.1) событие, имеющее вероятность 1, то, с учетом соотношения (5.2.12), получаем:

$$\begin{aligned} E(X_{n+1}) &= \sum_i P(X_n = i)E(X_{n+1} \mid X_n = i) = \\ &= \sum_i P(X_n = i)i = E(X_n). \end{aligned} \tag{П.3.2}$$

Следовательно, все случайные величины в мартингале имеют одинаковые математические ожидания. Это свойство уже отмечалось в формуле (8.2.3), но само по себе оно не имеет большого значения. Приведем теорему из мартингальной теории, на которой базируются рассматриваемые там и в § 8.7 применения. В ее формулировке используется понятие момента остановки из § 8.5.

Теорема. *Пусть мартингал ограничен, т. е. существует такая константа M, что $|X_n| \leqslant M$ для всех n. Тогда для произвольного момента остановки T справедливо равенство*

$$E(X_T) = E(X_0). \tag{П.3.3}$$

Для произвольного мартингала данное равенство остается верным, если момент остановки T ограничен.

В случае задачи 1 из § 8.1 при $p = 1/2$ имеем $|X_n| \leqslant C$; в примере 3 из § 8.3 выполняется ограничение $|X_n| \leqslant 2N$. Поэтому теорема применима, и вычисленные на ее основе вероятности таковы, как было показано в § 8.2.

Равенство (П.3.3), обобщающее (П.3.2), может не выполняться для неограниченных мартингалов и моментов остановки. В этом отношении приведенная теорема отличается от сильного марковского свойства, обсуждаемого в § 8.5. Вот тривиальный, но наглядный контрпример к формуле (П.3.3). Пусть блуждающая частица стартует из 0, и T обозначает момент первого попадания в 1. Тогда T конечен*) в силу теоремы 2 из § 8.2. Поэтому случайная величина X_T определена корректно

*) С вероятностью 1. — *Прим. перев.*

и обязана равняться 1 согласно своему определению. Таким образом, $E(X_T) = 1$, но $E(X_0) = 0$.

Теория мартингалов в значительной степени была развита в работах Дж. Л. Дуба (1910–2004). Она является одним из важнейших направлений в современной теории вероятностей. Введение в нее содержится в монографии [4, гл. 9]. Завершим наше краткое обсуждение следующим замечательным примером.

Санкт-Петербургский мартингал Бореля. Пусть $\{y_n, n \in \mathbb{N}\}$ — независимые случайные величины с $P(y_n = +1) = P(y_n = -1) = 1/2$, т. е. мы имеем дело с результатами бросания симметричной монеты. Зададим числовую последовательность следующим образом:

$$b_1 = 2; \quad \text{для } n \geqslant 2 \quad b_n = \sum_{j=1}^{n-1} b_j + 2^n.$$

Нетрудно проверить, что этим соотношениям удовлетворяют

$$b_n = (n+1) \cdot 2^{n-1}.$$

Далее положим

$$X_n = \sum_{j=1}^{n} b_j y_j.$$

Тогда $\{X_n, n \in N\}$ — мартингал. Действительно, любая последовательность сумм независимых случайных величин с нулевыми математическими ожиданиями образует мартингал (убедитесь). Теперь определим случайный момент T:

$$T(w) = \min\{n \in \mathbb{N} : y_n(w) = +1\},$$

т. е. T — количество испытаний до первого выпадения решки. Тогда T является марковским моментом с распределением $P(T = n) = 2^{-n}$ и математическим ожиданием $E(T) = 2$. Отсюда следует, что момент T конечен с вероятностью 1[*)]. В случае $T < \infty$ справедливо представление $X_T = 2^T$, поскольку при $T = n$ имеем равенство

$$X_n = -\sum_{j=1}^{n-1} b_j + b_n = 2^n.$$

Отсюда

$$E(X_T) = \sum_{n=1}^{\infty} 2^{-n} \cdot 2^n = \infty;$$

[*)] Другими словами, T — момент остановки. — *Прим. перев.*

в то время как $E(X_n) = 0$ для всех n. В итоге получаем, что невзирая на то, что игра честная, игрок гарантировано (точнее, почти наверное) выиграет сумму, имеющую бесконечное математическое ожидание. Этот результат носит название Санкт-Петербургского парадокса (см. задачу 28 в гл. 4), над которым в разные времена ломали голову многие математики. С математической точки зрения он заключается в том, что соотношение (П.3.3) (с заменой X_0 на X_1), выражающее справедливость игры, не выполняется. Теорему нельзя применить, так как ни сам мартингал, ни момент T не являются ограниченными. Давайте теперь для произвольного натурального числа t через $T \wedge t$ обозначим минимум из T и t. Эмиль Борель доказал, что

$$E(X_{T\wedge t}) = 0 \quad \text{для любого } t \geqslant 1, \tag{П.3.4}$$

путем прямых вычислений. Это — несложная задача, которая может послужить превосходным упражнением. Однако мы сошлемся на второе утверждение теоремы, из которого немедленно следует соотношение (П.3.4).

Равенство (П.3.4) выражает предпринятую Борелем попытку восстановления отношения к Санкт-Петербургскому парадоксу как к справедливой игре. В реальном мире, конечно, любая игра должна быть когда-то остановлена. Исторически идея об ограничении продолжительности игры или, что равносильно, размера долга игроков, неоднократно предлагалась в качестве практического разрешения парадокса. Однако только подход, развитый в работах Бореля, опубликованных в течение 1938–1949 гг., который опирается на понятия мартингала и ограниченного момента остановки, позволил свести парадокс к простому математическому утверждению, оказавшемуся верным для всех мартингалов.

Для доказательства теоремы введем обозначение

$$E(A; X) = E(I_A \cdot X)$$

для множества A и случайной величины X. Тогда мартингальное свойство (П.3.1) (при отбрасывании X_0) означает: для любого множества A, порожденного величинами $X_1, \ldots, X_n$, верно равенство

$$E(A; X_{n+1}) = E(A; X_n). \tag{П.3.5}$$

Ввиду того что T — момент остановки, $\{T \leqslant n\}$ и его дополнение $\{T > n\}$ являются именно такими множествами. Далее запишем:

$$E(X_{T\wedge(n+1)}) = E(T \leqslant n; X_T) + E(T > n; X_{n+1}),$$
$$E(X_{T\wedge n}) = E(T \leqslant n; X_T) + E(T > n; X_n).$$

Отсюда, используя соотношение (П.3.5) с $A = \{T > n\}$, устанавливаем равенство этих двух математических ожиданий. Следовательно, все $E(X_{T\wedge n})$ совпадают при всех $n \in \mathbb{N}$. Поэтому они равны $E(X_1)$, так как $T \wedge 1 = 1$. Если момент T ограничен, то он совпадает с $T \wedge n$ при некотором n. Тем самым, мы доказали второе утверждение теоремы. Для доказательства первого утверждения устремим n к бесконечности. Тогда $T\wedge n$ в пределе превращается в T благодаря тому, что момент остановки T конечен согласно определению. Поэтому $X_{T\wedge n}$ сходится к X_T. С учетом того, что в силу предположения случайные величины $X_{T\wedge n}$ ограничены при всех n, математические ожидания $E(X_{T\wedge n})$ стремятся к $E(X_T)$ по теореме Лебега о мажорируемой сходимости (см., например, [4, § 3.2]). Мы использовали более сложный математический аппарат, необходимый в данной ситуации. Дадим также поучительное прямое доказательство:

$$E(X_{T\wedge t}) = \sum_{n=1}^{t} E(T = n; X_T) + E(T > t; X_t).$$

Последнее математическое ожидание ограничено величиной

$$E(T > t; M) = P(T > t) \cdot M,$$

которая стремится к 0 при $t \to \infty$ в силу конечности момента T. Отсюда

$$\lim_{t\to\infty} E(X_{T\wedge t}) = \sum_{n=1}^{\infty} E(T = n; X_T) = E(X_T),$$

что и требовалось доказать.

Глава 9

Инвестирование на основе средних и дисперсий

В этой и следующей главах мы проиллюстрируем применение вероятностных понятий в области финансовой математики, начиная от выборочного пространства вплоть до стохастических процессов, включая мартингалы. В текущей главе изучается так называемая однопериодная модель, в которой используются случайные величины, значения которых наблюдаются в определенный момент времени. Следующая глава посвящена зависящим от времени (многопериодным) моделям, анализ которых базируется на теории случайных процессов. Две главы отличаются также в отношении их финансовых перспектив: первая служит введением в проблематику *сбалансированного инвестирования*, вторая описывает вероятностный подход к вычислению *справедливой цены опциона* на рынке без арбитража.

9.1. Финансовый букварь

Область финансов имеет собственную терминологию, отражающую как богатый концептуальный, так и практический аспекты. Этому предмету посвящены многочисленные вводные курсы. Укажем здесь только два из них, дающие, по нашему мнению, всестороннее и доступное описание финансовой практики: [21] и [3]. Книга [1] представляет собой живое воссоздание множества путей, пройденных учеными, которые развивали теорию финансов как количественную дисциплину и строгое направление в экономике. В текущем параграфе мы познакомимся с основными финансовыми понятиями, используемыми в этой книге.

Когда вы инвестируете, вы покупаете, продаете, занимаете или ссужаете финансовые инструменты, такие как акции или наличные деньги. Теория и практика финансов включают в себя ряд понятий, которые иногда весьма мало отличаются между собой. Ради упрощения мы будем употреблять некоторые термины как взаимозаменяемые. В частности, финансовые инструменты иначе называются активами или ценными бумагами. Говоря о богатстве или капитале инвестора, мы имеем в виду денежное выражение всех его активов.

По своей сути финансовые инструменты, с которыми мы будем иметь дело, являются либо ценными бумагами *типа акций*, либо ценными бумагами *типа долговых обязательств*. Инструменты типа акций отражают собственность компаний. В текущей главе для простоты будут рассматриваться акции, свободно продаваемые на финансовых рынках, подобных Нью-Йоркской фондовой бирже. Множество крупных и известных компаний платят дивиденды владельцам их акций. Выплаты дивидендов обычно производятся раз в квартал, иногда реже. К инструментам типа долговых обязательств относятся денежные займы. Инвесторы ссужают деньги корпорациям или правительству (федеральному или местному) путем приобретения облигаций. Как для всякого займа, указывается расписание выплат процентов (так называемых *купонных платежей*), а также устанавливается срок, по окончании которого заем должен быть возвращен (так называемый *срок погашения*). Например, инвестор, купивший за \$10 000 облигацию правительства США с 30-летним сроком погашения и ежегодной купонной выплатой в размере \$800, фактически одолжил правительству \$10 000 на 30 лет под 8 % годовых. Для некоторых видов облигаций купонные платежи не предусматриваются. Их называют «бескупонными». Покупая такие облигации, инвесторы в действительности ссужают деньги, которые возвращаются к ним с процентами в момент погашения. Другим примером может служить размещение депозита в некотором банке. Вы можете приобрести сертификат депозита, заплатив любую желаемую сумму, обычно ограниченную снизу некоторым минимумом, который устанавливается банком, скажем, \$5 000. При этом вы будете получать доход в виде процентов от суммы, положенной в банк. В сертификате указывается срок депозита — время, которое вам придется ждать момента возвращения банком ваших денег, — обычно 6 или 12 месяцев. Ради простоты мы будем рассматривать только одну разновидность активов типа долговых обязательств: инструмент финансового рынка, представляющий собой объединение нескольких займов с возможно различными (короткими) сроками погашения и разными процентными ставками. В случае его приобретения выплачивается фиксированный доход в виде процентов.

Базовые финансовые инструменты — акции и облигации — продаются и оцениваются в соответствии с законом спроса и предложения. Их цены также подвержены влиянию ожидаемых изменений уровня разнообразных экономических факторов, таких, как инфляция. Именно эти инструменты изучаются в текущей главе. К другой категории финансовых инструментов, рассматриваемой в данной книге, относятся так называемые *производные инструменты*. Их стоимость напрямую связана с ценами на базовые инструменты — акции и облигации. Примером производных инструментов типа акций служит *американский опцион-пут* на некоторые акции. Он представляет собой контракт между двумя сторонами, дающий держателю акций право продать их другой стороне по согласованной цене через определенный срок. Производные финансовые инструменты изучаются в следующей главе.

9.2. Доходность активов и риск

Допустим, что только что вы вложили сумму в x_0 долларов в какую-то ценную бумагу, скажем, акцию некоторой компании. Через год, скорее всего, цена этой бумаги будет отличаться от теперешней. Достаточно лишь взглянуть на финансовые страницы газет или войти в Интернет, чтобы убедиться в существовании такой изменчивости. На стоимость активов влияют многие факторы: общий экономический и политический климат, динамика цен в прошлом, деятельность руководства компаний, направленная на продвижение акций, погодные условия в случае ценных бумаг компаний, связанных с сельскохозяйственным производством, технологические новшества, природные катастрофы и т. п. Не представляется возможным указать все факторы, влияющие на стоимость определенной ценной бумаги. Более разумно рассматривать стоимость как случайную величину X. Согласно определению из § 4.2 случайной величиной называется числовая функция, заданная на выборочном пространстве $\Omega : \omega \to X(\omega)$, $\omega \in \Omega$, где все неизвестные факторы, могущие вызвать колебания наблюдаемого значения величины X сосредоточены в выборочной точке ω.

Можно интерпретировать каждую точку из Ω как некоторое «состояние экономического мира». К примеру, Ω может содержать только два элемента: «благоприятное» состояние экономики ω_1 и «неблагоприятное» состояние экономики ω_2. Предположим, что цена X нашей акции через год представляется как $X(\omega_1) = ux_0$ и $X(\omega_2) = dx_0$, где $u > 1$ и $0 \leqslant d < 1$. В данном примере ваш капитал увеличится, если экономика окажется в состоянии ω_1, и уменьшится в противном случае (цена акции не может быть отрицательной; условие $d \geqslant 0$ является математическим выражением того, что финансовые аналитики называют «ограниченной изменчивостью»). Рассмотренный пример — чрезмерное упрощение реальности. Обычно существуют другие акции, цены которых ведут себя прямо противоположным образом, т. е. падают в случае ω_1 и растут в случае ω_2. Это обстоятельство показывает, что нет необходимости в буквальной интерпретации точек выборочного пространства (см. также обсуждение понятия выборочного пространства в § 4.2). В самом деле можно было бы взять в качестве Ω пространство, используемое для описания эксперимента по бросанию, вообще говоря, несимметричной монеты, часто встречающееся в этой книге (впервые оно появилось в примере 8 из § 2.4). Пусть $\omega_1 = T$ соответствует выпадению решки, а $\omega_2 = H$ — выпадению орла*). Данное двухточечное пространство нетрудно обобщить на случай Ω, состоящего из счетного числа точек. Обычно Ω считается даже несчетным (подобно множеству действительных чисел) для отражения всего многообразия состояний экономики. У случайной величины X также допускаются произвольные действительные значения.

Давайте теперь определим случайную величину Δ формулой

$$\omega \in \Omega : \omega \to \Delta(\omega) = X(\omega) - x_0.$$

Величина Δ представляет собой изменение стоимости вашей ценной бумаги. Когда она положительна, ее называют доходом, а когда отрицательна — убытком. Далее, определим случайную величину

$$R = \frac{\Delta}{x_0}\,.$$

Это — *годовая доходность* вашего первоначального вложения x_0. Отметим, что мы здесь опустили аргумент ω у функции R. В данном случае можно также записать, что

$$X = (1 + R)x_0.$$

В денежном выражении x_0 и X представляются как

$$x_0 = np_0 \quad \text{и} \quad X = nP,$$

где n — количество акций, находящихся в вашем распоряжении, а p_0 и P — цены одной акции, соответственно, сегодня (константа) и через год (последняя, вообще говоря, не обязательно является постоянной случайной величиной). Отсюда имеем следующую формулу для R:

$$R = (P - p_0)/p_0.$$

Другими словами, не обязательно знать количество приобретенных вами акций для вычисления доходности вложения. Достаточно лишь владеть информацией об изменении цены акций данного вида.

---

*) Обозначения происходят от английских слов «tails» (решка) и «heads» (орел). — *Прим. перев.*

Однопериодная и многопериодная модели. В ситуации, рассматриваемой до сих пор, мы имели дело с единственной случайной величиной, обозначенной выше через R. В ее определении участвуют две даты: сегодня (дата t_0), когда стоимость ценной бумаги известна точно (константа x_0), и год спустя (дата t_1), когда стоимость, является, вообще говоря, случайной величиной.

Представим теперь, что нас интересует стоимость данной ценной бумаги в каждый из 3 последующих годов, или 12 месяцев, или 4 кварталов. Тогда говорят, что задан временной *горизонт*, обозначаемый через T, и даты $t_0, t_1, \ldots, t_N$, где $N = 3, 12, 4$ соответственно в каждом из трех приведенных выше случаев. Тем самым, интервал $[0, T]$ разбивается на N *периодов* $(t_{i-1}, t_i]$, $i = 1, \ldots, N$, где $t_i - t_{i-1} = T/N$. Если обозначить через P_i стоимость ценной бумаги в момент t_i, то можно определить доходности

$$r_i = \frac{P_i - P_{i-1}}{P_{i-1}} = \frac{P_i}{P_{i-1}} - 1 \quad \text{за период } i = 1, 2, \ldots, N. \tag{9.2.1}$$

Мы только что описали *многопериодную модель*. Однопериодная модель — ее частный случай при $N = 1$. Именно ей посвящена бо́льшая часть текущей главы.

Доходность ценной бумаги выражает относительное изменение ее стоимости в течение заданного периода, например, 1 % в месяц, 5 % в квартал, 17 % за два года. Согласно нашему определению, доходностью, строго говоря, называется скорость прироста стоимости, т. е. характеристика, связанная со временем, но мы не будем в дальнейшем подчеркивать это обстоятельство. Принято выражать доходности разных активов в единой временной шкале, % в год. Это упрощает их сравнение и позволяет использовать стандартные единицы в вычислительных формулах. Такая процедура называется *приведением*[*)] (к годовой доходности). Удобнее оценивать годовой прирост стоимости, работая с *валовой доходностью*[**)] $1 + r$ вместо r. Для периода i имеем $1 + r_i = P_i/P_{i-1}$. Отсюда, вычисляя валовую доходность P_N/P_0 за N периодов, приходим к формуле, связывающей ее с валовыми доходностями за периоды с 1 по N:

$$\frac{P_N}{P_0} = \frac{P_N}{P_{N-1}} \times \frac{P_{N-1}}{P_{N-2}} \times \ldots \times \frac{P_1}{P_0} = \prod_{i=1}^{N} (1 + r_i).$$

[*)] В оригинале используется термин «annualization». — *Прим. перев.*
[**)] «Gross return» в оригинале. — *Прим. перев.*

Если каждый из N периодов меньше года, причем все вместе они составляют ровно один год, то приведенная доходность $AR^{(N)}$ удовлетворяет соотношению

$$\frac{P_N}{P_0} = 1 + AR^{(N)} = \prod_{i=1}^{N}(1 + r_i).$$

Для малых значений r_i можно записать, что

$$AR^{(N)} \approx \sum_{i=1}^{N} r_i$$

(проверьте это, применив разложение в ряд Тейлора). В частности, наши примеры с 1 % в месяц и 5 % в квартал отвечают, соответственно, приведенным доходностям приблизительно в 12 % и 20 %.

С другой стороны, если каждый из периодов равен в точности одному году, то процедура приведения заключается в вычислении такой фиксированной ежегодной доходности $AR^{(N)}$, для которой выполняется равенство $P_N/P_0 = (1 + AR^{(N)})^N$. Отсюда

$$AR^{(N)} = \left(\frac{P_N}{P_0}\right)^{1/N} - 1.$$

Наш пример доходностью 17 % в течение двух лет соответствует ставке $\sqrt{1.17} - 1$ или 8.17 % годовых. Как видите, здесь прирост в 17 % за весь двухгодичный интервал преобразуется в доходность 8.17 % в каждый из годов. Часто, обладая информацией о доходностях $r_1, r_2, \ldots, r_N$ за N последовательных лет, мы хотим подсчитать «среднегодовую» доходность $AR^{(N)}$. Ввиду накопления процентов (получения, как говорят, «процентов на проценты») при сбережении денег в банке, корректный подход к ее вычислению должен опираться на соотношение

$$\frac{P_N}{P_0} = (1 + AR^{(N)})^N,$$

откуда

$$\frac{1}{N}\ln\left(\frac{P_N}{P_0}\right) = \ln\left(1 + AR^{(N)}\right).$$

Следовательно,

$$\ln\left(1 + AR^{(N)}\right) = \frac{1}{N}\ln\left[\prod_{i=1}^{N}(1 + r_i)\right] = \left[\frac{1}{N}\sum_{i=1}^{N}\ln(1 + r_i)\right].$$

Для малых значений r_i имеем аппроксимацию $AR^{(N)} \approx (1/N)\sum\limits_{i=1}^{N} r_i$ (убедитесь в этом, используя разложение в ряд Тейлора).

Риск ценных бумаг. Как мы только что видели из примера, приведенного выше, и как, вероятно, вы знаете из собственного опыта, инвестициям присущ определенный риск: вы можете лишиться части вложенных денег или даже потерять их полностью! Каким образом выразить (моделировать) этот риск математически? Нам уже встречалась подобная ситуация при изучении задачи о разорении игрока (см. задачу 1 из § 8.1). Правильно ли понимать под инвестиционным риском вероятность потери всей вложенной суммы, т. е. $P(X \leqslant 0)$? Или больше подходит такая характеристика, как вероятность потери половины первоначального капитала, т. е. $P(X < x_0/2)$? Заметим по ходу дела, что в первом случае мы использовали условие $X \leqslant 0$ вместо $X = 0$ ввиду того, что, вообще говоря, величина X (как мы увидим в дальнейшем) может оказаться отрицательной, если разрешается занимать деньги. Таким образом, существуют разные подходы к проблеме моделирования инвестиционного риска. Очевидно, мы хотим определить риск так, чтобы можно было решить, являются или нет для нас слишком рискованными определенные финансовые инструменты, скажем, акции, а также понять, какую сумму денег (если вообще что-то) нам следует вложить в них. Наиболее распространенным среди финансовых экономистов способом выражения отношения инвесторов к рискованным вложениям служит аппарат так называемых функций полезности, которые теоретически могут быть построены для каждого инвестора. В этой книге мы не станем применять указанный подход. В текущей главе используется другая мера степени риска, популярная как среди финансовых теоретиков, так и среди практиков.

Определение 1. *Риском ценной бумаги с доходностью R будем считать дисперсию* $\sigma^2(R)$ (см. § 6.3, где определены понятия дисперсии и стандартного отклонения случайной величины).

Замечание. Можно было бы использовать стандартное отклонение $\sigma(R)$ вместо его квадрата $\sigma^2(R)$. Преимущество такой меры заключается в том, что ее размерность (% за единицу времени) совпадает с размерностью R. Однако мы предпочитаем работать с дисперсией, так как она более удобна из математических соображений.

Безрисковые вложения. Один из типов вложений (активов) играет особую роль в финансовой теории. Он характеризуется тем, что при инвестировании в него мы точно знаем, какой доход получим. Другими словами, если X — стоимость ценной бумаги через год после ее покупки, то $X(\omega) = (1 + r_f)x_0$ для всех $\omega \in \Omega$, где годовая доходность r_f

является константой. Будем рассматривать данный тип вложений в качестве одного из инструментов финансового рынка, описанных в § 9.1. Его доходность есть

$$R(\omega) = \frac{X(\omega) - x_0}{x_0} = r_f, \quad \omega \in \Omega$$

(для простоты мы считаем, что $r_f \geqslant 0$). Дисперсия доходности равна

$$E\{(R - E(R))^2)\} = E\{(r_f - r_f)^2)\} = E\{0\} = 0.$$

Поскольку доходность имеет нулевую дисперсию, такие вложения называются безрисковыми.

9.3. Портфель инвестора

Предположим, что вы владеете $n_1 = 100$ акциями, скажем, компании AT&T, и $n_2 = 80$ акциями другой компании, скажем, Ford. Кроме того, допустим, что текущая цена одной акции первой компании есть $P_1 = \$60$, а второй — $P_2 = \$50$. Тогда текущий размер вашего капиталовложения составляет

$$x_0 = n_1 P_1 + n_2 P_2 = 6\,000 + 4\,000 = 10\,000 \text{ (долларов)}.$$

Величины

$$\alpha_1 = \frac{n_1 P_1}{x_0} = \frac{6\,000}{10\,000} = 0.6$$

$$\text{и} \quad \alpha_2 = \frac{n_2 P_2}{x_0} = \frac{4\,000}{10\,000} = 0.4$$

отражают доли вашего капитала, вложенные соответственно в AT&T и Ford. Иначе α_1 и α_2 выражают в процентах. В нашем примере имеем, что $\alpha_1 = 60\,\%$ объема инвестиций размещены в акциях первой компании, а $\alpha_2 = 40\,\%$ — в акциях второй компании. При этом

$$\alpha_1 + \alpha_2 = 1.$$

Определение 2. *В случае M видов ценных бумаг портфелем инвестора или стратегией капиталовложения называется набор $(\alpha_1, \alpha_2, \ldots, \alpha_M)$ из M действительных чисел, удовлетворяющих условию*

$$\sum_{i=1}^{M} \alpha_i = 1.$$

Для каждого $i \in \{1, 2, \ldots, M\}$ соответствующее α_i представляет собой долю от всего вложенного капитала, приходящуюся на бумаги вида i. Эту величину называют также весом активов вида i в портфеле инвестора.

Определение 3. *Доходностью портфеля* $(\alpha_1, \alpha_2, \ldots, \alpha_M)$ *называется (случайная) сумма*

$$\sum_{i=1}^{M} \alpha_i R_i,$$

где R_i — *однопериодная доходность (случайная величина) активов вида* i.

9.4. Диверсификация

Принятие инвестиционного решения теоретически направлено на достижение конкретной цели. Одной из таких целей может быть *уменьшение риска портфеля*, определенного нами как дисперсия доходности портфеля (аналогично ситуации, когда у нас имеются активы только одного вида, см. определение 1). Снижение риска может достигаться благодаря *диверсификации*, которая представляет собой стратегию инвестирования, определяемую следующим образом.

Определение 4. *Пусть* $(\alpha_1, \alpha_2, \ldots, \alpha_M)$ — *некоторый портфель, имеющий риск* $\sigma^2\Bigl(\sum\limits_{i=1}^{M} \alpha_i R_i\Bigr)$, *где* R_i *обозначает однопериодную доходность активов вида* i. *Портфель* $(\alpha'_1, \alpha'_2, \ldots, \alpha'_M)$ *называется диверсификацией первоначального портфеля, если*

$$\sigma^2\left(\sum_{i=1}^{M} \alpha'_i R_i\right) < \sigma^2\left(\sum_{i=1}^{M} \alpha_i R_i\right)^{*)}.$$

Пример 1. Рассмотрим $M = 2$ вида активов. Допустим, что

$$\sigma^2(R_1) = \sigma^2(R_2) > 0,$$

т. е. активы имеют одинаковый риск, а также, что $-1 \leqslant \rho_{12} < 1$, где ρ_{12} — коэффициент корреляции между R_1 и R_2 (понятия корреляции и ковариации определены в § 6.3). Покажем, что если первоначальным портфелем является $(\alpha_1, \alpha_2) = (1, 0)$, то его диверсификацией будет любой из портфелей вида

$$(\alpha'_1, \alpha'_2) = (\alpha, 1 - \alpha),$$

*) Другими словами, диверсификация — перераспределение долей, снижающее риск. — *Прим. перев.*

где $0 < \alpha < 1$. Действительно,

$$\sigma^2\left(\alpha R_1 + (1-\alpha)R_2\right) = \sigma^2(R_1) + 2\alpha(1-\alpha)\left[\operatorname{Cov}(R_1, R_2) - \sigma(R_1)\sigma(R_2)\right] <$$
$$< \sigma^2(R_1) = \sigma^2\left(\alpha_1 R_1 + \alpha_2 R_2\right), \tag{9.4.1}$$

где первое равенство вытекает из условия

$$\sigma^2(R_1) = \sigma^2(R_2) > 0,$$

а неравенство следует из того, что $0 < \alpha < 1$ и

$$\frac{\operatorname{Cov}(R_1, R_2)}{\sigma(R_1)\sigma(R_2)} = \rho_{12} < 1.$$

В частном случае, когда $\operatorname{Cov}(R_1, R_2) \leqslant 0$, можно дать следующую интерпретацию: если нам предлагается инвестировать в ценные бумаги двух видов, риск которых одинаков, а доходности либо некоррелированы ($\operatorname{Cov}(R_1, R_2) = 0$), либо отрицательно коррелированы ($\operatorname{Cov}(R_1, R_2) < 0$) (в этом случае доходности имеют тенденцию принимать значения противоположных знаков), то вложение в бумаги обоих видов менее рискованно, чем вложение только в один из видов.

Пример 2. При каком значении $\alpha \in (0, 1)$ портфель $(\alpha, 1-\alpha)$ из примера 1 обладает наименьшим риском? Другими словами, при каком значении α достигается минимум функции $V(\alpha)$, задаваемой формулой

$$V(\alpha) = \sigma^2\bigl(\alpha R_1 + (1-\alpha)R_2\bigr)?$$

Из соотношения (9.4.1) видим, что V — квадратичная функция относительно α, имеющая вторую производную

$$V''(\alpha) = 2\sigma^2(R_1) + 2\sigma^2(R_2) - 4\rho_{12}\sigma(R_1)\sigma(R_2).$$

При выполнении условий $-1 \leqslant \rho_{12} < 1$ и $\sigma(R_1) = \sigma(R_2) > 0$ имеем:

$$V''(\alpha) > 2\sigma^2(R_1) + 2\sigma^2(R_2) - 4\sigma(R_1)\sigma(R_2) = 2\bigl(\sigma(R_1) - \sigma(R_2)\bigr)^2 = 0.$$

Следовательно, $V''(\alpha) > 0$. Для поиска точки минимума α^* решим уравнение

$$V'(\alpha^*) = 2\alpha^*\sigma^2(R_1) - 2(1-\alpha^*)\sigma^2(R_2) + 2\rho_{12}\sigma(R_1)\sigma(R_2)(1-2\alpha^*) = 0.$$

В результате упрощения за счет сокращения $2\sigma^2(R_1)$ приходим к равенству

$$\alpha^* - (1-\alpha^*) + \rho_{12}\,(1-2\alpha^*) = 0,$$

из которого выводим, что

$$\alpha^* = \frac{1}{2}\,.$$

Подведем итоги. Если у нас имеются активы двух видов с одинаковым риском (одинаковой дисперсией доходности), то наиболее диверсифицированным портфелем, т. е. инвестиционной стратегией, обладающей наименьшей дисперсией доходности, является такая, при которой капитал вкладывается поровну в активы обоих видов. Этот результат интересен тем, что оптимальный портфель не зависит от величины коэффициента корреляции ($\rho_{12} \neq 1$) между доходностями активов. Минимальный риск

$$V(\alpha^*) = \frac{1}{2}\,(1 + \rho_{12})\,\sigma^2(R_1) < \sigma^2(R_1),$$

поскольку $\rho_{12} < 1$. Заметим, что если активы таковы, что $\rho_{12} = -1$, то $V(\alpha^*) = 0$. Доходность оптимального портфеля в этом случае имеет математическое ожидание $(E(R_1) + E(R_2))/2$, и мы почти наверное получим именно такой относительный доход.

Пример 3. Рассмотрим опять активы двух видов, у которых $\sigma(R_1) > 0$ и $\sigma(R_2) = 0$. Согласно ранее введенной терминологии, второй из активов является безрисковым вложением. Тогда для $0 \leqslant \alpha \leqslant 1$ имеем

$$V(\alpha) = \sigma^2\big(\alpha R_1 + (1-\alpha)R_2\big) = \alpha^2\sigma^2(R_1) \geqslant 0.$$

Отсюда заключаем, что точкой минимума служит $\alpha^* = 0$. При этом наименьший риск $V(\alpha^*) = 0$.

9.5. Оптимизация на основе средних и дисперсий

Как мы только что видели в приведенных выше примерах, за счет диверсификации можно выбрать благоразумную стратегию инвестирования. Однако не окажется ли при этом, что снижение риска повлечет существенное уменьшение ожидаемой доходности по сравнению с первоначальной? Обычно инвесторы готовы идти на некоторый дополнительный риск ради увеличения шансов на получение большей прибыли. Можно ли найти разумный компромисс? Экономист Гарри Марковиц был первым, кто предложил следующую формализацию компромисса, за которую он получил Нобелевскую премию: для заданного размера ожидаемой доходности требуется найти портфель, имеющий наименьший риск. Двойственная задача заключается в фиксации уровня риска и поиске портфеля, обладающего максимальной ожидаемой доходностью при данном ограничении. С математической точки зрения удобнее иметь дело с первой из постановок, которая на языке формул имеет следующий вид. Варьируя компоненты вектора $(\alpha_1, \alpha_2, \ldots, \alpha_M)$,

$$\text{минимизировать} \quad \frac{1}{2}\,\sigma^2\left(\sum_{i=1}^{M}\alpha_i R_i\right), \tag{9.5.1}$$

при выполнении условий

$$\sum_{i=1}^{M}\alpha_i = 1, \tag{9.5.2}$$

$$E\left(\sum_{i=1}^{M}\alpha_i R_i\right) = \mu, \tag{9.5.3}$$

где значение μ (ожидаемая доходность портфеля) считается заданным. На практике также накладываются дополнительные ограничения. Как правило, они имеют следующий вид:

$$l_i \leqslant \alpha_i \leqslant u_i, \quad 1 \leqslant i \leqslant M, \tag{9.5.4}$$

причем u_i обычно неотрицательны, а l_i могут быть отрицательными.

Короткие и длинные позиции. Если $\alpha_i \neq 0$, то говорят, что инвестор *имеет позицию* в активах вида i. При $\alpha_i < 0$ позицию называют короткой, при $\alpha_i > 0$ — длинной. Если $l_i < 0$ (обычно $l_i = -1$ или -0.5), то инвестору разрешается иметь короткую позицию в активах вида i. Это осуществляется за счет того, что он дает в долг находящиеся в его распоряжении активы вида i, получая деньги. Когда инвесторы имеют короткую позицию в безрисковых активах, они тем самым, занимают деньги, например, для вложения их в другие ценные бумаги для осуществления своей оптимальной стратегии. Когда вы кладете деньги на банковский счет, вы становитесь обладателем длинной позиции в безрисковых бумагах. Когда инвесторы имеют короткую позицию в акциях, они берут взаймы у брокеров какую-то часть акций и продают их на рынке ценных бумаг. Позже они возвращают заимствованные акции, покупая их на рынке.

Пример с активами трех видов (часть 1). В общем случае для задачи (9.5.1)–(9.5.4) не всегда удается найти решение явно. Обычно задача решается численно с помощью методов квадратичного программирования (см., например, [12] или [19]). Без ограничений (9.5.4) решение задачи (9.5.1)–(9.5.3), когда оно существует, можно выразить явно. Покажем это для случая $M = 3$, используя технику, соответствующую общему уровню сложности этой книги. Для $i = 1, 2, 3$ положим $\mu_i = E(R_i)$,

$\sigma_i = \sigma(R_i)$. Пусть ρ_{ij} обозначает коэффициент корреляции между R_i и R_j; $j = 1, 2, 3$. При заданном μ условия (9.5.2) и (9.5.3) могут быть записаны как

$$\alpha_3 = 1 - \alpha_2 - \alpha_1, \tag{9.5.5}$$

$$\alpha_1(\mu_1 - \mu_3) + \alpha_2(\mu_2 - \mu_3) = \mu - \mu_3. \tag{9.5.6}$$

Допустим теперь, что $\mu_2 \neq \mu_3$. Тогда из (9.5.5) и (9.5.6) имеем:

$$\alpha_2 = \frac{\mu - \mu_3}{\mu_2 - \mu_3} - \frac{\mu_1 - \mu_3}{\mu_2 - \mu_3}\,\alpha_1, \tag{9.5.7}$$

$$\alpha_3 = \frac{\mu - \mu_2}{\mu_3 - \mu_2} - \frac{\mu_1 - \mu_2}{\mu_3 - \mu_2}\,\alpha_1. \tag{9.5.8}$$

Отсюда (9.5.1) можно представить так:

$$\text{минимизировать } V(\alpha_1) = \frac{1}{2}\sum_{i,j=1}^{3} \alpha_i\alpha_j\rho_{ij}\sigma_i\sigma_j, \tag{9.5.9}$$

где α_2 и α_3 выражаются через α_1. Перепишем формулы (9.5.7) и (9.5.8) в следующем виде:

$$\alpha_i = a_i\alpha_1 + b_i\mu + c_i \quad \text{для } i = 2, 3,$$

где

$$a_2 = -\frac{\mu_1 - \mu_3}{\mu_2 - \mu_3}, \qquad b_2 = \frac{1}{\mu_2 - \mu_3}, \qquad c_2 = -\frac{\mu_3}{\mu_2 - \mu_3},$$

$$a_3 = -\frac{\mu_1 - \mu_2}{\mu_3 - \mu_2}, \qquad b_3 = \frac{1}{\mu_3 - \mu_2}, \qquad c_3 = -\frac{\mu_2}{\mu_3 - \mu_2}.$$

Тогда производная функции V задается формулой

$$V'(\alpha_1) = A\alpha_1 + B\mu + C,$$

в которой

$$A = \sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + 2a_2\rho_{12}\sigma_1\sigma_2 + 2a_3\rho_{13}\sigma_1\sigma_3 + 2a_3\rho_{23}\sigma_3\sigma_2,$$

$$B = a_2b_2\sigma_2^2 + a_3b_3\sigma_3^2 + b_2\rho_{12}\sigma_1\sigma_2 + b_3\rho_{13}\sigma_1\sigma_3 + (a_3b_2 + a_2b_3)\rho_{23}\sigma_2\sigma_3,$$

$$C = a_2c_2\sigma_2^2 + a_3c_3\sigma_3^2 + c_2\rho_{12}\sigma_1\sigma_2 + c_3\rho_{13}\sigma_1\sigma_3 + (a_3c_2 + a_2c_3)\rho_{23}\sigma_2\sigma_3.$$

Подставляя выражения для α_i в сумму (9.5.9), видим, что V является квадратичной функцией относительно α_1. Следовательно, решение уравнения $V'(\alpha_1) = 0$ является глобальной точкой минимума задачи (9.5.1)–(9.5.3) при $M = 3$, если $V''(\alpha_1) = A > 0$. Допустим, что параметры μ_i, σ_i и ρ_{ij}, $1 \leqslant i, j \leqslant 3$, таковы, что данное условие выполнено. Решая уравнение $V'(\alpha_1) = 0$, получаем

$$\alpha_i = A_i\mu + B_i, \quad \text{для } 1 \leqslant i \leqslant 3, \tag{9.5.10}$$

где $A_1 = -B/A$ и $B_1 = -C/A$; $A_i = a_i(B/A) + b_i$ и $B_i = -a_i(C/A) + c_i$ для $i = 2, 3$.

Дисперсия σ^2 доходности портфеля с весами α_i из (9.5.10), $i = 1, 2, 3$, как функция от μ имеет вид

$$\sigma^2 = a\mu^2 + b\mu + c, \tag{9.5.11}$$

где

$$a = \sum_{i,j}^{3} A_i A_j \rho_{ij} \sigma_i \sigma_j, \quad b = \sum_{i,j}^{3} (A_i B_j + A_j B_i) \rho_{ij} \sigma_i \sigma_j, \quad c = \sum_{i,j}^{3} B_i B_j \rho_{ij} \sigma_i \sigma_j.$$

Напомним, что мы пришли к функции (9.5.11) в результате фиксирования μ и поиска весов портфеля α_i, при которых достигается минимум дисперсии доходности при условии, что ее математическое ожидание равно μ. Повторим также, что на решенную задачу можно смотреть как на проблему поиска весов, максимизирующих ожидаемую доходность портфеля при заданной дисперсии σ^2. Уравнение (9.5.11) показывает, что в осях «ожидаемая доходность – дисперсия» графиком σ^2 как функции от μ служит парабола. Заметим, что единицы измерения μ (проценты за определенный период времени, скажем, 2 % в год) отличается от единиц измерения σ^2 (квадраты процентов за период времени). Более содержательным и распространенным на практике является представление связи между μ и σ^2 в осях «стандартное отклонение – ожидаемая доходность».

Пусть выполняются неравенства

$$a > 0, \quad b^2 < 4ac \quad \text{и} \quad b/2a < 0.$$

Тогда соотношение (9.5.11) можно представить как

$$\frac{\sigma^2}{(4ac - b^2)/4a} - \frac{(\mu - (-b/2a))^2}{(4ac - b^2)/4a^2} = 1, \tag{9.5.12}$$

откуда следует, что графиком зависимости между σ, стандартным отклонением доходности портфеля с весами α_i из (9.5.10), и μ, соответствующей ожидаемой доходностью, служит гипербола с асимптотами $\mu = -b/2a \pm \sigma/\sqrt{a}$ (см. рис. 37). Пара $(\bar{\sigma}, \bar{\mu})$ отвечает портфелю с наименьшей дисперсией ($\bar{\sigma}^2 = (4ac - b^2)/4a$) среди портфелей с ожидаемой доходностью $\bar{\mu} = -(b/2a)$.

Эффективная граница. Отметим, что при решении оптимизационной задачи (9.5.1)–(9.5.3) не накладывалось никаких ограничений на величину ожидаемой доходности μ. Иначе говоря, μ может принимать

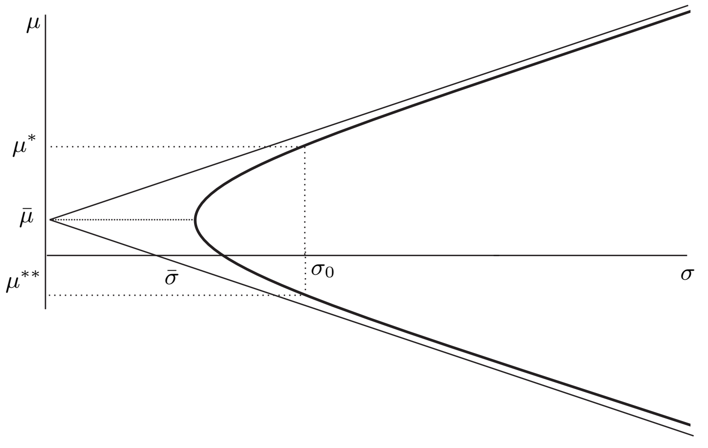


Рис. 37. Граница оптимального портфеля в координатах σ и μ.

любые действительные значения. Однако, взглянув на рис. 37, вы увидите, что для фиксированного уровня σ_0 ($\sigma_0 > \bar{\sigma}$) существуют ровно два значения: μ^* и μ^{**}, для которых риск σ_0 является минимальным. Для того, чтобы осуществить выбор между точками $A^* = (\sigma_0, \mu^*)$ и $A^{**} = (\sigma_0, \mu^{**})$, применим экономическую теорию, описывающую поведение инвесторов. Она, в частности, постулирует, что из двух вариантов с одинаковым риском инвесторы всегда предпочитают решение с наибольшей доходностью. Поэтому при уровне риска σ_0 типичный инвестор выберет портфель с ожидаемой доходностью μ^*, а не μ^{**}. Принимая во внимание эти соображения, достаточно рассматривать только часть графика, содержащую точки (σ, μ) с $\mu \geqslant \bar{\mu}$. Данный фрагмент графика, отдельно изображенный на рис. 38, называется *эффективной границей*. Название объясняется следующей интерпретацией. Любая точка (σ^*, μ^*) на рисунке 38, отвечающая оптимальному портфелю α^*, характеризуется тем, что для любого другого портфеля α' с ожидаемой доходностью μ' и стандартным отклонением σ' выполняются условия: если $\sigma' < \sigma^*$, то $\mu' < \mu^*$, а если $\mu' > \mu^*$, то $\sigma' > \sigma^*$. Другими словами, не существует портфеля, отличного от α^*, который бы обеспечивал ожидаемую доходность выше μ^* при риске, меньшем, чем σ^*.

Случай M видов активов. Подход, только что использованный для трех видов активов, позволяет получить аналогичный результат в общей ситуации, когда имеются ценные бумаги M разных видов: зависимость наибольшей ожидаемой доходности от риска графически представляет-

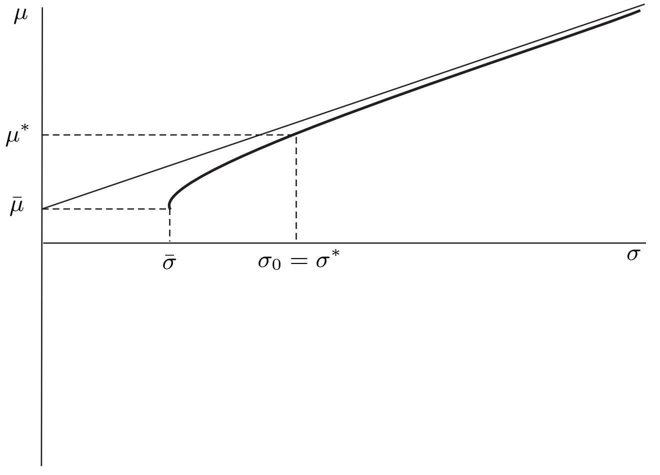


Рис. 38. Эффективная граница в координатах σ и μ.

ся гиперболой на плоскости с осями координат «стандартное отклонение — ожидаемая доходность». Напомним, что для при выводе данного результата делались некоторые предположения, относящиеся, в частности, ко второй производной дисперсии как функции от весов портфеля. Указанные предположения на самом деле проще выражаются на языке линейной алгебры в векторно-матричной форме. В общем случае M видов активов задача заключается в том, чтобы путем варьирования набора весов $(\alpha_1, \alpha_2, \ldots, \alpha_M)$, обозначаемого кратко через α и рассматриваемого как вектор размерности $M \times 1$,

$$\text{минимизировать} \quad \frac{1}{2}\,\alpha^{\mathrm{T}} W \alpha$$

при выполнении условий

$$\alpha^{\mathrm{T}} \varepsilon = \mu,$$
$$\alpha^{\mathrm{T}} \mathbf{1} = 1,$$

где W — ковариационная матрица M активов, т. е. $W_{ij} = \mathrm{Cov}(R_i, R_j)$, $1 \leqslant i, j \leqslant M$, ε — вектор размерности $M \times 1$ с компонентами $\varepsilon_i = \mu_i$, $i = 1, \ldots, M$, и $\mathbf{1}$ — M-мерный вектор, все компоненты которого равны 1.

Далее решение задачи поиска оптимального портфеля потребует знаний, немного выходящих за математический уровень данной книги. Заинтересовавшийся читатель может ознакомиться с гл. 3 книги [12], где приведены все детали в стиле изложения текущей главы. Матричное представление позволяет сформулировать в более удобном виде некоторые из условий, появившихся ранее в примере с тремя активами. В частности, там требовалось, чтобы вторая производная как функция от весов портфеля была положительной. На матричном языке данное условие выполняется тогда и только тогда, когда ковариационная матрица W является положительно определенной. Согласно определению, это означает, что для произвольного ненулевого вектора x справедливо неравенство $x^{\mathrm{T}}Wx > 0$. Предположим теперь, что среди ценных бумаг, участвующих в формировании портфеля, присутствует безрисковое вложение. Без потери общности можно считать, что оно имеет номер 1. Так как доходность такого вложения фиксирована, его стандартное отклонение σ_1 равно нулю. Коэффициент корреляции ρ_{1i} безрискового вложения с любой другой ценной бумагой также равен нулю. Другими словами, в данном случае все элементы первой строки и первого столбца матрицы W — нулевые. Таким образом, для того чтобы W была положительно определенной необходимо исключить из оптимизационной задачи безрисковые вложения. Ниже мы покажем, что даже если оставить их, то решение задачи может быть найдено, но зависимость между оптимальной доходностью и уровнем риска тогда будет иметь другой вид.

Влияние безрискового вложения. Дополним набор из M «рискованных» ценных бумаг, у которых $\sigma^2(R_i) > 0$, $1 \leqslant i \leqslant M$, безрисковым вложением с номером 0, т. е. $\sigma^2(R_0) = 0$. Будем искать набор $(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_M)$ из $M+1$ весов, минимизирующих соответствующий риск портфеля

$$\sigma^2\left(\sum_{i=0}^{M}\alpha_i R_i\right) \tag{9.5.13}$$

при ограничениях

$$\sum_{i=0}^{M}\alpha_i = 1, \tag{9.5.14}$$

$$E\left(\sum_{i=0}^{M}\alpha_i R_i\right) = \mu, \tag{9.5.15}$$

где μ — ожидаемая доходность портфеля. Кроме этого, сюда могут быть еще добавлены ограничения вида (9.5.4). С учетом равенства $\sigma(R_0) = 0$, получаем, что

$$\mathrm{Cov}(R_i, R_0) = \rho_{i0}\sigma(R_i)\sigma(R_0) = 0.$$

Так как $\alpha_0 = 1 - \sum\limits_{i=1}^{M}\alpha_i$ в силу (9.5.14), то задача сводится к нахождению набора $(\alpha_1, \alpha_2, \ldots, \alpha_M)$, минимизирующего дисперсию (9.5.13) и удовлетворяющего условию

$$E\left(\sum_{i=1}^{M}\alpha_i(R_i-R_0)\right)=\mu-R_0. \tag{9.5.16}$$

Пример с активами трех видов (часть 2). Рассмотрим случай, похожий на пример с активами трех видов (часть 1). Только теперь у нас имеются $M=2$ «рискованных» актива и одно безрисковое вложение. При этом, как и прежде, ограничения типа (9.5.4) не накладываются. Сохраняя ранее введенные обозначения, положим $\mu_0=R_0$. Из (9.5.16) выводим (предполагая, что $\mu_2\neq\mu_0$) формулу

$$\alpha_2=a+b\alpha_1, \tag{9.5.17}$$

где

$$a=\frac{\mu-\mu_0}{\mu_2-\mu_0}\quad\text{и}\quad b=-\frac{\mu_1-\mu_0}{\mu_2-\mu_0}\,.$$

Тем самым, задача сводится к поиску значения α_1, минимизирующего функцию

$$V(\alpha_1)=\alpha_1^2\sigma_1^2+(a+b\alpha_1)^2\sigma_2^2+2\alpha_1(a+b\alpha_1)\rho_{12}\sigma_1\sigma_2, \tag{9.5.18}$$

дифференцируя которую, получаем:

$$\begin{aligned}V'(\alpha_1)&=2\alpha_1\left(\sigma_1^2+b^2\sigma_2^2+2b\rho_{12}\sigma_1\sigma_2\right)+2a\rho_{12}\sigma_1\sigma_2,\\ V''(\alpha_1)&=2\left(\sigma_1^2+b^2\sigma_2^2+2b\rho_{12}\sigma_1\sigma_2\right)=2\sigma^2\left(R_1+bR_2\right),\end{aligned} \tag{9.5.19}$$

где в выражении (9.5.19) легко узнать дисперсию случайной величины R_1+bR_2 (см. свойства дисперсии из § 6.3, в частности, неравенство $\sigma^2(Z)\geqslant 0$ для произвольной случайной величины Z). Обратим снова внимание на то, что функция V (9.5.18) является квадратичной по α_1. Поэтому в случае $V''(\alpha_1)>0$ для определения точки минимума достаточно решить уравнение $V'(\alpha_1)=0$. Его решением служит

$$\alpha_1^*=-a\,\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2+b^2\sigma_2^2+2b\rho_{12}\sigma_1\sigma_2} \tag{9.5.20}$$

при условии, что

$$\sigma_1^2+b^2\sigma_2^2+2b\rho_{12}\sigma_1\sigma_2\neq 0. \tag{9.5.21}$$

Заметим, что это условие влечет неравенство $V''(\alpha_1)>0$. Предполагая выполнение неравенства (9.5.21), подставим выражение для α_1^* из (9.5.20) в (9.5.17) и найдем α_2^*. Наконец, α_0^* определяется из соотношения

$$\alpha_0^*=1-(\alpha_1^*+\alpha_2^*)\,.$$

Действуя так же, как в части 1 примера, установим вид зависимости между заданной ожидаемой доходностью μ и риском σ^2 оптимального портфеля с весами $(\alpha_0^*, \alpha_1^*, \alpha_2^*)$. Положим

$$C = -\frac{\rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + b^2\sigma_2^2 + 2b\rho_{12}\sigma_1\sigma_2}$$

и

$$K = \frac{[C^2\sigma_1^2 + (1+bC)^2\sigma_2^2 + 2C(1+bc)\rho_{12}\sigma_1\sigma_2]}{(\mu_2 - \mu_0)^2}.$$

Тогда из формул (9.5.18) и (9.5.20) имеем:

$$\sigma^2 = V(\alpha_1^*) = (\mu - \mu_0)^2 K, \tag{9.5.22}$$

где отметим, что

$$K = \frac{\sigma^2\,(CR_1 + (1+bC)R_2)}{(\mu_2 - \mu_0)^2} \geqslant 0$$

ввиду неотрицательности дисперсии произвольной случайной величины и неравенства $(\mu_2 - \mu_0)^2 > 0$, вытекающего из предположения $\mu_2 \neq \mu_0$. Соотношение (9.5.22) можно также записать в виде

$$\sigma = |\mu - \mu_0|\sqrt{K}, \tag{9.5.23}$$

откуда заключаем, что графиком зависимости минимального стандартного отклонения от ожидаемой доходности портфеля является объединение двух лучей, исходящих из точки $(0, \mu_0)$. Это очевидно из более подробной записи функции (9.5.22) в следующей форме:

$$\sigma = \begin{cases} \sqrt{K}(\mu - \mu_0), & \text{если } \mu \geqslant \mu_0, \\ -\sqrt{K}(\mu - \mu_0), & \text{если } \mu < \mu_0. \end{cases}$$

Что произойдет с ценами активов, если все инвесторы будут формировать портфели, опираясь на общий для всех оптимизационный подход на основе средних и дисперсий, т. е. станут вкладывать деньги в одинаковых оптимальных пропорциях на всех финансовых рынках? Не используя сложной терминологии, можно ответить, что в таком случае в результате действия закона спроса и предложения цены всюду будут колебаться вблизи уровней, соответстующих весам α^* оптимального портфеля. Данное утверждение о *ценовом равновесии* образует фундамент так называемой *модели ценообразования при инвестировании*, детально разработанной Уильямом Шарпом. За нее он получил Нобелевскую премию совместно с ранее упоминавшимся Гарри Марковицем.

Расчет оптимальных цен может быть выполнен с помощью математических методов (см., например, [8] или [12]). Теория позволяет предсказывать цены, которые установятся на рынке через определенный временной промежуток. До появления данной модели многие экономисты пытались вычислить справедливые цены, по которым следовало бы вести торговлю акциями. Часто они являлись представителями теоретической науки, и их выводы противоречили опыту тех, кто реально продавал акции на рынке. Время от времени экономистов высмеивали за их жесткие и убыточные рекомендации. Лишь немногие из них смогли успешно применить свои идеи в действительности и стать «богатыми и знаменитыми». Одним из таких исключений является Кейнс, которому удалось приобрести солидный капитал путем инвестирования. Он с презрением относился к тем, кто покупал исключительно краткосрочные акции, и называл их игроками.

Главная цель текущей главы — показать важность такого метода снижения риска, как диверсификация. Есть несколько путей, позволяющих применить эту идею на практике. Например, *паевые инвестиционные фонды* — активы, образуемые за счет объединения денег мелких инвесторов таких, как учителя или даже студенты, на которые приобретаются акции разных компаний*). С начала 1980-х инвестиционные фонды стали основным механизмом, используемым работниками для вложения денег и хранения сбережений на старость. К сожалению, некоторые служащие размещали бóльшую часть своих пенсионных денег только в акциях тех компаний, на которые работали. Ввиду отсутствия диверсификации успех такого подхода целиком зависит от судьбы единственной компании. Известным примером, когда данная стратегия с треском провалилась, затронув тысячи человек, служит компания Enron, которая в 2001 г. была одной из крупнейших в мире. В завершение приведем мнение выдающегося американского экономиста Фрэнка Х. Найта, высказавшего в своей книге [17] философскую сентенцию, согласно которой риск не только является неотъемлемой частью жизни, но он также необходим для предпринимательства и получения доходов от инвестиций. Кстати, его сын Фрэнк Б. написал труд по теории броуновского движения (см. гл. 8, § 2), которая в наше время широко используется для описания колебаний цен на финансовом рынке.

---

*) Существуют тысячи инвестиционных фондов, объединенных в семейства, но мы упомянем лишь один из них, Alger Group, в котором бывший (во время первого издания книги) второклассник, упоминавшийся в начале § 1.1 и скакавший верхом на карусельной лошадке в примере 14 из гл. 8 (см. рис. 34), в настоящее время служит главным распорядителем.

9.6. Распределения доходности активов

Установленные до сих пор результаты базируются почти без исключений на таких характеристиках, как средние, дисперсии и ковариации. О распределениях случайных величин говорилось мало. Понятно, что две случайные величины могут иметь одинаковые средние и дисперсии, но совершенно разные функции распределения. Очевидным примером служит пара, состоящая из целочисленной случайной величины, скажем, бернуллиевской, и величины, имеющей плотность распределения, скажем, нормальной. Если вы хотите убедиться на примере, в котором обе случайные величины обладают плотностью, то решите задачу 7.

Ввиду того, что доходности активов находятся в фокусе большинства исследований эффективности капиталовложений, не удивительно, что изучению их распределений уделяется много внимания. С одной стороны, анализ реальных данных (а их — огромное множество) позволяет высказывать гипотезы об их соответствии некоторым типам распределений. С другой, желание получать поддающиеся математическому инструментарию модели приводит к выдвижению дополнительных предположений. Предвосхищая следующую главу, обсудим распределения доходностей активов сразу в многопериодном контексте. Пусть имеется N периодов $(t_0, t_1]$, $(t_1, t_2]$, …, $(t_{N-1}, t_N]$. Обозначим через S_n стоимость ценной бумаги в конце n-го периода и определим однопериодную доходность r_n формулой

$$r_n = \frac{S_n - S_{n-1}}{S_{n-1}}, \quad n = 1, 2, \ldots, N.$$

Определим также N-периодную доходность

$$R_N = \frac{S_N - S_0}{S_0}$$

и введем случайные величины

$$\xi_n = \ln(1 + r_n) \quad \text{для } 1 \leqslant n \leqslant N \quad \text{и} \quad \Psi_N = \ln(1 + R_N).$$

Обычно предполагается, что случайные величины ξ_1, ξ_2, …, ξ_n, …, ξ_N независимы и одинаково распределены (в силу утверждения 6 из § 5.5, это равносильно независимости и одинаковой распределенности доходностей r_1, r_2, …, r_N). Данное предположение подтверждается в большинстве исследований реальных данных. Кроме того, обычно допускают, что случайные величины ξ_n нормально распределены. Ниже мы обсудим, можно ли считать это допущение приемлемым.

Когда речь идет о небольших временных периодах (например, r_n обозначают ежедневную или ежечасную доходность), исследования реальных данных позволяют заключить, что функция распределения величин ξ_n обладает более «тяжелыми» хвостами сравнительно с функцией нормального распределения. Говоря, что правый хвост распределения случайной величины Y «тяжелее» правого хвоста распределения величины Z, формально имеют в виду, что для всех достаточно больших x выполняется неравенство

$$P(Y > x) > P(Z > x).$$

Аналогично, левый хвост Y «тяжелее» левого хвоста Z, если

$$P(Y < x) > P(Z < x)$$

для всех $x < 0$ таких, что абсолютная величина $|x|$ достаточно велика. Таким образом, для коротких периодов, допущение, что ξ_n имеют нормальное распределение, скорее всего, ведет к недооценке вероятности получения больших доходов или потерь (соответствующих большим по абсолютной величине значениям ξ_n).

Распределение Коши. Распределением Коши с параметром сдвига μ и параметром масштаба $\sigma > 0$ называется распределение с плотностью

$$f(x) = \left[\pi\sigma\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)\right]^{-1}.$$

Логарифмическая шкала. Как мы видели, в однопериодной модели, изучаемой в текущей главе, предполагается, что дисперсия доходности конечна. Ввиду этого, распределение Коши не годится для использования в данной модели. В следующей главе рассматриваются многопериодные модели, в которых распределение однопериодных доходностей r_n или ξ_n, получаемых из них логарифмическим преобразованием, будет оцениваться на основе изучения случайной величины $\sum_{i=1}^{N} \xi_n$. Для понимания важности последней, напомним, что

$$1 + r_n = \frac{S_n}{S_{n-1}}, \tag{9.6.1}$$

откуда получаем формулу

$$\xi_n = \ln(1 + r_n) = \ln S_n - \ln S_{n-1}, \tag{9.6.2}$$

показывающую, что $\{\xi_n\}$ представляют собой последовательные изменения цены в логарифмической шкале. Величину ξ_n иногда называют логарифмической доходностью за n-й период (см. также примеры из § 9.2

для многопериодных моделей). Поэтому

$$\sum_{i=1}^{N} \xi_n = \Psi_N \tag{9.6.3}$$

представляет собой логарифмическую доходность за время от t_0 до t_N. Если предположить, как мы и собираемся, что величины $\xi_1, \xi_2, \ldots, \xi_N$ являются независимыми и одинаково распределенными, то станет возможным применение таких полезных инструментов, как модель случайного блуждания из гл. 8 и центральная предельная теорема из § 7.5.

Обратим внимание, что, поскольку все цены S_n должны оставаться положительными, значения доходностей r_n обязаны превосходить -1. Это условие ограничивает выбор распределения для r_n. В свою очередь, ξ_n могут принимать произвольные действительные значения, тем самым, предоставляя большую свободу в выборе подходящего для них распределения. Однако более важным является то, что логарифмическое преобразование позволяет работать с моделями, которые легко исследовать, т. е. для которых мы можем достаточно быстро получать результаты.

9.7. Устойчивые распределения

Пусть N — количество рабочих дней в данном месяце. Обозначим через $t_1, t_2, \ldots, t_N$ времена закрытия рынка акций в каждый из этих дней, а через t_0 — последнее время закрытия в предыдущем месяце. Тогда для $1 \leqslant n \leqslant N$ величины S_n, r_n и ξ_n представляют, соответственно, цену в момент закрытия, доходность и логарифмическую доходность за n-й период. Пусть S_0 обозначает цену акций в последний из рабочих дней предыдущего месяца. Рассмотрим теперь два подхода к оцениванию распределения ежемесячной логарифмической доходности Ψ_N. Первый заключается в оценивании распределения ежедневной логарифмической доходности ξ_n на основе данных о ежедневных доходностях, скажем, за 5 лет, и использовании затем (9.6.3). Другой — в том, чтобы оценить интересующее распределение непосредственно с помощью информации о ежемесячных доходностях, вычисляемых исходя из цен в конце каждого месяца этих 5 лет. Для простоты допустим, что в каждом из месяцев содержится ровно N (скажем, $N = 22$) рабочих дней. Тогда наши наблюдения за 5 лет представляют собой $5 \times 12 \times N$ ежедневных доходностей. Эти данные позволят установить вид распределения случайных величин r_n, а также их логарифмов ξ_n, и, наконец, найти распределение логарифмической доходности Ψ_N с помощью представления

$$\Psi_N = \sum_{n=1}^{N} \xi_n.$$

Второй подход состоит в том, чтобы основываясь на 5×12 наблюдениях ежемесячных доходностей в течение 5 лет прийти к заключению относительно формы их распределения (или, что эквивалентно, распределения их логарифма Ψ_N). Подчеркнем различие между двумя подходами. При первом используются реальные данные для оценивания распределения ежедневных доходностей. Затем распределение ежемесячных доходностей вычисляется как распределение суммы N независимых случайных величин (ежедневных доходностей). При втором подходе функция распределения ежемесячных доходностей оценивается прямо по данным, без использования ее представления в виде суммы N случайных величин (ежедневных доходностей). В идеале получаемое распределение должно мало отличаться от того, которое возникает в результате суммирования ежедневных доходностей. Это свойство, известное как устойчивость, является нашей следующей темой.

В силу теоремы 7 из § 7.4, если X_1 и X_2 — две независимые случайные величины, имеющие, соответственно, нормальные распределения $N(m_1, \sigma_1^2)$ и $N(m_2, \sigma_2^2)$, то случайная величина $X_1 + X_2$ также нормально распределена (со средним $m_1 + m_2$ и дисперсией $\sigma_1^2 + \sigma_2^2$). Другими словами, отдельные распределения величин X_1, X_2 и $X_1 + X_2$ отличаются только своими средними и дисперсиями (которые называются параметрами нормального распределения: когда они известны, распределение определено однозначно). Заметим, что

$$Z_i = \frac{X_i - m_i}{\sigma_i}, \quad i = 1, 2,$$

независимы и имеют стандартное нормальное распределение. Пусть

$$Z = \frac{X_1 + X_2 - (m_1 + m_2)}{\sigma},$$

где $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Тогда можно записать, что

$$\sigma Z = \sigma_1 \frac{X_1 - m_1}{\sigma_1} + \sigma_2 \frac{X_2 - m_2}{\sigma_2} = \sigma_1 Z_1 + \sigma_2 Z_2.$$

Обратите внимание, что каждая из величин Z_1, Z_2 и Z имеет стандартное нормальное распределение.

Определение. *Функцию распределения называют устойчивой, если для любых двух независимых случайных величин X_1 и X_2 с таким* *распределением и любых положительных констант c_1 и c_2 имеет место представление*

$$c_1X_1 + c_2X_2 = cX + d, \tag{9.7.1}$$

где $c > 0$ и d — константы, зависящие от c_1 и c_2, а X — случайная величина с тем же самым распределением, как у X_1 и X_2.

Пример. Как было замечено выше, если X_1 и X_2 независимые случайные величины, имеющие стандартное нормальное распределение, то случайная величина $Y = X_1 + X_2$ снова нормально распределена со средним 0 и дисперсией 2. Если существует представление для Y вида $Y = cX+d$, где $c > 0$ и d— действительные числа, а X имеет стандартное нормальное распределение, то должны выполняться равенства

$$\begin{aligned} &0 = E(Y) = cE(X) + d = d \\ \text{и}\quad &2 = \operatorname{Var}(Y) = c^2 \operatorname{Var}(X) + d = c^2. \end{aligned}$$

Отсюда $c = \sqrt{2}$ и $d = 0$.

При условии, что $c_1 = c_2 \neq 0$, (9.7.1) переписывается как

$$X_1 + X_2 = c'X + d',$$

где $c' = c/c_1$ и $d' = d/c_1$. Таким образом, устойчивость выражается в том, что сумма двух независимых и одинаково распределенных случайных величин имеет такое же распределение, как и слагаемые, *с точностью* до возможного изменения масштаба (за это отвечает c') и/или сдвига начала координат (за это отвечает d'). Нам уже встречалась данная форма устойчивости не только для отдельных распределений, как выше, но и для целых семейств распределений, а именно — в теореме 7 из § 7.4 для нормального семейства. Хотя свойство устойчивости наиболее полезно на уровне распределений, мы будем говорить об устойчивости как распределений, так и семейств (или типов)[*] распределений. Здесь необходимо сделать предупреждение. Согласно теореме 3 из § 7.2 сумма независимых и одинаково распределенных пуассоновских случайных величин снова имеет распределение Пуассона, однако данное распределение не является устойчивым. Чтобы убедиться в этом, прежде всего, вычислите характеристическую функцию распределения Пуассона $\pi(\alpha)$ (см. теорему 7 в § 6.5). Затем покажите, что если X_1 и X_2 независимы и имеют, соответственно, распределения $\pi(\alpha_1)$ и $\pi(\alpha_2)$, то их сумма $X_1 + X_2$ обладает распределением $\pi(\alpha_1 + \alpha_2)$. Поэтому $\pi(\alpha)$ не является устойчивым.

[*] Мы не будем различать семейства и типы.

Характеризация устойчивости, установленная Леви. П. Леви открыл устойчивые распределения (в 1923 г.) и получил явное представление для их характеристических функций. Обсудим кратко эту тему.

Пусть F обозначает устойчивый *закон* (это другое название для функции распределения или типа распределений), $\varphi(\theta)$ — его характеристическая функция (см. (6.5.17)). Для независимых случайных величин X_1 и X_2 с распределением F и заданных положительных констант c_1 и c_2 имеем:

$$E\left[e^{i\theta(c_1X_1+c_2X_2)}\right] = E\left[e^{i\theta c_1X_1}\right]\cdot E\left[e^{i\theta c_2X_2}\right] = \varphi(\theta c_1)\cdot\varphi(\theta c_2),$$

где первое равенство справедливо в силу независимости, а второе вытекает из одинаковости распределений X_1 и X_2 (сравните с теоремой 7 из § 6.5). Чтобы обеспечить представление $\mathrm{c}_1X_1 + \mathrm{c}_2X_2 = cX + d$ для случайной величины X с тем же самым распределением и некоторых констант $c > 0$ и d, необходимо, чтобы выполнялось равенство

$$E\left[e^{i\theta(c_1X_1+c_2X_2)}\right] = E\left[e^{i\theta(cX+d)}\right] = E\left[e^{i\theta cX}\right]\cdot e^{i\theta d}$$

или эквивалентная ему формула

$$\varphi(\theta c_1)\cdot\varphi(\theta c_2) = \varphi(\theta c)\cdot e^{i\theta d}. \tag{9.7.2}$$

Пример (*продолжение*). Поскольку характеристической функцией стандартного нормального закона служит $\varphi(\theta) = e^{-\theta^2/2}$ (см. (7.5.7)), то сотношение (9.7.2) превращается в

$$e^{-c_1^2\theta^2/2}\cdot e^{-c_2^2\theta^2/2} = e^{-c^2\theta^2/2}\cdot e^{i\theta d} \quad \text{или}$$
$$e^{-(c_1^2+c_2^2)\theta^2/2} = e^{-c^2\theta^2/2}\cdot e^{i\theta d},$$

где $c = \sqrt{c_1^2 + c_2^2}$ и $d = 0$. Тем самым, мы убедились, что стандартное нормальное распределение устойчиво.

Вообще, характеристическая функция нормального закона с математическим ожиданием m и дисперсией v имеет вид:

$$\varphi(\theta) = e^{-(v\theta^2/2)+im\theta}.$$

Другим примером может служить характеристическая функция закона Коши с параметрами μ и σ (см. § 9.6)

$$\varphi(\theta) = e^{-\sigma|\theta|+i\mu\theta}.$$

Симметричные распределения. Характеристические функции некоторых важных и полезных классов устойчивых законов имеют простой вид. Мы сосредоточимся сейчас на одном из таких классов. Распределение F называется симметричным, если

$$F(x) = 1 - F(-x - 0) \quad \text{для всех } x,$$

где $F(-x-0) = \lim\limits_{\varepsilon \to +0} F(-x-\varepsilon)$ (см. задачу 7 в гл. 4). Это определение на первый взгляд может показаться довольно неуклюжим. Дадим другое, более наглядное. Заметим, что если случайная величина X имеет функцию распределения F, то случайная величина $-X$ обладает функцией распределения $1 - F(-x - 0)$. Действительно,

$$P(X \leqslant x) = P(-X \geqslant -x) = 1 - P(X < -x) = 1 - F(-x - 0).$$

Таким образом, если X имеет функцию распределения F, то X и F называются симметричными, если X и $-X$ одинаково распределены. Нетрудно убедиться, что F симметрично тогда и только тогда, когда его характеристическая функция $\varphi(\theta) = \varphi(-\theta)$ для всех θ. Например, $\varphi(\theta) = e^{-\theta^2/2}$ для стандартного нормального закона и $\varphi(\theta) = e^{-|\theta|}$ для распределения Коши с $\mu = 0$ и $\sigma = 1$. Известно, что класс всех симметричных устойчивых законов характеризуется тем, что характеристические функции этих законов имеют вид

$$\varphi(\theta) = e^{-c|\theta|^\alpha},$$

где $0 < \alpha \leqslant 2$ и действительная константа $c \geqslant 0$.

Замечательно, что на основании показателя α в этом представлении можно однозначно определить, имеет ли закон конечное математическое ожидание и дисперсию. Если $\alpha = 2$, то и среднее, и дисперсия конечны; если $1 < \alpha < 2$, то только среднее конечно; если $0 < \alpha \leqslant 1$, то среднее и дисперсия бесконечны (см. [11, с. 182]).

Как мы увидим в следующей главе, особый интерес представляет распределение суммы $\sum\limits_{n=1}^{N} \xi_n$. Возникает вопрос, имеет ли место утверждение, подобное центральной предельной теореме (см. § 7.5), заключающееся в том, что данное распределение при достаточно больших N не зависит от вида распределения самих $\{\xi_n\}$. Мы уже упоминали в конце § 7.5, что существуют теоремы о сходимости к устойчивым законам (см. также дополнение 4). К сожалению, удается конструктивно охарактеризовать только хвосты распределения суммы $\sum\limits_{n=1}^{N} \xi_n$, и то в случае, когда хвосты распределения ξ_n удовлетворяют закону Парето (т. е. для больших $|x|$ вероятности $P\{\xi_n > x\}$ при $x > 0$ и $P\{\xi_n < x\}$ при $x < 0$ имеют вид $A/|x|^\beta$ с некоторыми константами A и β.) Для установления формы «средней части» (т. е. когда модуль $|x|$ не слишком велик) предельного распределения приходится (за исключением нормального закона) полагаться на численные процедуры. По этим причинам, а также ввиду ряда эмпирических свидетельств, предположение о нормальности распределения ξ_n стало стандартным в применяемых на практике финансовых моделях. В данном случае также говорят, что доходности r_n имеют *логнормальное распределение.*

Задачи

1. В примере, приведенном в начале § 9.2, выразите R явно через u и d. Чтобы получить представление о типичных значениях этой величины, рассмотрите также числовой пример с $u = 1.1$ и $d = 0.75$ и выразите соответствующее значение R в процентах.

2. Изучим теперь случай, когда R осуществляет отображение на всю действительную прямую. Допустим, что величина R нормально распределена со средним μ и дисперсией σ^2, см. § 7.4 и (7.4.6).

(a) Если $\mu = 0$, что является более вероятным: потерять деньги или приобрести деньги?

(b) Если $\mu < 0$, чему равна вероятность потери денег?

(c) Если $\mu > 0$, какова эта вероятность?

Данная задача показывает, что если принять за меру риска дисперсию доходности, то может оказаться, что разные активы с различными вероятностями потери будут считаться одинаково рискованными.

3. Покажите, что для любой ценной бумаги с положительной дисперсией найдется диверсифицированный портфель, содержащий ценную бумагу с меньшей дисперсией (в предположении, что вторая ценная бумага существует).

4. В примере с активами трех видов (часть 1) установите условия на μ_i, σ_i и ρ_{ij}, при которых функция $V''(\alpha_1) \leqslant 0$ при всех α_1.

5. Напомним, что σ_1 и σ_2 положительны. Найдутся ли такие значения b и ρ_{12}, при которых нарушается условие (9.5.21)?

6. Проверьте, что происходит, когда $\mu_1 = \mu_0$ или $\mu_2 = \mu_0$. Существуют ли при этом какие-нибудь решения у уравнения $V'(\alpha) = 0$?

7. Пусть X обозначает экспоненциально распределенную случайную величину со средним 1 и дисперсией 1, а Y — нормально распределенную величину, также со средним 1 и дисперсией 1.

Для произвольного $x > 0$ сравните $P(X < x)$ и $P(0 < Y < x)$. (Напомним, что экспоненциальная случайная величина X положительна, поэтому $P(0 < X < x) = P(X < x)$, см. пример 12 из § 4.5.)

8. Покажите, что распределение Парето (определенное в приложении 4) имеет более тяжелый правый хвост, чем стандартное нормальное распределение (распределение его хвоста приведено сразу после формулы (7.4.2)). (Надо проводить сравнение только при $x \to \infty$.)

9. (a) Установите, что у распределения Коши (см. § 9.6) математическое ожидание не определено. Что можно сказать о дисперсии этого закона?

(b) Докажите, что если X_1 и X_2 имеют распределения Коши и независимы, то их сумма $X_1 + X_2$ также распределена по закону Коши.

(c) Постройте на одном чертеже графики плотности распределения Коши при $\mu = 0$ и $\sigma = 1$ и стандартной нормальной плотности. Обратите внимание, что обе кривые симметричны, и что график плотности закона Коши лежит над графиком нормальной плотности при $|x| > 2$. Поэтому хвосты распределения Коши тяжелее.

10. (a) Покажите, что сумма двух независимых случайных величин, имеющих гамма-распределения (определение приведено в задаче 37(c) из § 6.5), снова имеет такое же распределение. [*Указание.* Это непосредственно вытекает из задачи 42 в § 6.5.]

(b) Установите, что распределение Парето не является устойчивым. [*Указание.* Проверьте, что для $\alpha = 1$ верно равенство

$$P\{X + Y > z\} = \frac{2}{z} + 2\,\frac{\ln(z-1)}{z^2}\,.]$$

11. Плотность случайной величины X с логнормальным распределением определяется формулой

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-(\ln x - \mu)^2/(2\sigma^2)} \quad \text{для } x > 0,$$

где μ и σ — заданные параметры. Покажите, что

(a) случайная величина $Y = \ln X$ имеет нормальное распределение с математическим ожиданием μ и дисперсией σ^2;

(b) все моменты величины X существуют и конечны;

(c) производящая функция моментов для X не определена.

Приложение 4

Распределение Парето и устойчивые законы

Здесь мы перечислим ряд основных фактов, касающихся устойчивых законов и распределения Парето, к которым в 1960-х и 1970-х гг. было привлечено особое внимание в связи с построением моделей для доходностей акций. Распределение получило свое название в честь Вилфредо Парето, инженера, ставшего экономистом, который в конце XIX в. изучал распределение богатства среди людей. Говорят, что случайная величина X имеет (стандартное) распределение Парето с параметром $\alpha > 0$, если

$$P\{X > x\} = \begin{cases} \dfrac{1}{x^{\alpha}}, & \text{если } x \geqslant 1, \\ 1, & \text{если } x < 1. \end{cases}$$

Если $0 < \alpha \leqslant 1$, то математическое ожидание величины X бесконечно.

Парето получил данное распределение примерно в 1895 г. в результате анализа статистических данных о налогах на доход в Англии за 1843 г. Он был поражен тем фактом, что доля индивидуумов, чей доход в два раза превосходил средний, была существенно больше, чем следовало из предположения о нормальности распределения. Изучая данные о доходах во всей Европе в конце 19-го столетия, Парето предложил в качестве приближения для α значение 1.5, которое интерпретировалось им как индикатор заметного неравенства людей в доходах.

Распределение Парето и устойчивые законы связаны тем, что хвосты последних, за исключением нормального закона, имеют приближенно степенную форму первого распределения. Точнее: для $0 < \alpha < 2$ существуют такие положительные константы C_1 и C_2, что при достаточно больших $|x|$ $P(X > x) \sim C_1/x^{\alpha}$, если $x > 0$, и $P(X < x) \sim C_2/|x|^{\alpha}$, если $x < 0$. Подобно результатам Парето, Мандельброт обнаружил, что предположение о нормальности распределения цен на хлопок не согласуется с реальными данными. По этой причине он рекомендовал использовать для описания финансового рынка устойчивые законы вместо нормального. В одном из недавних исследований по данному вопросу Фама, изучая распределения цен акций, получил для α оценку 1.8.

В теории вероятностей закон Парето использовался для обоснования необходимости обобщения центральной предельной теоремы из § 8.5.

Положим $S_n = \sum\limits_{i=1}^{n} X_i$, где X_i — независимые и одинаково распределенные случайные величины, $i = 1, 2, \ldots, n$. Напомним, что при условии конечности среднего μ и дисперсии σ^2 случайной величины X_1, величина

$$Y_n = \frac{S_n - n\mu}{\sigma n^{1/2}}$$

имеет приближенно стандартное нормальное распределение при достаточно больших n. Иначе говоря, при достаточно больших n

$$P\{Y_n > x\} \sim \int\limits_x^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy$$

(см. теорему 9 в § 7.5).

Но что происходит, когда μ и σ^2 не определены или бесконечны? Можно попытаться центрировать и масштабировать S_n другими константами, зависящими от n. Тогда при некоторых условиях наблюдается сходимость при $n \to \infty$ к закону, близкому к распределению Парето при больших $|x|$.

Для начала заметим, что если $X_1, \ldots, X_n$, где $n \geqslant 1$, — независимые случайные величины со стандартным нормальным распределением и если $S_n = \sum\limits_{i=1}^{n} X_i$, то

$$Y_n = \frac{1}{B_n}\, S_n - A_n \qquad (\text{П}.4.1)$$

также имеет стандартное нормальное распределение. Согласно § 9.7, нормальное распределение содержится в классе симметричных устойчивых законов, характеристические функции которых имеют вид

$$e^{-|\theta|^\alpha}, \quad 0 < \alpha \leqslant 2.$$

Можно обобщить (П.4.1) следующим образом. Если $X_1, \ldots, X_n$ — независимые случайные величины, одинаково распределенные согласно закону с характеристической функцией $e^{-|\theta|^\alpha}$, $0 < \alpha \leqslant 2$, и $S_n = \sum\limits_{i=1}^{n} X_i$, то

$$\frac{S_n}{n^{1/\alpha}}, \quad n \geqslant 1,$$

имеет такое же распределение. Этому свойству обязан своим происхождением термин «устойчивый закон». Для проверки данного результата достаточно убедиться, что характеристической функцией $\varphi_n(\theta)$ случайной величины $Y_n = S_n/n^{1/\alpha}$ служит $e^{-|\theta|^\alpha}$. Действительно,

$$\varphi_n(\theta) = E\left(e^{i\theta Y_n}\right) = E\left(e^{i\,(\theta/n^{1/\alpha})\sum\limits_{j=1}^{n} X_j}\right) = \varphi\left(\frac{\theta}{n^{1/\alpha}}\right)^n = $$
$$= e^{-n\left|\frac{\theta}{n^{1/\alpha}}\right|^\alpha} = e^{-|\theta|^\alpha}.$$

При доказательстве центральной предельной теоремы (см. теорему 9 в § 7.5) предельное стандартное нормальное распределение распознается благодаря известному виду его характеристической функции ($g(\theta) = e^{-\theta^2/2}$). Симметричные устойчивые законы с характеристической функцией $e^{-|\theta|^\alpha}$, $0 < \alpha < 2$, могут быть получены похожим образом, но с другой нормирующей последовательностью для сумм S_n при условии, что распределение независимых одинаково распределенных слагаемых не имеет конечного второго (или даже первого) момента. Хороший пример дает следующая теорема.

Теорема 1. *Пусть $\{X_k\}$ — последовательность независимых и одинаково распределенных случайных величин с плотностью $p(x)$, где*

$$p(x) = \begin{cases} \dfrac{\alpha}{2|x|^{\alpha+1}}, & \text{если } |x| > 1, \\ 0, & \text{если } |x| \leqslant 1. \end{cases}$$

Тогда распределение $\{S_n/n^{1/\alpha}\}$ сходится к симметричному устойчивому закону с характеристической функцией $e^{-|\theta|^\alpha}$.

Доказательство, которое требует слишком сложных для данного текста вычислений, приведено в [4, теорема 6.5.4].

Заметим, что последовательность независимых одинаково распределенных слагаемых в теореме 1 имеет хвостовую вероятность

$$P\{|X| > x\} = \frac{1}{x^\alpha} \quad \text{для } x > 1,$$

первый момент

$$E(X) = \begin{cases} \infty, & \text{если } \alpha \leqslant 1, \\ \dfrac{\alpha}{\alpha - 1}, & \text{если } \alpha > 1 \end{cases}$$

и второй момент

$$E(X^2) = \begin{cases} \infty, & \text{если } \alpha \leqslant 2, \\ \dfrac{\alpha}{\alpha - 2}, & \text{если } \alpha > 2. \end{cases}$$

Обратно, Леви установил, что такое же поведение хвостов асимптотически наблюдается для любого устойчивого закона с показателем $\alpha < 2$, т. е. существуют положительные константы C_1 и C_2, что

$$P\{X > x\} \sim C_1 x^{-\alpha} \quad \text{при} \quad x \to \infty;$$
$$P\{X < x\} \sim C_2 |x|^{-\alpha} \quad \text{при} \quad x \to -\infty.$$

За редким исключением явный вид устойчивых распределений неизвестен. Одним из таких исключений служит закон Коши, стандартная плотность которого задается формулой

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{для } -\infty < x < \infty.$$

Этот закон является симметричным устойчивым законом с показателем $\alpha = 1$. Нетрудно убедиться, что его хвосты имеют вид C/x при $x > 0$ и $C/|x|$ при $x < 0$, где c — положительная константа. Действительно, при $x > 0$ (случай $x < 0$ разбирается аналогично) запишем:

$$P\{X > x\} = \frac{1}{\pi} \int\limits_x^\infty \frac{1}{1+y^2}\, dy \sim \frac{1}{\pi} \int\limits_x^\infty \frac{1}{y^2}\, dy \text{ (при больших } x > 0) = \frac{C}{x},$$

где $C = 1/\pi$. Так как стандартное нормальное распределения является симметричным устойчивым законом, возникает предположение, что использование разных масштабирующих последовательностей, аналогичных последовательности из теоремы 1, может привести к обобщению центральной предельной теоремы и получению других предельных законов.

Для более полного изучения этого вопроса в контексте конструкции случайного блуждания S_n давайте прежде всего введем ряд обозначений. Пусть $F(x)$ — это функция распределения случайных величин X_i,

$$Y_n = \frac{1}{B_n} S_n - A_n \tag{П.4.2}$$

Таким образом, если $\mu < \infty$ и $\sigma < \infty$, то в случае центральной предельной теоремы

$$B_n = \sigma\sqrt{n} \quad \text{и} \quad A_n = \frac{\mu}{\sigma}\sqrt{n}.$$

Обозначим через $F_n(x)$ функцию распределения Y_n. Обобщение центральной предельной теоремы заключается в нахождении условий на F, B_n и A_n, при выполнении которых существует $\lim\limits_{n\to\infty} F_n(x)$. Например, для применимости центральной предельной теоремы достаточно, чтобы F было такой, что μ и σ конечны, а A_n и B_n такие, как указано выше. Доказательство обобщения во всей полноте выходит за рамки сложности этой книги. Ниже мы приводим лишь формулировки основных результатов.

Теорема 2. (См. теорему из § 33, с. 162 в [11].) *Если предельное распределение для Y_n, определяемое как*

$$\Phi(x) = \lim_{n\to\infty} F_n(x),$$

существует, то оно должно быть устойчивым законом.

Напомним, что в силу теоремы Леви, цитированной в § 9.7, устойчивые распределения идентифицируются с помощью показателя α, присутствующего в их характеристической функции. Это обстоятельство имеется в виду в следующих двух теоремах.

Теорема 3. (См. теорему 5, с. 181 в [11].) *Если функция распределения $F(x)$ случайных величин X_i удовлетворяет для больших $|x|$ условиям*

$$\begin{aligned} F(x) &\sim \frac{A}{|x|^\alpha}\,, \quad \textit{при } x < 0, \\ 1 - F(x) &\sim \frac{B}{|x|^\alpha}\,, \quad \textit{при } x > 0 \end{aligned}$$

для некоторых положительных констант A, B и $\alpha \in (0, 2)$, то $F_n(x)$ сходятся к устойчивому закону, обладающему характеристической функцией с показателем α.

Теорема 4. (См. теорему 1, с. 172 в [11].) *Для сходимости $F_n(x)$ к стандартному нормальному закону (имеющему показатель $\alpha = 2$) достаточно, чтобы дисперсия $\mathrm{Var}(X_i)$ была конечной.*

Указанные выше достаточные условия являются на самом деле и необходимыми, если потребовать, чтобы B_n из (A.4.2) имели вид $an^{1/\alpha}$, где a — положительная константа. В теоремах из [11], на которые даны ссылки, в действительности приведены эти достаточные и необходимые условия. В случае центральной предельной теоремы ($\alpha = 2$, дисперсия конечна) имеем $a = \sqrt{\mathrm{Var}(X_i)}$, при этом предельное распределение является стандартным нормальным. Изучим вопрос немного глубже. Один из путей к пониманию того, откуда для B_n появилось представление $an^{1/\alpha}$, основывается на следующей теореме.

Теорема 5. *Пусть $\{X_i\}$ независимы и одинаково распределены согласно устойчивому закону с характеристической функцией*

$$\varphi(\theta) = e^{-\gamma_\alpha|\theta|^\alpha + id\theta}.$$

Тогда для заданного положительного числа a случайная величина

$$Y_n = \frac{1}{a}\, n^{-1/\alpha} \sum_{k=1}^{n} (X_k - d)$$

также распределена согласно устойчивому закону с показателем α.

Доказательство. Для $1 \leqslant k \leqslant n$ положим

$$X_k^* = \frac{X_k - d}{a n^{1/\alpha}}\,.$$

Ее характеристической функцией является

$$\varphi^*(\theta) = E\left[e^{i\theta X_k^*}\right] = E\left[e^{i\theta(X_k - d)/(a n^{1/\alpha})}\right] =$$
$$= e^{-i\theta d/(a n^{1/\alpha})} E\left[e^{i\theta/(a n^{1/\alpha}) X_k}\right] = e^{-i\theta d/(a n^{1/\alpha})} \varphi\left(\frac{\theta}{a n^{1/\alpha}}\right).$$

При условии, что случайные величины $\{X_k\}$, а поэтому и $\{X_k^*\}$, независимы и одинаково распределены, теорема 7 из § 6.5 позволяет записать характеристическую функцию величины Y_n в виде

$$E\left[e^{i\theta Y_n}\right] = E\left[e^{i\theta \sum\limits_{k=1}^{n} X_k^*}\right] = \prod_{k=1}^{n} \varphi^*\theta) =$$
$$= e^{-in\theta d/(a n^{1/\alpha})} \left[\varphi\left(\frac{\theta}{a n^{1/\alpha}}\right)\right]^n = e^{-\gamma_\alpha / a|\theta|^\alpha}.$$

Данная теорема показывает, что (должным образом масштабированные и центрированные) суммы независимых случайных величин с одинаковым устойчивым распределением сходятся с распределению этого же типа.

Мы завершим обсуждение замечанием, что при $0 < \alpha < 2$ свойство иметь хвосты типа распределения Парето, приведенное в условиях теоремы 3, сохраняется и для предельного устойчивого закона. Этот результат был установлен Леви (см. (37) на с. 201 в монографии [18] и § 36 в книге [11]). В свете этого неудивительно, что если само F принадлежит устойчивому типу с показателем α, то и Φ должно, как утверждается в теореме 5, иметь тот же тип.

В случае центральной предельной теоремы указанное выше свойство не сохраняется. Для стандартной нормальной случайной величины Z при $z \geqslant 1$ имеем:

$$P\{|Z| > z\} = \int_z^\infty \frac{e^{-\zeta^2/2}}{\sqrt{2\pi}}\, d\zeta \leqslant \int_z^\infty \zeta\, \frac{e^{-\zeta^2/2}}{\sqrt{2\pi}}\, d\zeta = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}.$$

Это означает, что хвостовая вероятность убывает по меньшей мере со скоростью экспоненты от квадрата. Она сходится к нулю намного быстрее, чем экспонента e^{-z}, не говоря уже о степенной скорости вида $x^{-\alpha}$ ($\alpha > 0$), имеющей место для закона Парето. Это показывает, что оценка сверху для $P\{|Z| > z\}$, обеспечиваемая неравенством Чебышёва (см. (7.6.5)), в данном случае является очень грубой.

Заключительное замечание. Наше изложение теории устойчивых законов базируется в основном на материале из книг [18] и [11]. Однако аналогичные результаты с другими доказательствами содержатся также в монографии [10].

Глава 10

Расчет цены опциона

В предыдущей главе изучалась однопериодная модель, на которой базируется вычисление равновесных цен базовых финансовых инструментов таких, как акции. В текущей главе мы сконцентрируем внимание на так называемых *производных ценных бумагах* или *обусловленных требованиях*. Это — активы, стоимость которых зависит от цен на базовые финансовые инструменты. В этой главе мы проиллюстрируем применение наиболее сложного в этой книге аппарата из области теории случайных процессов. Представленные здесь идеи образуют фундамент многих направлений в финансовой математике, оказавших мощное воздействие как на теорию, так и на практику.

10.1. Основные понятия, относящиеся к опционам

Опцион — это контракт, дающий его обладателю право, но не обязанность, купить или продать ценные бумаги такие, как акции (*базовые активы*), по заранее установленной цене (*цене исполнения контракта* (exercise price или strike price). Если контракт гарантирует право купить бумаги, то он называется *опционом-колл* (option call). Если же предоставляется право продажи, то — *опционом-пут* (option put). В контракте устанавливается также *срок его действия* или *дата исполнения* (expiration или maturity date), которая может быть бесконечной. Опционы, исполняемые только в момент окончания их срока действия, называются опционами *европейского* типа, а те, которые допускают исполнение в произвольный момент вплоть до окончания срока действия — опционами *американского* типа. Оба типа распространены во всем мире, но опционы американского типа встречаются чаще. Опционы представляют собой один из наиболее используемых видов *обусловленных требований*. В качестве стандартных всесторонних источников, содержащих примеры реальных сделок и описание поведения рынка, мы рекомендуем книги [6] и [13].

Пример 1. Допустим, что сегодня вы заключили сделку, дающую вам право покупки одной акции компании General Electric (базовой ценной бумаги) за \$30 в любой из дней в течение следующих трех месяцев. Тогда вы являетесь владельцем американского опциона-пут на акции General Electric с ценой исполнения в размере \$30 за акцию и 3-месячным сроком действия.

Пример 2. Если вы заключили контракт, дающий вам право продавать один китайский юань за 16 японских иен в точности через 6 месяцев от текущей даты, то вы приобрели опцион европейского типа на обмен китайской и японской валют с ценой исполнения в 16 японских иен за юань и сроком действия длиной в 6 месяцев.

В обоих приведенных примерах цена, по которой базовые активы будут покупаться или продаваться, фиксирована и известна заранее. Такие опционы называются стандартными в отличие от «экзотических» опционов, где обычно оговаривается не сама цена, а некоторое правило, согласно которому эта цена будет определяться. Например, можно заключить сделку, предоставляющую право покупать акции General Electric по средней цене, вычисляемой на основании ежедневных цен на данные акции, которые будут наблюдаться в течение следующих трех месяцев. В оставшейся части текущей главы мы сконцентрируем внимание на стандартных опционах как европейского, так и американского типов. Для упрощения изложения будем рассматривать только опционы, в которых базовыми ценными бумагами являются не имеющие дивидендов акции, если особо не оговорено противное.

Опционы используются для нескольких целей. Цены стандартных опционов существенно ниже, чем соответствующих базовых активов. Поэтому спекулянты имеют возможность получать доход в случае благоприятного изменения цен на акции без вложения больших средств, которых потребовала бы покупка самих акций. Опционы можно также рассматривать как страховые договоры, ввиду того что они обеспечивают своим владельцам защиту от неблагоприятных изменений стоимости базовых активов. К примеру, инвестор, владеющий в настоящий момент акциями некоторой компании и опасающийся предполагаемого резкого снижения цены на них в ближайшие три месяца, может захотеть приобрести опцион-пут на данные акции, чтобы гарантировать себе минимальную цену, по которой он или она сможет продать акции через три месяца.

Для получения права на исполнение опциона, его владелец сначала должен заплатить некоторую сумму (премию) в момент подписания контракта. Эта премия, иначе называемая (начальной) *ценой опциона*, находится в фокусе теории. В случае опционов американского типа теория позволяет определить также оптимальный момент исполнения, т. е. самое выгодное время для предъявления контракта к исполнению.

Опционная выплата. Стандартный опцион характеризуется ценой исполнения K (\$30 в примере 1 выше), сроком действия T, выражаемым в годах (0.25 года в примере 1) и своим типом: колл или пут. Цена S_T на базовую ценную бумагу через T единиц времени от текущего момента считается случайной величиной, т. е. предполагается, что существует выборочное пространство Ω, такое, что стоимость базовой ценной бумаги в момент T есть $S_T(\omega)$ в случае выбора $\omega \in \Omega$. Давайте переместимся во времени до момента T и предположим, что реализовалась точка $\tilde{\omega}$. Тогда в случае, например, опциона типа колл сделка осуществляется только в случае, когда $S_T(\tilde{\omega}) > K$. Причина заключается в том, что нет смысла покупать актив по цене K, если он продается на рынке по цене $S_T(\tilde{\omega}) \leqslant K$ (для удобства мы будем считать, что когда $S_T(\tilde{\omega}) = K$ опцион также не реализуется (исполняется)). Таким образом, при условии, что $S_T(\tilde{\omega}) > K$, владелец опциона с ценой исполнения K может реализовать опцион, т. е. купить акции по K долларов за штуку, а затем продать их на рынке, получая прибыль в размере $S_T(\tilde{\omega}) - K$ за акцию. Если $S_T(\tilde{\omega}) \leqslant K$, то владелец опциона его не реализует и не получает никакой прибыли. Обобщая вышесказанное, дадим следующее

Определение 1. *Пусть для заданного T величина S_T обозначает цену акций в момент T. Рассмотрим стандартный опцион европейского типа с ценой исполнения K и сроком действия T. Тогда опционной выплатой называется случайная величина $g = \max(S_T - K, 0)$ для опциона-колл, либо $g = \max(K - S_T, 0)$ для опциона-пут.*

Как уже отмечалось в начале главы, опционы являются одним из видов широкого класса финансовых инструментов, называемых производными. Ниже мы приводим общее определение и несколько примеров.

Определение 2. *Производным финансовым инструментом называется контракт между двумя сторонами, в котором указываются:*

(a) *базовый актив, на предмет которого он заключен,*

(b) *дата исполнения (срок действия) контракта (конечный или неограниченный),*

(c) *условия соглашения (необязательный для одной из сторон или обязательный для обеих, сделка осуществляется в любой момент вплоть до окончания срока действия или только при наступлении даты исполнения),*

(d) *согласованный сторонами размер выплаты.*

Вы познакомились в примерах 1 и 2 с конкретными формами производных инструментов (опционы на акции, валютные опционы) с конечными сроками действия, касающихся определенных базовых активов (акции, валюта). Одна сторона является покупателем (владельцем) контракта, другая сторона его подписывает. Условия договора дают владельцу право осуществить сделку либо при наступлении даты исполнения (европейский стиль), либо в любое время вплоть до этой даты (американский стиль) и получить выплату g, указанную в определении 1.

Бывают и более сложные типы опционов. Например,

- опцион «с памятью» (lookback) с ценой исполнения K и сроком действия T — это опцион с выплатой
$$\max\big(\max_{0\leqslant t\leqslant T} S_t - K, 0\big);$$
- усредненный (average) опцион с ценой исполнения K имеет выплату
$$\max\left(\frac{1}{N}\sum_{k=1}^{N} S_{t_k} - K, 0\right),$$
где
$$1 \leqslant k \leqslant N, \quad 0 = t_1 \leqslant t_2 \leqslant \ldots \leqslant t_N = T;$$
- ограниченный (barrier) опцион — стандартный опцион (такой, как мы описывали в примерах 1 и 2), который отменяется («knocked-out») или вступает в силу («knocked-in»), если цена на базовый актив выходит за заранее оговоренный уровень (границу). В последнем случае в момент заключения контракта его владелец еще не получает право на осуществление сделки и обязан ждать, пока граница не будет нарушена.

Существуют производные финансовые инструменты и более простого вида. Например, форвардным контрактом называется обязательство поставить определенный актив (скажем, валюту, нефть, зерно) по заранее согласованной цене (K) к определенной дате T. Выплатой владельцу служит величина $S_T - K$, где S_T — это цена на базовый актив в момент исполнения контракта. Отметим, что при применении данного инструмента обе стороны берут на себя обязательства, которые они должны выполнить, когда наступит дата T.

Цена D_t на производный финансовый инструмент в момент t зависит от поведения цены на базовый актив вплоть до момента t. В случае стандартного опциона можно записать, что $D_t = f(t, S_t)$, где S_t — цена базового актива в момент t, которая одна только и требуется из всей последовательности цен, наблюдавшихся вплоть до момента t. Важнейшим результатом в финансовой математике, появившимся в основополагающей статье Блэка и Шоулза (1973), стало открытие дифференциального уравнения в частных производных, характеризующего f в предположении, что $\log S_t$ является броуновским движением (определение броуновского движения приведено в § 8.2, обоснование выбора в качестве изучаемого процесса $\log S_t$ содержится в § 9.6). В случае функций выплат, указанных в определении 1, Блэк и Шоулз нашли в явном виде решение этого уравнения и вывели свою знаменитую формулу, которую мы кратко рассмотрим ниже.

С точки зрения спекулянта или рыночного игрока, ценой опциона является сумма, которую он или она готовы заплатить в момент 0 для того, чтобы получить опционную выплату через время T.

Как правило, покупка опциона считается менее дорогостоящим способом извлечения дохода из колебаний цен, позволяющим не платить вперед полную стоимость акций. Например, в 8 марта 2002 г. цена одной акции компании IBM составляла приблизительно \$105, в то время как требовалось заплатить всего около \$11 за опцион-колл на эти акции, имеющий цену исполнения \$100 и срок действия до июля 2002 г. В свою очередь, в то же самое время, необходимо было заплатить \$6 для приобретения опциона-пут на акции IBM с ценой исполнения \$100 и тем же самым сроком действия.

Необходимо также не забывать, что опцион предполагает наличие двух сторон, подписывающих контракт: покупателя (или владельца) опциона и продавца, которые имеют разные представления о будущем поведении цены базового актива. Часто продавец опциона-колл в действительности не располагает соответствующими ценными бумагами, надеясь, что покупатель никогда не предъявит опцион к исполнению. В истории известен ряд ярких эпизодов, когда спекуляция приводила к колоссальному росту цен и в результате к крупному финансовому краху. Один из наиболее знаменитых примеров — «тюльпаномания» в Нидерландах в 1630-х гг. Изначально тюльпаны выращивались в Турции и были завезены в Голландию в начале XVI в. К XVII в. садоводами было выведено множество новых сортов тюльпанов. Эти красивые и экзотические цветы выросли в цене и стали играть роль символа, отражающего статус покупателя. Как только на них появился спрос, торговцы поняли, что путем спекуляции на их ценах можно получать большую прибыль. Люди с умеренными доходами покупали и продавали луковицы тюльпанов. Для защиты от беспрестанного роста цен многие спекулянты приобретали опционы-колл (с высокими ценами исполнения) у других спекулянтов, которые надеялись, что им не придется предоставлять владельцам опционов луковицы по этим ценам. В конце концов, покупатели перестали верить в то, что контракты будут выполнены. Многие уже не могли позволить себе покупать по таким высоким ценам, и рынок обрушился.

Похожие случаи спекуляции с производными финансовыми инструментами происходили и в недавнем прошлом. В частности, в 1990-х гг. заключались рискованные пари относительно валютных курсов и доходностей активов. Их предлагали инвесторы, желающие тем самым увеличить известность своих организаций. В результате многие из организаций обанкротились. Barings, один из старейших банков, сыгравший важную роль в расширении Британской Империи; Orange County, одна из крупнейших компаний в Калифорнии; Long-Term Capital Management, занимавшаяся управлением инвестициями, которую консультировали два нобелевских лауреата по экономике, и в акции которой вложили деньги многие очень богатые люди, крупные организации и университеты[*] — все они либо отказались выполнять свои контракты, либо выкарабкались из долговой ямы с огромным трудом.

Дисконтирование. В задаче определения цены опциона необходимо учитывать изменение во времени покупательной способности денег. Обычным в финансах подходом является введение временного коэффициента для цен; т. е. вещь, которую сегодня можно купить за один доллар, через год в результате инфляции будет стоить больше одного доллара. Другими словами, реальная стоимость доллара через год может отличаться от его сегодняшнего номинала. Чтобы производить справедливое сравнение стоимостей наличных денег в разные времена, финансовые экономисты предложили понятие, позволяющее сводить будущие цены к теперешним, называемое *учетной ставкой*[**]. В нашем случае она совпадает с доходностью r безрискового вложения, введенного в § 9.2. Ради простоты будем считать, что в нашей экономике существует только одно безрисковое вложение, и мы сводим цены, наблюдаемые через интервалы длиной в один день, к цене в момент 0. Если у нас есть ценная бумага, имеющая цену C_n в момент t_n, то ей соответствует дисконтированная цена

$$\frac{C_n}{(1+r)^n}$$

в момент 0.

[*] В феврале 2000 г. телевизионная сеть PBS показала документальный фильм с участием главных действующих лиц этих событий и ряда экономических обозревателей. Так же, как и рукопись данной книги, стенографическая запись фильма («Пари на триллион долларов») доступна на сайте www.pbs.org/wgbh/nova/stockmarket.

[**] В оригинале используется термин «discount rate». — *Прим. перев.*

Субмартингалы и супермартингалы. Покажем кратко, что определение цены опциона тесно связано с вычислением математических ожиданий случайных величин. Мартингалы, введенные в приложении 3, дают модель для честной игры, при которой шансы на выигрыш и проигрыш уравновешиваются. Однако встречаются ситуации, когда шансы, определяемые, скажем, вероятностной мерой, по которой вычисляются условные математические ожидания из (П.3.1), либо благоприятны, либо неблагоприятны для игрока. При использовании обозначений из приложения 3 первый случай выражается как

$$E(X_{n+1}|X_0, X_1, \ldots, X_n) \geqslant X_n, \tag{10.1.1}$$

а второй случай как

$$E(X_{n+1}|X_0, X_1, \ldots, X_n) \leqslant X_n. \tag{10.1.2}$$

Процесс $\{X_n\}$ называется *субмартингалом*, если он удовлетворяет условию (10.1.1), и *супермартингалом*, если выполняется неравенство (10.1.2). К сожалению, существует несоответствие между словесным выражением этих понятий и их смыслом. Вообще говоря, слова с приставкой «супер» обычно ассоциируются с благоприятным исходом. Но в данном случае супермартингал соответствует невыгодной игре.

Теорема 1. *При выполнении условий теоремы из приложения 3 для произвольного момента остановки T выполняются неравенства*

$$\begin{aligned} & E(X_T) \geqslant E(X_0) \quad \textit{для субмартингала} \\ \textit{и} \quad & E(X_T) \leqslant E(X_0) \quad \textit{для супермартингала.} \end{aligned}$$

Для любого субмартингала (соответственно, супермартингала) неравенства имеют место, если момент T ограничен.

Доказательство. Достаточно заменить равенства в доказательстве из приложения 3 на соответствующие неравенства (убедитесь).

Проиллюстрируем диапазон применимости данной теоремы на следующем примере.

Пример жадности. Рассмотрим двух людей, имеющих одинаковый начальный капитал X_0. Первый человек не очень жадный. Он согласен закончить процесс инвестирования, как только получит больше, чем одну единицу прибыли (единицей может служить как \$1, так и \$10 000). Второй — более агрессивный. Он не собирается останавливаться до тех пор, пока не увеличит свой начальный капитал более чем в 10 раз. Для первого инвестора положим

$$T = \min\{n\colon X_n > X_0 + 1\},$$

где X_n — размер капитала в момент n. Для второго инвестора момент T определяется формулой

$$T = \min\{n\colon X_n > 10X_0\}.$$

Согласно законам природы, любому человеку отведено лишь конечное число лет жизни (некоторые, как Королева-мать в Великобритании, доживают до 101 года). Давайте обозначим через $\bar{t}$ верхнюю границу продолжительности жизни. Положим $\overline{T} = T \wedge \bar{t}$, тогда момент $\overline{T}$ ограничен. Если процесс $\{X_n\}$ является супермартингалом, то выполняется неравенство

$$E\big(X_{\overline{T}}\big) \leqslant E(X_0).$$

В любом случае его/ее ожидает уменьшение капитала! Все, что мы можем вычислить с помощью теории вероятностей, — это *ожидаемый* (предполагаемый) результат. Люди могут возразить: «Но я знаю, что Майк купил акции компании Enron по 8.52, а продал по 88.87». На это лорд Кейнс давным давно дал ответ, цитированный в § 5.3 после теоремы 2. Он привел его в своей докторской диссертации по философии прежде, чем погрузился в денежные вопросы (эта его знаменитая работа по экономике озаглавлена «Money»). Ссылка на его философский труд дана в списке литературы.

Можно также подсчитать шансы, что наши инвесторы достигнут своих целей. Согласно неравенству Чебышёва из (7.6.5) (точнее — его разновидности из задачи 21 в § 7.6) для заданной положительной константы b имеем:

$$P\big(X_{\overline{T}} > b\big) \leqslant \frac{E(X_{\overline{T}})}{b}\,.$$

В силу приведенной выше теоремы 1 верно неравенство $E\big(X_{\overline{T}}\big) \leqslant E(X_0)$. Отсюда

$$P\big(X_{\overline{T}} > b\big) \leqslant \frac{E(X_0)}{b}\,.$$

Допустим, что X_0 равняется константе C, а $b = a \times C$ для некоторого $a > 0$. Тогда

$$P\big(X_{\overline{T}} > aC\big) \leqslant \frac{1}{a}\,.$$

Например, для $a = 10$ шансы достижения нашим агрессивным инвестором поставленной цели составляют менее 10 %.

Колебания цен акций и их ожидания. Если следить за эволюцией цены определенной акции с помощью средств информации таких, как газеты, радио, телевидение или Интернет, вы убедитесь, что она то растет, то падает, или имеет тенденцию к росту или падению. Это поведение, возможно, напомнит вам случайное блуждание (определенное в § 8.1). Для формализации данного наблюдения мы фактически будем опираться на аппарат, введенный в § 9.2.

Напомним главный вывод из проведенного ранее обсуждения темы дисконтирования: если требуется сравнить цену акций S_n в момент n с ценой S_m в момент m, надо посмотреть на их дисконтированные значения. Поэтому введем

$$Y_n = \frac{1}{(1+r)^n}\, S_n \quad \text{для} \quad n = 0, 1, 2, \ldots, N.$$

В дальнейшем любые ожидания, касающиеся перспектив акций, станем вычислять на основе дисконтированных цен $\{Y_n\}$.

Напомним, что главной причиной заинтересованности инвесторов в опционах служит их представление о возможном изменении стоимости ценных бумаг. Они могут рассматривать опцион как страховку, например, если собираются в будущем приобретать акции. Если ожидается увеличение стоимости, то им стоит зафиксировать покупную цену (цену исполнения) сегодня, путем приобретения опциона. За право купить в будущем акции по зафиксированной цене инвесторы должны заплатить страховую премию (цену опциона).

В оставшейся части главы будем предполагать, что однопериодные доходности $\{r_n\}$, $n = 1, 2, \ldots, N$, независимы и одинаково распределены. Тогда, следуя обозначениям, введенным в приложении 3, вычислим условные математические ожидания величин Y_n при условии всего «прошлого» процесса вплоть до момента $n-1$:

$$\begin{aligned}
E[\,Y_n \mid Y_0, Y_1, \ldots, Y_{n-1}\,] = E[\,Y_n \mid Y_{n-1}\,] = E\left[\frac{1}{(1+r)^n}\, S_n \mid S_{n-1}\right] &= \\
= \frac{E[S_n \mid S_{n-1}]}{(1+r)^n} = \frac{E[(1+r_n)S_{n-1} \mid S_{n-1}]}{(1+r)^n} &= \\
= \frac{1}{(1+r)^n}\,[1+E(r_n)]\,S_{n-1} = \frac{1+E(r_n)}{1+r}\,\frac{S_{n-1}}{(1+r)^{n-1}}&.
\end{aligned}$$

Отсюда выводим равенство

$$E[\,Y_n \mid Y_0, Y_1, \ldots, Y_{n-1}\,] = \frac{1+E(r_n)}{1+r}\, Y_{n-1},$$

показывающее, что процесс дисконтированных цен является мартингалом, субмартингалом или супермартингалом, соотвественно, если $E(r_n) = r$, $E(r_n) \geqslant r$ или $E(r_n) \leqslant r$.

Дискретные и непрерывные модели. Для определения цены опциона необходимо изучить эволюцию стоимости базового актива, рассматриваемую как стохастический процесс. До сих пор в этой книге мы встречались с двумя типами стохастических процессов: теми, которые изменяют значение только в дискретные моменты времени, например, случайное блуждание из § 8.1, и теми, которые эволюционируют непрерывно, например, броуновское движение из § 8.2. Исторически важнейший результат, относящийся к проблеме определения цены опциона, был получен в контексте более сложного случая процессов с непрерывным временем. Он представляет собой явную формулу (называемую формулой Блэка—Шоулза), о которой мы собираемся кратко рассказать. Как признание их заслуг в области экономики, Майрон Шоулз и Роберт Мертон получили Нобелевскую премию за эту формулу и ее фундаментальные обобщения (Фишер Блэк уже скончался к тому времени, когда в 1997 г. было принято решение о награждении). С другой стороны, модель с дискретным временем, которую мы будем изучать детально, проще для понимания, но имеет тот недостаток, что для нее не удается выписать формулу решения в замкнутой форме. В этой главе мы изложим подход, позволяющий аппроксимировать явную формулу, полученную для непрерывной модели, с помощью моделей с дискретным временем.

Формула Блэка—Шоулза. Рассмотрим акцию, имеющую среднюю мгновенную доходность μ и дисперсию мгновенной доходности σ^2, рынок, на котором мгновенная доходность безрискового вложения равна r, и опцион с датой исполнения T. Допустим, что на данную акцию дивиденды не выплачиваются. Блэк и Шоулз вывели следующие формулы для вычисления цен C_0 и P_0, соответственно, опциона-колл и опционапут европейского типа на эту акцию:

$$C_0 = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

и

$$P_0 = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1),$$

где

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T},$$

а Φ — функция распределения стандартного нормального закона. Обратите внимание, что средняя мгновенная доходность не участвует ни в одной из формул. Мы покажем в случае дискретной модели, что данное обстоятельство имеет отношение к определению цены в отсутствие арбитража (это понятие будет определено ниже) и связанному с ним понятию эквивалентной мартингальной меры.

10.2. Цена опциона при отсутствии арбитража: 1-периодная модель

В контексте дискретного времени однопериодная модель служит строительным блоком для многопериодной модели, используемой для определения цены опциона. Для вычисления последней необходимо знать природу стохастического поведения стоимости соответствующей базовой ценной бумаги. В простой однопериодной ситуации наше выборочное пространство Ω состоит только из двух элементов H и T таких, что $P(H) = p$ и $P(T) = 1 - p$, $p \in (0,1)$. Обозначим через S_0 начальную стоимость акций, на которые заключается опцион (S_0 — постоянная случайная величина: $S_0(H) = S_0(T)$). Если выбрана точка $\omega = H$, то стоимость $S_1(H)$ акций в момент 1 равна uS_0, если выбрана $\omega = T$, то $S_1(T) = dS_0$, причем u и d таковы, что

$$0 < d < 1 + r < u, \tag{10.2.1}$$

где $r \geqslant 0$ — однопериодная доходность безрискового вложения. Напомним, что однопериодная доходность акций определяется как $(S_1 - S_0)/S_0$. Отсюда

$$\frac{S_1(\omega) - S_0}{S_0} = \begin{cases} u - 1, & \text{если } \omega = H, \\ d - 1, & \text{если } \omega = T. \end{cases}$$

Можно переписать неравенство (10.2.1) в виде

$$d - 1 < r < u - 1.$$

Интерпретируя $u - 1$ и $d - 1$, соответственно, как благоприятную (т. е. когда мы получаем прибыль) и неблагоприятную ставки доходности акций, покажем, что доходность безрискового вложения обязана удовлетворять (10.2.1). В противном случае, если $d \geqslant 1 + r$, то

$$\frac{S_1(\omega) - S_0}{S_0} \geqslant r \quad \text{для всех} \quad \omega \in \Omega.$$

Следовательно, доходность акций всегда не ниже доходности безрискового вложения. В такой ситуации нет резона инвестировать в безрисковое вложение, т. е. само существование этого актива теряет смысл. Аналогично, условие $1+r \geqslant u$ влечет неравенство

$$\frac{S_1(\omega) - S_0}{S_0} \leqslant r \quad \text{для всех} \quad \omega \in \Omega,$$

делая непривлекательным инвестирование в акции.

Обозначим через V_0 и V_1 цены опциона европейского типа, соответственно, в моменты 0 и 1. Как мы отмечали выше, V_1 также называется выплатой g в момент исполнения 1. Для определения цены опциона V_0 в момент 0, предположим, что инвестиционные возможности представлены на рынке только следующими тремя видами активов: опционами, соответствующими им акциями и безрисковым вложением. Последнее мы будем считать пакетом облигаций, чем оно обычно и является в действительности. Тем самым, инвестиционный портфель будет включать в себя три параметра (α, β, γ), где α — количество опционов, β — количество акций, γ — количество облигаций. Определим B_n и S_n как цены, соответственно, облигаций и акций в моменты $n = 0, 1$.

Рынки без арбитража. Очевидно, что нельзя рассчитывать на применение такой инвестиционной стратегии (α, β, γ), при которой, стартуя с начального капитала

$$W_0 = \alpha V_0 + \beta S_0 + \gamma B_0 = 0,$$

мы придем в момент 1 к капиталу

$$W_1 = \alpha V_1 + \beta S_1 + \gamma B_1,$$

такому, что $W_1(\omega) > 0$ при всех $\omega \in \Omega$, т. е. начиная с нулевого богатства мы гарантировано получим положительный доход в момент 1. Дело в том, что такая стратегия экономически неустойчива. Если бы она существовала, то это было бы быстро обнаружено многими маклерами. Из-за их желания придерживаться в точности таких (α, β, γ), цены на опционы, акции или облигации в момент 0 выросли или упали бы так, что равенство $W_0 = 0$ сразу перестало бы выполняться. Более строгое рассуждение, опирающееся на так называемые теоремы отделимости, содержится в [8]. Формально определим *арбитражную возможность* или *арбитражный портфель* как стратегию (α, β, γ), при которой выполняются следующие условия:

(i) $\alpha V_0 + \beta S_0 + \gamma B_0 = 0$,

(ii) $\alpha V_1(\omega) + \beta S_1(\omega) + \gamma B_1(\omega) \geqslant 0$ при всех $\omega \in \Omega$,

(iii) $E\{\alpha V_1 + \beta S_1 + \gamma B_1\} > 0$.

Как мы скоро увидим, могут существовать такие арбитражные возможности, что (ii) выполняется строго, т. е.

$$\alpha V_1(\omega) + \beta S_1(\omega) + \gamma B_1(\omega) > 0 \quad \text{при всех} \quad \omega \in \Omega.$$

Это означает, что стратегия (α, β, γ) обеспечивает гарантированный выигрыш: стартуя с нулевого капитала вы обязательно получите прибыль, независимо от того, какая ω реализуется.

Что такое *арбитраж*? Большинство финансовых книг сразу описывают механику арбитража, не останавливаясь подробно на происхождении самого термина. Если вы посмотрите в словарь, то, вероятно, у вас останется чувство неудовлетворенности, вызванное неполнотой определения. Нам удалось обнаружить один источник [23], точнее — гл. 1 из этой книги, где автор отслеживает некоторую эволюцию значения данного слова за несколько изданий словаря. Понятие арбитража, как мы здесь его определяем, является в действительности расширенным толкованием его первоначального значения, которое употребляется в финансовой области, где оно используется в ситуации, когда некоторое лицо покупает ценную бумагу на одном рынке (например, на Нью-Йоркской фондовой бирже) и сразу продает ее на другом рынке (например, на Парижской бирже), где за нее дают более высокую цену. Оксфордский английский словарь приводит эту интерпретацию и, кроме того, указывает на французские корни термина, конкретно — глагол «arbitrer», который переводится как «быть арбитром» или «судить». В этом смысле, арбитраж описывает человека (на самом деле, спекулянта), который пытается определить стоимость ценной бумаги или судить о ее цене, для того, чтобы обнаружить недооценку (или переоценку), т. е. различие между ценой, по которой бумага продается, и той, по которой она должна продаваться. Рынком без арбитража называется такой рынок, на котором не существует арбитражных возможностей. В дальнейшем мы будем предполагать, что рынок имеет именно такой тип.

Лемма. *Пусть $\Omega = \{H, T\}$. Тогда существует портфель $(\alpha_0, \beta_0, \gamma_0)$ такой, что $\alpha_0 = 0$ и*

$$\beta_0 S_1(\omega) + \gamma_0 B_1(\omega) = g(\omega) \quad \textit{для всех} \quad \omega \in \Omega.$$

Доказательство очевидно, так как достаточно просто найти β_0 и γ_0 из системы двух линейных уравнений. Действительно, так как $\Omega = \{H, T\}$, то β_0 и γ_0 должны удовлетворять соотношениям

$$\beta_0 S_1(H) + \gamma_0 B_1(H) = g(H) \quad \text{и}$$
$$\beta_0 S_1(T) + \gamma_0 B_1(T) = g(T)$$

или эквивалентным им равенствам

$$\beta_0 u S_0 + \gamma_0(1+r)B_0 = g(H) \quad \text{и}$$
$$\beta_0 d S_0 + \gamma_0(1+r)B_0 = g(T),$$

из которых выводим, что

$$\beta_0 = \frac{g(H) - g(T)}{(u-d)S_0}, \quad \gamma_0 = \frac{1}{(1+r)B_0}\left[\frac{ug(T) - dg(H)}{u-d}\right]. \tag{10.2.2}$$

Эта лемма показывает, что используя портфель, состоящий только из акций и облигаций (без опционов), мы можем воспроизвести цену опциона в момент исполнения контракта. Такая стратегия называется *воспроизводящей**). Следующее утверждение показывает, что цена опциона V_0 должна совпадать с ценой данного портфеля $(\beta_0 S_0 + \gamma_0 B_0)$ в момент 0.

Утверждение 1. *Цена в момент 0 опциона европейского типа на акции с ценой S и выплатой g в момент 1 задается формулой*

$$V_0 = \frac{1}{1+r}\left[\tilde{p}g(H) + (1-\tilde{p})g(T)\right], \tag{10.2.3}$$

где

$$\tilde{p} = \frac{1+r-d}{u-d}. \tag{10.2.4}$$

Доказательство. Прежде всего покажем, что

$$V_0 = \beta_0 S_0 + \gamma_0 B_0,$$

где β_0 и γ_0 определяются формулами (10.2.2). Допустим, что

$$\varepsilon = V_0 - (\beta_0 S_0 + \gamma_0 B_0) > 0.$$

Тогда рассмотрим следующий портфель:

$$\left(-1, \beta_0, \gamma_0 + \frac{\varepsilon}{B_0}\right).$$

В момент 0 цена этого портфеля есть

$$W_0 = -V_0 + \beta_0 S_0 + \left(\gamma_0 + \frac{\varepsilon}{B_0}\right)B_0 = 0.$$

В момент 1 его цена есть

$$W_1 = -V_1 + \beta_0 S_1 + \left(\gamma_0 + \frac{\varepsilon}{B_0}\right)B_1 = V_1 + \beta_0 S_1 + \gamma_0 B_1 + \varepsilon\left(\frac{B_1}{B_0}\right).$$

*) В оригинале используется термин «replicating strategy». — *Прим. перев.*

Вспоминая, что $V_1 = g$, и применяя приведенную выше лемму, имеем:

$$-V_1(\omega) + \beta_0 S_1(\omega) + \gamma_0 B_1(\omega) = 0, \quad \text{для всех} \quad \omega \in \Omega.$$

Поскольку B_0 и B_1 положительны, получаем, что

$$W_1(\omega) = \varepsilon\left(\frac{B_1}{B_0}\right) > 0$$

для всех $\omega \in \Omega$. Тем самым, мы показали, что $(-1, \beta_0, \gamma_0 + \varepsilon/B_0)$ — это арбитражный портфель. Аналогичное доказательство с использованием портфеля $(1, -\beta_0, -\gamma_0 - \varepsilon/B_0)$ применимо в случае

$$\varepsilon = V_0 - (\beta_0 S_0 + \gamma_0 B_0) < 0.$$

С учетом (10.2.2), теперь мы можем записать цену опциона в виде

$$V_0 = \beta_0 S_0 + \gamma_0 B_0 = \frac{g(H) - g(T)}{u - d} + \frac{1}{1+r}\left(\frac{ug(T) - dg(H)}{u - d}\right) =$$
$$= \frac{1}{1+r}\left[\frac{1+r-d}{u-d}\,g(H) + \frac{u-(1+r)}{u-d}\,g(T)\right],$$

где, как мы помним,

$$g(H) = \begin{cases} \max(uS_0 - K,\ 0) & \text{для опциона-колл,} \\ \max(K - uS_0,\ 0) & \text{для опциона-пут,} \quad \text{и} \end{cases}$$

$$g(T) = \begin{cases} \max(dS_0 - K,\ 0) & \text{для опциона-колл,} \\ \max(K - dS_0,\ 0) & \text{для опциона-пут.} \end{cases}$$

Замечания

1. В силу условия (10.2.1) имеем неравенство $0 < \tilde{p} < 1$. Ввиду этого величина

$$\tilde{p}g(H) + (1 - \tilde{p})g(T)$$

может *интерпретироваться* как математическое ожидание случайной величины g, вычисленное по вероятностной мере $\tilde{P}$, определенной на $\Omega = \{H, T\}$ следующим образом:

$$\tilde{P}\{H\} = \tilde{p} \quad \text{и} \quad \tilde{P}\{T\} = 1 - \tilde{p}$$

(так как в случае $\Omega = \{0,1\} \times \{0,1\} \times \ldots \times \{0,1\}$ это рассуждение появится снова, читателю, возможно, захочется перечитать сейчас § 2.4, чтобы вспомнить, как задаются вероятностные меры на счетных пространствах). Вероятностную меру $\tilde{P}$ часто называют *нейтральной по отно-*

шению к риску, однако мы будем использовать краткий термин *мартингальная мера**). Итак, в однопериодной модели цена в момент 0 опциона европейского типа с выплатой g в момент 1 представляет собой дисконтированное ожидаемое значение этой выплаты по мартингальной мере. Это ожидание — условное, потому что $g(T)$ и $g(H)$ зависят от информации в момент 0, а именно — от S_0.

2. Говорят, что мартингальная мера $\tilde{P}$ *эквивалентна* мере P. Данное свойство формально определяется так: $P(\omega) > 0$ тогда и только тогда, когда $\tilde{P}(\omega) > 0$ для всех $\omega \in \Omega$. Это означает, что меры P и $\tilde{P}$ сосредоточены на одних и тех же точках выборочного пространства. В рассматриваемой сейчас случае $\Omega = \{H, T\}$ это свойство является тривиальным, поскольку неравенства $0 < p < 1$ и $0 < \tilde{p} < 1$ означают, что

$$
\begin{aligned}
&P\{H\} = p > 0, \quad P\{T\} = 1 - p > 0,\\
&\tilde{P}\{H\} = \tilde{p} > 0, \quad \tilde{P}\{T\} = 1 - \tilde{p} > 0.
\end{aligned}
$$

Однако для более сложных пространств Ω проверка может оказаться не такой легкой.

3. Цена V_0 не зависит от первоначальной вероятностной меры, заданной на Ω. Давайте вычислим условное ожидание по мере $\tilde{P}$:

$$
\begin{aligned}
\tilde{E}\,[\,S_1 \mid S_0\,] &= \tilde{p}uS_0 + (1-\tilde{p})dS_0 = \\
&= \frac{1+r-d}{u-d}\;uS_0 + \frac{u-(1+r)}{u-d}\;dS_0 = (1+r)S_0,
\end{aligned}
$$

которое можно переписать как

$$
\tilde{E}[Y_1|Y_0] = Y_0, \tag{10.2.5}
$$

где мы определили

$$
Y_k = \frac{1}{(1+r)^k}\;S_k, \quad k = 0, 1.
$$

Равенство (10.2.5) выражает то, что по отношению к мартингальной мере $\tilde{P}$ дисконтированные цены акций $\{Y_k\}_{k=0}^{1}$ образуют мартингал в нашей достаточно тривиальной ситуации однопериодной модели. В свою очередь, по отношению в первоначальной вероятностной мере данный процесс может быть как субмартингалом, так и супермартингалом, в зависимости от величины математического ожидания однопериодной доходности. На самом деле, по отношению к мартингальной мере

*) В оригинале – «pricing probability». Происхождение названия объясняется в замечании 3. — *Прим. перев.*

все процессы, образованные дисконтированными ценами как облигаций, так и опционов, суть мартингалы. Для облигаций этот результат очевиден, потому что процесс, состоящий из их цен, не является случайным. Для опционов достаточно заметить, что (10.2.3) можно переписать как

$$\frac{1}{(1+r)^0}\ V_0 = \frac{1}{1+r}\ \tilde{E}\,[\,V_1 \mid V_0\,].$$

Действительно, согласно определению, $V_1(\omega) = g(\omega)$ для всех $\omega \in \Omega$, и $g(\omega)$ есть функция от S_0 и ω, что ввиду соотношения $V_0 = \beta_0 S_0 + \gamma_0 B_0$ влечет равенство

$$\tilde{E}[V_1 \mid S_0] = \tilde{E}[\,V_1 \mid V_0\,].$$

4. *Принцип относительной инвариантности.* При нашем арбитражном оценивании опциона мы проводили рассуждения, в которых цены измерялись в долларах, т. е. доллар использовался в качестве единицы измерения. При таком подходе результат выражался в виде суммы наличных. Другими словами, мы рассматривали наличные как актив, имеющий постоянную стоимость 1. Но мы могли бы использовать в качестве единицы измерения цену другого актива, которая всегда положительна. Например, возьмем цену акций и определим новые (относительные) цены для облигаций как

$$\overline{B}_0 = \frac{B_0}{S_0} \quad \text{и} \quad \overline{B}_n = \frac{B_n}{S_n}\,,$$

в предположении, что $S_0 > 0$ и $S_n > 0$ для всех n и $\omega \in \Omega$. Конечно, $\overline{S}_n \equiv 1$.

Используя акции в качестве единицы измерения, перепишем (10.2.3) в виде

$$\frac{V_0}{S_0} = \frac{1}{(1+r)S_0}\left[\tilde{p}\,S_1(H)\ \frac{g(H)}{S_1(H)} + (1-\tilde{p})\,S_1(T)\ \frac{g(T)}{S_1(T)}\right]$$

или

$$\overline{V}_0 = \left[\tilde{p}\ \frac{S_1(H)}{(1+r)S_0}\ \bar{g}(H) + (1-\tilde{p})\ \frac{S_1(T)}{(1+r)S_0}\ \bar{g}(T)\right],$$

где

$$\overline{V}_0 = \frac{V_0}{S_0} \quad \text{и} \quad \bar{g}(\omega) = \frac{g(\omega)}{S_1(\omega)}$$

при $\omega \in \{H, T\}$. Таким образом, в данном случае можно определить еще одну вероятностную меру $\overline{P}$, эквивалентную мере P, с помощью формул

$$\overline{P}(H) = \tilde{p}\ \frac{S_1(H)}{(1+r)S_0} \quad \text{и}$$

$$\overline{P}(T) = (1-\tilde{p})\ \frac{S_1(T)}{(1+r)S_0}\,.$$

Тем самым, имеем

$$\overline{V}_0 = [\overline{P}(H)\bar{g}(H) + \overline{P}(T)\bar{g}(T)].$$

Это означает, что цена опциона в таких единицах измерения представляет собой еще одно математическое ожидание (в качестве упражнения проверьте, что $\overline{P}(H) + \overline{P}(T) = 1$). На самом деле можно доказать, что новый рынок (с ценами $\overline{S}_n$ и $\overline{B}_n$) не имеет арбитража тогда и только тогда, когда нет арбитража на первоначальном рынке (с ценами S_n и B_n). Этот результат обобщается и на многопериодную модель (попробуйте установить его самостоятельно и найдите, например, в [8] его аналог для непрерывного случая).

10.3. Цена опциона при отсутствии арбитража: N-периодная модель

Для общей модели, имеющей $N \geqslant 1$ периодов, рассуждения, опирающиеся на отсутствие арбитража, которые были проведены для однопериодной модели, можно применить к каждому из N периодов. При этом свойства, указанные в списке замечаний в конце предыдущего параграфа, сохраняют свою силу. Из-за того, что прошлый анализ ограничивался только одним периодом, мы не могли проиллюстрировать еще два важных понятия, которые появляются в общем случае и имеют смысл для смежных периодов: (i) *динамическое воспроизведение* и (ii) *самофинансируемая стратегия*. Давайте познакомимся с ними.

Владелец опциона не обязан хранить его у себя вплоть до даты исполнения. Он или она может продать его на рынке (например, на Чикагской бирже опционов). Тогда новый владелец опциона столкнется с той же самой проблемой оценивания, т. е. вычисления справедливой цены опциона в момент его покупки. В результате приходим к заключению, что на самом деле в теории оценивания опционов интерес представляет определение цены в произвольный момент времени. Для многопериодной модели существует стратегия, позволяющая воспроизвести цену опциона в любой из моментов (в однопериодном случае они сводятся к начальному моменту и дате исполнения). Процесс, происходящий при применении данной стратегии, называется *динамическим воспроизведением*. Можно при определенных ограничениях утверждать, что такая стратегия является *самофинансируемой*: в каждый из моментов она не приводит к увеличению капитала, но и не требует дополнительных денежных вложений. Например, предположим, что во время n-го периода (т. е. на интервале $(t_{n-1}, t_n]$) вы имели 10 акций и 20 облигаций. Получив информацию о ценах в момент t_n этих двух активов, вы решили, что на следующий период $(t_n, t_{n+1}]$ у вас должны сохраниться только 8 акций, а число облигаций следует увеличить до 30. Если данная стратегия является самофинансируемой, то на покупку десяти дополнительных облигаций вы потратите всю сумму, вырученную от продажи двух акций (мы, конечно, игнорируем докучливые проблемы, связанные с необходимостью платить за проведение сделок, налоги и т. п.). Другими словами, сумма, получаемая в результате продажи одного из активов, полностью уходит на покупку другого.

Будем предполагать, что доходности акций во всех N периодах независимы и одинаково распределены. Простейшим обобщением рассмотренной выше однопериодной бернуллиевской модели является N-периодная модель, эквивалентная схеме бросания несимметричной монеты, подробно изучаемой в этой книге (см. пример 8 в § 2.4, § 3.2 и пример 9 в § 4.4). Для однопериодной модели мы также могли взять вместо $\Omega = \{H, T\}$ выборочное пространство $\Omega = \{0, 1\}$. Для строгости изложения, как и в случае бросания монеты, будем в N-периодной модели использовать в качестве выборочного пространства $\Omega = \{0, 1\}^N$. При этом точки выборочного пространства (также называемые траекториями) $\omega \in \Omega$ имеют вид $(\omega_1, \omega_2, \ldots, \omega_N)$. Для каждого $n \leqslant N$ цена S_n, реализующаяся в момент t_n, зависит только от наблюдаемых значений компонент вектора $\hat{\omega}_n = (\omega_1, \omega_2, \ldots, \omega_n)$. Поэтому можно записать тождество $S_n(\omega) \equiv S_n(\hat{\omega}_n)$, в котором

$$S_n(\hat{\omega}_n) = \begin{cases} uS_{n-1}(\hat{\omega}_{n-1},\ 1) & \text{с вероятностью } p, \\ dS_{n-1}(\hat{\omega}_{n-1},\ 0) & \text{с вероятностью } 1-p, \end{cases}$$

где u, d и p определены в (10.2.1). При этом вероятностная мера P на Ω определяется формулой

$$P\{(\omega_1, \omega_2, \ldots, \omega_N)\} = p^{\sum\limits_{i=1}^{N} \omega_i} (1-p)^{N - \sum\limits_{i=1}^{N} \omega_i},$$

задающей совместное распределение N испытаний Бернулли.

Пусть в моменты $0, 1, 2, \ldots, N$ (что эквивалентно фактическим временам $t_0, t_1, \ldots, t_N$) соответствующие величины $V_0, V_1, \ldots, V_N$ обозначают цены опциона европейского типа с ценой исполнения K на акции, имеющие стоимости $S_0, S_1, \ldots, S_N$. Как и для однопериодной модели, цена V_N опциона в момент исполнения является случайной величиной, которая задается следующими явными формулами:

$$V_N = \begin{cases} \max(S_N - K,\ 0) & \text{для опциона-колл,} \\ \max(K - S_N,\ 0) & \text{для опциона-пут.} \end{cases}$$

Можно даже выразить V_N как функцию от выборочной траектории $\omega \in \Omega$:

$$V_N = \begin{cases} \max\left(u^{\sum\limits_{i=1}^{N}\omega_i} d^{N-\sum\limits_{i=1}^{N}\omega_i} S_0 - K,\ 0\right) & \text{для опциона-колл,} \\ \max\left(K - u^{\sum\limits_{i=1}^{N}\omega_i} d^{N-\sum\limits_{i=1}^{N}\omega_i} S_0,\ 0\right) & \text{для опциона-пут.} \end{cases}$$

В случае N периодов будем определять цены опциона последовательно, двигаясь вспять по времени: сначала, зная V_N, найдем V_{N-1}, затем — V_{N-2} и так далее, пока не получим V_0. Говоря об определении цен V_n, $0 \leqslant n \leqslant N-1$, мы не имеем в виду, что они обязательно фиксированы; на самом деле цены V_n являются такими же случайными величинами, как V_N (вспомните случай $N = 1$ в предыдущем параграфе). Для любого момента $1 \leqslant n \leqslant N$, зная V_n, находим V_{n-1} точно так же, как в однопериодной модели. К моменту $n-1$ (принадлежащему n-му периоду) цены акций $S_0, S_1, \ldots, S_{n-1}$ и облигаций $B_0, B_1, \ldots, B_{n-1}$ считаются уже известными. Сами $V_0, V_1, \ldots, V_{n-1}$ также уже реализовались, однако целью теории является нахождение таких цен опциона, которые не допускают, аналогично однопериодному случаю, арбитражных возможностей.

Стратегия формирования портфелей для N периодов. Свойство отсутствия арбитража, введенное для однопериодной модели, легко обобщается на N-периодный контекст: оно должно выполняться каждого из периодов в отдельности.

Определение. *N-периодная стратегия представляет собой набор троек (портфелей)*

$$\{(\alpha_n, \beta_n, \gamma_n), \quad n = 0, 1, \ldots, N\},$$

где α_n, β_n и γ_n при $0 \leqslant n \leqslant N$ являются случайными величинами, определенными на $\Omega = \{0, 1\}^N$. (Например, для $\omega = (\omega_1, \omega_2, \ldots, \omega_N) \in \Omega$ значение $\alpha_n(\omega) \in (-\infty, \infty)$.)

Для $0 \leqslant n \leqslant N$ также требуется, чтобы портфель $(\alpha_n, \beta_n, \gamma_n)$ формировался только на основе информации, известной вплоть до n-го периода. К примеру, мы не допускаем «внутреннюю торговлю»: в n-й период вам лично сообщается некоторая информация, которая становится общеизвестной, скажем, в $(n+2)$-й период. Мы не определяем формально, что такое «информация». Для наших целей достаточно понимать под ней необходимую для формирования портфеля историю цен акций. Как отмечалось выше, наблюдаемая в момент t_n цена S_n зависит только от реализовавшихся значений компонент вектора $\hat{\omega}_n = (\omega_1, \omega_2, \ldots, \omega_n)$. Аналогичным образом, портфель $(\alpha_n, \beta_n, \gamma_n)$ в момент t_n зависит только от $\hat{\omega}_n$, т. е. можно записать, что $\alpha_n(\omega) \equiv \alpha_n(\hat{\omega}_n)$.

Лемма. *В n-й период существует такой портфель*

$$\big(\alpha_{n-1}(\hat{\omega}_{n-1}),\ \beta_{n-1}(\hat{\omega}_{n-1}),\ \gamma_{n-1}(\hat{\omega}_{n-1})\big),$$

что $\alpha_{n-1}(\hat{\omega}_{n-1}) = 0$ и

$$\beta_{n-1}(\hat{\omega}_{n-1})S_n(\hat{\omega}_{n-1}, \omega_n) + \gamma_{n-1}(\hat{\omega}_{n-1})B_n(\hat{\omega}_{n-1}, \omega_n) = V_n(\hat{\omega}_{n-1}, \omega_n)$$

для $\omega_n \in \{0, 1\}$.

Доказательство. Введем краткие обозначения:

$$\alpha_{n-1} \equiv \alpha_{n-1}(\hat{\omega}_{n-1}), \quad \beta_{n-1} \equiv \beta_{n-1}(\hat{\omega}_{n-1}), \quad \gamma_{n-1} \equiv \gamma_{n-1}(\hat{\omega}_{n-1}).$$

Как в лемме 10.2, будем искать β_{n-1} и γ_{n-1}, удовлетворяющие уравнениям (напомним, что в N-периодной модели мы заменили H на 0 и T на 1)

$$\begin{aligned} &\beta_{n-1}S_n(\hat{\omega}_{n-1}, 0) + \gamma_{n-1}B_n(\hat{\omega}_{n-1}, 0) = V_n(\hat{\omega}_{n-1}, 0) \\ \text{и}\quad &\beta_{n-1}S_n(\hat{\omega}_{n-1}, 1) + \gamma_{n-1}B_n(\hat{\omega}_{n-1}, 1) = V_n(\hat{\omega}_{n-1}, 1). \end{aligned}$$

Данную систему линейных уравнений можно переписать в виде

$$\begin{aligned} &\beta_{n-1}uS_{n-1}(\hat{\omega}_{n-1}) + \gamma_{n-1}(1+r)B_{n-1}(\hat{\omega}_{n-1}) = V_n(\hat{\omega}_{n-1}, 0) \\ \text{и}\quad &\beta_{n-1}dS_{n-1}(\hat{\omega}_{n-1}) + \gamma_{n-1}(1+r)B_{n-1}(\hat{\omega}_{n-1}) = V_n(\hat{\omega}_{n-1}, 1), \end{aligned}$$

откуда получаем искомое решение

$$\begin{aligned} \beta_{n-1} &= \frac{V_n(\hat{\omega}_{n-1}, 0) - V_n(\hat{\omega}_{n-1}, 1)}{(u-d)S_{n-1}(\hat{\omega}_{n-1})}, \\ \gamma_{n-1} &= \frac{1}{(1+r)B_{n-1}(\hat{\omega}_{n-1})}\left[\frac{uV_n(\hat{\omega}_{n-1}, 1) - dV_n(\hat{\omega}_{n-1}, 0)}{u-d}\right]. \end{aligned} \tag{10.3.1}$$

Утверждение 2. *Цены $V_0, V_1, \ldots, V_N$ опциона европейского типа с датой исполнения t_N удовлетворяют рекуррентным соотношениям*

$$V_{n-1}\left(\hat{\omega}_{n-1}\right) = \frac{1}{1+r}\left[\tilde{p}\, V_n\left(\hat{\omega}_{n-1}, 0\right) + (1-\tilde{p})\, V_n\left(\hat{\omega}_{n-1}, 1\right)\right],$$

выполняющимся для $1 \leqslant n \leqslant N$, где $\tilde{p}$ определяется как в (10.2.4).

Доказательство. Прежде всего, аналогично доказательству утверждения 1, необходимо проверить, что

$$V_{n-1} = \beta_{n-1} S_{n-1} + \gamma_{n-1} B_{n-1},$$

где β_{n-1} и γ_{n-1} задаются формулами (10.3.1). Для этого достаточно повторить проведенные там рассуждения (убедитесь).

Замечание 1. Как только что мы видели, для любого периода $1 \leqslant n \leqslant N$ не допускающая арбитража стратегия удовлетворяет соотношениям

$$\begin{aligned} &\beta_{n-1} S_n + \gamma_{n-1} B_n = V_n \\ \text{и} \quad &\beta_{n-1} S_n + \gamma_{n-1} B_n = \beta_n S_n + \gamma_n B_n. \end{aligned}$$

Первое из уравнений выражает динамическое воспроизведение. Второе уравнение иллюстрирует понятие самофинансируемой стратегии. Его левая часть задает цену портфеля до проведения сделки в момент t_n, а правая — цену портфеля после проведения сделки в момент t_n.

Обобщая однопериодный случай, определим N-периодную мартингальную меру формулой

$$\tilde{P}\{(\omega_1, \omega_2, \ldots, \omega_N)\} = \tilde{p}^{\sum\limits_1^N \omega_i} (1 - \tilde{p})^{N - \sum\limits_1^N \omega_i}.$$

Еще немного об условных ожиданиях. Понятие условного математического ожидания относительно единственного события было введено в § 5.2. На самом деле, можно определить условное ожидание случайной величины относительно *совокупности* событий. Мы неявно уже использовали его. А именно, при изучении мартингалов (см. приложение 3), а также субмартингалов и супермартингалов (см. (10.1) и (10.2)), мы неявно имели дело с (прошлыми) событиями, относящимися к истории процесса $X_0, X_1, \ldots, X_n$ вплоть до момента n.

Читатель знает, что формально событие определяется как множество точек выборочного пространства. Следовательно, совокупностью событий называется набор множеств (состоящих из точек выборочного пространства). Мы уже встречались с некоторыми семействами множеств, в частности, с сигма-алгеброй, введенной в приложении 1, которая обладает определенными свойствами (см. (a) и (b) в приложении 1). Именно эти свойства необходимы для того, чтобы корректно определить условное ожидание относительно совокупности событий.

В дальнейшем нам понадобится следующее свойство: пусть $\mathfrak{F}_1$ и $\mathfrak{F}_2$ — две совокупности событий, удовлетворяющих условиям (a) и (b) из приложения 1. Допустим, что $\mathfrak{F}_1 \subset \mathfrak{F}_2$, а Y — случайная величина, у которой существует математическое ожидание. Тогда

$$E\big\{\{Y \mid \mathcal{F}_2\} \mid \mathcal{F}_1\big\} = E\{Y \mid \mathcal{F}_1\}. \tag{10.3.2}$$

Если вы внимательно посмотрите на эту формулу, то увидите, что в левой части присутствуют два условных ожидания: первым является $Z = E\{Y \mid \mathfrak{F}_2\}$, а вторым — $X = E\{Z \mid \mathfrak{F}_1\}$. Мы не станем доказывать здесь эту формулу (называемую «башенным» свойством условных ожиданий), но заинтересованный читатель может обратиться к книгам [4,5,24]. Несмотря на ее довольно пугающий вид, она выражает тот факт, что при усреднении относительно двух информационных совокупностей, определяющим является усреднение относительно наименьшей из совокупностей*). В более сложных книгах совокупность множеств $\{\mathfrak{F}_n\}$, таких что $\mathfrak{F}_n \subset \mathfrak{F}_{n+1}$, называют *фильтрацией.*

Утверждение 3. *Цена в момент 0 опциона европейского типа с выплатой g в момент N на акции, цены которых образуют определенный выше процесс S, удовлетворяет соотношению*

$$V_0 = \frac{1}{(1+r)^N}\ \tilde{E}\{g \mid (S_0, B_0)\},$$

где $\tilde{E}$ обозначает математическое ожидание по мартингальной мере $\tilde{P}$. Для опциона с ценой исполнения K имеем следующие конкретные представления:

$$V_0 = \frac{1}{(1+r)^N} \sum_{n=0}^{N} \binom{N}{n} \tilde{p}^n (1-\tilde{p})^{N-n} \max\big(u^n d^{N-n} S_0 - K, 0\big)$$

для опциона-колл,

$$V_0 = \frac{1}{(1+r)^N} \sum_{n=0}^{N} \binom{N}{n} \tilde{p}^n (1-\tilde{p})^{N-n} \max\big(K - u^n d^{N-n} S_0, 0\big)$$

для опциона-пут.

Доказательство. Аналогично однопериодному случаю, можно записать для $1 \leqslant n \leqslant N$, что

$$V_{i-1} = E\left\{\frac{1}{1+r}\, V_i \,\middle|\, (S_{i-1}, B_{i-1})\right\}.$$

*) Говорят, что «маленький съедает большого». — *Прим. перев.*

Тогда

$$\begin{aligned} V_{i-2} &= E\left\{\frac{1}{1+r}\,V_{i-1}\,\middle|\,(S_{i-2},B_{i-2})\right\} = \\ &= E\left\{\frac{1}{(1+r)^2}\,E\left\{\frac{1}{1+r}\,V_i\,\middle|\,(S_{i-1},B_{i-1})\right\}\,\middle|\,(S_{i-2},B_{i-2})\right\} = \\ &= E\left\{\frac{1}{(1+r)^2}\,V_i\,\middle|\,(S_{i-2},B_{i-2})\right\}, \end{aligned}$$

где при получении последнего равенства применялось «башенное» свойство. Остальная часть доказательства проводится просто и оставлена в качестве упражнения. [*Указание.* Используйте биномиальное распределение, как в (4.4.15).]

10.4. Фундаментальные теоремы оценивания опционов

Для рассмотренной выше биномиальной модели удается получить формулы, выражающее единственное решение для цен как опциона-пут, так и опциона-колл. Эти формулы выводятся на основе предположения об отсутствии арбитража, которое, в свою очередь, приводит к построению вероятностной меры, относительно которой все ценовые процессы становятся мартингалами. На самом деле приведенные результаты служат иллюстрациями следующих фундаментальных теоремы оценивания опционов:

1. Конечная модель рынка свободна от арбитража тогда и только тогда, когда существует эквивалентная вероятностная мера, по отношению к которой все процессы, образуемые ценами акций, являются мартингалами.
2. Свободная от арбитража модель рынка является полной (т. е. такой, где каждый опцион европейского типа может быть оценен) тогда и только тогда, когда эквивалентная вероятностная мера единственна.

Заметим, что в данных теоремах под моделями рынка понимаются определенные вероятностные модели (т. е. выборочные пространства, вероятности и т. д.), представляющие рынок. Их устройство играет решающую роль. К примеру, при замене Ω в однопериодном случае на пространство, состоящее из трех точек, мартингальная мера уже не будет существовать (предлагаем проверить это в качестве упражнения). Мы отсылаем к работе [14] для получения более полного представления о применении этих теорем в контексте конечных вероятностных пространств с конечным временным горизонтом. Несмотря на пугающее название, данная статья вполне доступна для читателей, математический уровень которых соответствует принятому в этой книге. В ней вводятся несколько новых понятий, и задача авторов заключается в том, чтобы продемонстрировать, как некоторые глубокие результаты могут быть получены элементарными методами. Эти результаты сохраняются для так называемых конечных моделей рынка, т. е. таких, что пространство Ω конечно, $P(\omega) > 0$ для всех $\omega \in \Omega$; на рынке имеется только конечное число видов активов (у нас было три), горизонт конечен и количество моментов, в которые проводятся сделки, также конечно.

В свою очередь, непрерывный случай (когда время непрерывно и/или пространство состояний несчетно) намного сложнее. Любопытно, что непрерывная модель Блэка—Шоулза и рассматриваемая здесь биномиальная модель тесно связаны. Если длины интервалов в N-периодной модели убывают к нулю при стремлении N к бесконечности, то цена V_0, полученная в утверждении 3, сходится к цене из модели Блэка—Шоулза (несложное доказательство, не выходящее за уровень этой книги, приводится, например, в гл. 5 книги [6]).

Задачи

1. Убедитесь, что выражение $\max(K - S_T, 0)$ действительно задает выплату опциона-пут европейского типа, с помощью тех же рассуждений, которые мы использовали для опциона-колл.

2. Докажите, что процесс $\{Y_n\}_{n=0}^{N}$, образованный дисконтированными ценами акций, также является мартингалом относительно мартингальной меры $\tilde{P}$. (Это непосредственно вытекает из рассмотренного ранее однопериодного случая и из доказательства утверждения 3.)

3. Анализируя формулу Блэка и Шоулза, покажите, что цена опциона-пут P_0 является убывающей функцией от первоначальной цены акций S_0, а цена опциона-колл C_0 — возрастающей. Другими словами, опцион-пут оценивается ниже, чем опцион-колл при высоких ценах на акции (подумайте, в чем смысл этого, исходя из вида функций, определяющих выплаты).

4. Покажите, что и P_0, и C_0 возрастают как функции от σ (изменчивости). Это означает, что цены как опциона-пут, так и опциона-колл увеличиваются при увеличении неопределенности перспектив цены на акции.

5. Убедитесь, что C_0 возрастает с увеличением срока исполнения T, т. е. опцион-колл европейского типа с более поздней датой исполнения стоит дороже. А что в данной ситуации можно сказать об опционе-пут?

6. Соответствие опционов-пут и опционов-колл (часть 1).

Проверьте, что $C_0 = P_0 + S_0 - Ke^{-rT}$. Это означает, что при одинаковых ценах исполнения и сроках действия цена опциона-колл может быть получена из цены опциона-пут. Данное свойство важно ввиду того, что в некоторых случаях легче, например, доказать некоторые математические результаты именно для опционов-пут. Последнее происходит благодаря ограниченности выплаты для опционов-пут (убедитесь).

7. Соответствие опционов-пут и опционов-колл (часть 2).

Чтобы вывести приведенную выше формулу соответствия, рассмотрим портфель, состоящий (в начальный момент) из опциона-колл с положительным весом (имеющего длинную позицию), опциона-пут с отрицательным весом (имеющего короткую позицию), акций с короткой позицией и занимаемой суммой Ke^{-rT}. Здесь оба опциона имеют цену исполнения K и срок действия T.

Покажите, что цена данного портфеля в момент T равна 0. Для выполнения условия отсутствия арбитража необходимо, чтобы цена портфеля в начальный момент также равнялась 0. Отсюда следует формула соответствия.

Ответы к задачам

Глава 1

7. $(A \cup B)(B \cup C) = ABC + ABC^c + A^cBC + A^cBC^c + AB^cC$;
$A \backslash B = AB^cC + AB^cC^c$; {множество точек ω, которые принадлежат в точности одному из множеств A, B, C} $= AB^cC^c + A^cBC^c + A^cB^cC$.

10. Двойственная формула верна.

14. Определите $A \# B = A^c \cup B^c$ или $A^c \cap B^c$.

19. $I_{A \backslash B} = I_A - I_A I_B; I_{A-B} = I_A - I_B$.

20. $I_{A \cup B \cup C} = I_A + I_B + I_C - I_{AB} - I_{AC} - I_{BC} + I_{ABC}$.

Глава 2

4. $P(A + B) \leqslant P(A) + P(B)$.

5. $P(S_1 + S_2 + S_3 + S_4) \geqslant P(S_1) + P(S_2) + P(S_3) + P(S_4)$.

11. Возьмите $AB = \varnothing, P(A) > 0, P(B) > 0$.

13. 17.

14. 126.

15. $|A \cup B \cup C| = |A| + |B| + |C| - |AB| - |AC| - |BC| + |ABC|$.

16. $P(A \,\Delta\, B) = P(A) + P(B) - 2P(AB) = 2P(A \cup B) - P(A) - P(B)$.

17. Равенство выполняется, когда числа m и n взаимно простые.

20. $p_n = 1/2^n$, $n \geqslant 1$; $p_n = 1/(n(n+1))$, $n \geqslant 1$.

22. 14/60.

24. Если A не зависит само от себя, то $P(A) = 0$ или $P(A) = 1$; если A и B несовместны и независимы, то $P(A)P(B) = 0$.

28. $p_1p_2q_3p_4q_5$, где p_k — вероятность того, что k-я монета выпадет «орлом», $q_k = 1 - p_k$. Вероятность того, что среди 5 монет окажутся в точности 3 «орла» равна $\sum p_{k_1}p_{k_2}p_{k_3}q_{k_4}q_{k_5}$, где сумма распространяется на все 10 неупорядоченных троек (k_1, k_2, k_3) из $(1, 2, 3, 4, 5)$, и (k_4, k_5) обозначает оставшуюся неупорядоченную пару.

Глава 3

1. $3 + 2$; $3 + 2 + (3 \times 2)$.

2. 3^2, $\binom{3+2-1}{2}$.

3. Три рубашки помещаются в двух различных упаковках, каждая из которых может содержать от 0 до 3 рубашек. Если рубашки различимы, то 2^3; если нет, то $\binom{2+3-1}{3}$.
4. $3 \times 4 \times 3 \times 5 \times 3$; $3 \times 4 \times (3+1) \times (2+1) \times 3$.
5. $26^2 + 26^3$; 100.
6. 9^7.
7. $\binom{12}{6}$.
8. $4!$; $2 \times 4!\,4!$.
9. $\binom{20}{3}$; $(20)_3$.
10. 35 (цена 0 не включена); 23.
11. 1/2, если отсутствующие шурупы могут равновероятно быть как одинаковых, так и разных размеров; 2/3, если каждый отсутствующий шуруп равновероятно может иметь любой из размеров.
12. 2/3; $4!/6!$ или $(2 \times 4!)/6!$, в зависимости от того, испытываются ли два ключа в одном или обоих порядках (как учесть потерянный ключ?).
13. 20/216 (ответ получается прямым вычислением); некоторые интерпретируют «увеличиваются постоянно» как «образуют арифметическую прогрессию», если вам известно, что это означает.
14. (a) $1/6^3$; (b) $\{6 \times 1 + 90 \times 3 + 120 \times 6\}/6^6$.
15. $\binom{6}{4}4!$; $\binom{6}{3}\binom{4}{3}3!$.
16. $1 - \left\{\binom{5}{0}\binom{5}{4} + \binom{5}{1}\binom{4}{3} + \binom{5}{2}\binom{3}{2} + \binom{5}{3}\binom{2}{1} + \binom{5}{4}\binom{1}{0}\right\} \Big/ \binom{10}{4}$.
17. Из внешнего ряда: 3/8; из внутреннего ряда: 11/16.
18. $\left(\dfrac{m-1}{m}\right)^n$; $\dfrac{(m-1)_n}{(m)_n}$.
19. 1/6, 4/6, 1/6.
20. (a) $4 \Big/ \binom{18}{15}$; (b) $\binom{14}{11} \Big/ \binom{18}{15}$.
21. В предположении, что ни одна из стопок не пуста: (a) стопки и книги различимы: $2^{10}-2$; (b) книги различимы, а стопки — нет: $(2^{10}-2)/2$; (c) стопки различимы, а книги — нет: 9; (d) ни те, ни другие не различаются: 5.
22. $\dfrac{10!}{3!\,3!\,2!\,2!}$; $\dfrac{10!}{3!\,3!\,2!\,2!} \times \dfrac{4!}{2!\,2!}$; $\dfrac{10!}{3!\,3!\,2!\,2!} \times \dfrac{6!}{2!\,2!\,2!\,2}$.
23. (a) $\binom{31}{15}^7\binom{30}{15}^4\binom{29}{15}(180)!/(366)_{180}$,
(b) $(305)_{30}/(366)_{30}$.
24. $\binom{29}{10} \Big/ \binom{49}{30}$.

25. $\binom{n-100}{93}\binom{100}{7}\Big/\binom{n}{100}$.

27. Разделите следующие числа на $\binom{52}{5}$:

(a) $4\times\binom{13}{5}$; (b) 9×4^5; (c) 4×9;
(d) 13×48; (e) $13\times 12\times 4\times 6$.

29. Разделите следующие числа на 6^6:

$$6;\quad 6\times 5\times\frac{6!}{5!\,1!};\quad 6\times 5\times\frac{6!}{4!\,2!};\quad 6\times\binom{5}{2}\times\frac{6!}{4!};$$

$$\binom{6}{2}\times\frac{6!}{3!\,3!};\quad (6)_3\times\frac{6!}{3!\,2!};\quad 6\times\binom{5}{3}\times\frac{6!}{3!};$$

$$\binom{6}{3}\times\frac{6!}{2!\,2!\,2!};\quad \binom{6}{2}\binom{4}{2}\times\frac{6!}{2!\,2!};\quad \binom{6}{1}\binom{5}{4}\times\frac{6!}{2!};\quad 6!.$$

Проверьте ответы сложением; для этого используйте калькулятор, если он у вас есть.

30. Сначала для прояснения ситуации решите задачу при $n=2$ перебором. В общем случае, допустим, что коробок в правом кармане оказался пустым, а в коробке, находящемся в левом кармане, осталось k спичек. Для $0\leqslant k\leqslant n$ вероятность такого события равна $\frac{1}{2^{2n-k}}\binom{2n-k}{n}\frac{1}{2}$. Ее надо увеличить вдвое, так как правый карман можно заменить на левый. Красивым следствием данного результата служит следующая формула:

$$\sum_{k=0}^{n}\frac{1}{2^{2n-k}}\binom{2n-k}{n}=1.$$

Глава 4

2. $P\{X+Y=k\}=1/3$ для $k=3,4,5$; то же самое для $Y+Z$ и $Z+X$.

3. $P\{X+Y-Z=k\}=1/3$ для $k=0,2,4$;
$P\{\sqrt{(X^2+Y^2)Z}=x\}=1/3$ для $x=\sqrt{13},\sqrt{15},\sqrt{20}$;
$P\{Z/|X-Y|=3\}=1/3$, $P\{Z/|X-Y|=1\}=2/3$.

4. Пусть $P(\omega_j)=1/10$ для $j=1,2$; $=1/5$ для $j=3,4$; $=2/5$ для $j=5$; $X(\omega_j)=j$ для $1\leqslant j\leqslant 5$; $Y(\omega_j)=\sqrt{3}$ для $j=1,4$; $=\pi$ для $j=2,5$; $=\sqrt{2}$ для $j=3$.

5. Пусть $P(\omega_j)=p_j$, $X(\omega_j)=v_j$, $1\leqslant j\leqslant n$.

6. $\{X+Y=7\}=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$.

8. $P\{Y=14\,000+4n\}=1/5\,000$ для $1\leqslant n\leqslant 5\,000$; $E(Y)=24002$.

9. $P\{Y=11\,000+3n\}=1/10\,000$ для $1\leqslant n\leqslant 1\,000$;
$P\{Y=10\,000+4n\}=1/10\,000$ для $1001\leqslant n\leqslant 10\,000$;
$E(Y)=41052.05$.

10. $E(Y) = 29\,000 + 7\,000 \cdot e^{-2/7}$.
11. $\lambda e^{-\lambda x}$, $x > 0$.
12. $2xf(x^2), x > 0$; $2x/(b-a)$ для $\sqrt{a} \leqslant x \leqslant \sqrt{b}$.
13. (i) $f\left((x-b)/a\right)/|a|$, если $a \neq 0$. (ii) $\dfrac{1}{2\sqrt{x}}\{f(\sqrt{x}) + f(-\sqrt{x})\}, x > 0$.
15. $c = 1/(1-q^m)$.
16. $P(Y = j) = \dbinom{n}{\frac{n+j}{2}} \dfrac{1}{2^n}$ для таких $-n \leqslant j \leqslant n$, что $n + j$ четно.
    $E(Y) = 0$.
17. $P(X = j) = \dbinom{11}{j}\dbinom{539}{25-j}\Big/\dbinom{550}{25}$, $0 \leqslant j \leqslant 25$.
18. Если всего в мешке находятся n яблок, из которых r — гнилые, то при случайном выборе k яблок ожидаемое число гнилых среди них равно kr/n.
19. $P(X \geqslant m) = \dfrac{1}{m}, E(X) = +\infty$.
20. 1.
21. Возьмите $v_n = (-1)^n 2^n/n$, $p_n = 1/2^n$.
23. В соответствии с предположениями примера 11 из § 4.5: (1) $\sqrt{3}/2$; (2) $3/4$; (3) $2/3$.
24. 2.
26. $F_R(r) = r^2/100, f_R(r) = r/50$ для $0 \leqslant r \leqslant 100$; $E(R) = 20/3$.
27. $Y = d \operatorname{tg}\theta$, где d — расстояние от дула до стены, а θ — угол между направлением выстрела и горизонталью.
    $P(Y \leqslant y) = \operatorname{arctg}(y/d)$; $E(Y) = +\infty$.
28. $E(2^X) = +\infty$.
29. Если допускается всего m бросаний, то его математическое ожидание выигрыша составляет m центов.
31. $P((X,Y) = (m,m')) = \dbinom{n}{2}^{-1}$ для $1 \leqslant m < m' \leqslant n$;
    $P(X = m) = (n-m)\dbinom{n}{2}^{-1}$; $P(Y = m') = (m'-1)\dbinom{n}{2}^{-1}$;
    $P(Y - X = k) = (n-k)\dbinom{n}{2}^{-1}$, $1 \leqslant k \leqslant n-1$.
32. Плотность (X,Y) есть $f(u,v) = \begin{cases} 2, & \text{если } 0 \leqslant u < v \leqslant 1, \\ 0, & \text{в противном случае.} \end{cases}$

Глава 5

1. $1050/6145, 95/1095$.
2. $\dfrac{18826}{19400}$.

3. 5/9.
4. (a) 1/2; (b) 1/10.
5. 1/4; 1/4.
6. 1/4.
7. 1/2.
8. $2\beta(1-\alpha+\beta)^{-1}$.
9. 6/11, 3/11, 2/11.
10. 400/568.
17. 1/2.
18. $p^3 + (3/2)\times p^3(1-p)$.
19. 379/400.
20. P(нет зонта | идет дождь) = 2/9; P(нет дождя | есть зонт) = 5/12.
21. 27/43.
22. $[p^2+(1-p)^2]\,/\,[3p^2+(1-p)^2]$.
23. (a) 3/8; (b) 3/4; (c) 1/3.
25. (a) $\frac{1}{6}\sum\limits_{n=1}^{6}\binom{n}{k}\frac{1}{2^n}$; (b) $\binom{n}{3}\frac{1}{2^n}\left\{\sum\limits_{n=3}^{6}\binom{n}{3}\frac{1}{2^n}\right\}^{-1}$ для $3\leqslant n\leqslant 6$.
26. Вероятности, что число частиц равно 1, 2, 3, 4, 5, 6, равны, соответственно: (1) p_1^2; (2) $p_1p_2+p_1^2p_2$; (3) $p_1p_3+2p_1p_2^2+p_1^3p_3$; (4) $2p_1p_2p_3+p_2^3+3p_1^2p_2p_3$; (5) $2p_2^2p_3+3p_1p_2^2p_3+3p_1^2p_3^2$; (6) $p_2p_3^2+p_2^3p_3+6p_1p_2p_3^2$; (7) $3p_1p_3^3+3p_2^2p_3^2$; (8) $3p_2p_3^3$; (9) p_3^4. Слишком хлопотно? См. пример 20 и задачу 23 в гл.. 8, где изложен общий метод.
27. $\left(\frac{4}{6}\right)^n - 2\left(\frac{3}{6}\right)^n + \left(\frac{2}{6}\right)^n$.
28. $\sum\limits_{n=0}^{\infty} p_n\left(\sum\limits_{k=0}^{n} p_k\right)$; $\sum\limits_{n=0}^{\infty} p_n^2$.
29. 2/7.
30. $P(\text{maximum} < y \mid \text{minimum} < x) = y^2/(2x-x^2)$, если $y\leqslant x$; $= (2xy-x^2)/(2x-x^2)$, если $y>x$.
31. 1/4.
33. (a) $(r+c)/(b+r+2c)$; (b) $(r+2c)/(b+r+2c)$; (c), (e), (f): $(r+c)/(b+r+c)$; (d) вероятность такая же, как в (b).
34. $\{b_1(b_2+1)r_1+b_1r_2(r_1+1)+r_1b_2(r_1-1)+r_1(r_2+1)r_1\}/(b_1+r_1)^2(b_2+r_2+1)$.
35. $\left(\sum\limits_{k=1}^{N} k^{n+1}\right)\Big/ N\left(\sum\limits_{k=1}^{N} k^n\right)$.
39. $(1+p)^2/4$; $(1+pq)/2$.

40.

	0	1	2
0	q	p	0
1	$q/2$	$1/2$	$p/2$
2	0	q	p

Глава 6

1. \$0.1175; \$0.5875.
2. \$94 000; \$306 000.
3. $2\left(\frac{3}{13}+\frac{2}{12}+\frac{4}{13}+\frac{3}{14}+\frac{4}{14}\right)$.
4. 21; 35/2.
5. 0.5; 2.5.
6. $13/4$; $4\left\{1-\binom{39}{13}\Big/\binom{52}{13}\right\}$.
7. $(6/7)^{25}$; $7\left\{1-\left(\frac{6}{7}\right)^{25}\right\}$.
8. (a) $1-(364/365)^{500}-500(364)^{499}/(365)^{500}$;
   (b) $500/365$;
   (c) $365\left\{1-\left(\frac{364}{365}\right)^{500}\right\}$;
   (d) $365p$, где p — вероятность из (a).
9. Ожидаемое число ящиков, в которые попадут в точности k жетонов, равно $m\binom{n}{k}\frac{(m-1)^{n-k}}{m^n}$; среднее число жетонов, которые не будут делить свой ящик ни с одним другим жетоном, равно $n\left(\frac{m-1}{m}\right)^{n-1}$.
10. $P(n_j$ жетонов в j-м ящике, $1\leqslant j\leqslant m) = \frac{n!}{n_1!\cdots n_m!}\,m^{-n}$, где $n_1+\ldots+n_m=n$.
11. 49.
12. 7/2.
13. $100p$; $10\sqrt{p(1-p)}$.
14. 46/5.
15. (a) $N+1$; (b) $\sum\limits_{n=1}^{N+1}\frac{(N)_{n-1}}{N^{n-1}}$.
16. Пусть M обозначает максимум. Тогда в случае возвращения:
    $P(M=k)=\frac{k^n-(k-1)^n}{N^n}$, $1\leqslant k\leqslant N$;
    $$E(M)=\sum_{k=1}^{N}\left\{1-\left(\frac{k-1}{N}\right)^n\right\};$$
    в случае без возвращения:

$P(M=k) = \binom{k-1}{n-1} \Big/ \binom{N}{n}$, $n \leqslant k \leqslant N$;
$E(M) = n(N+1)/(n+1)$.

17. (a) $nr/(b+r)$; (b) $(r^2+br+cnr)/(b+r)$.

19. $1/p$.

22. $E(X) = 1/\lambda$.

23. $E(T) = \dfrac{a}{\lambda} + \dfrac{1-a}{\mu}$; $\sigma^2(T) = \dfrac{2a}{\lambda^2} + \dfrac{2(1-a)}{\mu^2} - \left(\dfrac{a}{\lambda} + \dfrac{1-a}{\mu}\right)^2$.

24. $E(T \mid T > n) = 1/\lambda$.

25. (a) $1/5\lambda$; (b) $137/60\lambda$.

26. $0.4\,\%$.

27. $E(aX+b) = aE(X)+b, \sigma^2(aX+b) = a^2\sigma^2(X)$.

28. Вероятность того, что он закончит игру победителем, выиграв \$1, равна 127/128. С вероятностью 1/128 ему придется прекратить игру после очередного проигрыша из-за недостатка денег для удвоения ставки. При этом его проигрыш составит составит \$127. Математическое ожидание равно 0. Так стоит ли играть? Во втором случае вероятность проигрыша \$150 равна 1/256, такова же вероятность проигрыша \$104, а вероятность выигрыша \$1 равна 127/128. Математическое ожидание, по-прежнему, нулевое.

29. $E(\text{maximum}) = n/(n+1)$; $E(\text{minimum}) = 1/(n+1)$;
$E(\text{range}) = (n-1)/(n+1)$, где range — размах.

30. $g(z) = \prod_{j=1}^n (q_j + p_j z)$; $g'(1) = \sum\limits_{j=1}^{n} p_j$.

31. $u_k = P\{S_n \leqslant k\}$; $g(z) = (q+pz)^n/(1-z)$.

32. $g(z) = (1-z^{2N+1})/(2N+1)z^N(1-z)$; $g'(1) = 0$.

33. $\dbinom{2n}{n}\dfrac{1}{4^n}$.

34. $g(z) = z^N \prod\limits_{j=0}^{N-1} \dfrac{N-j}{N-jz}$; $g'(1) = N \sum\limits_{j=0}^{N-1} \dfrac{1}{N-j}$.

35. $m_1 = g'(1)$; $m_2 = g''(1) + g'(1)$; $m_3 = g'''(1) + 3g''(1) + g'(1)$; $m_4 = g^{(\mathrm{iv})}(1) + 6g'''(1) + 7g''(1) + g'(1)$.

36. $(-1)^n L^{(n)}(0)$.

37. (a) $(1-e^{-c\lambda})/c\lambda, \lambda > 0$; (b) $2(1-e^{-c\lambda} - c\lambda e^{-c\lambda})/c^2\lambda^2$, $\lambda > 0$;
(c) $L(\mu) = \lambda^n/(\lambda+\mu)^n$.

38. Преобразование Лапласа для S_n есть $\lambda^n/(\lambda+\mu)^n$;

$$P(a < S_n < b) = \frac{\lambda^n}{(n-1)!} \int\limits_a^b u^{n-1} e^{-\lambda u}\, du.$$

Глава 7

1. $1 - \dfrac{5}{3}\, e^{-2/3}$.
2. $\left(1 - \dfrac{4}{100}\right)^{25} \approx e^{-1}$.
3. $e^{-\alpha}\alpha^k/k!$, где $\alpha = 1000/324$.
4. $e^{-20} \sum\limits_{k=20}^{30} (20)^k/k!$.
5. Пусть $\alpha_1 = 4/3, a_2 = 2$. Тогда
   $P\{X_1 = j \mid X_1 + X_2 = 2\} = \dfrac{2!}{j!\,(2-j)!}\,\dfrac{\alpha_1^j \alpha_2^{2-j}}{(\alpha_1+\alpha_2)^2}$ для $j = 0, 1, 2$.
6. Если $(n+1)p$ не является целым, наибольшая вероятность среди $B_k(n;p)$ достигается при $k = [(n+1)p]$, где $[x]$ обозначает ближайшее целое число, не превосходящее x; если $(n+1)p$ — целое, то существуют две максимальные вероятности при $k = (n+1)p - 1$ и $(n+1)p$.
7. Если α не является целым, то максимум вероятностей $\pi_k(\alpha)$ достигается при $k = [\alpha]$; если α — целое, то максимум достигается при $k = \alpha - 1$ и $k = \alpha$.
8. $e^{-\lambda c + \alpha(e^{-\lambda h} - 1)}$.
9. $\pi_k(\alpha + \beta)$.
11. $e^{-50} \sum\limits_{k=50}^{60} (50)^k/k!$.
12. $\dfrac{1}{(n-1)!\,2^n} \int\limits_N^\infty u^{n-1} e^{-u/2}\,du$.
13. $\Phi\left(3\sqrt{\dfrac{12}{35}}\right) - \Phi\left(-2\sqrt{\dfrac{12}{35}}\right)$.
14. Найдите такое n, что $2\Phi\left(\sqrt{n}/10\right) - 1 \geqslant 0.95$. Можно считать, что $p > 1/2$ (если используется обычная кнопка).
15. 537.
16. 475.
24. $(1/(2\pi x))^{-1/2}\ e^{-x/2}$.
27. $P\{\delta'(t) > u\} = e^{-\alpha u}$; $P\{\delta(t) > u\} = e^{-\alpha u}$ при $u < t$; $= 0$ при $u \geqslant t$.
28. Нет!

Глава 8

1. (a) $p^4 + 3p^3q + 2p^2q^2$; (b) $p^4 + 2p^3q + 2pq^3 + q^4$; (c) $1 - (p^4 + p^3q)$; (d) $1 - (p^4 + p^3q + pq^3 + q^4)$.
2. Для Y_n: I — множество четных целых чисел;
   $p_{2i,2i+2} = p^2, p_{2i,2i} = 2pq, p_{2i,2i-2} = q^2$;
   для Z_n: I — множество нечетных целых чисел;

$p_{2i-1,2i+1} = p^2, p_{2i-1,2i-1} = 2pq$, $p_{2i-1,2i-3} = q^2$;
$P\{Z_0 = 1\} = p$, $P\{Z_0 = -1\} = q$.

3. Для X_n: I — множество неотрицательных целых чисел; $p_{i,i} = q, p_{i,i+1} = p$.
Для $Y_n: I$ — множество всех целых чисел, $p_{i,i-1} = q$, $p_{i,i+1} = p$.

4. $P\{|Y_{2n+1}| = 2i+1 \parallel Y_{2n}| = 2i\} = (p^{2i+1} + q^{2i+1})/(p^{2i} + q^{2i})$,
$P\{|Y_{2n+1}| = 2i-1 \parallel Y_{2n}| = 2i\} = (p^{2i}q + pq^{2i})/(p^{2i} + q^{2i})$.

5. (a)

n	1	2	3
$f_{11}^{(n)}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{9}$
$f_{12}^{(n)}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$
$g_{12}^{(n)}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{9}$

(b)

n	1	2	3
$f_{11}^{(n)}$	p_1	0	$q_1q_2q_3$
$f_{12}^{(n)}$	q_1	p_1q_1	$p_1^2q_1$
$g_{12}^{(n)}$	q_1	q_1p_2	$q_1p_2^2$

6. $\begin{bmatrix} 1 & 0 \\ \alpha & 1-\alpha \end{bmatrix}$; $f_{21}^{(n)} = (1-\alpha)^{n-1}\alpha$; $p_{21}^{(n)} = 1 - (1-\alpha)^n; m_{21} = 1/\alpha$.

9. $I = \{0; 2^i,\ 0 \leqslant i \leqslant n-1\}$,
$p_{2^i,2^{i+1}} = 1/2$, $p_{2^i,0} = 12$ для $0 \leqslant i \leqslant n-1$.

10. 1/13.

11.

	U	H	D
U	0	1	0
H	p	0	q
D	0	1	0

$w_U = p/2, w_H = 1/2, w_D = q/2$.

12. Та же самая вероятность, как в (8.1.9) и (8.1.10).

13. $e_j = \dfrac{l(1-r^j)}{(p-q)(1-r^l)} - \dfrac{j}{p-q}$, где $r = q/p$.

15. $p_{i,i-1} = (i/N)^2$, $p_{i,i} = 2i(N-i)/N^2$, $p_{i,i+1} = ((N-i)/N)^2$;
$w_i = \dbinom{N}{i}^2 \Big/ \dbinom{2N}{N}$; $0 \leqslant i \leqslant N$.

16. $w_s = (1-\beta)/(2-\alpha-\beta)$; $w_f = (1-\alpha)/(2-\alpha-\beta)$.

18. $p_{j,j+1} = p$, $p_{j,0} = 1-p$; $w_j = p^jq$, $0 \leqslant j \leqslant \infty$.

19. Пусть $r = pq$, $A^{-1} = 1 + p^{-1}\sum\limits_{k=1}^{c-1} r^k + r^{c-1}$. Тогда $w_0 = A$; $w_k = p^{-1}r^kA$, $1 \leqslant k \leqslant c-1$; $w_c = r^{c-1}A$.

20. $f_{21}^* = .721$; $f_{31}^* = 0.628$.

21. $f_{i,2N}^* = i/2N$, $f_{i,0}^* = 1 - (i/2N)$, $0 \leqslant i \leqslant 2N$.

23. Коэффициент при z^j в $g(h(z))$, где $h(z) = \sum\limits_{j=0}^{\infty} b_jz^j$.

24. Пусть e_j обозначает ожидаемое количество дополнительных импульсов, которые зарегистрирует прибор до момента появления l, если текущим максимумом является j. Тогда $e_j = l$ для $1 \leqslant j \leqslant l-1$, $e_l = 0$.

25. $e_m = \sum\limits_{j=1}^{m} \dfrac{1}{j}$.

26.

	(123)	(132)	(213)	(231)	(312)	(321)
(123)	0	0	q	p	0	0
(132)	0	0	0	0	q	p
(213)	q	p	0	0	0	0
(231)	0	0	0	0	p	q
(312)	p	q	0	0	0	0
(321)	0	0	p	q	0	0

27.

	1	2	3
1	0	q	p
2	1	0	0
3	0	p	q

32. $p_{ij} = 1/(i+1)$ для $s-i \leqslant j \leqslant s$ и $=0$ в противном случае;

$w_j = 2(j+1)/(s+1)(s+2)$ для $0 \leqslant j \leqslant s$; $\displaystyle\sum_{j=0}^{s} jw_j = 2s/3$.

33. $p_{ij} = \dbinom{i}{s-j} p^{s-j}(1-p)^{i-s+j}$ для $s-i \leqslant j \leqslant s$ и $=0$ в противном случае;

$w_j = \dbinom{s}{j}\left(\dfrac{1}{1+p}\right)^{j}\left(\dfrac{p}{1+p}\right)^{s-j}$ для $0 \leqslant j \leqslant s$; $\sum\limits_{j=0}^{s} jw_j = s/(1+p)$.

44. $P\{S = k \mid X_0 = i\} = p_{ii}^k(1-p_{ii}^k)$, $k \geqslant 1$.

45. $\tilde{p}_{ij} = p_{ij}/(1-p_{ii})$ для $i \neq j$; $\tilde{p}_{ii} = 0$.

46. $P\{(X_n, X_{n+1}) = (k, 2k-j+1) \mid (X_{n-1}, X_n) = (j,k)\} = p$,
$P\{(X_n, X_{n+1}) = (k, 2k-j) \mid (X_{n-1}, X_n) = (j,k)\} = q$.

Пусть $H_n^{(3)} = \sum\limits_{v=1}^{n} H_v^{(2)}$. Тогда $\{H_n^{(3)}\}$ является цепью Маркова порядка 3 и т. д.

Глава 9

1. $R(\omega_1) = u-1$, $R(\omega_2) = d-1$. При $u = 1.10$, $R(\omega_1) = 0.10$, т. е. для доходности в 10 % требуется, чтобы реализовалась ω_1. При $d = 0.75$ $R(\omega_2) = -0.25$, т. е. потери составят 25 %, если реализуется ω_2.

2. Мы хотим сравнить $P\{R \leqslant 0\}$ с $P\{R > 0\}$. $P\{R \leqslant 0\} = P\{(R-\mu)/\sigma \leqslant -\mu/\sigma\}$. Вспомним, что $(R-\mu)/\sigma$ имеет стандартное нормальное распределение. Поэтому когда $\mu = 0$, $P\{R \leqslant 0\} = 1/2 = P\{R > 0\}$;
если $\mu < 0$, то $-\mu/\sigma > 0$ и, следовательно, $P\{R \leqslant 0\} > 1/2 > P\{R > 0\}$;
если $\mu > 0$, то $-\mu/\sigma < 0$, отсюда $P\{R \leqslant 0\} < 1/2 < P\{R > 0\}$.

3. Пусть R_1 и R_2 — доходности первого и второго актива, соответственно. Наше предположение заключается в том, что $\sigma(R_2) < \sigma(R_1)$. Для упрощения обозначений, положим $\sigma_i \equiv \sigma(R_i)$, $i = 1, 2$. Тогда для $\alpha \in (0, 1)$ имеем:
$$\begin{aligned}\sigma^2(\alpha R_1 + (1-\alpha)R_2) &= \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 + 2\alpha(1-\alpha)\rho_{12}\sigma_1\sigma_2 < \\ &< \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)\rho_{12}\sigma_1\sigma_2 < \\ &< \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)\sigma_1^2 = \\ &= (\alpha\sigma_1 + (1-\alpha)\sigma_1)^2 = \sigma_1^2,\end{aligned}$$
где первое неравенство вытекает из условия $\sigma_2 < \sigma_1$, а второе — из условий $|\rho_{12}| < 1$, $\sigma_2 < \sigma_1$ и $0 < \alpha < 1$.
4. $\sigma_1 = 0$ и $\mu_1 = \mu_2 = \mu_3$. В этом случае нет смысла пытаться диверсифицировать активы, поскольку все они приносят одинаковый средний доход, и один из них — безрисковый.
5. Если такие b и ρ_{12} существуют, то b есть решение квадратного (относительно b) уравнения (9.5.21). Для отрицательности его дискриминанта необходимо, чтобы $|\rho_{12}| \geqslant 1$. Поскольку, согласно определению, $|\rho_{12}| \leqslant 1$, возможны только два случая: $\rho_{12} = 1$ или $\rho_{12} = -1$. В любом из них должно выполняться равенство $b = -\rho_{12}\sigma_1/\sigma_2$.
Если $\rho_{12} = 1$, то $(\mu_1 - \mu_0)/\sigma_1 = (\mu_2 - \mu_0)/\sigma_2$. Мы получили равенство двух величин, имеющих вид отношений. Они называются *отношениями Шарпа*. На их основе сравниваются активы, обладающие риском, при условии доступности безрискового вложения. Отношение выражает ожидаемое превышение доходности актива над доходностью безрискового вложения, т. е. $\mu_i - \mu_0$, относительно к риску актива, т. е. σ_i. В качестве другой интерпретации данного отношения можно предложить рассматривать его как меру ожидаемого превышения доходности актива над безрисковой ставкой в шкале, единицей которой служит стандартное отклонение доходности данного актива. Когда $\rho_{12} = 1$, доходности активов 1 и 2 линейно связаны. Кроме того, они имеют одинаковое отношение Шарпа. Таким образом, мы видим, что нет смысла проводить диверсификацию по этим двум активам.
Если $\rho_{12} = -1$, то $(\sigma_1 + \sigma_2)\mu_0 = -(\sigma_1\mu_2 + \sigma_2\mu_1)$. Отсюда выводим, что μ_0, μ_1 и μ_2 не могут все быть положительными. Поэтому не существует диверсификации, включающей все три актива.

Глава 10

1. Так как опцион-пут дает право продавать в момент T, вы сделаете так, если $K > S_T$. В таком случае, вы сначала купите акции по цене S_T и продадите их по цене K, получая чистую прибыль (выплату) размера $K - S_T$. Если $K \leqslant S_T$, нет смысла продавать по цене K, когда рыночная цена равна S_T.
2. $$\begin{aligned}\tilde{E}[Y_{k+1} \mid S_0, \ldots, S_k] &= \frac{1}{(1+r)^{k+1}}\,\tilde{E}[S_{k+1} \mid S_0, \ldots, S_k] = \\ &= \frac{1}{(1+r)^{k+1}}\,\tilde{E}[S_{k+1} \mid S_k] = \frac{(1+r)S_k}{(1+r)^{k+1}} = Y_k.\end{aligned}$$

3. Для опциона-колл $\partial C_0/\partial S_0 = \Phi(d_1) > 0$, где $\Phi(x)$ — функция распределения стандартного нормального закона. Для опциона-пут $\partial P_0/\partial S_0 = -\Phi(-d_1) < 0$.
4. $\partial C_0/\partial\sigma = S_0\sqrt{T}\phi(d_1) > 0$ и $\partial P_0/\partial\sigma = S_0\sqrt{T}\phi(-d_1) > 0$, где $\phi(x)$ — плотность стандартного нормального закона.
5. Для опциона-колл $\partial C_0/\partial T = -S_0\sigma\phi(d_1)/\left(2\sqrt{T}\right) - rKe^{-rT}\Phi(d_2) < 0$.
Для опциона-пут $\partial C_0/\partial T = -S_0\sigma\phi(-d_1)/\left(2\sqrt{T}\right) + rKe^{-rT}\Phi(-d_2)$, которая может иметь любой знак.
7. Цена этого портфеля в момент 0 равна $V_0 = C_0 - P_0 - S_0 + Ke^{-rT}$. В момент T его цена составит $V_T = C_T - P_T - S_T + K$, где $C_T = \max(S_T - K, 0)$ и $P_T = \max(K - S_T, 0)$.
Случай 1: $K > S_T$. Когда $C_T = 0$, $P_T = K - S_T$ и
$V_T = -K + S_T - S_T + K = 0$.
Случай 2: $K < S_T$. Когда $C_T = S_T - K$, $P_T = 0$ и
$V_T = S_T - K - S_T + K = 0$.
Случай 3: $K = S_T$. Когда $C_T = P_T = 0$ и $V_T = 0$.

Литература

1. Bernstein, P. L. *Capital Ideas: The Improbable Origins of Modern Wall Street.* The Free Press, New York, 1993.

2. Black, F. and M. Scholes. The pricing of options and corporate liabilities, *Journal of Political Economy*, 1973.

3. Bodie, Z., A. Kane and A. J. Marcus. *Investments*, Fifth ed. McGraw-Hill/Irwin, Boston, 2002.

4. Chung, Kai Lai. *A Course in Probability Theory*, Third enlarged ed. with supplement of «Measure and Integral». Academic Press, New York, 2001.

5. Chung, Kai Lai. *Markov Chains with Stationary Transition Probabilities*, Second ed. Springer-Verlag, New York, 1967.

6. Cox, J. C. and M. Rubinstein. *Options Markets.* Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

7. David, F. N. *Games, Gods and Gambling.* Hafner Publishing Co., New York, 1962.

8. Duffie, J. D. *Dynamic Asset Pricing Theory*, Third ed. Princeton University Press, 2001.

9. Feller, William. *An Introduction to Probability Theory and Its Applications*, Vol. 1, Third ed. John Wiley & Sons, New York, 1968. (Имеется перевод: Феллер В. *Введение в теорию вероятностей и ее приложения*, В 2-х томах, Т. 1 – М: Мир, 1984, 528 с.)

10. Feller, William. *An Introduction to Probability Theory and its Applications*, Vol. 2, Second ed. John Wiley & Sons, New York, 1971. [Имеется перевод: Феллер В. *Введение в теорию вероятностей и ее приложения*, В 2-х томах, Т. 2 – М: Мир, 1984, 738 с.]

11. Гнеденко Б. В., Колмогоров А. Н. *Предельные распределения для сумм независимых случайных величин*, – Л., М.: Гостехиздат, 1949.

12. Huang, C. F. and R. H. Litzenberger. *Foundations for Financial Economics.* Prentice–Hall, Englewood Cliffs, New Jersey, 1993.

13. Hull, J. C. *Options, Futures and Other Derivatives.* Fourth ed. Prentice-Hall, Englewood Cliffs, New Jersey, 2000.

14. Jacod, J. and R. Shiryaev. Local martingales and the fundamental asset pricing theorems in the discrete-time case, *Finance and Stochastics*, 259–273, 1998.

15. Karlin, Samuel. *A First Course in Stochastic Processes.* Academic Press, New York, 1966. [Имеется перевод: Карлин С. *Основы теории случайных процессов* – М: Мир, 1971, 536 с.]

16. Keynes, John Maynard. *A Treatise on Probability.* Macmillan Co., London, 1921.

17. Knight, Frank H. *Risk, Uncertainty and Profit.* Houghton Mifflin Company, Boston, 1921.

18. Lévy, Paul. *Théorie de l'Addition des Variables Aléatoires.* Gauthiers-Villars, Paris, 1937.

19. Luenberger, D. G. *Investment Science.* Oxford University Press, New York, 1998.

20. Råde, Lennart, et al. *The Teaching of Probability and Statistics.* Proceedings of the First CSMP International Conference, Edited by Lennart Råde. Almqvist & Wiksell Förlag AB, Stockholm, 1970.

21. Sharpe, W. F., G. J. Alexander and J. V. Bailey. *Investments*, Sixth ed. Prentice-Hall, Upper Saddle River, New Jersey, 1999.

22. Uspensky, J. V. *Introduction to Mathematical Probability.* McGraw-Hill Book Co., New York, 1937.

23. Weisweiller, R. (ed.). *Arbitrage.* John Wiley and Sons, New York, 1986.

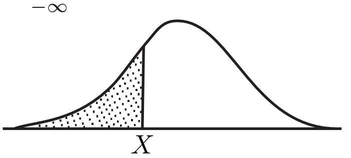
24. Williams Williams, D. *Probability with Martingales.* Cambridge University Press, 1991.

Функция стандартного нормального распределения

Таблица 1

Функция стандартного нормального распределения

$$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, du = P(X \leqslant x)$$



x	0	1	2	3	4	5	6	7	8	9
$-3.$	.0013	.0010	.0007	.0005	.0003	.0002	.0002	.0001	.0001	.0000
-2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0020	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0238	.0233
-1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0300	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0570	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611

Таблица 1. *Продолжение*

x	0	1	2	3	4	5	6	7	8	9
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5363	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000

Предметный указатель

Оглавление

Учебное электронное издание

Чжун Кай Лай
АитСахлиа Фарид

ЭЛЕМЕНТАРНЫЙ КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ. СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ И ФИНАНСОВАЯ МАТЕМАТИКА

Ведущий редактор *М. Стригунова*
Художник *С. Инфантэ*
Оригинал-макет подготовлен *М. Копаницкой* в пакете LaTeX 2_ε

Подписано 27.02.14. Формат $60 \times 90/16$.
Усл. печ. л. 28,44.

Издательство «БИНОМ. Лаборатория знаний»
125167, Москва, проезд Аэропорта, д. 3
Телефон: (499) 157-5272
e-mail: binom@Lbz.ru, http://www.Lbz.ru

Минимальные системные требования определяются соответствующими требованиями программы Adobe Reader версии не ниже 10-й для операционных систем Windows, Android, iOS, Windows Phone и BlackBerry